



## OPEN Discovering hidden satellites in Mn $K\alpha$ using novel high-accuracy fluorescence, with PCA revealing evolution of quantum processes

Ramesh Rijal<sup>1</sup>, Jack A. Stephens<sup>1</sup>, Nicholas T. T. Tran<sup>1</sup>, Daniel Sier<sup>1</sup>, Jonathan W. Dean<sup>1</sup>, Truong V. B. Nguyen<sup>1</sup>, Tony Kirk<sup>2</sup>, Chanh Q. Tran<sup>2</sup>, Sofia Diaz-Moreno<sup>3</sup> & Christopher T. Chantler<sup>1</sup>✉

We report the first experimental discovery of Hidden Satellites within the  $K\alpha_{1,2}$  emission lines of manganese metal (Mn,  $Z = 25$ ) with a total integrated statistical significance exceeding  $270 \sigma_{s.e.}$  (standard error), far beyond the discovery threshold. Experimental data were collected at the I20-Scanning beamline at the Diamond Light Source using our new eXtended-Range High-Energy-Resolution Fluorescence Detection (XR-HERFD) technique. The Hidden Satellites, embedded in the core emission structure, represent novel quantum many-body processes that evolve systematically as the incident photon energy increases. Principal Component Analysis (PCA) was applied to extract the major separable physical processes and validate the significance of the observed Hidden Satellites. The application of physical insight to the PCA method allowed us to isolate the satellites, and measure the evolutionary profile. Our paper reveals that the total intensity of shake-off satellites can reach as high as 20–25%. Although these are hidden, they are very significant. These results directly challenge the traditional treatment of the many-body reduction factor,  $S_0^2$ , as a constant in the standard XAFS equation. Our findings demonstrate that this term must be modelled as an energy-dependent function, reflecting its variation with incident photon energy and highlighting its role in many-body interactions. This deeper understanding of fundamental atomic processes directly impacts relativistic quantum mechanics, in theory and application. Also, this develops the two most popular experimental techniques at synchrotrons: X-ray absorption and X-ray emission spectroscopy, responsible for some 12,000 papers per annum, and all applications of these techniques in chemistry, physics, and biology. It offers insights into the evolution of satellites and underscores the broader implications of hidden features in X-ray spectra.

**Keywords** XR-HERFD,  $K\alpha$ , Hidden satellites, XES, Principal component analysis

X-ray Absorption Spectroscopy (XAS)<sup>1</sup> has offered ever-improving structural insights on elements and local chemical structure allowing high resolution and real-time measurements<sup>2–4</sup>. Advances in XAS, like resonant inelastic X-ray scattering (RIXS)<sup>5</sup> and high-energy-resolution fluorescence detection (HERFD), overcome limitations of conventional XAS by providing sharper and more precise spectra, enabling the identification of highly dilute compounds with greater accuracy<sup>6–8</sup>. These experimental techniques serve as valuable tools for probing complex systems and are mainstays at synchrotrons<sup>9</sup>. Whilst XAS is concerned with X-ray absorption, X-ray emission spectroscopy (XES) involves collecting emitted or fluorescence X-rays<sup>10</sup>. These two axes of information provide critical structural determination for some 12 000 publications per annum, and are a foundation of modern materials science<sup>11</sup>.

Recent synchrotron RIXS and HERFD experiments collect data along both the emission (XES) and absorption (XAS) axes<sup>12,13</sup>. In our investigation, we focus on the  $K\alpha$  transition, which originates from the  $2p \rightarrow 1s$  decay following the photoionization of a  $1s$  core electron. While two main peaks ( $K\alpha_1$  and  $K\alpha_2$ ) are expected due to the spin-orbit splitting of the  $2p$  level, the observed spectra often exhibit additional complexity in the form of

<sup>1</sup>School of Physics, University of Melbourne, Melbourne 3010, Australia. <sup>2</sup>Department of Chemistry and Physics, La Trobe University, Melbourne 3086, Australia. <sup>3</sup>Diamond Light Source, Harwell Science and Innovation Campus, Didcot OX11 0DE, UK. ✉email: chantler@unimelb.edu.au

asymmetric line shapes or sub-structure. Although several theories have been proposed over the past few decades to explain these features, a definitive understanding remains elusive<sup>14–19</sup>. The most widely accepted explanation attributes this additional substructure to shake processes<sup>20–22</sup>, in which the sudden ionization of a core electron results in the ejection of a secondary electron. This alters the potential experienced by the remaining electrons, causing the relaxation process to emit X-ray photons with slightly shifted, non-degenerate energies. These multi-electron effects give rise to spectral asymmetries and contribute to the complexity of the observed features<sup>23</sup>.

These additional features, commonly known as satellite lines, typically appear close to the strong parent (or diagram) line. They may be well-resolved peaks, such as  $K\alpha_{3,4}$  spectra, displaced by about 20 eV from the primary  $K\alpha_{1,2}$  peak for manganese<sup>22</sup>; or as unresolved satellites embedded very close to or hiding within the main spectral profile, which is the primary target of this paper. With our recent development of advanced techniques of XR-HERFD<sup>22,24–26</sup>, these subtle spectral signatures are now accessible, providing new insights into challenging regions of electronic structure and revealing novel physical processes and chemical phenomena.

A key area of investigation in the current work is the energy-dependence of physical processes in the XES and XAS spectra, leading to a corresponding energy dependence of many-body processes and the well-known many-body reduction factor ( $S_0^2$ ). Most work of the last few decades in XAS has assumed that this parameter is constant; that it is either unity ( $S_0^2 = 1$ ) or 0.9 ( $S_0^2 = 0.9$ ) or a free but constant value. Some authors have suggested that this constant is between 0.6 and 1.0<sup>27</sup>, and others have commented upon observing values of 0.4–1.2<sup>28,29</sup>, or 0.7–1.0<sup>30</sup>. It is generally acknowledged that the value should not be larger than unity within three standard errors of uncertainty.

XES measurements provide valuable insights into material properties. However, the interpretation of complex spectral features, such as diagram lines and (hidden) satellite peaks<sup>31,32</sup>, presents a significant challenge in accurately quantifying the spectra. This requires a careful systematic approach. To extract deeper structural insights, Principal Component Analysis (PCA)<sup>33</sup> was employed as a powerful tool for deconstructing and interpreting the spectral information. This paper aims to offer a deeper understanding of X-ray emission spectroscopy and a presentation of the success of new experimental techniques. It provides a particular focus on uncovering new, previously hidden or unobservable, phenomena and their implications for future synchrotron experiments, as well as the analysis and modelling of many-body interactions in condensed matter systems for the first time.

## Experimental setup

Measurements were carried out at the Diamond Light Source on the I20-Scanning beamline in the UK with the setup in Fig. 1a<sup>34</sup>. The purity of Mn metal foil was 98.7% and its thickness was 25  $\mu\text{m}$ . A Si(111) four-bounce monochromator was used, providing an energy range of 4.5–20 keV with an energy resolution of  $\Delta E/E \approx 1.3 \times 10^{-4}$ . Our experimental energy calibration is based on the well-characterized experimental values from Kraft et al.<sup>35</sup>. The beam size at the sample was approximately 400  $\mu\text{m}$  (horizontal) by 300  $\mu\text{m}$  (vertical), ensuring a well-defined and stable illumination of the sample. The incident photon flux at 10 keV exceeded  $10^{12}$  photons  $\text{s}^{-1}$ , enabling high signal-to-noise measurements. Typical data-acquisition times ranged from minutes to hours, depending on the scan modality.

The four-bounce crystal monochromator, designed in-house at Diamond, features two pairs of counter-rotating crystals, with advantages of high stability and energy resolution of transmitted photons, independent of the incident beam divergence<sup>36</sup>.

This experiment used a MAXIPIX (multichip area X-ray detector based on a photon-counting pixel array) TAA22PC detector and is a development of Medipix2 single-photon counting technology<sup>37</sup>. Compared with conventional detectors such as silicon drift detectors, scintillators, and ion chambers, the MAXIPIX detector (hereafter referred to as Medipix) offers multiple advantages, including superior spatial resolution, high readout speed, minimal electronic noise, no dark current, a wide dynamic range, and built-in energy discrimination. Figure 1b illustrates the three independent measurements for each crystal region, which allows identification of issues and inconsistencies in the data individually. Each fluorescence measurement is well-aligned and deliberately separated from each other to avoid overlap. Our analysis, outlined in Methods, included isolation of each crystal region, background subtraction, and consistency checks before merging data, which overall improves quality, reduces noise, and enhances sensitivity to underlying phenomena.

