Discovery of an energy-dependent many-body process in the $K\beta$ spectrum of manganese metal using extended-range high-energy-resolution fluorescence detection with advanced structural insights from principal component analysis

Jack Stephens, Ramesh Rijal, Daniel Sier, Nicholas T. T. Tran, Jonathan W. Dean, Paul Di Pasquale, Tony Kirk, Minh Dao, Chanh Q. Tran, Shusaku Hayama, Sofia Diaz-Moreno and Christopher T. Chantler

IUCr Journals

CRYSTALLOGRAPHY JOURNALS ONLINE

research papers

Discovery of an energy-dependent many-body process in the *Kβ* spectrum of manganese metal using extended-range high-energy-resolution fluorescence detection with advanced structural insights from principal component analysis

Jack Stephens,[a] Ramesh Rijal,[a] Daniel Sier,[a] Nicholas T. T. Tran,[a] Jonathan W. Dean,[a] Paul Di Pasquale,[b] Tony Kirk,[b] Minh Dao,[b] Chanh Q. Tran,[b] Shusaku Hayama,[c] Sofia Diaz-Moreno[c] and Christopher T. Chantler[a]*

[a]School of Physics, University of Melbourne, Melbourne, Victoria, Australia, [b]Department of Chemistry and Physics, La Trobe University, Melbourne, Victoria, Australia, and [c]Diamond Light Source, Harwell Science and Innovation Campus, Didcot OX11 0DE, UK. *Correspondence e-mail: chantler@unimelb.edu.au

The discovery of the novel $n = 2$ satellite transition in the $K\beta$ emission spectrum of manganese and its evolution with incident photon energy are presented. Using the XR-HERFD (extended-range high-energy-resolution fluorescence detection) technique, we conclusively demonstrate the existence of this phenomenon with a statistical significance corresponding to 652 σ_{se} across the measured spectrum, far above the discovery threshold of 3–6 σ_{se}. We apply principal component analysis (PCA) to the XR-HERFD data to extract advanced structural insights. The evolution of this novel spectral feature and physical process are quantified by incorporating regression, revealing the increase in intensity over a wide range of incident photon energies. We validate these findings through independent test data. These results directly challenge the conventional treatment of the many-body reduction factor S_0^2 as a constant independent of incident photon energy in the standard XAFS (X-ray absorption fine structure) equation. Thereby, these results present compelling evidence that S_0^2 should be modelled as a varying function of incident photon energy, marking the first observation of this behaviour in $K\beta$ spectra. This facilitates a greater quantitative understanding of HERFD spectra and a comprehensive representation of many-body effects in condensed matter systems.

1. Introduction

Core–hole spectroscopy describes the measurement of fluorescent photons emitted during the decay of a hole state created by the photoionization of a 1*s* electron (Glatzel *et al.*, 2001; de Groot & Kotani, 2008; Ament *et al.*, 2011; Glatzel *et al.*, 2013). Many applications of core–hole spectroscopy centre around *Kβ* transitions, which arise when a 3*p* electron fills the core vacancy. The 3*d* transition metals, such as manganese, garner a great deal of contemporary interest in *Kβ* spectroscopy due to their strong coupling between unpaired electrons in the 3*d* subshell and the 3*p* hole state which gives rise to distinctive fine structure (Gamblin & Urch, 2001; Glatzel *et al.*, 2003; Glatzel & Bergmann, 2005; Beckwith *et al.*, 2011; Bauer, 2014). As a result, recent applications of X-ray emission spectroscopy (XES) have employed *Kβ* transitions to investigate structure, spin state and oxidation number in transition metals, leading to applications in materials science (Vankó *et al.*, 2006; Lafuerza *et al.*, 2020) and bioinorganic chemistry (Pollock & DeBeer, 2011; Rees *et al.*, 2016; Martin-Diaconescu
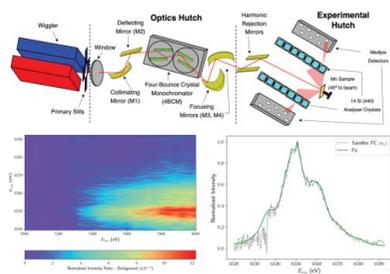
et al., 2016) and investigations of the reaction mechanisms of catalysts and batteries *in operando* (Lancaster *et al.*, 2011; Günter *et al.*, 2016).

Recent research efforts in core–hole spectroscopy centre on many-body phenomena, now referred to as shake processes (Millikan, 1917; Åberg, 1967). The study of shake-off events, wherein two electrons are ejected simultaneously from an atomic system following interaction with a singular incident photon, are a particular recent area of theoretical and experimental investigation (Kheifets, 2022; Pedersen *et al.*, 2023; Kavčič & Žitnik, 2024; Sier *et al.*, 2024; Dean *et al.*, 2024). Fluorescence transitions following shake-off events are referred to as satellites. Due to the presence of an additional electron hole, the characteristic fluorescent photon emitted in a satellite transition is non-degenerate to the single-body diagram counterpart, and hence satellite transitions can explain high emission energy (E_{em}) peaks and asymmetries in X-ray emission spectra (Hölzer *et al.*, 1997).

Shake-off events have a significant impact on the X-ray absorption fine structure (XAFS). The fine structure in absorption spectra reveals crucial structural information, including the type and location of constituent atoms (Stern, 1978). In modern condensed-matter physics, it is modelled by an equation of the form (Rehr & Albers, 2000)

$$\chi(k) = \sum_j S_0^2 N_j \frac{|f(k)|}{kR_j^2} \sin(2kR_j + 2\delta_c + \Phi) \exp\left[-2R_j/\lambda(k)\right]$$
$$\times \exp\left(-2\sigma^2 k^2\right). \tag{1}$$

The meaning of these parameters is discussed in great detail elsewhere (Stern, 1974; Lee & Pendry, 1975; Rehr *et al.*, 1978; Rehr & Albers, 2000; Gaur *et al.*, 2013; Chantler *et al.*, 2024). Most important to this work is S_0^2, which represents an overlap (damping) factor that quantifies how the atomic potential experienced by passive electrons changes between the initial and final states due to the photoionization of a core electron, quantifying the probability that only a single photoelectron is ejected (Lee & Beni, 1977; Rehr *et al.*, 1978; Roy *et al.*, 2001). Physically, S_0^2 is intended to represent a reduction in the amplitude of XAFS oscillations due to many-body phenomena, and is thus referred to as the many-body reduction factor.

In the literature S_0^2 is typically assigned a constant value between 0.6 and 0.9 (Rehr *et al.*, 1978), derived from the total proportion of many-body events in the limit of high photon energy (the sudden approximation), or unity. However, this assumption often results in suboptimal fits to experimental data, particularly in transition metals (Roy *et al.*, 2001). Experimental observations show that the proportion of many-body events increases with incident photon energy (E_{inc}) beyond a certain threshold (a phenomenon we refer to henceforth as evolution) (Carlson & Krause, 1965; Stöhr *et al.*, 1983). This threshold, known as the onset, occurs when E_{inc} exceeds the sum of the core and shake electron binding energies and the satellite transition becomes energetically allowed. These findings have led to the proposal that S_0^2 should be corrected to take some energy-dependent form (Lee &

Beni, 1977; Tran *et al.*, 2023; Sier *et al.*, 2024). A comprehensive reformulation of S_0^2 requires precise measurement of many-body processes in a given material. No studies to date have reported measurements of evolving shake-off satellites in $K\beta$ spectra, highlighting a significant gap in the literature. To interrogate the nature of S_0^2 in the $K\beta$ fluorescence channel, we employ the XR-HERFD technique to collect two-dimensional spectra over an extended range of E_{inc} and observe new many-body processes in the $K\beta$ spectrum of manganese. By isolating prominent components of the data, we present the first quantified amplitude profile of such a process in a $K\beta$ spectrum, showcasing quantum evolution.

2. The XR-HERFD technique

Measuring the variance in intensity of an atomic transition with incident photon energy requires the use of a two-dimensional spectroscopic technique such as RIXS (resonant inelastic X-ray scattering) that employs a combination of X-ray absorption spectroscopy (XAS) and X-ray emission spectroscopy (XES) (Eisenberger *et al.*, 1976; Guo *et al.*, 1995; Glatzel & Bergmann, 2005; Ament *et al.*, 2011; Schmitt *et al.*, 2014; Gel'mukhanov *et al.*, 2021). RIXS requires high-resolution detection apparatus and high-flux photon sources (Sier *et al.*, 2025), and so the advent of HERFD (high-energy-resolution fluorescence detection) spectroscopy in the late twentieth century has catalysed its experimental realization (Hämäläinen *et al.*, 1991). HERFD substantially improves experimental resolution by isolating a specific transmission line as a fluorescence channel, minimizing the effect of core–hole lifetime broadening on the achievable energy resolution of the fluorescence spectrum (Glatzel *et al.*, 2013; Günter *et al.*, 2016). Often, this allows for an energy resolution finer than the core–hole lifetime, which has facilitated a boom in core–hole spectroscopy (Hämäläinen *et al.*, 1991) and enabled detailed investigations into shake-off satellites (Deutsch *et al.*, 1996), which are often suppressed by several orders of magnitude relative to their single-body analogues (Dean *et al.*, 2024).

RIXS scans, which typically focus on resonant transitions in the pre-edge region (Eisenberger *et al.*, 1976; Glatzel & Bergmann, 2005; Ament *et al.*, 2011), are often insufficient to investigate shake-off events. To address this limitation, the XR-HERFD (extended-range HERFD) technique was developed to investigate the structure and evolution of novel processes well above a given transition edge (Tran *et al.*, 2023; Sier *et al.*, 2024; Sier *et al.*, 2025). In this work, the XR-HERFD technique is employed to uncover the structure and evolution of a previously undetected satellite transition in the $K\beta$ spectrum of manganese metal.

3. Experimental measurement

Following Tran *et al.* (2023) and Sier *et al.* (2024), measurements were conducted on the I20-Scanning beamline at Diamond Light Source. The experimental apparatus, depicted in Fig. 1, ensures high incident photon flux ($\sim 10^{12}$ s^{-1}) and beam uniformity and monochromaticity through a combina-
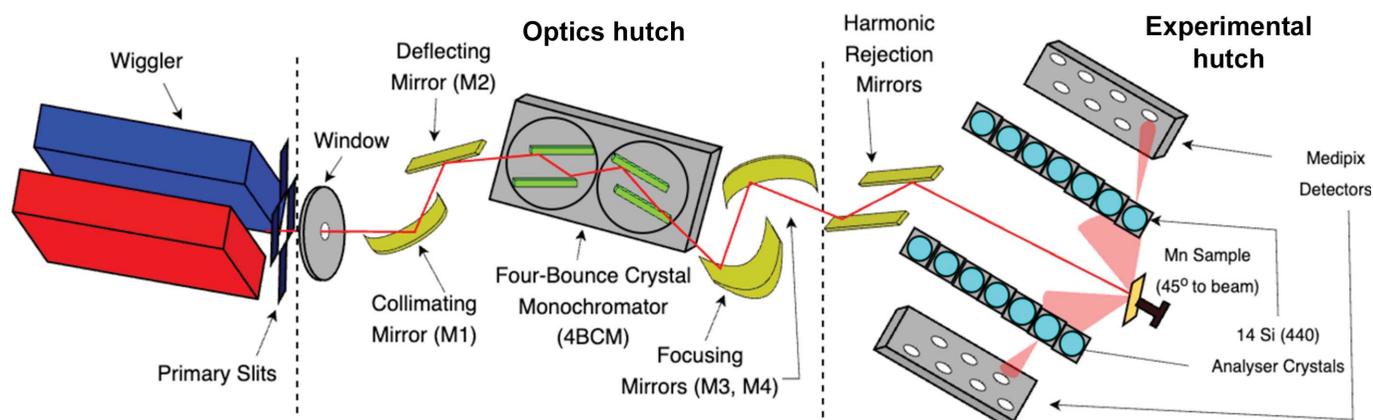
Figure 1
Schematic diagram of the beamline apparatus on I20-Scanning, which includes several advanced optical components. X-rays delivered by the wiggler source are deflected upwards by a vertical collimating mirror with rhodium and platinum stripes (M1) to maximize flux, before being downward deflected by the deflecting mirror (M2) to restore the beam's horizontal trajectory and isolate any mirror adjustments. The X-rays then pass through the custom-built Si(111) four-bounce crystal monochromator (4BCM), ensuring the incident X-ray energy is well defined. Vertical and horizontal focusing mirrors (M3, M4) after the monochromator focus the X-rays on the sample position, while a set of harmonic rejection mirrors with rhodium and silicon stripes maintain spectral purity by reducing the content of higher-order harmonics in the X-ray beam before it is incident on the manganese sample. Fluorescent X-rays from the sample are Bragg-reflected by two sets of seven spherically bent and cylindrically sliced Si(440) analyser crystals, angled close to back-scattering. The analysers lie on a Rowland circle of 1 m diameter in the vertical plane, based on Johann-type geometry. Fluorescent photons are directed by the analysers to seven independent regions on each of the two respective Medipix detectors. This setup allows 14 separable and independent measurements of the XR-HERFD signal.

