Dissecting complex nanoparticle heterostructures via multimodal data fusion with aberration-corrected STEM spectroscopy

Thomas Thersleff,† Serhiy Budnyk,‡ Larissa Drangai,‡ and Adam Slabon†
†Stockholm University, Department of Materials and Environmental Chemistry, 10691 Stockholm, Sweden
‡Austrian Centre of Competence for Tribology, AC2T research GmbH, Viktor-Kaplan-Straße 2, 2700, Wr. Neustadt, Austria
E-mail: thomas.thersleff@mmk.su.se

Abstract

With nanostructured materials such as catalytic heterostructures projected to play a critical role in applications ranging from water splitting to energy harvesting, tailoring their properties to specific tasks requires an increasingly comprehensive characterization of their local chemical and electronic landscape. Although aberration corrected electron spectroscopy currently provides sufficient spatial resolution to study this space, an approach to concurrently dissect both the electronic structure and full composition of buried metal/oxide interfaces remains a considerable challenge. In this manuscript, we outline a statistical methodology to jointly analyze simultaneously-acquired STEM EELS and EDX datasets by fusing them along their shared spatial factors. We show how this procedure can be used to derive a rich descriptive model for estimating both transition metal valency and full chemical composition from encapsulated morphologies such as core-shell nanoparticles. We demonstrate this on a heterogeneous Co-P thin film catalyst, concluding that this system is best described as a multi-shell phosphide structure with a P-doped metallic Co core.

Abbreviations

EELS, EDX, STEM, TEM, PCA, NMF-QMV, ELNES, GIF, SSS, LMM, HAADF, HER

Keywords

Data fusion, Electron Energy-Loss Spectroscopy, Energy-Dispersive X-Ray Spectroscopy, Nanostructured catalyst, Scanning Transmission Electron Microscopy

Introduction

The rise of nanostructured materials encompassing the control over the particle interfaces has resulted in significant advances in applications including energy harvesting, energy storage, (opto)electronics and heterogeneous catalysis.¹–³ The latter is one of the key technologies to ensure sustainable processes for the industrial production of chemicals. Besides tuning the elemental composition of nanostructured catalysts,⁴,⁵ the tailoring the electronic structure of their interfaces toward heterostructured entities, such as core-shell⁶ or Janus nanoparticles,⁷ is a general strategy for increasing their activity and/or selectivity.

Similar to bulk materials, the drive towards efficient nanostructured catalysts requires an
intimate understanding of the chemical, morphological, and electronic landscape of these materials; however, unlike bulk materials, these properties must be understood on the nanometer length scale. Consequently, many well-established characterization techniques are difficult to employ, as they lack the necessary spatial resolution and are thus restricted to probing the bulk (global) behavior. This places a strong emphasis on techniques with a high spatial resolution, inviting innovative solutions for their application to nanoscale materials systems. This is particularly pronounced for nanoparticles exhibiting a core-shell morphology, as these can essentially be modelled as buried junctions, providing access to novel materials properties that are unavailable in bulk materials.