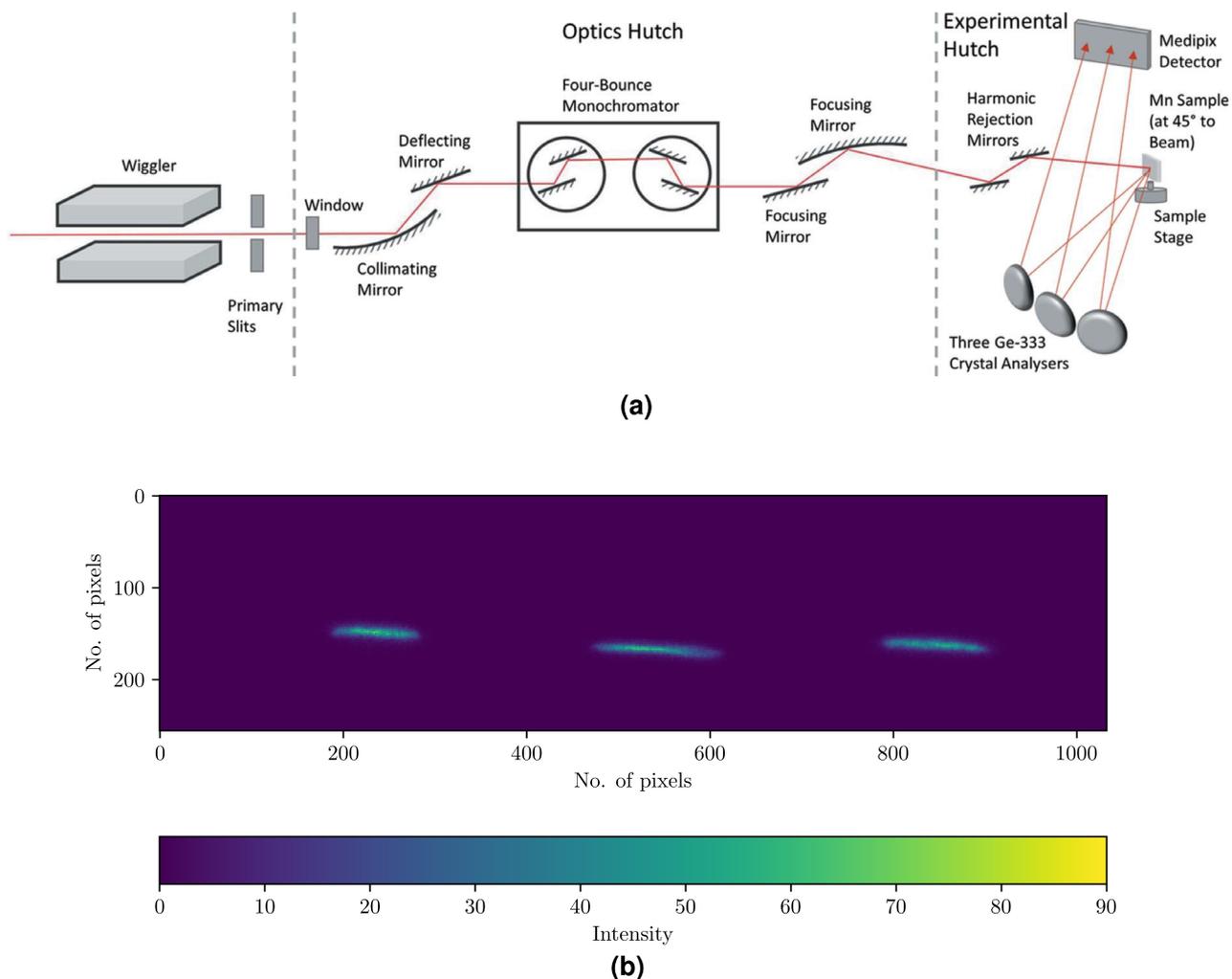
## Data processing

Part of the new data processing methodology has been presented by us<sup>22,26</sup>. Further details of the data processing procedure are outlined in Methods (Background) and Methods (Consistency).

## Results

### experimental observation: dominant manganese $K\alpha$ profile

Our datasets consist of multiple XES slices ranging from an incident energy ( $E_{in}$ ) of 6539 eV to 6754 eV with a step size of 5 eV. For each  $E_{in}$  value, the  $K\alpha_1$  and  $K\alpha_2$  peaks in the XES emission spectra are observed at nearly identical emission energy ( $E_{em}$ ) levels, typically differing by no more than 0.1 eV. Figure 2b indicates the good quality of crystals, the clear definition of the Mn  $K\alpha_1$  and  $K\alpha_2$  peaks, and the consistency across energy of the apparent signal. These contour scans are extremely useful, yet can obscure much of the relevant structure. Also, the general XES plot from all three crystal regions does not show any indication of additional feature, even when plotted on a natural logarithmic scale of intensity (Fig. 2a). In our previous XR-HERFD analysis particularly designed for emission energy range (5880 to 5935 eV) and incident energies (7200–8000 eV), a stack plot was effective in revealing the onset and evolution of satellite features located a few eV from the main diagram line<sup>22,38</sup>. Here, combined XES spectra collected across a range of incident energies ( $E_{in}$ ) from 6539 to 6754 eV (Fig. 3)



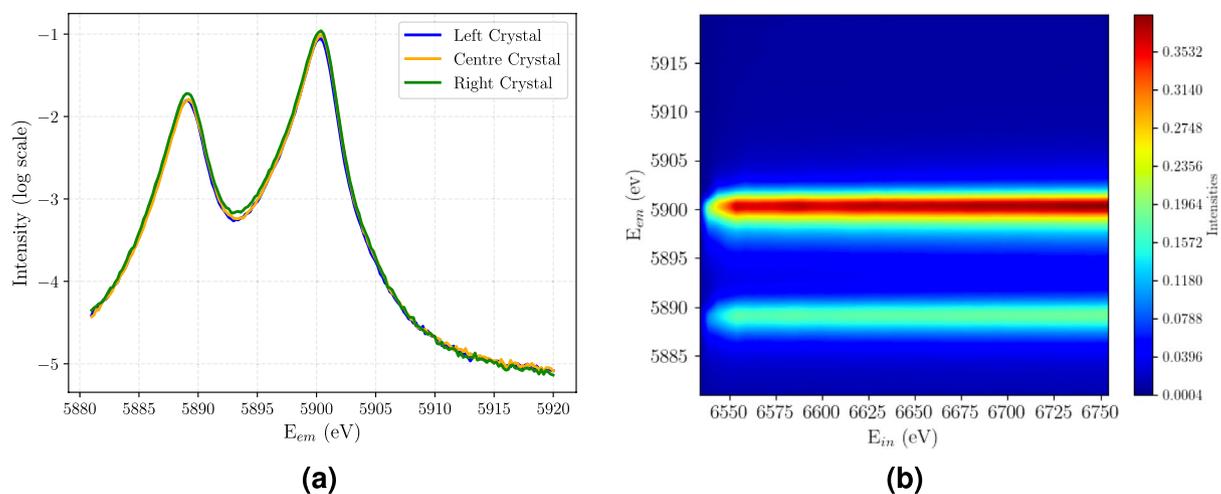
**Fig. 1.** (a) Experimental setup: Three Ge-333 crystal analysers lie adjacent to each other on the Rowland circle and reflect X-ray emissions to three separate regions on the Medipix detector. (b) Medipix detector depicting the reflection of analyser crystal images, represented as left, centre, and right crystal respectively. A  $4 \times 1$  detector setup was used, with total effective pixel grid of  $1024 \times 256$ .

exhibit excellent quality, and display the gradual emergence and intensity growth of two main peaks particularly evident at the first three lower incident energy points, but no visible evidence of other phenomena.

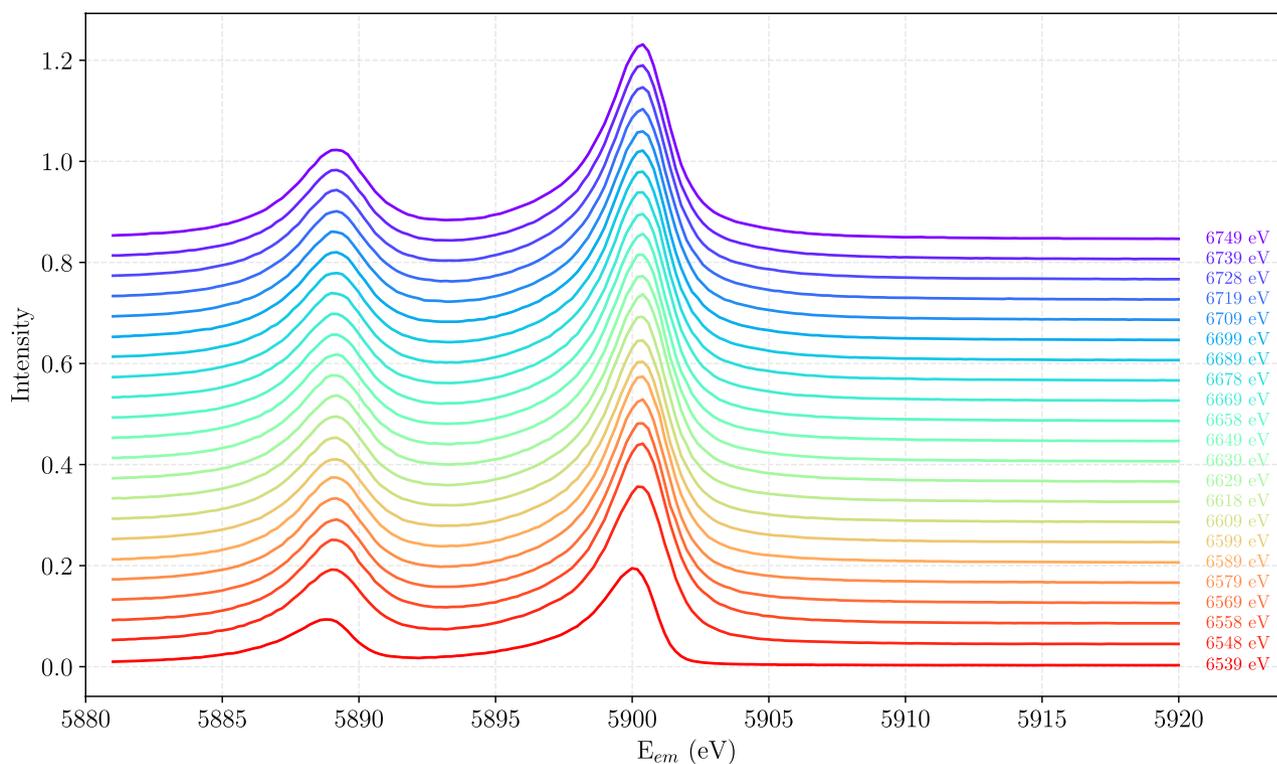
### Isolation and extraction of hidden satellites

Those plots and processing typically fail to reveal new physical processes such as satellites which are either hidden or unresolved within the main peaks. However, by differencing normalised spectra with references based upon the physics of the processes, we can reveal changes between spectra of similar appearance. Hence, subtracting the XES spectra at lower incident energies from the emission spectra at higher incident energies from the XR-HERFD datasets reveals satellites that were not apparent in the XES spectrum. We define these as 'Hidden Satellites'. Figure 4 subtracts the XES slice corresponding to a lower incident energy, just a few eV away from the binding energy (6536.67 eV) of the  $1s$  electron at the K-edge for manganese. To clarify and quantify these Hidden Satellites, we normalize the XES slices based on the maximum intensity of the first 10 channels at the lowest XES energies (Fig. 4).