tion of focusing, collimating and harmonic rejection mirrors alongside a purpose-built four-bounce monochromator (Diaz-Moreno *et al.*, 2009; Diaz-Moreno, 2012; Hayama *et al.*, 2021; Sier *et al.*, 2025). Fourteen Si(440) Bragg analyser crystals in the Johann geometry were employed, which recorded the $K\beta_{1,3}$ maximum at a Bragg angle $\theta_B \simeq 84.2°$. This configuration significantly increases the total subtended detector solid angle relative to previous results in manganese $K\alpha$ that employed at most three Bragg analyser crystals, hypothetically improving the statistical quality by a factor of $(14/3)^{1/2} \simeq 2.2$.

Fluorescent photons of the allowed energy are Bragg-reflected by the analyser crystals onto two MAXIPIX (multichip area X-ray detector based on a photon-counting pixel array) TAA22PC detectors, an advanced iteration of the Medipix2 single-photon counting pixel detector. The detectors are configured in a 4×1 arrangement, creating a total pixel grid of 1024 × 256 pixels, and are collectively referred to as Medipix detectors for simplicity. Each analyser is arranged independently in the Johann geometry, lying on a Rowland circle with a diameter of 1 m with the sample and respective Medipix detector.

3.1. Data processing

The processing procedure for experimental data follows a methodology similar to that used in the analysis of manganese $K\alpha$ spectra (Sier *et al.*, 2025). A region of interest (RoI) is defined around the image of each analyser crystal on the Medipix detector to minimize false counts from electronic noise and stray background photons that are not Bragg-reflected by the analyser crystals. A corner of the detector region well removed from any RoI is selected to define inherent background counts at each incident energy, which were subtracted from the intensities recorded in each RoI. The

resulting intensities were then normalized by upstream ion counts.

The next processing step, known as binary data splicing, addresses variations in the Bragg angle θ_B seen in the dispersion direction of the detector apparatus, corresponding to the vertical plane of the analyser crystals. These variations arise as the Bragg condition is precisely met only at the centre of each analyser crystal, due to their finite size and the geometry of the detector apparatus (Moretti Sala *et al.*, 2018). To correct for this, the RoI of each analyser crystal was divided into ten narrow horizontal slices. A Gaussian profile is then fitted to the peak of the $K\beta_{1,3}$ transition of the isolated emission spectrum of each slice, aligning them to a reference spectrum.

Following alignment, the fluorescent intensities are extracted for each analyser crystal by aggregating across the ten slices. A sum is performed across each of the analyser crystals, weighted by standard error counts, to obtain the final processed XR-HERFD data. By incorporating binary data splicing, intrinsic instrumental broadening is minimized, increasing the limiting resolution of the data by approximately 10%, resulting in a fluorescent energy bandwidth of 0.61 eV. As a result, any remaining broadening in the XR-HERFD spectrum can be attributed predominantly to physical phenomena and investigated by advanced theory. A quantified description of the improvements afforded by an application of binary data splicing is given by Sier *et al.* (2025) and Rijal *et al.* (2025).

Fig. 2(*a*) displays the fully processed XR-HERFD intensity map. Two characteristic peaks are evident and presented in further detail in Fig. 2(*b*). The first peak at $E_{em} = 6490.4$ eV corresponds to the $K\beta_{1,3}$ transition, arising from electric dipole transitions when a $3p$ electron fills the $1s$ core hole. The second peak, at $E_{em} = 6535$ eV, represents the $K\beta_{2,5}$ transition,

primarily composed of quadrupole transitions from the $3d$ to $1s$ orbitals. Due to electronic selection rules, this peak is suppressed by several orders of magnitude relative to $K\beta_{1,3}$.

In line with our objective to resolve novel satellites in manganese $K\beta$, we examine the high E_{em} region of the XR-HERFD map in more detail in Fig. 3(a). A faint structure emerges past $E_{inc} \simeq 7300$ eV in the region between $E_{em} = 6540$ and 6560 eV. This previously unobserved physical phenomenon in manganese $K\beta$ highlights the effectiveness of XR-

HERFD in resolving processes well above the absorption edge. Due to the separation of the analyser crystal images on our detector apparatus, we can display standard error uncertainties (σ_{se}) as in Fig. 3(b), quantifying the precision of our results.

4. Empirical isolation of the satellite

Whilst Fig. 3(a) appears to reveal a new satellite in manganese $K\beta$, garnering any insights on its structure or characteristics appears at this stage to be incredibly difficult and imprecise. Shake-off processes are generally weak compared with their single-body counterparts, although the meaning of this remains quite unclear in the experimental literature and theoretical investigations are only now beginning to make verifiable estimates. Here, the isolation of the satellite is exacerbated due to the intrinsically lower intensity of $K\beta$

(a)

(b)
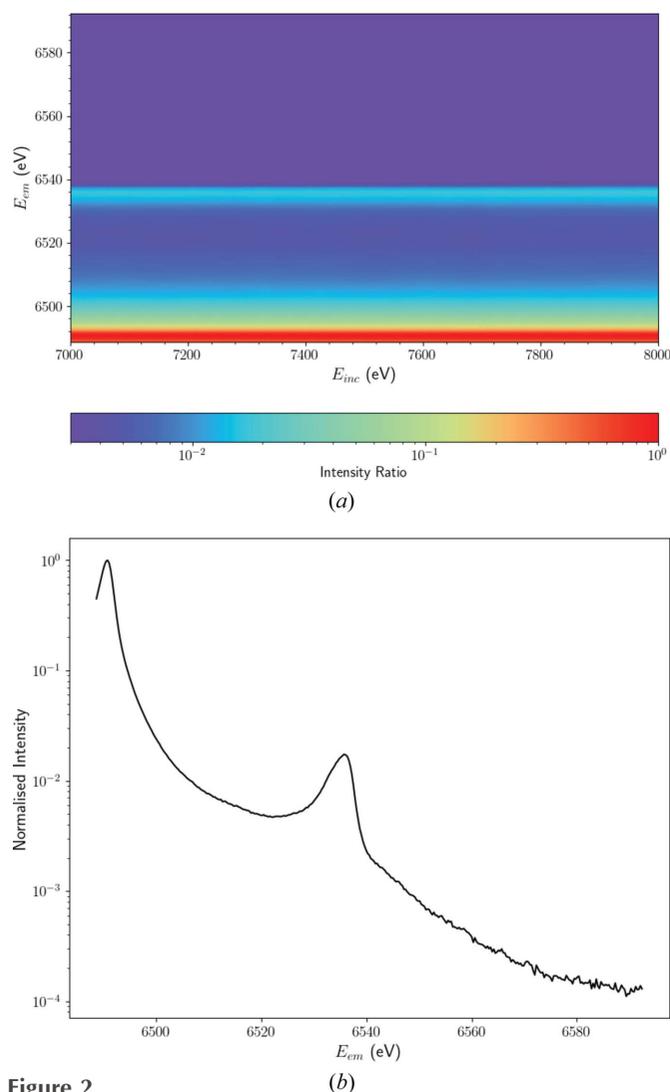
Figure 2
(a) XR-HERFD contour plot of the average ratio of fluorescent counts to upstream ion-chamber counts of the crystal analysers in Mn metal. Two spectral features are observed. The first is the $K\beta_{1,3}$ transition at an emission energy (E_{em}) of 6490.4 eV, which arises due to a $3p$ electron filling the vacancy of the core photoelectron. The second feature is the $K\beta_{2,5}$ transition centred at $E_{em} = 6535$ eV, which arises predominantly due to $3d$ electrons filling the core vacancy. The incident energy (E_{inc}) ranges are well above the K edge and hence the relative intensity of the two processes observed remains consistent. The structure of these two spectral features is evident in more detail in panel (b), which portrays an aggregated HERFD-XES slice of the contour plot from $E_{inc} = 7000$ to 7160 eV. A logarithmic scale is required on the vertical axis to depict the $K\beta_{2,5}$ transition in detail, which is suppressed by several orders of magnitude relative to $K\beta_{1,3}$ due to its primary origin from quadrupole transitions. Notably, there are no spectral features evident in the high-energy ($E_{em} > 6540$ eV) region.

(a)

(b)
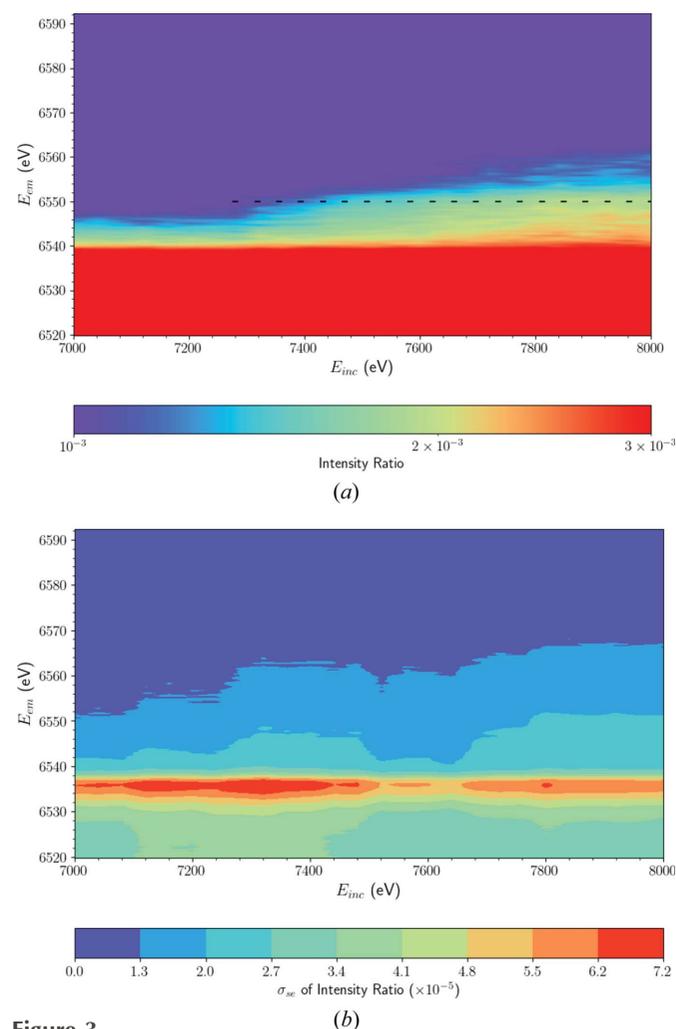
Figure 3
(a) XR-HERFD contour plot, analogous to Fig. 2, with restricted domain to focus on the high-energy region. The emergence of a spectral feature is evident past an incident energy $E_{inc} = 7280$ eV, marked by the dashed line at an emission energy $E_{em} = 6550$ eV. We attribute this spectral feature to a $K\beta$ satellite of Mn. (b) Propagated standard error (σ_{se}) for the XR-HERFD map of intensity ratios in panel (a).