One nanoscale characterization technique that can address these questions on the nanoscale is aberration-corrected Scanning Transmission Electron Microscopy (STEM). Aberration-corrected STEM involves focusing a beam of spatially and temporally coherent electrons down to the lateral dimensions of an atom and rapidly scanning it across a thin sample suspended in a vacuum or on a thin carbon film. While this is traditionally used to produce a two-dimensional Z-Contrast image via Rutherford scattering, the experimental geometry of STEM can readily be exploited to additionally probe the local chemical composition and electronic structure of a material via both absorption and emission spectroscopy. Absorption spectroscopy in the electron microscope is referred to as Electron Energy-Loss Spectroscopy (EELS) and can be used to study the local electronic structure of a material via both absorption and emission spectroscopy. Absorption spectroscopy in the electron microscope is referred to as Electron Energy-Loss Spectroscopy (EELS) and can be used to study the local electronic structure of a material via both absorption and emission spectroscopy. Absorption spectroscopy in the electron microscope is referred to as Electron Energy-Loss Spectroscopy (EELS) and can be used to study the local electronic structure of a material via both absorption and emission spectroscopy. Absorption spectroscopy in the electron microscope is referred to as Electron Energy-Loss Spectroscopy (EELS) and can be used to study the local electronic structure of a material via both absorption and emission spectroscopy. Absorption spectroscopy in the electron microscope is referred to as Electron Energy-Loss Spectroscopy (EELS) and can be used to study the local electronic structure of a material via both absorption and emission spectroscopy. Absorption spectroscopy in the electron microscope is referred to as Electron Energy-Loss Spectroscopy (EELS) and can be used to study the local electronic structure of a material via both absorption and emission spectroscopy. Absorption spectroscopy in the electron microscope is referred to as Electron Energy-Loss Spectroscopy (EELS) and can be used to study the local electronic structure of a material via both absorption and emission spectroscopy. Absorption spectroscopy in the electron microscope is referred to as Electron Energy-Loss Spectroscopy (EELS) and can be used to study the local electronic structure of a material via both absorption and emission spectroscopy.

An important example how buried heterostructures can impact materials properties includes transition metal phosphide thin films, which have emerged as powerful catalysts for the hydrogen evolution reaction (HER) and catalytic hydrogen generation from hydrolysis of boron-hydrides. In comparison to monodisperse and single-crystalline metal phosphide nanoparticles obtained from colloidal synthesis, cobalt-phosphorus thin films can exhibit a more complex microstructure as a result of multiple deposition steps. For instance, Cho et al. synthesized thin films by electroplating cobalt and subsequent conversion into an amorphous Co-P thin film. Another example of Co-P thin films containing buried junctions are crystalline cobalt nanoparticles embedded in a matrix of amorphous cobalt phosphide. Critically, the high performance of these Co-P systems is suspected to result from charge separation between the Co and P at buried metal/oxide interfaces, emphasizing the importance of fully characterizing these features.

A comprehensive understanding of the chemical and electronic heterogeneity in this system requires two key questions need to be addressed. First, do these thin films exhibit a core-shell morphology and, if so, how is P distributed? Second, what is the metallicity of the Co and how does this correlate to compositional and morphological features? The first question is challenging using standard STEM spectroscopy approaches because P is present in low concentrations and thus challenging to reliably detect with EELS, while EDX datasets suffer from very low count rates resulting in noisy maps. The second question requires EELS to address but, due to overlap of multiple distinct phases in 2D projection, it is difficult to separate the pure spectral components from their mixed observables.

In this work, we describe an experimental design and statistical methodology to address these questions by dissecting both the local chemistry and electronic structure of these complex nanoparticle heterostructures using simultaneously-acquired STEM-EELS-EDX. We achieve this by combining two statistical techniques. First, we detail how performing
low-level block-weighted data fusion\textsuperscript{29,30} between the EELS and EDX datasets prior to multivariate decomposition can directly elucidate correlations between EDX composition and EELS near-edge fine structure (ELNES) features. We also show how this procedure can exploit descriptive statistics such as the data variance to inpaint sparse data,\textsuperscript{31,32} allowing us to map the P distribution using EDX despite its low collection efficiency in this experiment as well as very weak impurity phases that were not detected in the EELS data block. Second, we demonstrate that a blind-source separation procedure based on a minimum-volume geometric unmixing algorithm\textsuperscript{33,34} can be performed on the fused dataset to estimate both the local composition and metallicity of Co-rich segregations buried within a complex Cobalt phosphide shell structure. This is despite the absence of pure spectral signatures from this core in the raw data due to its encapsulation within a series of oxide shells in the 2D projective geometry of the microscope. We subsequently argue that this evidences alloying between the metallic Co phase of this heterostructure and P, despite the absence of a P edge in the EELS dataset that defines its multivariate geometry. We conclude with a brief discussion on the prospects of this approach for the wider research community in nanostructured materials.