Now, a spectrum of Hidden Satellites is clearly distinct, well-separated and broadly consistent, with a distribution across the main spectrum and a gradual reduction in the tail regions. Figure 4 shows structure, or a change of structure, involving what appears to be two clear peaks, with the suggestion of additional structure spread around the  $K\alpha$  profile. Whilst exciting, the magnitude is unclear. The statistical, point-wise and correlated noise appear quite large, and might question the significance and incisiveness of the purported result. The detailed structure observed is also dependent on the choice of the XES background profile chosen for subtraction.



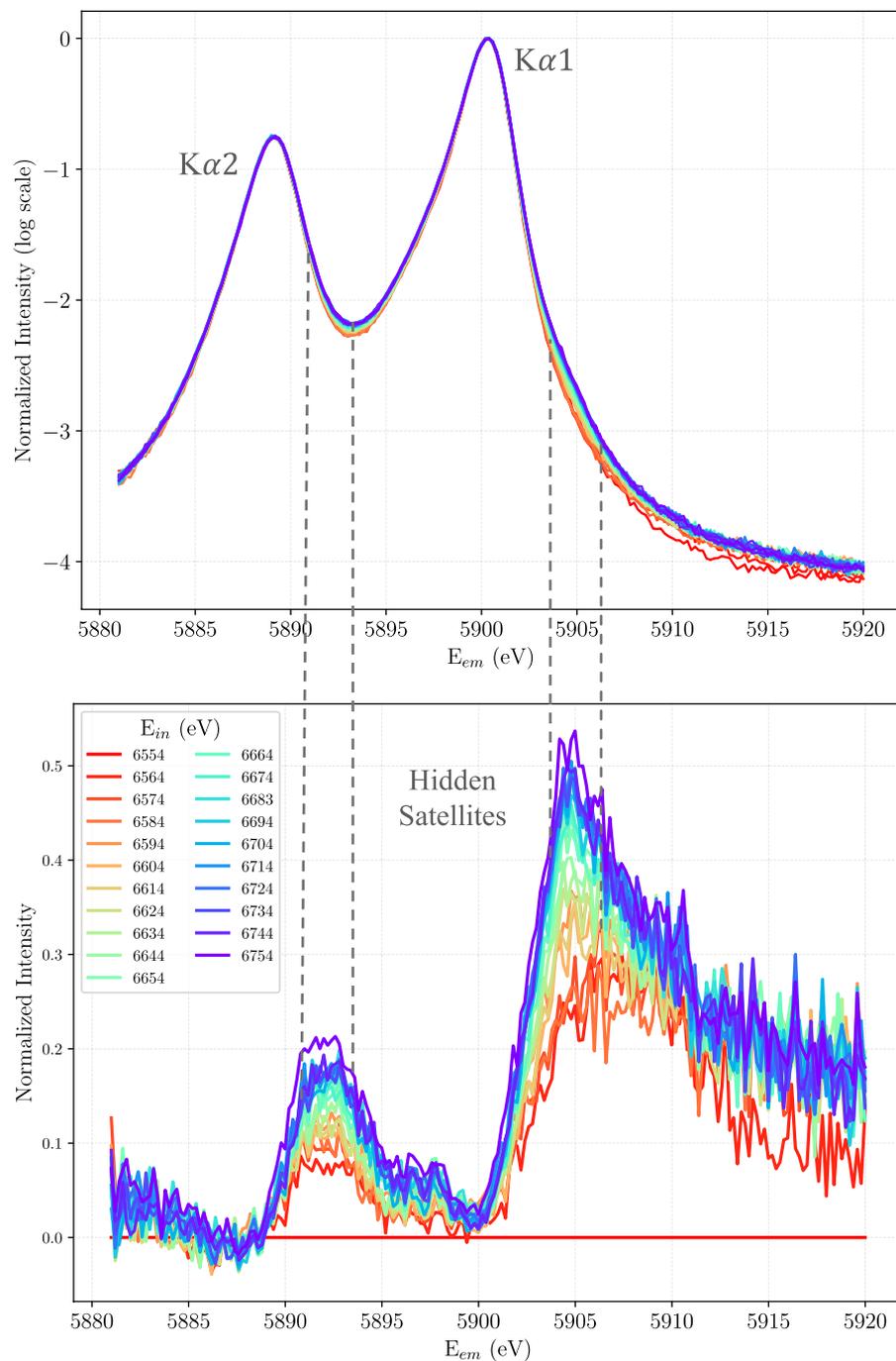
**Fig. 2.** (a) XES spectra of Mn showing the  $K\alpha_1$  and  $K\alpha_2$  peaks measured by three crystal analysers at an incident energy of 6654 eV, with the  $y$ -axis plotted as a normalized natural logarithm of intensity; (b) Contour map displaying the range of measured incident photon energies  $E_{in}$  and emission energies  $E_{em}$ . The peaks of  $K\alpha_1$  and  $K\alpha_2$  are clear and distinct, with no visible sign of additional sub-structure.



**Fig. 3.** Combined XES spectra with labels representing  $E_{in}$  for individual XES slices, indicating the overall growth of signal strength with energy. The spectra are offset vertically for clarity.

### Principal component analysis (PCA): initial estimates

PCA is a powerful statistical technique in modern data analysis, often employed to reduce the dimensionality of complex datasets while preserving the most significant orthogonal contributors to variance<sup>39–41</sup>. It has numerous applications across various fields, from drug discovery<sup>42</sup> to forensic science<sup>43</sup>, sports science<sup>44</sup>, and beyond<sup>45–47</sup>. Often PCA is used in a machine-learning or abstract mathematical sense with a package to extract dominant orthogonal components contributing to the variance between an experimental spectrum or image and a developing model. Sometimes, this deconstruction is largely void of physical meaning. Here, however, we intend to search for specific physical meaning of significant components and coefficients and apply this methods to the XES characteristic spectra.



**Fig. 4.** Normalized XES slices (top) presenting log scaled diagram spectrum from  $E_{in}$  of 6554 eV to 6754 eV and the spectra of the new Hidden Satellites (bottom) revealed by the differencing and normalisation of total XES slices from reference  $E_{in}$  of 6554 eV.

Although the quantum mechanical processes underlying both diagram and satellite lines have been the subject of extensive investigation, a comprehensive and quantitative understanding remains elusive<sup>48,49</sup>. The emergence of novel hidden satellite features introduces an even greater challenge, complicating efforts to achieve a full spectral profile. Despite these complexities, we demonstrate that the complete XES at various  $E_{in}$  can be accurately represented as a linear superposition of two distinct spectral components: the conventional diagram lines and the experimentally isolated hidden satellite structure. A simple model, Eq. (1),

$$X(E_{in}, E_{em}) = \alpha(E_{in}) \cdot \mu(E_{em}) + \beta(E_{in}) \cdot X_{HS}(E_{em}) \quad (1)$$

can be used to observe the energy dependent behaviour of our hidden satellites as observed in Fig. 4.

In this analysis, the complete XES datasets are stored on the left-hand side (LHS) of Eq. (1), while the right-hand side (RHS) components are fixed at a reference index of 6554 eV. This reference energy or index corresponds to a single incident energy chosen from the set of multiple  $E_{in}$  values. The spectrum  $\mu(E_{em})$  is the XES profile measured at this reference energy. This low-energy point is used as a baseline to subtract the total XES spectrum, which allows the hidden satellite features to be revealed and is represented as  $X_{HS}$ . The model is constructed as a weighted linear combination of these two components, parameterized by the energy-dependent coefficients  $\alpha(E_{in})$  and  $\beta(E_{in})$ . The initial step was to calculate the coefficient  $\alpha(E_{in})$ , which modulates the relative contributions of the  $K\alpha$  diagram lines and the hidden satellite features as a function of incident energy. Figure 5 demonstrates that this coefficient decreases with increasing incident energy, indicating a diminishing contribution from the  $K\alpha$  background and a growing prominence of Hidden Satellites. This trend aligns with the expectation that higher excitation energies enhance secondary processes responsible for satellite emission. The energy-dependent variation of  $\alpha(E_{in})$  enables a dynamic decomposition of the XES spectrum, reflecting the evolving balance between direct and satellite contributions. By fitting this model across a range of incident energies, we can quantitatively track the emergence of Hidden Satellite structures.

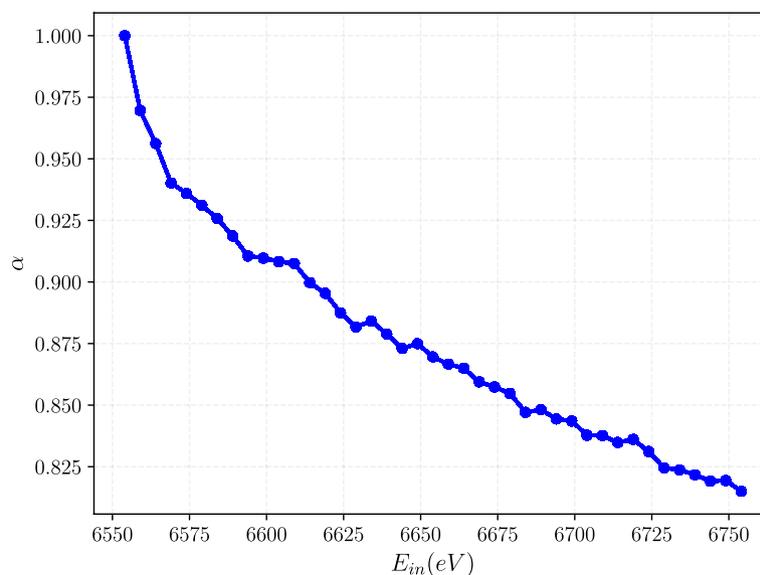
### Advanced structural insights from PCA

At this stage, we incorporated the scikit-learn (sklearn) library in Python<sup>33,50</sup>. Further details are given in Methods (PCA).