Jack Stephens *et al.* · Energy-dependent many-body process in Mn $K\beta$ spectrum

compared with $K\alpha$ and the anticipated dominance of Auger decay in the shake-off channel (Dean *et al.*, 2024).

We address this issue using data processing with a background subtraction. Here, the reference background is defined as processes which are stable with respect to increasing E_{inc} at some reference energy well above the K edge. To define this background spectrum, we first posit a lower bound for the onset of the satellite process. Based on observations of Fig. 3(*a*) this occurs prior to $E_{inc} = 7280$ eV, and so we define our background as an aggregate of the normalized HERFD-XES slices from $E_{inc} = 7000$ to 7160 eV. This corresponds to the emission spectrum seen in Fig. 2(*b*), incorporating the $K\beta_{1,3}$ and $K\beta_{2,5}$ transitions and their high-energy tails. Note we do not assume that this is stable, only that it is relatively and sufficiently stable.

By subtracting this (scaled) background spectrum from the normalized XR-HERFD data, we are left with a clear depiction of spectral features that emerge only well above the K edge and reference incident energies E_{inc}. Fig. 4 presents a collection of the resulting HERFD-XES slices after subtraction. The signal-to-noise ratio is improved significantly and we can clearly see the emergence of a consistently shaped dual-peak substructure past $E_{inc} = 7280$ eV. The level of noise in Fig. 4 still appears large, with quite significant statistical uncertainty on a given pixel point in the map, but we note that the structure appears to be both consistent and becoming

stronger with increasing energy E_{inc}. The apparent improvement is a consequence of the elimination of electronic noise in the XR-HERFD data (Appendix *A*), selective use of the eight highest-resolution analyser crystals (Appendix *B*) and combining statistics. That we are able to resolve the satellite's structure and evolution, given these circumstances and the relatively low intensity of the satellite, is a triumph for XR-HERFD.

The background-subtracted XR-HERFD map is depicted in Fig. 5(*a*). Here, the satellite is clearly significant and distinct from residual noise and more clearly identified.

The number of standard errors for each pixel in Fig. 5(*a*) is depicted in Fig. 5(*b*). The significance increases as the satellite
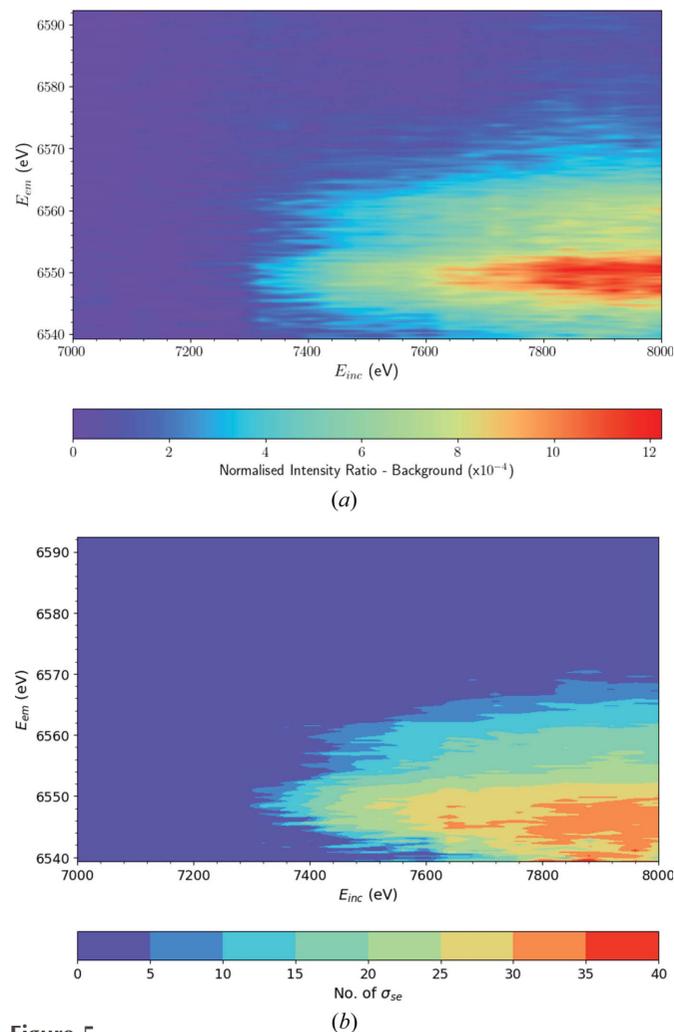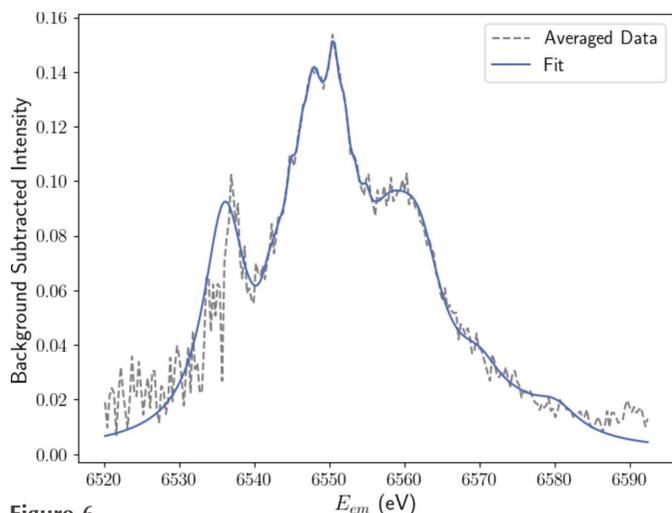

Figure 4
Stack plot of a collection of HERFD-XES slices following the subtraction of the $K\beta$ background. The background, which corresponds to the stable transitions in this incident energy (E_{inc}) range, is evident in Fig. 2(*b*). The signal-to-noise ratio is improved due to the elimination of electronic noise in the data (Appendix *A*), selective use of the eight highest-resolution analyser crystals (Appendix *B*) and the naturally suppressed intensity of the $K\beta$ satellite. While only noise fluctuations are evident at low E_{inc}, two distinct and prominent peaks emerge and remain evident in the substructure beyond 7280 eV. This is definitive evidence for the evolution of the satellite.

Figure 5
(*a*) XR-HERFD map beyond $K\beta_{2,5}$ after background subtraction using eight analyser crystals. A dual-peak substructure clearly emerges beyond an incident energy $E_{inc} = 7280$ eV and is distinct from the noise fluctuations seen at higher emission energies (E_{em}). (*b*) XR-HERFD contour plot of the significance of the satellite region, displaying the number of standard errors (σ_{se}) following subtraction of the background. The number of standard errors per pixel around the peak E_{em} of the satellite, ~6550 eV, increases from effectively 0 prior to the onset to over 30 at the maximum E_{inc} for just a single pixel. This is well beyond the usual threshold for discovery (3–6 σ_{se} for the total integrated signal) and corresponds to a total signature of 652 σ_{se} in the satellite region. This is a new satellite.

Figure 6
Extracted satellite, obtained by averaging across the background-subtracted spectra for incident energies from 7840 to 8000 eV. This is then fitted empirically with a sum of Lorentzians, concentrated around prominent features in the data, to obtain a detailed initial guess for the structure of the satellite. Notable features include a main double peak and prominent low- and high-energy shoulders. Note that this extraction does not use the full data set but uses the strongest statistical representation of the new spectrum. Empirical extraction is prone to the propagation of systematic errors or noise which may distort the structure of novel spectral features. We thus turn to statistically rigorous methodologies in Section 5 to extract detailed structural insights.

evolves and is above 20 at each pixel along the satellite peak. This lies far above the typical level for discovery (σ_{se} = 3 to 6) and results in a total integrated significance of 652 σ_{se}.

We need to do better, so we average the background-subtracted spectrum from E_{inc} = 7840 to 8000 eV to yield a better defined representation of the satellite structure. Across this higher energy range, the satellite intensities begin to stabilize. This representation is evident in Fig. 6. For visual clarity and to suppress noise, we fitted this structure with a sum of Lorentzians weighted around prominent features in the spectrum.

The satellite appears to consist of a main peak with both low- and high-energy shoulders, while the main peak itself exhibits an intriguing double apex. This detailed fine structure suggests that the satellite has origins in multiple physical phenomena, as has been concluded for the $K\alpha$ satellite (Sier *et al.*, 2024). Note that this does not use the full data set but uses the strongest statistical representation of the new spectrum. It might be recognized that the background subtraction technique is prone to propagating systematic errors or noise into the XR-HERFD data, particularly given the low intensities investigated here, which could mislead researchers or conceal key structural insights. Hence, we turn to a more statistically rigorous and complete approach.

5. Investigation with principal component analysis

To extract detailed structural information on the satellite, we employ the machine-learning processing technique known as principal component analysis (PCA). By transforming the original data set into a new set of uncorrelated variables (the 'principal components', PCs) through linear combinations that maximize variance of the next component, PCA enables dimensionality reduction without significant information loss, making complex data sets easier to analyse and interpret (Jolliffe, 2002; Ramsay & Silverman, 2006; Witten *et al.*, 2016; Jolliffe & Cadima, 2016; Hao & Ho, 2019). To perform PCA we use the *scikit-learn* (*sklearn*) library in Python (Pedregosa *et al.*, 2011; Buitinck *et al.*, 2013) with some minor modifications. A complete description of our PCA procedure is presented in Appendix C.