Results and discussion

Elemental maps and PCA analysis

A thin film of CoP was synthesized with electrochemical deposition as detailed in the experimental section. A scanning electron microscopy image of the catalyst surface is presented in figure 1. The nickel-plated steel substrate is observed below a porous Co-P thin film.

A sample of the thin film was prepared for TEM analysis following the procedure outlined in the experimental section. A suitable region of the material was scanned with a finely-focused electron probe (see the methodology section for details). The probe position was advanced across the sample in discrete steps and, at each position (denoted a ”pixel position” in this work), the movement was paused for approximately 2 ms. During this time, one EDX and two EEL spectra (corresponding to core-loss and low-loss EELS) were acquired. This yields three independent, complimentary datasets having dimensions $N_x \times N_y \times N_E$ where $N_x$ and $N_y$ represent the number of pixels in the $x$ and $y$ spatial dimensions and $N_E$ represents the number of energy channels for the given dataset. Critically, while $N_E$ varies arbitrarily between all three datasets, their synchronous acquisition permits coupling along their shared spatial dimensions $N_x$ and $N_y$, a property we exploit in the data fusion section below.

The EELS datasets were used to map the elemental distribution of Co and O, and the results of this analysis are presented in figure 2. In figure 2a, the High-Angle Annular Dark Field (HAADF) survey image of the catalyst region under investigation is presented. This figure is generated by collecting electrons scattered to high angles due to Rutherford scattering. It is immediately evident that this region of the catalyst increases in thickness towards the top of the scanning region. However, in the thinner regions towards the bottom, a clear inhomogeneity in the scattering contrast is observed, revealing strong variations in the average atomic weight. Cursory inspection of the core-loss EELS and EDX datasets suggests that this inhomogeneity can be largely explained by Co and O chemical mixing.

Further insight into the nature of this chemical mixing can be extracted by mapping the chemical distribution of Co and O. Figures 2c,
Figure 2: (a) Z-contrast map of the catalyst generated with the HAADF detector. (b) Comparison between the raw core-loss EELS spectrum and the PCA model from a single pixel, denoted in (a). (c, d) Areal density maps of the Co and O elemental distributions from the raw data, respectively. (e, f) Areal density maps of the Co and O elemental distributions from the PCA model, respectively. (g) Quantitative ratio of Co to O. (h) White-line ratio of the Co L$_{2,3}$ edges.

d show this elemental distribution as captured by the core-loss EELS. These maps were generated from the raw data by first subtracting off the pre-edge background contribution and subsequently removing plural scattering by Fourier ratio deconvolution of the the core-loss EELS spectra with the simultaneously-acquired low-loss EELS spectra on a pixel-by-pixel basis. The Co map closely mirrors the Z-contrast observed in figure 2a; however, the O map is very noisy by comparison, complicating the extraction of the elemental ratio between Co and O. Moreover, as evidenced by the raw spectrum in figure 2b, the high noise level for each individual pixel hampers the reliable estimation of the Co white-line ratio.

To improve the estimation of the elemental distribution, we build an empirical data model using a dimensionality reduction technique known as Principal Component Analysis (PCA). PCA consolidates the observational variance of a dataset into a small number of empirical spectra (referred to as components) and then projects these back onto the raw data (yielding so-called ”score maps”). The components are obtained by orthogonalizing the covariance matrix of the entire dataset via eigenvalue decomposition. For non-sparse datasets with highly redundant observations, such as EELS data cubes, the PCA model can typically describe the vast majority of the signal variance within the first few components, while the remaining components largely describe the noise response of the hyperspectral scene. However, the PCA model itself does not provide any real explanatory power to the system under investigation; it merely allows for a separation between the signal subspace (SSS) and noise response to be estimated. Thus, a second processing step is required to construct a physically or chemically meaningful model of the system. In most cases, this second step involves the application of standard elemental mapping procedures such as EELS cross-section modelling or EDX peak integration to the PCA model itself rather than the raw, noisy data. Provided that the PCA decomposition faithfully captured all of the significant variance of the chemical species under investigation, this physical model results in a very powerful noise reduction that does not rely on local sampling or blurring, thereby maintaining the original spatial resolution. This has been applied to a wide variety of systems in the STEM-EELS and...
EDX literature to, for example, map the spatial distribution of low-intensity elements, visualize EELS fine structure variations that can be interpreted as modifications to the local electronic properties and even map magnetic moments. A comparison between the PCA model and the raw data is presented in figure 2b, where the red curve represents the rank 6 PCA model at the shown pixel position, the light gray shows the raw spectra, and the dark gray presents the residuals. The first 6 components of the PCA model as well as further mathematical details of the PCA approach used in this manuscript are provided in the supplementary information.