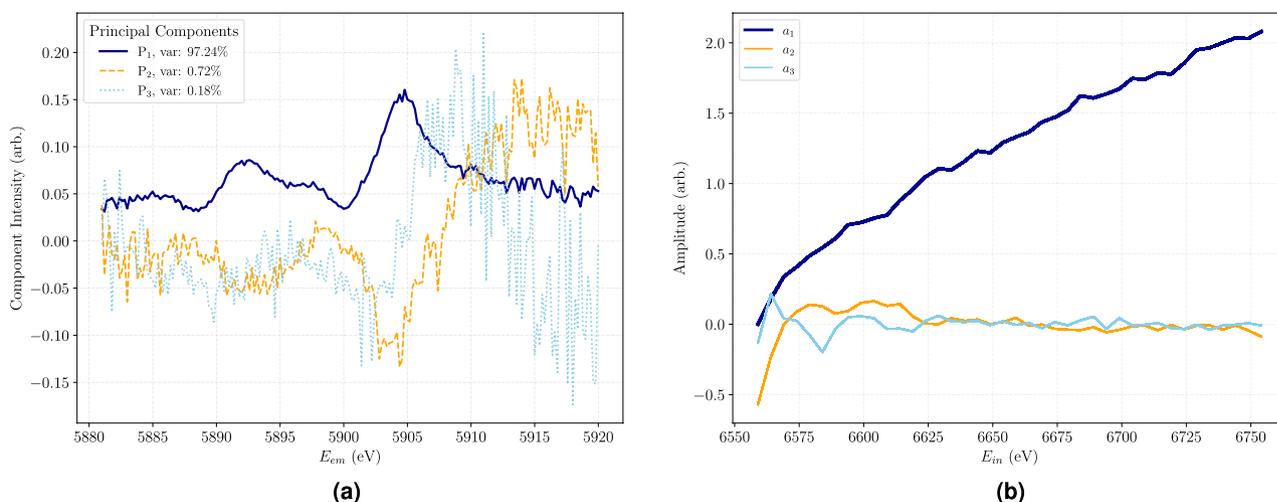
Figure 6a presents the spectral distribution of first three principal components across each emission energies  $E_{em}$ . The first component ( $P_1$ ) alone explains 97.24% of the total variance, indicating that the dominant structure in the data is effectively captured by this component. Notably,  $P_1$  closely mirrors the hidden satellite features previously identified in Fig. 4. With much of the correlated noise removed through PCA, the satellite structure becomes significantly clearer and more visually distinct. This enhanced representation reveals multiple substructures, perhaps six or more separated peaks, across different emission energies that were previously masked by noise, reinforcing the physical interpretation of  $P_1$  as the satellite signature. The second and third components,  $P_2$  and  $P_3$ , account for a very small fraction of the total variance. Their irregular patterns and lack of consistent features suggest they primarily capture noise or subtle effects that require deeper theoretical insight to interpret meaningfully.

In PCA,  $P_0$  refers to the pre-defined mean of the total XES data. On using this default mean from the package we incorporated, the amplitude of  $P_1$  shifted from negative to positive rather than following an ideal evolutionary profile. This indicates  $P_0$  was not aligned with the stationary background of the XES profile. To correct this behaviour, we modified the PCA functional to subtract a predefined mean before extracting the principal components. This is sometimes termed regression, and is equivalent to a small refinement of the background definition.

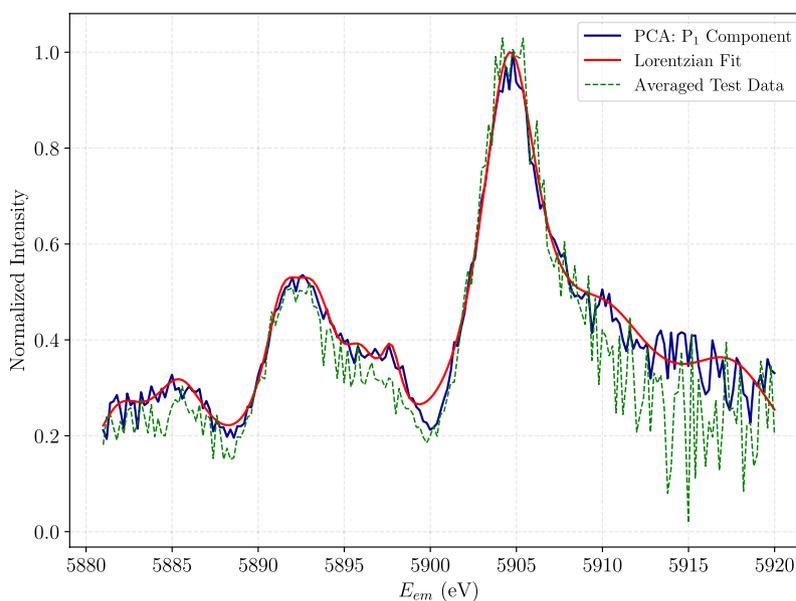
The refined mean was obtained by adjusting  $P_0$  using the minimum amplitude  $a_{1,min}$  of the satellite component. Specifically, the background XES spectrum  $\mu_{BG}$  was reconstructed by subtracting from  $P_0$  the contribution of the first principal component weighted by this minimum amplitude:  $\mu_{BG} = P_0 + a_{1,min} \cdot P_1$ . Since  $a_{1,min} < 0$ , this adjustment re-centres the mean, removing residual contributions from the satellite structure and isolating the stationary background signal and improves the reliability (and physical meaning) of the subsequent analysis.



**Fig. 5.** Energy-dependent behaviour of the coefficient  $\alpha(E_{in})$  for incident energies ranging from 6554 eV to 6754 eV. The observed decreasing trend of  $\alpha$  with increasing  $E_{in}$  reflects a relative enhancement in the intensity of hidden satellite features, indicating their growing contribution to the XES as the excitation energy increases. This indicates that the hidden satellite influence is significant and responsible for 20% or more of the overall spectra.



**Fig. 6.** (a) Spectral distributions ('loadings') of the first three Principal Components (PCs) as a function of  $E_{em}$ . The first component ( $P_1$ ) accounts for over 97% of the total variance and exhibits a shape that closely resembles the hidden satellite structure (Fig. 4) but with increased detail and much reduced statistical noise. In contrast, the second and third components ( $P_2$  and  $P_3$ ) primarily capture correlated noise. (b) Amplitudes (weights or 'scores') of the first three principal components,  $a_1$ ,  $a_2$ , and  $a_3$ , respectively. The amplitude  $a_1$  shows a consistent growth with increasing  $E_{in}$ , reflecting the systematic evolution of the hidden satellite spectral shape (Fig. 5). In contrast,  $a_2$  and  $a_3$  display irregular fluctuations, indicating correlated noise.



**Fig. 7.** The first principal component ( $P_1$ ) derived from PCA, fitted with a sum of Lorentzians, is compared with averaged hidden satellite structure derived from the top five maximum incident energies of the XES spectrum. Both plots exhibit similar growth patterns and peak structures, with multiple substructures emerging at nearly identical emission energies. This consistency clarifies the hidden satellite structure formation and demonstrates the measurement accuracy and noise.

Figure 6b displays the corresponding amplitudes of the first three principal components as a function of incident energy  $E_{in}$  after the modification. The amplitude of  $a_1$  shows a clear and systematic increase with  $E_{in}$ , consistent with the gradual evolution of the satellite structure - an observation that aligns well with prior observation (Fig. 5). In contrast, the amplitudes  $a_2$  and  $a_3$  exhibit no discernible trends, further supporting the notion that these components are dominated by correlated noise.

To quantitatively characterize the hidden satellite structure and its associated spectral features, we fit a sum of Lorentzian functions to the first principal component ( $P_1$ ) (Fig. 7). We used the `lmfit` package to perform nonlinear least-squares fitting of multiple Lorentzian peaks to the PCA-derived spectral data. The details of their

peak energies and widths are detailed in Methods. This empirical fitting allows for extraction of peak positions, widths, and intensities, and a quantitative representation of hidden satellite contributions, without attributing physical significance of its cause and origin. The resulting fit closely matches the satellite profile obtained earlier by averaging the XES spectra at the five highest incident energies. Those five high energy points were chosen to improve noise suppression and clarity, as subtraction using low-energy points tends to increase noise and suppress the hidden satellite structure. The satellite structure derived through PCA exhibits significantly reduced noise compared with the raw spectra, with improved feature distinction that align well. This highlights the strength of PCA: it enhances signal clarity by suppressing noise, and can isolate key evolving spectral features, enabling a more accurate and interpretable physical model of the system.