Fig. 7(*a*) depicts the weight (loading) of each E_{em} in the first two PCs. This reveals clear physical interpretations for both. Those of the first PC (p_1) correspond to the satellite, closely resembling the structure extracted in Fig. 6. We shall thus refer to p_1 as the satellite (component). The second PC (p_2) and above possess similar structures, dominated by correlated residual noise in the $K\beta$ tail, as indicated by the higher intensity around E_{em} = 6535 eV, where the peak of $K\beta_{2,5}$ resides. When extended to include the entire range of E_{em}, each of these higher-order components reveals dominant features at ∼6490 eV, corresponding to the $K\beta_{1,3}$ peak. Given that their only significant features emerge within characteristic spectral profiles, we can attribute p_2 and above to correlated noise. Additional physical features may be hiding within these components, but show no prominent contributions to satellite spectra.
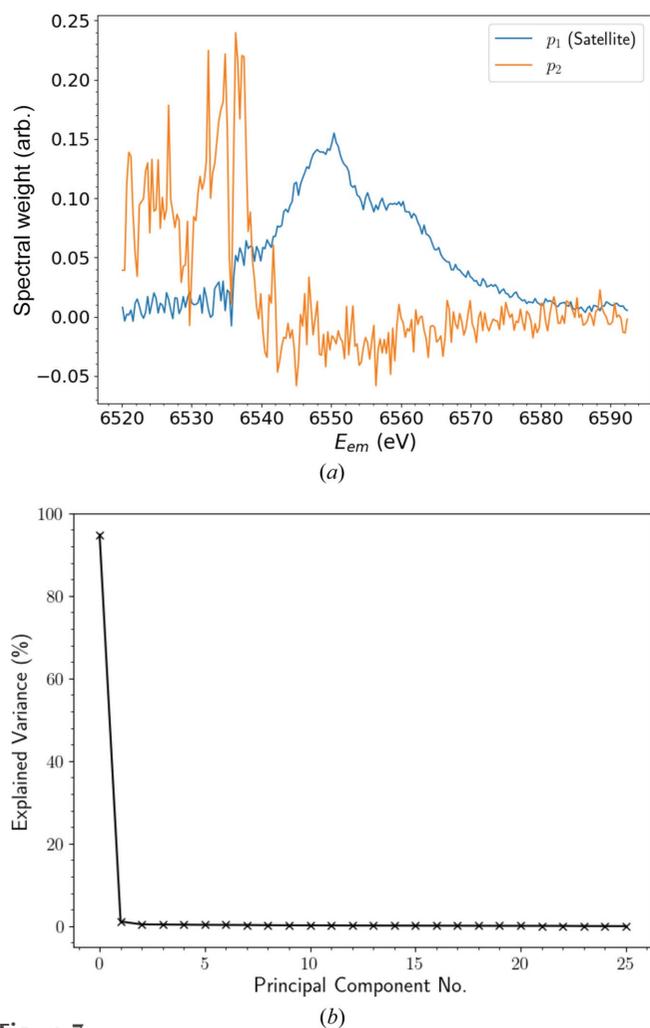
The impact of pre-processing artefacts is evident in the variation of component weights based on the selected range of E_{em}. The choice of normalization can substantially affect the apparent variance within characteristic spectral features, which contribute less to the overall variance given their stability in the high E_{inc} region of our data set. Notably, correction for Bragg angle variation through binary data splicing and removal of intensity aberrations (Appendices *A* and *B*) have enhanced the physical interpretability of our components, which may otherwise have been dominated by experimental systematics rather than physical spectral variation.

Fig. 7(*b*) reveals the proportion of variance across the background-subtracted data explained by each PC. Clearly, the satellite is the most prominent component, explaining >95% of the total variance, while each of the noise components contributes 1% or less. In the remainder of our analysis, we can thus neglect or reject p_2 and above and focus solely on the satellite component without significant information loss.

We garner further structural insights from the satellite component by fitting a sum of Lorentzians (Fig. 8). Note that in Fig. 6 we were able to extract a detailed representation of the most important p_1 component using our 'manual' approach, yet also note that there are limitations of interpretation with any such approach including PCA. The resulting structure closely resembles our fit of the averaged empirical satellite in Fig. 6 but with significantly less noise, in part because the whole data set can be used. This demonstrates the effectiveness of applying PCA to XR-HERFD data sets, as it allows us to extract key evolutionary components and remove almost all of the (randomly) distributed residual

Jack Stephens *et al.* · Energy-dependent many-body process in Mn $K\beta$ spectrum   **553**

noise, allowing for physical interpretation of the resulting structure.

As discussed in Section 4, the satellite component possesses significant fine structure. A prominent double peak stands at $E_{em} \simeq 6550$ eV, straddled by high- and low-energy shoulders at 6560 and 6535 eV, respectively. The low-energy shoulder was not consistently resolved in the background-subtracted data in Figs. 4, 5(a) and 6 due to noise fluctuations in the $K\beta_{2,5}$ tails. Even in the satellite PC, the shoulder remains somewhat ambiguous. The dominance of correlated noise in each of the p_2 and above components in the region of the low-energy shoulder is a likely cause for this discrepancy. Intensities are two orders of magnitude greater in the $K\beta_{2,5}$ profile than in the satellite and hence 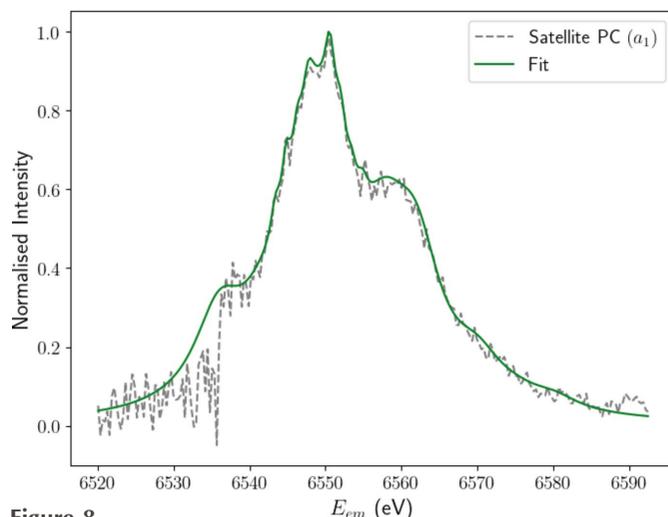any fluctuations in this region are bound to obscure the resolution of evolving satellite features. Regardless, the entirety of the structure is resolved in much greater detail than in our earlier analysis, due to the combination of statistics.

The level of fine structure evident in the satellite component suggests it has multiple near-degenerate physical origins, typical of a shake-off satellite. Notable features offer insights into the origin of the process. The satellite is displaced by ~60 eV in E_{em} from the $K\beta_{1,3}$ peak and the onset is ~700 eV past the K edge. In addition to the low intensity of the process, these factors indicate that the physical origin of the satellite is $n = 2$ shake-off events in the $K\beta$ spectrum. Specifically, we expect either a $2p$ or $2s$ electron to be 'shaken-off' during the photoionization of a core electron, before a $3p$ electron fills the core hole in a high-energy environment, resulting in the emission of a photon non-degenerate to the characteristic $K\beta_{1,3}$ spectrum. This aligns with theoretical predictions for manganese (Dean et al., 2024).

The weights in Fig. 7(a) are applied to the background-subtracted data to produce linear combinations which are representative of the amplitude of a given PC at each E_{inc}. Depicted in Fig. 9, the amplitude of the satellite component (a_1) increases monotonically from 0 to a maximum as E_{inc} of the HERFD-XES slices passes 7200 eV, indicating that the satellite feature p_1 is growing stronger, evolving with increasing E_{inc}. This is equivalent to the growth of a consistent process observed in Fig. 4, but with a reduction in noise and dimensionality. The significance of this observation is highlighted by the amplitudes of the a_2 component and above, which oscillate around zero without any discernible pattern. This confirms that these components arise from correlated noise contributions and not from physical phenomena that contribute prominent variance to the $K\beta$ background spectra. An application of k-means clustering to the data following

Figure 7
(a) Weights of the first two principal components (PCs) as a function of emission energy E_{em}. A visual inspection makes it immediately apparent that the first PC can be interpreted physically as the satellite (p_1) and the second (and higher) PCs correspond to (correlated) noise (p_2). The satellite closely resembles the initial estimate in Fig. 6 but with increased structural detail. The first component of noise has a higher magnitude in the vicinity of the $K\beta_{2,5}$ peak, a consequence of the larger intensities and hence noise in that region. (b) Fraction of the total variance explained by each subsequent PC. The satellite component accounts for over 95% of the variance across the background-subtracted XR-HERFD data set. Hence we can focus on the satellite component without sacrificing any loss of information.

Figure 8
The satellite principal component (p_1) fitted with a sum of Lorentzians to emphasize key features. The structure appears similar to that derived in Section 4, but with greater detail, revealing a main dual peak with shoulders on the high- and low-energy sides. The level of fine structure depicted suggests that numerous components contribute to the overall satellite structure, which is indicative of a shake process.

PCA projection reveals much the same insights, but separates the HEFD-XES slices rather arbitrarily with regard to a_1 and a_2 amplitudes, limiting physical interpretability.

6. Evolution of the satellite

The form of the amplitudes in Fig. 9 clearly indicates the evolving intensity of the satellite past an onset at approximately 7200 eV. Near the onset, E_{inc} is approximately equal to the combined binding energies of the two ejected electrons, and hence the probability of ejection for the secondary photoelectron is low. Since the primary photoelectron exits the atomic system with low kinetic energy, the timescale for its ejection from the atom is large and thus the mechanisms behind the shake-off event must be considered through Coulomb interactions, necessitating solutions to the time-dependent Schrödinger equation in an incredibly complex atomic system (Mukoyama *et al.*, 2009; Valenza *et al.*, 2017).

As the incident photon energy increases, the timescale of the primary photoelectron ejection continues to decrease as a result of its greater kinetic energy and velocity, resulting in a gradual increase in shake-off probability, as was observed experimentally through the increase in fluorescence intensity. Well past the satellite onset, the kinetic energy of the primary photoelectron becomes so large that its ejection can be treated as instantaneous and without interaction with the remaining atomic system in what is known as the sudden approximation (Dean *et al.*, 2024). In the sudden approximation, the mechanism of the shake process can be derived entirely through the relaxation of atomic orbitals following ionization of the primary electron (Kochur & Popov, 2006).
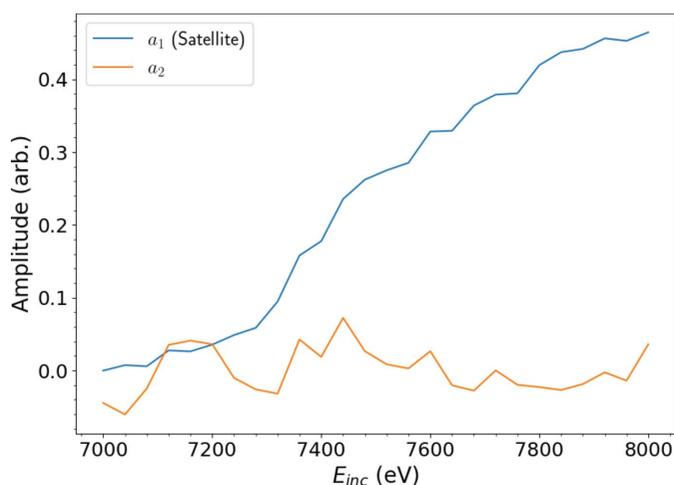

Figure 9
Amplitudes of the first two principal components of the background centred data. These correspond to linear combinations of the background-subtracted XR-HERFD data weighted by the satellite and noise components defined in Fig. 7. The monotonic increase of the satellite amplitude past an onset around an incident energy E_{inc} = 7200 eV is indicative of many-body process evolution. This begins to slow past 7800 eV, as was observed with an analysis of empirical data in Section 4. The amplitude of the noise component oscillates without any clear pattern around 0, confirming that it does not correspond to any significant or persistent structure.

Under the sudden approximation, the shake-off probability approaches a maximum value, referred to as $P(\infty)$. This limiting shake probability could be interpreted as a basis for the current representation of S_0^2 (Roy *et al.*, 2001). Contemporary XAFS models and analysis consider a constant probability of shake events, which is fundamentally in disagreement with experimental results observed here and elsewhere (Tran *et al.*, 2023; Sier *et al.*, 2024). To demonstrate this, we analyse and model the evolution of the satellite.