Elemental distribution maps generated from the PCA model are presented in figure 2e, f for Co and O, respectively. While a decent improvement in the Co map is observed, the most striking improvement occurs in the Oxygen map due to its lower signal to background ratio in the raw data. The structure revealed in this map appears to be largely anti-correlated to the Co map, consistent with the interpretation that the chemical inhomogeneity observed in the HAADF survey image originates from strong nanoscale mixing between Co and O.

Deeper insight into the nature of the chemical mixing can be achieved by employing two additional processing steps to the PCA model. The first step is to visualize the Co:O stoichiometry. This can be performed in a quantitative manner on the EELS data since plural scattering has been removed and the maps presented in figure 2 are normalized to the inelastic scattering cross sections to have units of atomic density [atoms / nm²]. We directly observe strong variations in the Co:O ratio that suggest the presence of Co-rich inclusions within the bulk of this catalyst. The second step involves estimation of the oxidation state of the encapsulated Co from the white-line ratio of the Co \( \text{L}_{2,3} \) edges. This ratio can be calculated by fitting the Co edges and integrating the area under the \( \text{L}_{2,3} \) and \( \text{L}_{3} \) peaks (similar to the approach taken for Fe in Thersleff et al.'). A map of this ratio for each individual spectrum is presented in figure 2. This map not only implies that the catalyst is chemically heterogeneous but that this heterogeneity mirrors variations in the valence state of the Co.

Limitations of the PCA analysis

While both the elemental and physical property maps derived from the PCA model in figure 2e-h provide considerable insight into the complexity and chemical heterogeneity of this system, we note that it is insufficient to answer the two main questions posed at the outset of this study. First, the PCA model presented in figure 2 arises entirely from the core-loss EELS datacube. This dataset only captures variations in Co and O and, thus, excludes the remainder of elements in the material including the impurities of C, F, Si, S, and Cl but, critically, also the active element P, whose distribution is a primary objective of the study. While the exact same PCA approach described here can be applied to the EDX datacube, the rapid scanning results in an extremely low count rate, yielding a highly sparse dataset. Consequently, the covariance matrix fails to faithfully describe the SSS and its orthogonalization results in a model where only one component is retrieved, effectively describing nothing more than a change in the total photon emission resulting from thickness variations. The full PCA decomposition for both EELS and EDX is presented in the supplementary information.

The sparsity problem of rapidly-scanned STEM-EDX datasets is well documented and a common way to overcome it is to fill in missing values by asserting strong spatio-spectral smoothness constraints. This can be achieved by either rebinning the datacube or applying a blurring kernel to the energy and/or spatial dimensions both prior to mapping of the raw data in commercial software packages as well as a preparatory step prior to PCA decomposition. While this permits three components to be retrieved by inpainting missing values in the EDX data, it also deteriorates the spatial resolution due to the blurring effect of local sampling.