### Evolution of hidden satellites

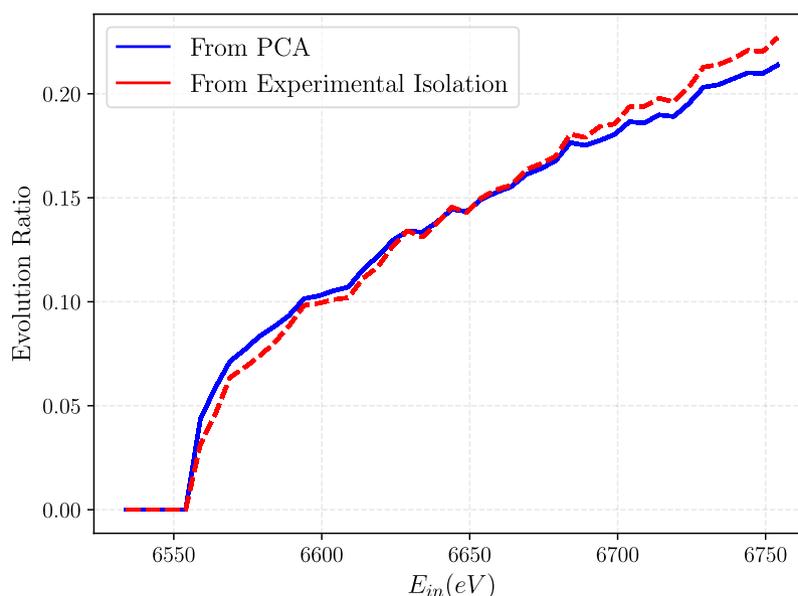
PCA enhances the hidden satellite structure by isolating it from correlated noise. It provides an initial estimate of its evolution (Fig. 6b). Our initial simple model function, Eq. (1), is now slightly modified to incorporate ( $P_1$ ), representing the evolving hidden satellite structure:

$$X(E_{in}, E_{em}) = \gamma_1(E_{in}) \cdot P_1(E_{em}) + \gamma_{11}(E_{in}) \cdot \mu_{BG}(E_{em}) \quad (2)$$

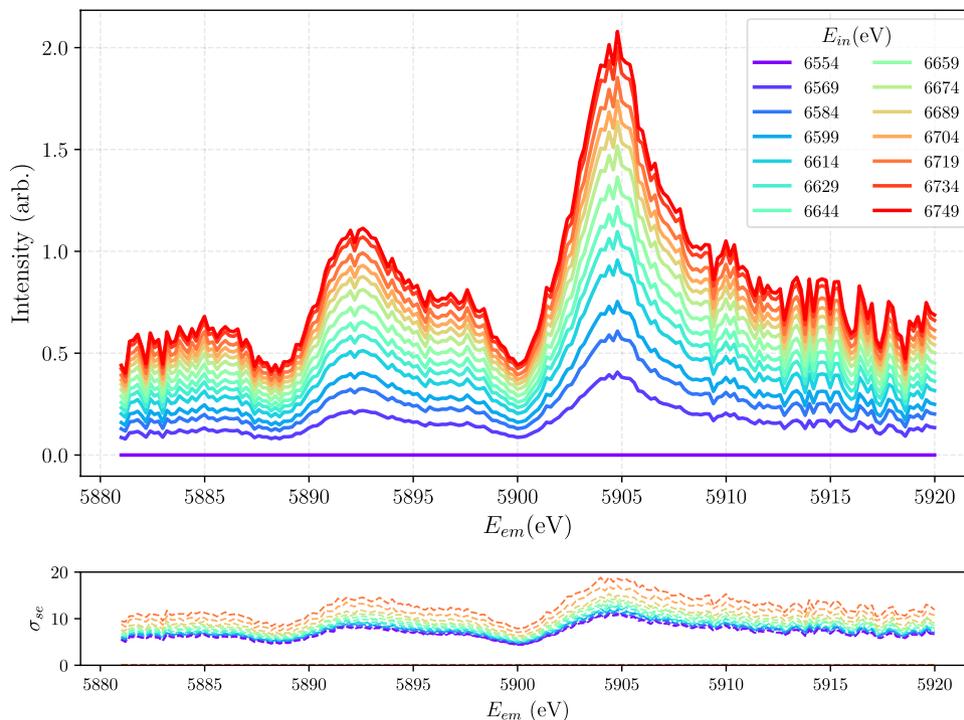
This functional model is expressed as a weighted linear combination of the first principal component  $P_1$  and the (regressed) background profile  $\mu_{BG}$ , with coefficients  $\gamma_1(E_{in})$  and  $\gamma_{11}(E_{in})$ , respectively.

We examine the evolution ratio of Hidden Satellites by defining the ratio of the intensity of the satellite component to the  $K\alpha$  spectrum,  $\frac{I_{sat}}{I_{K\alpha}}$ , as a function of  $E_{in}$ . We simply take the ratio of the satellite component area to the total  $K\alpha$  spectrum area:  $\frac{\gamma_1(E_{in}) \cdot A_{P_1}}{\gamma_{11}(E_{in}) \cdot A_{K\alpha}}$ . Here,  $A_{P_1}$  is the area of the first principal component or satellite component and  $A_{K\alpha}$  is the area of the PCA-extracted background  $\mu_{BG}$  representing the full  $K\alpha$  spectrum. Figure 8 presents the evolution ratio of the hidden satellite for both the first principal component (in blue) and the previously experimentally isolated hidden satellite structure (in red). Both are in close agreement, beginning at an onset near 6554 eV and growing monotonically as the incident energy increases. The total intensity of shake-off satellites reaches 20 - 25%, with an onset occurring above the absorption edge.

Figure 9 demonstrates the growth of the hidden satellite structures as a function of increasing  $E_{in}$  with reference to the structure generated by PCA model. The emergence of these Hidden Satellites, along with their distinct onset and evolution, is evident at multiple emission energies. In contrast with Fig. 4, Fig. 9 now shows a detailed structured spectrum with enhanced noise suppression. The evolving nature of the satellite features is clearer and better defined. In our dataset, the uncertainty arises from repeated measurements and background fitting, both dominated by statistical noise. Because of the multiple energy scans and the use of separate analyser crystals, the standard error (s.e.,  $\sigma$ ) associated with each data point can be easily determined<sup>24,51</sup>. A single point in the incident–emission energy space exhibits a statistical significance between  $5\sigma$  and  $20\sigma$  above background. However, because the measurement is a two-dimensional map of the new process with over 45 separate data points, each with 5 to  $20\sigma$  statistical significance, a total of  $271.83\sigma$  is obtained when summing the signal



**Fig. 8.** The evolution ratio of the  $K\alpha$  Hidden Satellites, extracted from the first principal component of PCA (shown in blue) and from experimental isolation and analysis (shown in red). Both components exhibit a similar trend, beginning at an onset around 6554 eV and increasing monotonically with incident energy. Both methods reveal the same underlying trend, showing the onset and gradual evolution of the satellite structure as the incident energy increase.



**Fig. 9.** Evolution of the first principal component, the hidden satellite structures  $P_1$ , with increasing  $E_{in}$ . The spectral peaks are clear, following the PCA and emphasising the distinct structures at relevant emission energies, the estimated point noise level with respect to any feature, and the growth of the peaks with increasing  $E_{in}$ . The integrated significance of the Hidden Satellite signature reaches 272 standard errors (s.e. or  $\sigma_{se}$ ) across the 45 data points of the measured spectra. This level of significance far exceeds the conventional discovery threshold of 3–6  $\sigma_{se}$  per data point for typical scientific discoveries.

over all 45 points. This cumulative level far exceeds the conventional discovery threshold of 3–6  $\sigma_{se}$ , further supporting the existence of the Hidden Satellites.

### Comparison with models from relativistic quantum mechanics for quantum evolution

To interpret the evolution of the hidden satellite observed in our experimental data, we compare it with three models describing the evolution of quantum mechanical processes. A commonly cited model, the Thomas model<sup>52</sup> is based on a quantum mechanical description of a bound-bound transition:

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

Here,  $P(\infty)$  is the limiting shake probability in high-energy limit from a sudden approximation,  $R$  is the shell radius corresponding for example to a shake orbital in  $\text{\AA}$ ,  $E_B$  is the computed shake electron binding energy,  $E_P = E_{inc} - E_{ons}$  is the energy of the shake electron after ejection.  $E_{ons} = E_{edge} + E_B - \Delta E$  is the onset energy of the shake electron, and  $E_{edge}$  the threshold energy of the photoelectron.  $\Delta E$  is the difference between the fitted and predicted ( $E_{edge} + E_B$ ) onset. The constant (15.32) arises from  $\frac{m_e e^4}{2\hbar^2}$  in units of eV and Angstrom.

Roy et al. (2001) introduced a more generic approach to describe the evolution of transition probabilities using atomic orbital wavefunctions in the sudden approximation for bound-free transitions<sup>53</sup>, expressed as:

$$P(E_P) = P(\infty) \cdot \left[\frac{2^{2n}(n-1)!(n+1)!}{\pi(2n-1)!}\right] E_B^{n+1/2} \cdot \int_0^{E_P - E_B} \frac{E^{1/2}}{(E_B + E)^{n+2}[1 + \tau_0^2(E_B + E)^2]} dE \quad (4)$$

Here  $n$  is the principal quantum number of the shake shell,  $\tau_0 = \frac{R}{(2E_P)^{1/2}}$  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 ( $m_e, \hbar, e, 4\pi\epsilon_0 = 1$ ). We convert these to eV (1 hartree =  $\frac{m_e e^4}{\hbar^4(4\pi\epsilon_0)^2} = 27.21$  eV) and  $\text{\AA}$  (1 bohr = 0.5219  $\text{\AA}$ ) for ease of comparison.

A third model by Mukoyama et al.<sup>54</sup> modifies the Roy and Thomas models to account for bound-bound transitions, where an electron is excited into an unoccupied bound state. This model differs from the Thomas model in the treatment of the time-dependence of the electron within the Hamiltonian:

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

We use the same K-edge reference (1s binding energy) for manganese metal as that employed during our experimental energy calibration, 6537.67 eV<sup>35,55</sup>. For comparison, we also considered a theoretical value obtained using the multi-configuration Dirac-Hartree-Fock (MCDHF) method with a relativistic QED approach, implemented in the General Relativistic Atomic Structure Package (GRASP)<sup>56–58</sup>, for which we have obtained a value of 6536.11 eV. The theoretical value is defined for  $T = 0K$  and may not be fully converged; and yet the experimental value obtained by Kraft is based on the extremum of the spectral derivative, rather than the ionisation energy. Hence it is a remarkable achievement that these agree within about an eV.