6.1. Extracting a functional form for many-body evolution

By applying PCA to our XR-HERFD data set, we have obtained a representation for the structure of the satellite significantly bereft of correlated noise [Fig. 7(*a*)] and an initial guess for the functional form of the satellite evolution (Fig. 9). To test the robustness of the PC satellite structure, we implement a form of principal component regression, which attempts to express a set of data as a weighted linear combination of its PCs (Witten *et al.*, 2016). As Fig. 7(*b*) depicts, the satellite (p_1) explains >95% of the variance from the refined $K\beta$ background profile (μ'). Hence we can ignore contributions from the correlated noise components and attempt to recreate the XR-HERFD data (X) using p_1 and our refined $K\beta$ background profile μ' (defined in Appendix C),

$$X(E_{inc}, E_{em}) = \alpha_1(E_{inc}) p_1(E_{em}) + \beta(E_{inc}) \mu'(E_{em}) + \epsilon(E_{inc}, E_{em}), \tag{2}$$

where ϵ is a residual term. To test the effectiveness of this representation (which we term the PCA model), we compare the performance of an alternative model. This model replaces the relevant terms in equation (2) with the forms of the background [Fig. 2(*b*)] and satellite (Fig. 6) we determined in Section 4 by aggregating relevant segments of the HERFD-XES slices in X. We denote this model as the aggregate model.

Fig. 10 presents the results of this comparison. While the aggregate model provides a better fit at low E_{inc}, the PCA model frequently outperforms it as E_{inc} progresses beyond the satellite onset (~7200 eV). This is demonstrated by the PCA model yielding an average goodness-of-fit χ_r^2 of 9.2, which is approximately 25% lower than that of the aggregate model (12.0). Hence the parameters extracted using PCA yield an improved and accurate representation of the XR-HERFD data. Therefore, we adopt α_1 from equation (2) as the functional form to describe the evolution of the satellite intensity.

6.2. Comparison of evolutionary models

We investigate the evolution of the satellite (component) from our observed data compared with theoretical predictions from three literature models for satellite or quantum mechanical process evolution. The first, the most commonly cited quantum evolution model (Thomas, 1984), is

$$P(E_P) = P(\infty) \exp\left[\frac{-R^2 E_B^2}{15.32(E_P)}\right], \tag{3}$$

where $P(\infty)$, defined here as the change in potential due to the ejected core electron divided by the binding energy

squared, is the limiting shake probability in the sudden approximation, R is the atomic size corresponding to the radius of the shake-off orbital in ångströms, E_B is the binding energy of the shake electron, $E_P = E_{inc} - E_{edge} - E_B$ is the energy of the shake electron and the constant is given by $m_e/2\hbar^2$ in units of electronvolts and ångströms.

Roy *et al.* (2001) developed an alternative form for quantum evolution, employing the Slater form of one-electron wavefunctions and assuming the time dependence carries an exponential form,
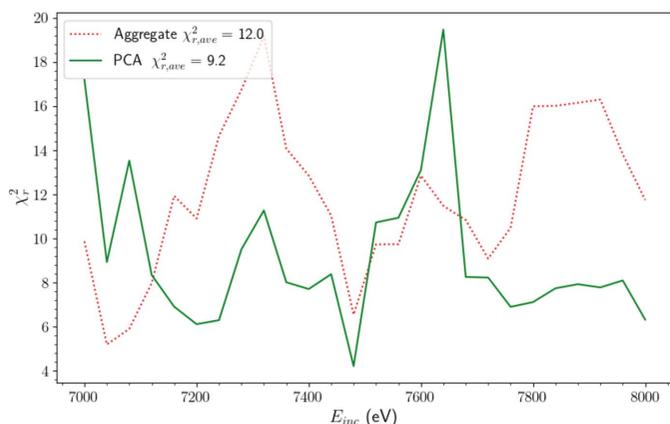
$$P(E_P) = P(\infty)\left[\frac{2^{2n}(n-1)!(n+1)!}{\pi(2n-1)!}\right]E_B^{n+1/2}$$
$$\times \int_0^{E_P-E_B} \frac{E^{1/2}}{(E_B+E)^{n+2}[1+\tau_0^2(E_B+E)^2]}\,dE, \quad (4)$$

where n is the principal quantum number of the shake shell, $\tau_0 = R/(2E_P)^{1/2}$ is the characteristic time for the interaction and $P(\infty)$ is obtained by performing the high-energy limit of the integral and setting $\tau_0 = 0$. All variables in the Roy formalism are expressed in Hartree atomic units ($m_e, \hbar, e, 4\pi\epsilon_0 = 1$), which we convert to electronvolts and ångströms when presenting our results.

Mukoyama *et al.* (2009) adapted the Thomas model using the exponential time dependence from Roy to obtain

$$P(E_P) = \frac{P(\infty)}{1 + [R^2 E_B^2/(15.32 \times E_P)]}. \quad (5)$$

While Mukoyama and Thomas considered bound–bound transitions (shake-up) and Roy bound–free (shake-off), each model is sufficient to describe the evolution of shake prob-abilities. We examine the evolution of $I_{sat}/I_{K\beta}$ as a function of E_{inc}, which reflects the probability of an $n = 2$ satellite transition relative to the entire $K\beta$ spectrum. We determine this by extracting α_1 from equation (2) and multiplying it by the satellite area as a fraction of the total $K\beta$ spectrum $[A_{P_1}/\beta(E_{inc})A_{K\beta}]$. A_{P_1} is the area of the satellite PC defined in Fig. 8, $A_{K\beta}$ is the area of the PCA-extracted background μ' extended to incorporate the full $K\beta$ spectrum and $\beta(E_{inc})$ is the scaling parameter of the background from equation (2).

We fitted the models to our evolution data while leaving R, $P(\infty)$ and E_B as free parameters and extracted statistical uncertainties for each. The results are depicted in Fig. 11, alongside data for the $K\alpha$ satellite from Sier *et al.* (2024). Clearly, many-body transitions do not display a sharp jump in intensity after their onset, deviating significantly from the well documented step-function-like increase claimed in the near-edge region of single-body spectra. Instead, they gradually evolve after the onset incident energy has been reached, slowly increasing in intensity over a spread of incident energies before approaching a maximum intensity.

The performance of each model is quantified by the fitting parameters (Table 1). Reference values were calculated using the multi-configuration Dirac–Hartree–Fock (MCDHF) method with a relativistic quantum electrodynamics approach in the *General Relativistic Atomic Software Package* (*GRASP*) by Sier *et al.* (2024). While these were originally derived for analysis of the $K\alpha$ satellite, the parameters E_B, $P(\infty)$ and R
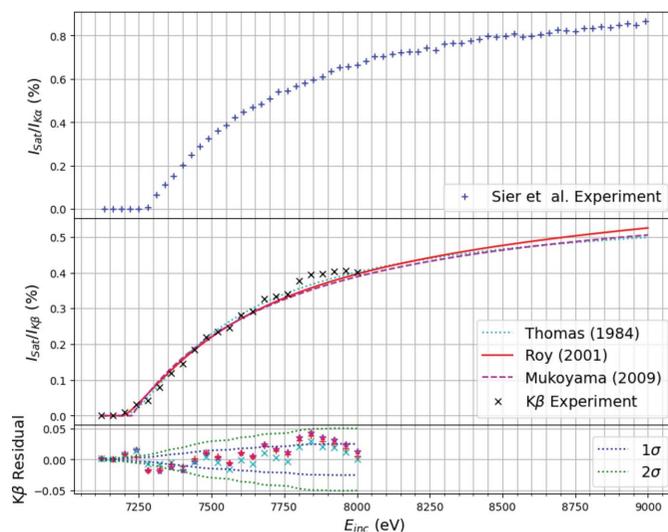
Figure 10
Performance of two separate models fitted to HERFD-XES slices in the XR-HERFD data evaluated across the entire range of incident energies E_{inc}. The PCA model follows the functional form in equation (2), employing the satellite (p_1) and background (μ') extracted using PCA in Section 5. The aggregate model substitutes the relevant terms in equation (2) with the satellite (Fig. 6) and background [Fig. 2(*b*)] we obtained by averaging over relevant HERFD-XES slices in Section 4. The PCA model frequently outperforms the aggregate model, especially past $E_{inc} \simeq$ 7200 eV, yielding an average χ_r^2 25% lower than the aggregate model. This is a strong indicator that the PCA algorithm accurately captures the satellite structure.

Figure 11
(Top) Evolution of the $K\alpha$ satellite intensities from Sier *et al.* (2024). (Middle) Evolution of the relative intensity of the $K\beta$ satellite extracted from experimental data after PCA processing. The line shape of the $K\beta$ evolutionary curve is remarkably similar to that of the $K\alpha$ satellite. Note, however, the relative intensities are consistently a factor of 1.5–2 times lower and the onset appears to be 50–100 eV earlier. The results for $K\beta$ are fitted to three models describing the evolution of quantum probability using the parameters in equation (2). (Bottom) Residuals of the $K\beta$ satellite. The Thomas model consistently returns the lowest residual to the experimental data. However, evolution of the experimental intensity ratios begins to slow past an incident energy $E_{inc} = 7800$ eV, which disagrees with the models.

Table 1
Fitting parameters and associated statistical uncertainties extracted from the Roy, Thomas and Mukoyama models for the evolution of the $K\beta$ satellite.

Ab initio calculations alongside fitted parameters for the Thomas model (determined the best performing) for the $K\alpha$ satellite in Sier *et al.* (2024) are included as a reference point. $P(\infty)$ is the limiting shake probability relative to the entire $K\beta$ spectrum in the sudden limit, R is the radius of the $2p$ (shake) orbital and ΔE is the difference in $2p$ binding energy relative to the computed value in *GRASP*.

Model	Thomas (1984)	Roy *et al.* (2001)	Mukoyama *et al.* (2009)	Sier *et al.* (2024)	Thomas (1984) ($K\alpha$)
$P(\infty)$ (%)	0.60 ± 0.052	0.85 ± 0.086	0.66 ± 0.12	1.11	1.041
R (Å)	0.105 ± 0.0076	0.129 ± 0.016	0.131 ± 0.021	0.098	0.1074
ΔE (eV)	-104.9 ± 16.2	-53.5 ± 5.08	-19.9 ± 7.11	0	-62.13
χ_r^2	3.46	3.94	6.55	–	3.39

describe an $n = 2$ shake-off event independent of the subsequent decay channel. Hence, they serve as suitable comparative estimates for the $K\beta$ satellite. The quantitative performance of the three models can be compared with their results for the $K\alpha$ satellite.

ΔE is the difference between the fitted binding energy (E_B) of the $n = 2$ shake orbital and that computed using *GRASP*. *GRASP* yields estimates of 706.53 eV and 721.05 eV for the $2p_{3/2}$ and $2p_{1/2}$ orbitals, respectively.

Resolving the onsets of individual satellite components with statistical robustness is challenging due to low signal-to-noise ratios and near-degeneracy. Consequently, we analyse the evolution of the satellite structure as a whole based on the lowest binding energy component ($2p_{3/2}$). Several estimates for the K edge exist in the literature. Computationally, *GRASP* returns a value of 6547.1 eV and Huang *et al.* (1976) predict 6549.8 eV. The most up-to-date experimentally measured value (at 77 K) obtained by Kraft *et al.* (1996) and Chantler *et al.* (2025) is 6537.67 eV, which we take as the $1s$ binding energy in our calculations for the onset. Intriguingly, the theoretical predictions are $\sim$10 eV higher than accurate experimental measurements, which may be a result of neglecting pre-edge structure.

Sier *et al.* (2024) predicted an onset at 7244.2 eV for the $2p$ shake-off, while the Thomas, Roy and Mukoyama models return 7139.3 eV, 7190.7 eV and 7224.2 eV, respectively, using fitted values for the binding energy in $K\beta$, all noticeably lower than *GRASP*.