The Hidden Satellites observed in the emission spectra of our  $K\alpha_{1,2}$  transition potentially arise from an  $n = 3$  shake-off process, typically involving the  $3s$ ,  $3p$ , and  $3d$  states. Hence we use  $n = 3$  in the second model. Figure 8 clearly suggest the onset of 6554 eV. The  $3d$  orbital is predicted to lie approximately 15.41 eV above the K-edge ( $6537.67 + 15.64 = 6553.31$  eV) according to our new theoretical computation. This consistency within 1 eV forms the basis of our consideration in the current analysis of the models.

We then fitted three theoretical models to our experimental evolution data, treating  $P(\infty)$ ,  $R$ , and an energy offset term ( $\Delta E$ ) as free parameters.  $\Delta E$  is the difference in the fitted onset of the satellite to that we obtained using values for  $E_B$  from GRASP and  $E_{edge}$  from experiment<sup>35</sup>.

Defining a precise onset from our data set is difficult, yet the Thomas, Roy and Mukoyama models return the plausible and modest offsets ( $\Delta E$ ) of  $-7.8$ ,  $-1.34$ , and  $-2.0$  eV respectively (Table 1). The Roy and Mukoyama models predict a similar onset within 2 eV. All three models show a consistent agreement with experiment (Fig. 10).

Our fitted models yield  $R$  values of 1.89 Å (Thomas), 2.09 Å (Roy), and 2.40 Å (Mukoyama) (Table 1). Although there is some agreement in the fitted  $R$  between models, there is significant uncertainty and further theoretical investigation is suggested here especially in relation to quantum expectation values.

The goodness-of-fit parameter,  $\chi_r^2$ , indicates that the Roy model performs best across the entire parameter space.

Both other quantum evolution models also achieve agreement with experiment. Taken together, all three models exhibit consistent behaviour.

Recent investigations into  $3d$  transition metals have highlighted the significant role of unresolved shake-off satellites in shaping the  $K\alpha_{1,2}$  emission profiles. For scandium, the  $P(\infty)$  or the intensity of these unresolved shake-off satellites is stated to account for up to 34.7% of the total  $K\alpha_{1,2}$  spectrum<sup>59</sup>, while for copper, this contribution is stated to be of order 25.5%<sup>20</sup>. For manganese  $K\alpha_{1,2}$ , theoretical calculations of these probabilities have reported values of 32.03%<sup>60</sup> and 25.86%<sup>61</sup>. Our experimental evolution ratio reaches 25% and more of the total intensity, significant and yet plausibly within 2–3% with available prior computations.

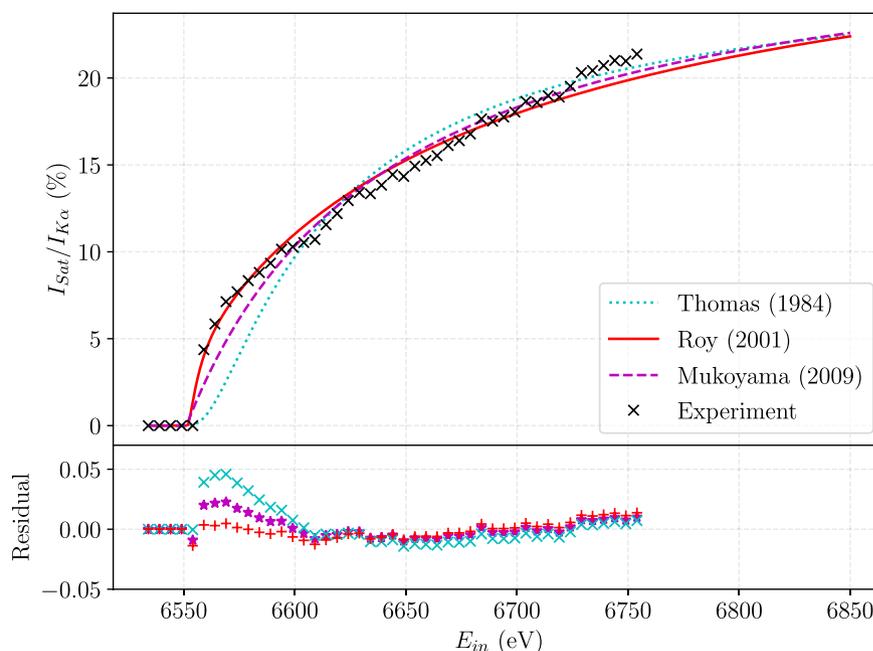
Each many-body process has an onset energy determined by its binding energy relative to the K-edge. As the energy increases, the intensity of each shake-off satellite evolves from zero at the onset energy, gradually reaching its maximum value, as demonstrated by our new processes (Fig. 10). This evolution introduces a clear energy dependence for the many-body reduction factor, which can reach as high as 32.03% for manganese  $K\alpha_{1,2}$ , according to Kochur et al.<sup>60</sup> and Dean et al.<sup>62</sup>. Our results demonstrate that many-body processes account for 25% or more of the total intensity. This development of the many-body processes calls for a new approach and insight on analysing X-ray fine structure and in general in X-ray spectroscopy. Given the pronounced energy dependence of this contribution, the conventional assumption of a constant  $S_0^2$  term in the XAFS equation becomes increasingly concerning<sup>24,25</sup>. Such energy dependence calls for a fundamental revision of traditional models, which typically treat the reduction factor as constant, underscoring the importance of incorporating the incident photon energy to more accurately represent the complexity of many-body interactions.

## Conclusion

Synchrotron data using the XR-HERFD technique has revealed the spectral distribution profile of novel Hidden Satellites inside the  $K\alpha$  spectrum of manganese metal. Through multiple XES scans, we confirmed the presence of these satellites and were able to isolate their structure with excellent statistics and resolution. The clarity of the dataset was enhanced by isolating each crystal region, combining them, checking their consistency, and removing the background noise signal. The existence of the Hidden Satellites is very clear, defining a precise spectrum and components with small noise and uncertainty. Principal Component Analysis (PCA) was then

Model	Thomas <sup>52</sup>	Roy et al. <sup>53</sup>	Mukoyama et al. <sup>54</sup>
$P(\infty)$ (%)	$27.0 \pm 1.65$	$32.1 \pm 1.16$	$29.41 \pm 1.48$
$R$ (Å)	$1.89 \pm 0.18$	$2.09 \pm 0.56$	$2.40 \pm 0.16$
$\Delta E$ (eV)	$-7.8 \pm 6.41$	$-1.34 \pm 0.06$	$-2.0 \pm 2.65$
$\chi_r^2$	11.27	2.18	3.30

**Table 1.** Fitting parameters with uncertainties obtained using the Roy, Thomas, and Mukoyama models for the evolution of the  $K\alpha$  Hidden Satellites.  $P(\infty)$  denotes the fitted limiting shake probability relative to the entire  $K\alpha$  spectrum in the sudden limit.



**Fig. 10.** Comparisons of theoretical models with XR-HERFD experimental data, plotted against  $E_{in}$ . All models predict the growth of the new spectral features.

able to unfold the satellite profile, accounting for more than 97% of the residual variance in the  $K\alpha$  regime while minimising any interference from noise. The evolution of the overall Hidden Satellite structure is well-defined, with statistical significance far exceeding the conventional discovery threshold of 3–6 standard errors (s.e.), reaching an integrated significance of up to  $271.83 \sigma_{se}$  across multiple peaks of the measured spectrum for the extended data set. This discovery allowed for a definitive comparison with conventional theoretical models for quantum mechanical process evolution. Our paper confirms that the total probability of shake-off processes during evolution can reach 20–25%. Results emphasise the need to redefine the many-body reduction factor, particularly in terms of its variation with incident photon energy, to effectively describe the dynamics of shake-off processes. This work highlights the strengths of the XR-HERFD technique, of effective signal separation and analysis, and of the use of PCA when its components have clear physical meaning. It demonstrates the potential of advanced relativistic quantum theory to enhance our understanding of XAFS, XES, HERFD, and many-body effects in applications to fundamental research and applied physics and chemistry. Continued theoretical and experimental efforts will further improve our knowledge of atomic and condensed matter systems and the role of relativistic quantum mechanics in these advanced X-ray spectroscopies.

## Methods

### Isolation of crystal region

Figure 1b notes that there are three locations observing fluorescent counts ( $N_F$ ) from each Bragg analysers. The total fluorescence counts for the crystal are divided by the upstream count ( $I_0$ ), which is a single point count. The ratio of fluorescent photons ( $N_{F_i}$ ) entering the detector from individual crystal analysers to incident photons ( $N_{I_0}$ ) entering the upstream ion chamber determines the measured intensity. Here, we isolate each analyser by separating the images by aligning the analyser crystals, and by choosing an appropriate Region of Interest (RoI) around the detector image, allowing for the signal of each analyser to be well-separated. The counts of each crystal region is presented as  $N_{F_i}$  where  $F_i$  represents  $i^{th}$  crystal region.

If we integrate the total pixel counts, the standard deviation of total counts is represented as:

$$\sigma_{F_i} = \sqrt{N_{F_i}} = \sqrt{N_1 + N_2 + \dots + N_n} \quad (6)$$

ie

$$\sigma_{F_i} = \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2} \quad (7)$$

where,  $n$  is the total number of pixels. Figure 2a clearly displays the plot showing three separate crystal regions where the intensity slightly differs from each other.

### Background subtraction from each crystal analyser

A region of interest (RoI) was defined around the image of each analyser crystal to minimise false counts and background arising from electronic noise or scattered photons that were not Bragg-reflected by the analyser

crystals. To estimate the inherent background, a region away from any ROI was selected for each incident energy, and the corresponding background counts per pixel were subtracted from the recorded intensities. The resulting intensities were then normalized by the upstream ion counts to correct for potential variations in signal intensity and to assess crystal defects and misalignment.

The background signal ( $N_{bg}$ ) for three different crystals is chosen using an area of the detector well above from any of the ROI's and the area formed of the region is  $A_{(bg)_i}$  where  $i$  represents different regions of background signal. To ensure that background signal formed is uniform for all regions, we chose three different sources of background in rectangular shape from the topmost region of each crystal region.

The uncertainty involved in background is written as:  $\sigma_{se(bg)} = \sqrt{N_{bg}}$ . Then, the total background signal for each crystal region can be summarized as:

$$N_{F_i(bg)} = \frac{N_{bg} * A_{F_i}}{A_{(bg)_i}} \quad (8)$$

The total counts after background subtraction can now be expressed as:

$$N'_{F_i} = N_{F_i} - N_{F_i(bg)} \quad (9)$$

with error as

$$\sigma_{F'_i}^2 = \sigma_{F_i}^2 + \sigma_{F_i(bg)}^2 \quad (10)$$

### Consistency test and combination of crystals

After background subtraction, we perform a consistency check. If the spectra from each crystals are consistent, we sum the crystals otherwise weighted average is considered. We begin by scaling all the crystals analysers to the strongest one. The scaling factor ( $\alpha_i$ ) for the strongest crystal is assumed to be 1. The scaling factors are generated using the formula

$$\alpha_i = \frac{\max(N_{F_i})_{i=1}^3}{N_{F_i}} \quad (11)$$

Also, the weight per crystal region ( $w_i$ ) is:

$$w_i = \frac{1}{\sqrt{N_{F_i} + N_{F_i(bg)}}} = \frac{1}{\sigma_{F'_i}} \quad (12)$$

Following the weights per individual crystal, the weighted sum now becomes:

$$\bar{N}_w = \frac{\sum_i \alpha_i N'_{F_i} \cdot w_i^2}{\sum_i w_i^2} \quad (13)$$

If the datasets and standard deviation of the fluorescent counts of individual crystals are consistent, the weighted sigma or total uncertainty is:

$$\sigma_{total(w)} = \frac{1}{\sqrt{\sum_{i=1}^3 w_i^2}} = \sqrt{\sum_{i=1}^3 \sigma_{F'_i}^2} \quad (14)$$

But if they are inconsistent to each other, following expression comes into account:

$$\sigma_{total(w)} = \sqrt{\frac{\sum_{i=1}^3 (\alpha_i N'_{F_i} - \bar{N}_w)^2 \cdot w_i^2}{\sum_{i=1}^3 w_i^2}} \quad (15)$$

If the data are consistent, the total counts can then be summed as:

$$N_F = \sum_{i=1}^3 N'_{F_i} \quad (16)$$

where

$$\sigma_{total} = \sqrt{N_F}$$

which is the uncertainty of the total counts.

Peak	Energy (eV)	Width (eV)
1	5881.85 ± 1.42	2.09 ± 3.55
2	5885.45 ± 1.89	2.29 ± 1.97
3	5891.38 ± 0.46	1.46 ± 0.79
4	5895.94 ± 0.41	1.38 ± 1.47
5	5897.72 ± 0.20	0.78 ± 0.42
6	5904.59 ± 0.05	2.20 ± 0.11
7	5910.25 ± 0.28	4.00 ± 39.72
8	5917.57 ± 0.31	4.00 ± 0.88

**Table 2.** Peak positions and widths of Hidden Satellites derived from fitting.

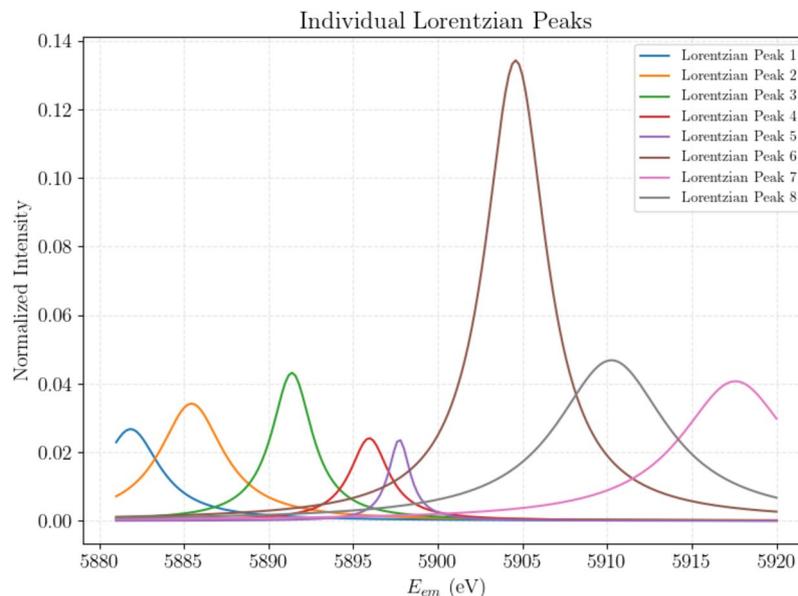
### Method for principal component analysis

Typical PCA utilizes Singular Value Decomposition (SVD) rather than a more traditional least-squares analysis. By transforming the original dataset into a new set of orthogonal spectral components, known as Principal Components (PCs), the analysis and interpretation of complex datasets is simplified. To implement PCA, we compute the covariance matrix  $S$  of the XR-HERFD dataset  $I$ , and then perform eigenvalue decomposition to obtain the eigenvectors  $a_i$ , which are ordered by the magnitude of their corresponding eigenvalues. The eigenvector associated with the largest eigenvalue,  $a_1$ , represents the first principal component. The vector  $a_1$  has a size of  $N_{E_{in}}$ , corresponding to the number of columns in  $I$ , or the number of incident energies. Each entry in  $a_1$  reflects the contribution of the corresponding incident energy to the first principal component. By projecting the dataset  $I$  onto  $a_1$ , we obtain the vector  $P_1$ , of size  $N_{E_{em}}$ , corresponding to the number of emission energies. These are linear combinations of the XES spectra that capture the largest variance across the incident energy dimension of the dataset. This projection isolates the dominant structure in the data.

The first and most important question that arises is: how many components are necessary to reproduce the observed spectra in the dataset? Whilst there are several approaches to address this, the most straightforward and informative method is to examine the explained variance ratios and assess the contribution of each component. In our case, it was evident that the first principal component ( $P_1$ ) accounted for over 97% of the total variance, clearly dominating the dataset and making the remaining components significantly less informative. For the sake of completeness and robustness, we considered the first three components in our analysis. However, the second and third components carried minimal physical significance and were largely associated with noise.

### Information on fitting parameters

As shown in Fig. 7, we modelled the PCA-derived spectrum using nonlinear least-squares fitting implemented in the `lmfit` package. An initial inspection suggested the presence of up to eight Lorentzian features, which were subsequently refined through the fitting procedure. The optimized centre energies and widths are listed in Table 2. To illustrate how each Lorentzian component contributes to the reconstructed spectrum, the individual fitted peaks are shown separately in Fig. 11.



**Fig. 11.** Individual Lorentzian peaks of the Hidden Satellites.

### Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 23 May 2025; Accepted: 21 December 2025

Published online: 29 December 2025

### References

- Chantler, C. T., Bunker, G., D'Angelo, P. & Diaz-Moreno, S. X-ray absorption spectroscopy. *Nat. Rev. Methods Primers* **4**, 1–25 (2024).
- Smith, J. W. & Saykally, R. J. Soft X-ray absorption spectroscopy of liquids and solutions. *Chem. Rev.* **117**, 13909–13934 (2017).
- Jiang, K. et al. Single platinum atoms embedded in nanoporous cobalt selenide as electrocatalyst for accelerating hydrogen evolution reaction. *Nat. Commun.* **10**, 1743 (2019).
- Kravtsova, A. N. Synchrotron-based X-ray absorption spectroscopy for the study of geological materials. *J. Surf. Invest.* **14**, 135–149 (2020).
- de Groot, F. M. et al. Resonant inelastic X-ray scattering. *Nat. Rev. Methods Primers* **4**, 45 (2024).
- Glatzel, P. et al. Hard X-ray photon-in-photon-out spectroscopy with lifetime resolution—of XAS, XES, RIXSS and HERFD. In *AIP Conference Proceedings 1731–1734* (American Institute of Physics, 2007).
- Niskanen, J. et al. Sulphur  $k\beta$  emission spectra reveal protonation states of aqueous sulfuric acid. *Nat. Sci. Rep.* **6**, 21012 (2016).
- Castillo, R. G. et al. High-energy-resolution fluorescence-detected X-ray absorption of the Q intermediate of soluble methane monooxygenase. *J. Am. Chem. Soc.* **139**, 18024–18033 (2017).
- Wernet, P. et al. The structure of the first coordination shell in liquid water. *Science* **304**, 995–999 (2004).
- Soldati, A. L., Jacob, D. E., Glatzel, P., Swarbrick, J. C. & Geck, J. Element substitution by living organisms: The case of manganese in mollusc shell aragonite. *Nat. Sci. Rep.* **6**, 22514 (2016).
- Greczynski, G. & Hultman, L. X-ray photoelectron spectroscopy: Towards reliable binding energy referencing. *Prog. Mater. Sci.* **107**, 100591 (2020).
- Singh, J., Lamberti, C. & van Bokhoven, J. A. Advanced X-ray absorption and emission spectroscopy: In situ catalytic studies. *Chem. Soc. Rev.* **39**, 4754–4766 (2010).
- Chiu, C.-C. et al. Spectroscopic characterization of electronic structures of ultra-thin single crystal  $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ . *Nat. Sci. Rep.* **11**, 5250 (2021).
- Asada, S. & Sugano, S. Satellites in X-ray photoelectron spectra of transition-metal compounds. *J. Phys. Soc. Jpn.* **41**, 1291–1299 (1976).
- Chandesris, D., Lecante, J. & Petroff, Y. Two-electron resonances in transition metals. *Phys. Rev. B* **27**, 2630 (1983).
- Nath, K. G., Haruyama, Y. & Kinoshita, T. Observation of the satellite signal in Co 2 p photoemission spectra: Evidence of a localized electronic structure in thin films. *Phys. Rev. B* **64**, 245417 (2001).
- Glatzel, P. & Bergmann, U. High resolution 1s core hole X-ray spectroscopy in 3d transition metal complexes—electronic and structural information. *Coord. Chem. Rev.* **249**, 65–95 (2005).
- Diamant, R. et al.  $K^h\alpha_{1,2}$  hypersatellites of 3d transition metals and their photoexcitation energy dependence. *Phys. Rev. A* **79**, 062511 (2009).
- Ikeno, H., de Groot, F. M. F., Stavitski, E. & Tanaka, I. Multiplet calculations of  $L_{2,3}$  X-ray absorption near-edge structures for 3d transition-metal compounds. *J. Phys.: Condens. Matter* **21**, 104208 (2009).
- Nguyen, T. V. B., Melia, H. A., Janssens, F. I. & Chantler, C. T. Theory of copper  $K\alpha$  and  $K\beta$  diagram lines, satellite spectra, and ab initio determination of single and double shake probabilities. *Phys. Lett. A* **426**, 127900 (2022).
- Dean, J. W., Pushkarna, P., Melia, H. A., Nguyen, T. V. B. & Chantler, C. T. Theoretical calculation of characteristic radiation: Multiconfiguration Dirac-Hartree-Fock calculations in scandium  $K\alpha$  and  $K\beta$ . *J. Phys. B: At. Mol. Opt. Phys.* **55**, 075002 (2022).
- Sier, D. et al. High-accuracy high-resolution measurements of fluorescence in manganese using extended-range high-energy-resolution fluorescence detection. *J. Appl. Crystallogr.* **58**, 42–59 (2025).