While experimentally defining a precise onset is difficult, Figs. 9 and 11 suggest a range of 7200–7280 eV. The onset for $K\alpha$ appears to be $\sim$50–100 eV higher. This may be a consequence of the lower signal-to-noise ratio in $K\beta$ contributing to intensities in the satellite region. Both the Roy and Mukoyama models are relatively consistent with experimental observations, but the Thomas model significantly underestimates the onset, falling below even a lower bound estimate derived from the unperturbed binding energy of the $2p_{3/2}$ orbital (638.7 eV; Thompson *et al.*, 2001). This discrepancy highlights the limitations of the Thomas model in predicting the form of satellite evolution near the onset.

GRASP returned a value of 0.098 Å for the expectation value of the $2p$ shake orbital effective radius (R). Among the models considered, the Thomas model (0.105 Å) exhibits the closest agreement, as was also observed for the $K\alpha$ satellite. The Roy (0.129 Å) and Mukoyama (0.131 Å) models return significantly larger values.

GRASP yielded a limiting shake-off probability $P(\infty)$ of 1.108%. Fig. 11 suggests a lower experimental value for the limit of $I_{sat}/I_{K\beta}$, with evolution slowing around 0.4% beyond $E_{inc} = 7800$ eV. The three models predict $P(\infty)$ within the range of 0.6%–0.8%, contrasting with the $K\alpha$ satellite where all models were within 5% of the *GRASP* estimate. The most significant contribution to this discrepancy is the narrow range of E_{inc} recorded here (7000–8000 eV) which extends only $\sim$700 eV beyond the onset, as opposed to thousands of electronvolts in data for $K\alpha$ (7000–10000 eV). The models are thus unreliable in predicting the evolutionary behaviour of the satellite well beyond the onset into the sudden limit, as shown by the relatively large statistical uncertainty (10%–20%) on $P(\infty)$ in each of the models. An extended-range investigation of the $K\beta$ satellite may reveal closer agreement with predictions for $P(\infty)$ at higher incident energies.
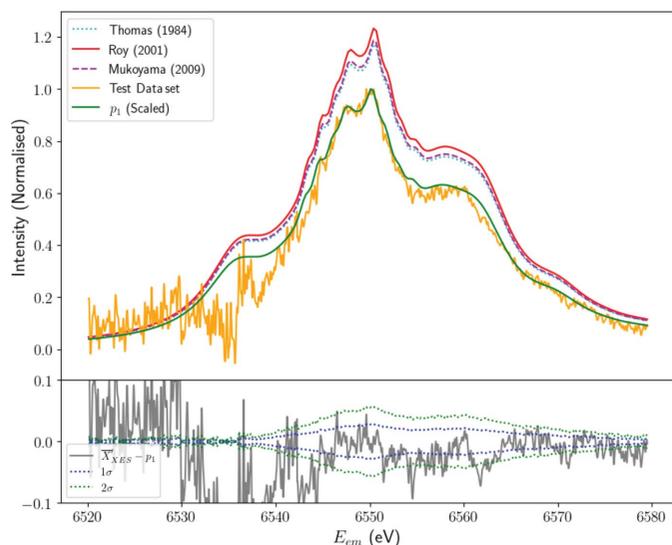
While $P(\infty)$ is defined in *GRASP* as the total $n = 2$ shake-off probability in the sudden limit, experimentally it is inferred from the limiting fluorescent intensity relative to the entire $K\beta$ spectrum ($I_{sat}/I_{K\beta}$). Additional physical mechanisms may be present that limit the utility of this comparison for $K\beta$. As Fig. 11 depicts, experimental satellite intensities are consistently a factor of 1.5–2 times lower in $K\beta$ than in $K\alpha$. As hypothesized by Melia *et al.* (2023) and Dean *et al.* (2024), this may result from differences in non-radiative Auger decay rates between single-body and shake-off decay channels in manganese $K\beta$, reducing the relative yield of fluorescent satellite transitions within the total $K\beta$ spectrum.

The χ_r^2 values in Table 1 indicate that the Thomas model provides an improved fit within the experimental energy range. This aligns with findings for the $K\alpha$ satellite. The parameters in Table 1 are consistent in the Thomas model for $K\alpha$ and $K\beta$, with the only significant deviations emerging in ΔE and $P(\infty)$. This aligns with trends for the onset and limiting intensity observed experimentally. Poor performance near the onset is evident for each of the models, exemplified by the large statistical uncertainty in ΔE, which totals 10%–35%, and significant deviation in ΔE. Statistical uncertainties are notably lower for R than ΔE for each of the models, perhaps indicating negative correlation between the two parameters, as each of the models possesses some degree of E_B–R dependence. The models may thus compensate for larger ΔE with smaller R and *vice versa*, indicating that the radius parameter is somewhat arbitrary, hence our observations of its relatively poor consistency here.

6.3. Application to independent test data

Predictions made by the three literature models are prone to fitting bias when trained solely on the XR-HERFD data set, which measures only a relatively small sample of E_{inc} past the satellite onset. To assess the feasibility of the predicted satellite intensity (I_{sat}) and the inferred satellite structure from PCA (p_1), we require independent data at higher E_{inc}, where the satellite intensities will 'saturate' and thus the structure will be fully evolved. An appropriate data set was recorded on I20-scanning, a high dwell time and low spacing HERFD-XES scan at E_{inc} = 9000 eV. To extract the satellite structure from the test data, we subtract the PCA-derived background [μ', illustrated in Fig. 15(b) in Appendix C] with appropriate normalization and alignment corrections.

Fig. 12 illustrates the test data and its comparison with the models. Intensities were extracted by extrapolating the intensity ratio ($I_{sat}/I_{K\beta}$) at E_{inc} = 9000 eV for each model from Fig. 11 to determine the predicted satellite intensity (I_{sat}) relative to the measured $K\beta$ intensity in the test data. All three models overestimate the measured satellite intensity by >20%. Notably, the satellite intensity in the test data is only marginally higher than it was in the high E_{inc} HERFD-XES slices from the training data, reinforcing our earlier observation that the $K\beta$ satellite intensity begins to asymptote around 7800 eV. While this may highlight intrinsic limitations of the

models, it would be premature to draw any definitive conclusions given the limited scope of the test data (a single XES scan). The lower-than-expected satellite intensities may stem from undetected systematic errors or noise which could contribute to errors of >10% in the measured intensity ratio.

The structural agreement between the PCA model and the test data is confirmed, demonstrated by the scaled PCA satellite profile (from Fig. 8). The test data exhibit the same dual-pronged main peak and high-energy shoulder, with fine structure in these regions aligning closely. Comparing the two structures above E_{em} = 6540 eV, where systematic errors are minimized, yields a χ_r^2 of 7.81, indicating strong agreement, especially given the low data uncertainty in the high-statistics test data set.
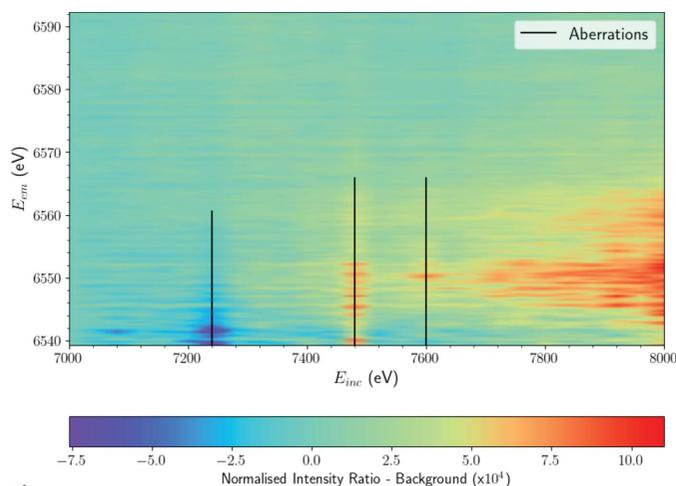
Below 6540 eV, significant deviations emerge. The low-energy shoulder appears sharper and more distinctly separated from the main satellite peak than predicted by the training data. This discrepancy may result from noise artefacts or misalignment causing a shift in the $K\beta_{2,5}$ peak location. Such behaviour is not unprecedented, as a similar sharp dip in intensity is visible in the PCA satellite component in Fig. 7(a). While an initial analysis of energy-eigenvalue calculations for the $K\beta$ satellite reported by Dean et al. (2024) supports the presence of the low-energy shoulder, the scope of our data leaves it unclear whether it originates from a real physical phenomenon or merely from fluctuations in noise that have contaminated the first PC.

7. Conclusion

This work presents XR-HERFD measurements of the manganese $K\beta$ spectrum across an extended range of E_{inc}, enabling the discovery of a new satellite in the high-energy tail of $K\beta_{2,5}$. This marks the first observation of such a satellite in $K\beta$ spectra. The high precision of these results allowed the satellite to be resolved with a standard error signature of 35 σ_{se} per pixel along the satellite peak and 652 σ_{se} across the full spectrum, far exceeding conventional discovery thresholds. This underscores the value of XR-HERFD in detecting new or known physical processes.

Application of principal component analysis (PCA) extracted detailed structural information on the satellite, depicting fine structure that suggests multiple physical origins, as is typical of shake processes. The significance of the satellite was further underscored by its contribution to over 95% of variance from the $K\beta$ background, with correlated noise accounting for less than 5%. Given the onset and structure of the satellite, we attribute its physical origin to $n = 2$ ($2s$ and $2p$) shake-off events following core photoionization, producing a spectrum non-degenerate to the characteristic $K\beta_{1,3}$ when a $3p$ electron fills the core hole.

Regression analysis following PCA allowed scaling parameters for the satellite (α_1) and $K\beta$ background (β_1) to be determined at each E_{inc} in the experimental energy range, and thus the relative intensity $I_{sat}/I_{K\beta}$ could be defined. This provided a highly accurate depiction of the satellite's evolution, the first such example for $K\beta$ spectra.

Figure 12
Predictions of satellite structure and intensity using the Thomas, Roy and Mukoyama models for the evolution of shake probabilities. The models are evaluated against the test data set, a background-subtracted HERFD-XES scan collected at an incident energy of 9000 eV, beyond the range of our experimental training data. Visual inspection reveals that all three models overestimate observed satellite intensities. This discrepancy may stem from intrinsic limitations of the models or, alternatively, from systematic errors or noise in the test data that reduce the measured satellite intensity. Comparison with the satellite component (p_1), scaled to match the intensity of the test data, indicates excellent agreement with the structure predicted using PCA, particularly in the fine structure surrounding the main dual peak at 6550 eV and high-energy shoulder at 6560 eV. This strong consistency underscores the efficacy of PCA as a tool for probing suppressed satellite structures.

Figure 13
Uncorrected background-subtracted intensities using 14 analyser crystals. Vertical intensity spikes are evident, corresponding to fluctuations in the HERFD-XES slices at incident energies E_{inc} = 7240, 7480 and 7600 eV. These are best attributed to electronic noise spikes in the Medipix detectors and significantly decrease the signal-to-noise ratio of the satellite. To correct for this, affected slices are replaced by the average of the surrounding data points. The improvements can be seen in Fig. 14.

Near-degeneracy and low intensities near the onset complicate the separation of individual shake components. Consequently, the satellite and its evolution were considered through a single component. Future work will focus on resolving the physical components of the satellite fine structure through a detailed statistical analysis using theoretical models, allowing for the extraction of the independent evolution of each satellite component.