23. Lowe, J. A., Chantler, C. T. & Grant, I. P. Ab initio determination of satellite intensities in transition-metal photoemission spectroscopy using a multiconfiguration framework. *Phys. Rev. A* **83**, 060501 (2011).
24. Tran, N. T. T. et al. A new satellite of manganese revealed by extended-range high-energy-resolution fluorescence detection. *J. Synchrotron Radiat.* **30**, 605–612 (2023).
25. Sier, D. et al. High-accuracy measurement, advanced theory and analysis of the evolution of satellite transitions in manganese  $K\alpha$  using XR-HERFD. *IUCr J.* **11**, 620–633 (2024).
26. Rijal, R. et al. Hyper-resolution in X-ray emission spectroscopy: integrating XR-HERFD and multiple crystal spectrometry with advanced binary data splicing. *J. Synchrotron Radiat.* **TBD**, TBD (2025).
27. Rehr, J. J., Stern, E. A., Martin, R. L. & Davidson, E. R. Extended X-ray absorption fine structure amplitudes - wave-function relaxation and chemical effects. *Phys. Rev. B* **17**, 560–565 (1978).
28. Lee, P. A. & Beni, G. New method for the calculation of atomic phase shifts: Application to extended X-ray absorption fine structure (EXAFS) in molecules and crystals. *Phys. Rev. B* **15**, 2862–2883 (1977).
29. Rehr, J. J., de Leon, J. M., Zabinsky, S. I. & Albers, R. C. Theoretical X-ray absorption fine-structure standards. *J. Am. Chem. Soc.* **113**, 5135–5140 (1991).
30. Newville, M. *Fundamentals of XAFS* (CARS. University of Chicago, Chicago IL, 2004).
31. de Groot, F. M. F. et al. 2p X-ray absorption spectroscopy of 3d transition metal systems. *J. Electron. Spectrosc. Relat. Phenom.* **249**, 147061 (2021).
32. Kalha, C. et al. Lifetime effects and satellites in the photoelectron spectrum of tungsten metal. *Phys. Rev. B* **105**, 045129 (2022).
33. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
34. Diaz-Moreno, S. et al. I20; The versatile X-ray absorption spectroscopy beamline at diamond light source. In *Journal of Physics: Conference Series*, vol. 190, pp. 012038 (IOP Publishing, 2009).
35. Kraft, S., Stümpel, J., Becker, P. & Kuetgens, U. High resolution x-ray absorption spectroscopy with absolute energy calibration for the determination of absorption edge energies. *Rev. Sci. Instrum.* **67**, 681–687 (1996).
36. Hayama, S. et al. The scanning four-bounce monochromator for beamline I20 at the diamond light source. *J. Synchrotron Radiat.* **25**, 1556–1564 (2018).
37. Ponchut, C. et al. MAXIPIX, a fast readout photon-counting X-ray area detector for synchrotron applications. *J. Instrum.* **6**, C01069 (2011).
38. Tran, N. T. T. et al. A new satellite of manganese revealed by extended-range high-energy-resolution fluorescence detection. *Synchrotron Radiat.* **30**, 605–612 (2023).
39. Jolliffe, I. T. & Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. Royal Soc. A: Math. Phys. Eng. Sci.* **374**, 20150202 (2016).
40. Gewers, F. L. et al. Principal component analysis: A natural approach to data exploration. *ACM Comput. Surv. (CSUR)* **54**, 1–34 (2021).
41. Beattie, J. R. & Esmonde-White, F. W. L. Exploration of principal component analysis: Deriving principal component analysis visually using spectra. *Appl. Spectrosc.* **75**, 361–375 (2021).
42. Giuliani, A. The application of principal component analysis to drug discovery and biomedical data. *Drug Discov. Today* **22**, 1069–1076 (2017).
43. Lee, L. C. & Jemain, A. A. On overview of PCA application strategy in processing high dimensionality forensic data. *Microchem. J.* **169**, 106608 (2021).
44. Federolf, P., Reid, R., Gilgien, M., Haugen, P. & Smith, G. The application of principal component analysis to quantify technique in sports. *Scand. J. Med. Sci. Sports* **24**, 491–499 (2014).
45. Ghosh, D. & Chattopadhyay, P. Application of principal component analysis (PCA) as a sensory assessment tool for fermented food products. *J. Food Sci. Technol.* **49**, 328–334 (2012).
46. Tadić, L., Bonacci, O. & Brleković, T. An example of principal component analysis application on climate change assessment. *Theoret. Appl. Climatol.* **138**, 1049–1062 (2019).
47. Chernyshov, D., Dovgaliuk, I., Dyadkin, V. & van Beek, W. Principal component analysis (PCA) for powder diffraction data: Towards unblinded applications. *Crystals* **10**, 581 (2020).
48. De Groot, F. High-resolution X-ray emission and X-ray absorption spectroscopy. *Chem. Rev.* **101**, 1779–1808 (2001).
49. Rankine, C. D. & Penfold, T. J. Progress in the theory of X-ray spectroscopy: From quantum chemistry to machine learning and ultrafast dynamics. *J. Phys. Chem. A* **125**, 4276–4293 (2021).
50. Buitinck, L. et al. API design for machine learning software: Experiences from the scikit-learn project. *arXiv preprint arXiv:1309.0238* (2013).
51. Stephens, J. et al. 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. *IUCr J.* **12**, (2025).
52. Thomas, T. D. Transition from adiabatic to sudden excitation of core electrons. *Phys. Rev. Lett.* **52**, 417 (1984).
53. Roy, M., Lindsay, J., Louch, S. & Gurman, S. J. Multiple-electron excitation in X-ray absorption: A simple generic model. *J. Synchrotron Radiat.* **8**, 1103–1108 (2001).
54. Mukoyama, T. et al. Evolution of satellite lines. *X-Ray Spectrometry: Int. J.* **38**, 138–143 (2009).
55. Chantler, C., Bunker, B. & Boscherini, F. *International Tables for Crystallography, Volume I: X-ray Absorption Spectroscopy and Related Techniques 1* (Wiley, 2024).
56. Jönsson, P., Gaigalas, G., Bieroń, J., Fischer, C. F. & Grant, I. New version: Grasp2K relativistic atomic structure package. *Comput. Phys. Commun.* **184**, 2197–2203 (2013).
57. Lowe, J. A., Chantler, C. T. & Grant, I. P. Self-energy screening approximations in multi-electron atoms. *Radiat. Phys. Chem.* **85**, 118–123 (2013).
58. Nguyen, T., Lowe, J., Pham, T., Grant, I. & Chantler, C. Electron self-energy corrections using the Welton concept for atomic structure calculations. *Radiat. Phys. Chem.* **204**, 110644 (2023).
59. Dean, J. W., Melia, H. A., Nguyen, T. V. B. & Chantler, C. T. Scandium  $K\alpha$  and  $K\beta$  X-ray spectra with ab initio satellite intensities and energy eigenvalues. *Phys. Rev. A* **109**, 022809 (2024).
60. Kochur, A. G., Dudenko, A. I. & Petrini, D. Shake process probabilities for outer-shell electrons in atoms with  $Z_Z \leq 71$ . *J. Phys. B: At. Mol. Opt. Phys.* **35**, 395 (2002).
61. Mukoyama, T. & Taniguchi, K. Atomic excitation as the result of inner-shell vacancy production. *Phys. Rev. A* **36**, 693 (1987).
62. Dean, J. W., Thompson, S. N. & Chantler, C. T. Ab initio manganese  $K\alpha$  and  $K\beta$  energy eigenvalues, shake-off probabilities, Auger rates, with convergence tests. *Molecules* **29**, 4199 (2024).

## Acknowledgements

The experiment for this work was performed with the invaluable support of the I20-Scanning Beamline Team at Diamond Light Source, UK.

### Author contributions

C.T.C., D.S., N.T.T.T., J.W.D., T.K., C.Q.T., S.D.M. conducted the experiments, R.R. and J.A.S. analysed the results, T.V.B.N. calculated theoretical values, R.R., J.A.S. and C.T.C. guided the analysis and edited the manuscript. All authors edited the manuscript.

### Funding

The following funding is acknowledged: Australian Research Council (Grant No. DP210100795).

### Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to C.T.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025