Fitting literature models to the evolutionary data enabled quantification of key physical parameters. As was the case for the $K\alpha$ satellite, the Thomas model yielded the lowest χ_r^2 of
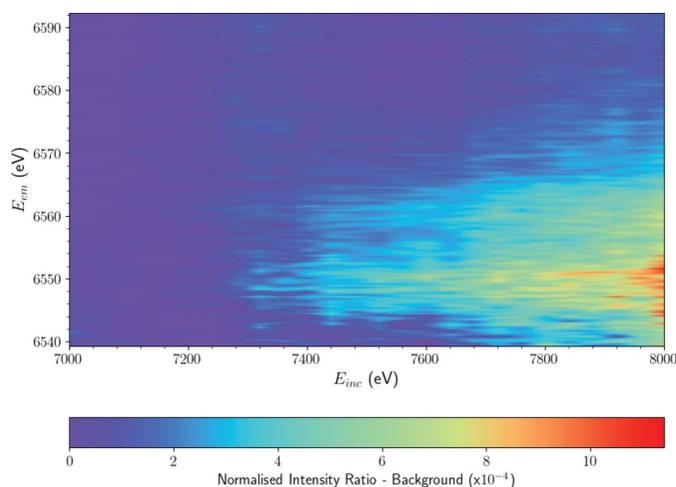
Figure 14
XR-HERFD map after subtraction of the $K\beta$ background using all 14 analyser crystals. The satellite structure and evolution are difficult to resolve. A closer analysis of the fluorescent intensities from each analyser crystal revealed that six of the analysers delivered a signal of poor statistical quality, significantly increasing the signal-to-noise ratio. These six crystals where excluded from the final processed data seen in Fig. 5, resulting in a much clearer image of the satellite's structure and evolution.

3.46. The presence of additional physical mechanisms such as Auger suppression in the satellite channel may affect the consistency of the process amplitude between $K\beta$ and $K\alpha$, warranting further investigation through *ab initio* calculations.

Further evidence for the evolving intensities in shake-off processes reinforces the need to refine the many-body reduction factor S_0^2 as a varying function of incident photon energy. While the n = 2 satellite has a limiting contribution to overall fluorescent intensities in manganese, shake-off processes with onsets closer to the absorption edge may contribute upwards of 20–30% (Dean *et al.*, 2024; Nguyen *et al.*, 2022), providing a significant energy dependency to S_0^2. A fully comprehensive representation of S_0^2 should incorporate the evolution of all many-body processes in a given material and continue to decrease past the absorption edge. Such a framework will provide deeper insights into many-body interactions and facilitate significant improvements in the extraction of structural parameters in condensed-matter materials from XAFS measurements.

APPENDIX A
Removal of intensity aberrations

As outlined in Section 4, the satellite intensities in $K\beta$ are several orders of magnitude lower than the spectrum and so even minor fluctuations in detector counts can significantly reduce the signal-to-noise ratio. Fig. 13 illustrates this issue, displaying several vertical intensity spikes that correspond to fluctuations in three individual HERFD-XES slices within the XR-HERFD array. These spikes are unlikely to originate from physical processes and are instead better attributed to false counts in the Medipix detectors caused by electronic noise.

To ensure the satellite is well resolved, these aberrations are removed from the final data and replaced by the average intensity of adjacent data points. For instance, in the affected HERFD-XES scan at E_{inc} = 7240 eV, the fluorescent intensity at each E_{em} is replaced by the average of the corresponding intensities at E_{inc} = 7200 and 7280 eV. While this correction alters the raw experimental data, its primary purpose is to enhance the statistical accuracy of the satellite measurement. The effectiveness of this approach is demonstrated by a comparison of Figs. 13 and 14.

APPENDIX B
Selection of analyser crystals

As the experimental apparatus in Fig. 1 depicts, the 14 analyser crystals are arranged such that their fluorescence signals are completely separable on the two detector apparatuses, allowing the output from each crystal to be analysed independently. As discussed in Section 4, this allowed us to identify that six of the 14 crystals produced fluorescence data with significantly lower statistical quality, to the point that their inclusion actively reduced the signal-to-noise ratio of the overall data set. To rectify this, fluorescent counts from only the eight crystals with the highest statistical quality were

included in the final data. A comparison of Fig. 14, which uses all 14 crystals, and Fig. 5, which uses only eight, clearly demonstrates that this processing step significantly improves the resolution of the satellite region's structure and evolution.

The cause of the statistical inconsistencies between the 14 analyser crystals is difficult to determine precisely, but is most likely due to a combination of systematic errors including misalignment, calibration issues or non-uniform analyser curvature. These inconsistencies stem largely from the pioneering experimental setup, as the 14-analyser apparatus had been implemented only a few months prior to our experimental run time on I20-Scanning. Recent upgrades have rectified these issues, ensuring all 14 analysers consistently deliver high-quality data, significantly enhancing the resolving power of future investigations.

Our ability to resolve the satellite despite these systematic issues highlights the advantages of independently projecting signals from each analyser crystal. If intensities had instead been homogenized onto a single region of the photon detector array, any faults in an individual analyser would have introduced irreversible distortions into the data set. Furthermore, failing to separate analyser signals would hinder the effective application of the splicing methodology, leaving experimental results vulnerable to spectral broadening and misalignment artefacts.

APPENDIX C
Principal component analysis in XR-HERFD

The application of PCA discussed in Section 5 uses a singular value decomposition algorithm (SVD) (rather than a least-squares or Bayesian analogue). SVD-based PCA applies to many fields in the current era of big data (Wang & Zhu, 2017), including geochemistry (Praus, 2005a), gene expression (Wall et al., 2003), facial recognition (Asiedu et al., 2016) and hydrology (Praus, 2005b). Using SVD, any real $n \times m$ data matrix Y of rank r [$r \leq \min(n, m)$] can be factorized as $Y = BSC^T$ (Wang & Zhu, 2017). S is a diagonal matrix whose entries, the 'singular values', correspond to the eigenvalues of YY^T and are ordered by decreasing magnitude (Kim & You, 2023).

In SVD-based PCA, the singular values correspond to the ordered eigenvalues of the covariance matrix Σ, obtained by performing SVD on the mean-centred data: $\overline{Y} = Y - 1_n \times \mu^T$, where μ is the mean of the data set. Since $\mathrm{Tr}(\Sigma)$ is representative of the total variance across Y, the magnitude of each eigenvalue is indicative of the proportion of the total variance represented by the corresponding PC.

Our XR-HERFD data can be interpreted as an $N_{E_{\mathrm{inc}}} \times N_{E_{\mathrm{em}}}$ matrix X of rank $N_{E_{\mathrm{inc}}}$, where $N_{E_{\mathrm{inc}}}$ and $N_{E_{\mathrm{em}}}$ are the numbers of incident energies and emission energies, respectively, in the XR-HERFD data set.

Unlike in standard applications of PCA, we wish to isolate the key components of variance from the stationary background profile, as opposed to the mean μ of the HERFD-XES slices in X, which incorporates not only the $K\beta$ background but also the satellite.

Application of the PCA algorithm using the uncorrected mean μ is depicted in Fig. 15(a). The amplitude of the satellite component increases from negative to positive, behaviour inconsistent with the expected characteristics of an evolutionary profile.

To improve this and measure the appropriate variance, we modified the PCA functionality to allow the subtraction of a pre-defined mean before extracting the PCs. This mean was determined by correcting μ using the minimum amplitude $a_{1,\mathrm{min}}$ of the satellite component in Fig. 15(a). This allows us to extract the refined background profile μ' by subtracting the PC satellite (p_1) weighted by $a_{1,\mathrm{min}}$,
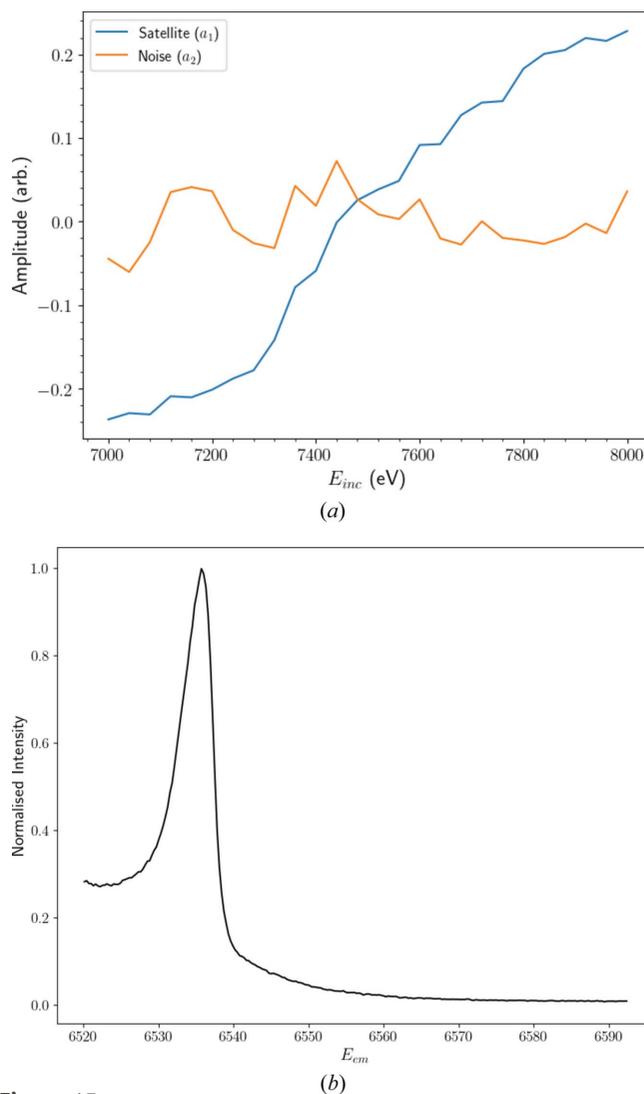
Figure 15
(a) Amplitude of the satellite and correlated noise PCs obtained without specifying a predefined mean in PCA. While the form of a_1 closely resembles that shown in Fig. 9, the amplitudes are shifted, evolving from negative to positive values. This is a poor representation of the satellite's evolution. (b) Stationary background profile (μ') of the PCA data set. In the results presented in Section 5, variance across X was measured from μ', as opposed to the true mean μ, to measure the evolution of satellite intensities accurately beyond the onset. μ' constitutes a significant improvement on the background profile depicted in Fig. 2(b), as there is a significant reduction in noise within the satellite region.

$$\mu' = \mu - |a_{1,\min}|p_1. \tag{6}$$

This ensures that the contribution of the satellite to the background is 0. The resulting form of μ' is depicted in Fig. 15(b) and corresponds to the form of the $K\beta_{1,3}$ and $K\beta_{2,5}$ background profiles bereft of any satellite contribution. This process ensures a significant reduction in noise in the high-E_{em} region of μ', where we expect to find the satellite, compared with the background we extracted by averaging across relevant HERFD-XES slices [Fig. 2(b)].

We then perform SVD on the refined background-subtracted data $\overline{X}' = X - 1_{N_{E_{\text{inc}}}}\,\mu'^{\text{T}}$ by

$$\overline{X}' = ASP^{\text{T}}. \tag{7}$$

The columns of P, the right singular values p_i, represent the loadings (which we refer to as 'weights' in Section 5) of each of the emission energies in the ith principal component. The entry $p_{i,k}$ is indicative of the degree to which intensities at the kth emission energy vary across the ith PC axis. The loadings of the first two PCs are discussed in detail in Section 5 (Fig. 7). After correction of the mean, the structure of p_1, the satellite loadings, remains largely unchanged yet accounts for 10% more variance across the XR-HERFD data set as a result of the improved accuracy of the reference point.

The columns of A (the left singular values a_i) each represent one of the PCs and are equivalent to a linear combination of the entries in each of the background-subtracted HERFD-XES slices in $\overline{X}'$ weighted by the PC loadings, $a_i = p_i \overline{X}'$. The entries of a_i ($a_{i,j}$) thus represent the 'score' of the jth HERFD-XES slice in the ith PC.

We refer to the scores as 'amplitudes' in Section 5. Upon correction of the background, the scores of the satellite component (a_1) correspond to the appropriate form for the satellite evolution profile (Fig. 9).

Conflict of interest

The authors declare no conflict of interest.

Data availability

The data supporting the findings of this study are available upon request.

References

Åberg, T. (1967). *Phys. Rev.* **156**, 35–41.

Ament, L. J., van Veenendaal, M., Devereaux, T. P., Hill, J. P. & van den Brink, J. (2011). *Rev. Mod. Phys.* **83**, 705–767.

Asiedu, L., Adebanji, A., Oduro, F. & Mettle, F. (2016). *Br. J. Math. Comput. Sci.* **12**(6), 1–23.

Bauer, M. (2014). *Phys. Chem. Chem. Phys.* **16**, 13827–13837.

Beckwith, M. A., Roemelt, M., Collomb, M.-N., DuBoc, C., Weng, T.-C., Bergmann, U., Glatzel, P., Neese, F. & DeBeer, S. (2011). *Inorg. Chem.* **50**, 8397–8409.

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B. & Varoquaux, G. (2013). *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pp. 108–122.

Carlson, T. A. & Krause, M. O. (1965). *Phys. Rev.* **140**, A1057–A1064.

Chantler, C. T., Bunker, G., D'Angelo, P. & Diaz-Moreno, S. (2024). *Nat. Rev. Methods Primers* **4**, 1–25.

Chantler, C. T., Smale, L. F. & Hudson, L. T. (2025). *International Tables for Crystallography*, Vol. C, 4th ed., edited by T. R. Welberry. International Union of Crystallography. In the press.

de Groot, F. & Kotani, A. (2008). *Core level spectroscopy of solids.* CRC Press.

Dean, J. W., Thompson, S. N. & Chantler, C. T. (2024). *Molecules* **29**, 4199.

Deutsch, M., Gang, O., Hämäläinen, K. & Kao, C. (1996). *Phys. Rev. Lett.* **76**, 2424–2427.

Diaz-Moreno, S. (2012). *J. Synchrotron Rad.* **19**, 863–868.

Diaz-Moreno, S., Hayama, S., Amboage, M., Freeman, A., Sutter, J. & Duller, G. (2009). *J. Phys. Conf. Ser.* **190**, 012038.

Eisenberger, P., Platzman, P. & Winick, H. (1976). *Phys. Rev. B* **13**, 2377–2380.

Gamblin, S. D. & Urch, D. S. (2001). *J. Electron Spectrosc. Relat. Phenom.* **113**, 179–192.

Gaur, A., Shrivastava, B. D. & Nigam, H. (2013). *Proc. Indian Natl. Sci. Acad.* **79**, 921–966.

Gel'mukhanov, F., Odelius, M., Polyutov, S. P., Föhlisch, A. & Kimberg, V. (2021). *Rev. Mod. Phys.* **93**, 035001.

Glatzel, P. & Bergmann, U. (2005). *Coord. Chem. Rev.* **249**, 65–95.

Glatzel, P., Bergmann, U., de Groot, F. & Cramer, S. (2003). *AIP Conf. Proc.* **652**, 250–255.

Glatzel, P., Bergmann, U., de Groot, F. M. & Cramer, S. P. (2001). *Phys. Rev. B* **64**, 045109.

Glatzel, P., Weng, T.-C., Kvashnina, K., Swarbrick, J., Sikora, M., Gallo, E., Smolentsev, N. & Mori, R. A. (2013). *J. Electron Spectrosc. Relat. Phenom.* **188**, 17–25.

Günter, T., Doronkin, D. E., Carvalho, H. W. P., Casapu, M. & Grunwaldt, J. (2016). *J. Phys. Conf. Ser.* **712**, 012071.

Guo, J., Skytt, P., Wassdahl, N., Nordgren, J., Luo, Y., Vahtras, O. & Ågren, H. (1995). *Chem. Phys. Lett.* **235**, 152–159.

Hämäläinen, K., Siddons, D., Hastings, J. & Berman, L. (1991). *Phys. Rev. Lett.* **67**, 2850–2853.

Hao, J. & Ho, T. K. (2019). *J. Educ. Behav. Stat.* **44**, 348–361.

Hayama, S., Boada, R., Chaboy, J., Birt, A., Duller, G., Cahill, L., Freeman, A., Amboage, M., Keenan, L. & Diaz-Moreno, S. (2021). *J. Phys. Condens. Matter* **33**, 284003.

Hölzer, G., Fritsch, M., Deutsch, M., Härtwig, J. & Förster, E. (1997). *Phys. Rev. A* **56**, 4554–4568.

Huang, K.-N., Aoyagi, M., Chen, M. H., Crasemann, B. & Mark, H. (1976). *At. Data Nucl. Data Tables* **18**, 243–291.

Jolliffe, I. (2002). *Principal component analysis.* New York: Springer-Verlag.

Jolliffe, I. T. & Cadima, J. (2016). *Phil. Trans. R. Soc. A* **374**, 20150202.

Kavčič, M. & Žitnik, M. (2024). *International Tables for Crystallography*, Vol. I, ch. 5.5, edited by C. T. Chantler, F. Boscherini & B. Bunker. International Union of Crystallography.

Kheifets, A. S. (2022). *Atoms* **10**, 89.

Kim, D. & You, K. (2023). arXiv:2307.15213.

Kochur, A. & Popov, V. (2006). *J. Phys. B At. Mol. Opt. Phys.* **39**, 3335–3344.

IUCrJ (2025). **12**, 548–562   Jack Stephens *et al.* · Energy-dependent many-body process in Mn $K\beta$ spectrum   **561**

electronic reprint

Kraft, S., Stümpel, J., Becker, P. & Kuetgens, U. (1996). *Rev. Sci. Instrum.* **67**, 681–687.

Lafuerza, S., Retegan, M., Detlefs, B., Chatterjee, R., Yachandra, V., Yano, J. & Glatzel, P. (2020). *Nanoscale* **12**, 16270–16284.

Lancaster, K. M., Roemelt, M., Ettenhuber, P., Hu, Y., Ribbe, M. W., Neese, F., Bergmann, U. & DeBeer, S. (2011). *Science* **334**, 974–977.

Lee, P. A. & Beni, G. (1977). *Phys. Rev. B* **15**, 2862–2883.

Lee, P. A. & Pendry, J. B. (1975). *Phys. Rev. B* **11**, 2795–2811.

Martin-Diaconescu, V., Chacón, K. N., Delgado-Jaime, M. U., Sokaras, D., Weng, T.-C., DeBeer, S. & Blackburn, N. J. (2016). *Inorg. Chem.* **55**, 3431–3439.

Melia, H., Dean, J., Nguyen, T. & Chantler, C. (2023). *Phys. Rev. A* **107**, 012809.

Millikan, R. (1917). *The electron: Its isolation and measurement and the determination of some of its properties.* University of Chicago Press.

Moretti Sala, M., Martel, K., Henriquet, C., Al Zein, A., Simonelli, L., Sahle, C., Gonzalez, H., Lagier, M.-C., Ponchut, C., Huotari, S., Verbeni, R., Krisch, M. & Monaco, G. (2018). *J. Synchrotron Rad.* **25**, 580–591.

Mukoyama, T., Uda, M., Kövér, L., Berényi, Z., Cserny, I. & Drube, W. (2009). *X-ray Spectrom.* **38**, 138–143.

Nguyen, T., Melia, H., Janssens, F. & Chantler, C. (2022). *Phys. Rev. A* **105**, 022811.

Pedersen, J., Decleva, P., Coriani, S. & Tenorio, B. N. C. (2023). *J. Chem. Phys.* **159**, 024121–1–12.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. (2011). *J. Mach. Learn. Res.* **12**, 2825–2830.

Pollock, C. J. & DeBeer, S. (2011). *J. Am. Chem. Soc.* **133**, 5594–5601.

Praus, P. (2005a). *Open Chem.* **3**, 731–741.

Praus, P. (2005b). *Water SA* **31**, 417–422.

Ramsay, J. & Silverman, B. W. (2006). *Functional data analysis.* Springer.

Rees, J., Wandzilak, A., Maganas, D., Wurster, N., Hugenbruch, S., Kowalska, J., Pollock, C., Lima, F., Finkelstein, K. & DeBeer, S. (2016). *J. Biol. Inorg. Chem.* **21**, 793–805.

Rehr, J. J. & Albers, R. C. (2000). *Rev. Mod. Phys.* **72**, 621–654.

Rehr, J. J., Stern, E. A., Martin, R. L. & Davidson, E. R. (1978). *Phys. Rev. B* **17**, 560–565.

Rijal, R., Stephens, J., Sier, D., Tran, N. T. T., Nguyen, T. V. B., Dean, J. W., Bowman, P., Dao, M., Di Pasquale, P., Kirk, T., Tran, C. Q., Hayama, S., Aramini, M., Ramanan, N., Diaz-Moreno, S. & Chantler, C. T. (2025). *J. Synchrotron Rad.* **32**, 994–1009.

Roy, M., Lindsay, J. D., Louch, S. & Gurman, S. J. (2001). *J. Synchrotron Rad.* **8**, 1103–1108.

Schmitt, T., de Groot, F. M. F. & Rubensson, J.-E. (2014). *J. Synchrotron Rad.* **21**, 1065–1076.

Sier, D., Dean, J. W., Tran, N. T. T., Kirk, T., Tran, C. Q., Mosselmans, J. F. W., Diaz-Moreno, S. & Chantler, C. T. (2024). *IUCrJ* **11**, 620–633.

Sier, D., Tran, N. T. T., Kirk, T., Tran, C. Q., Mosselmans, J. F. W., Diaz-Moreno, S. & Chantler, C. T. (2025). *J. Appl. Cryst.* **58**, 42–59.

Stern, E. (1978). *Contemp. Phys.* **19**, 289–310.

Stern, E. A. (1974). *Phys. Rev. B* **10**, 3027–3037.

Stöhr, J., Jaeger, R. & Rehr, J. J. (1983). *Phys. Rev. Lett.* **51**, 821–824.

Thomas, T. D. (1984). *Phys. Rev. Lett.* **52**, 417–420.

Thompson, A., Attwood, D., Gullikson, E., Howells, M. K., Kim, K., Kirz, J., Kortright, J., Lindau, I., Liu, Y., Pianetta, P., Robinson, A., Scofield, J., Underwood, J., Williams, G. & Winick, H. (2001). *X-ray data booklet.* Lawrence Berkeley National Laboratory, USA.

Tran, N. T. T., Sier, D., Kirk, T., Tran, C. Q., Mosselmans, J. F. W., Diaz-Moreno, S. & Chantler, C. T. (2023). *J. Synchrotron Rad.* **30**, 605–612.

Valenza, R. A., Jahrman, E. P., Kas, J. J. & Seidler, G. T. (2017). *Phys. Rev. A* **96**, 032504.

Vankó, G., Neisius, T., Molnár, G., Renz, F., Kárpáti, S., Shukla, A. & de Groot, F. M. (2006). *J. Phys. Chem. B* **110**, 11647–11653.

Wall, M. E., Rechtsteiner, A. & Rocha, L. M. (2003). *Singular value decomposition and principal component analysis.* In *A practical approach to microarray data analysis*, edited by D. P. Berrar, W. Dubitzky & M. Granzow, pp. 91–109. Springer.

Wang, Y. & Zhu, L. (2017). *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, pp. 471–475. IEEE.

Witten, I. H., Frank, E., Hall, M. A. & Pal, C. J. (2016). *Data mining: Practical machine learning tools.* Morgan Kaufmann Series in Data Management Systems. Amsterdam: Elsevier.