Second, since the PCA model itself is nothing more than an empirical model of the original data, it cannot predict physically meaning-
ful information such as the presence or shape of pure spectral features that are highly mixed in the raw data. This can be understood by considering TEM spectroscopy measurements on core-shell structures, as is suspected here. Since the data are acquired in a 2D projection, an observation which would include spectra originating from the core are also mixed with spectra originating from the shell. Thus, when estimating the valence state of the Co, one has to consider that the pure spectrum from the core will be mixed with the pure spectra from the encapsulating medium. The limitations of this assumption become particularly pronounced when interpreting the white-line map in figure 2h. The white-line ratio of transition metal oxides is known to monotonically decrease as the valence state increases.\textsuperscript{44,45,48} CoO, with a valence of +2, and mixed-valence Co\textsubscript{3}O\textsubscript{4} with an average valence of +2.6 follow this trend, exhibiting white-line ratios of 4.5 and 2.4, respectively.\textsuperscript{49} However, as noted by Zhao et al, metallic Co has a white-line ratio of 3.2, thus sitting between the two oxide values.\textsuperscript{49} In the system investigated here, this distinction between a Co-oxide and a metallic phase is crucial to the understanding of the catalyst. Thus, for white-line values close to 3 in figure 2h, it is impossible to conclude whether these regions consist of pure metallic Co or some mixture of two Co-oxide phases containing different valence states without additional information.

One method for distinguishing between metallic Co and some Co-oxide phase is to inspect the entire Co ionization edge. It is known that metallic Co exhibits an increased width of the 3d orbital near the Fermi-level compared to CoO and Co\textsubscript{3}O\textsubscript{4}, leading to a broader energy tail in the experimental \textit{L}\textsubscript{3} peak.\textsuperscript{49} Additionally, first-principles studies show that transitions from 2p to 4s states contribute additional intensity to the metallic Co edge in the energy range between \textit{L}\textsubscript{3} and \textit{L}\textsubscript{2}.\textsuperscript{50} This results in an overall edge shape for metallic Co that is quite distinct from the cobalt oxides in that the scattering cross section underneath the \textit{L}\textsubscript{2} edge increases and, subsequently, flattens out in the post-edge region. One might hope to see such features in the raw data, and this is attempted in figure 2b where an EELS spectrum is taken from a highly Co-rich pixel. Even though a reliable model of the signal is provided by the PCA analysis, a significant contribution from the O edge still clearly remains. Moreover, the shape of the Co edge cannot be easily separated into either metallic Co or a Co oxide. This precludes a definitive conclusion to be drawn on the nature of the core: is it metallic Co, or some oxide / phosphide composition? Clearly this materials question requires more than a simple denoising approach to reasonably test this hypothesis.

Two related strategies to address these shortcomings constitute the remainder of this manuscript. First, we explore how the rich descriptive power of the core-loss EELS dataset can be used to predict the behavior of the sparse EDX datacube by coupling the two datasets, yielding reliable maps of the distributions of elemental P as well as the additional impurities of C, Si, S, and Cl. Second, we investigate how pure constituent spectra can be estimated from the fused dataset by blind source separation using geometric unmixing.

Data Fusion

Low-level data fusion methodology

Since the core-loss EELS and EDX datasets can, to a first approximation, be considered to have been simultaneously acquired, the individual EELS and EDX spectra emerge from (nearly) the same illuminated region of the sample. This implies that the observation mode of the joint data set is shared. This is equivalent to considering the multiple spectroscopy techniques as extensions of each other with complimentary information, and that this information is strongly correlated between the disparate methods via their joint spatial factors. Thus, any correlations between the detected spectra (also referred to as the variables) can be extracted and scrutinized through use of a variance consolidation algorithm such as PCA described above. The only precondition is that the disparate data blocks can somehow be linked in a meaningful manner. The process
of linking that has been previously established in the literature is broadly referred to as data fusion.\textsuperscript{30}

Data fusion is commonly used in a wide range of scientific disciplines ranging from the chemometrics to the internet of things,\textsuperscript{51–55} but has only been sparingly applied to multimodal hyperspectral electron microscopy datasets.\textsuperscript{56,57} The ultimate goal of data fusion is the discovery and quantification of variations between disparate data blocks containing complementary information. This goal is distinctly different from the goal of PCA on a single data blocks, which is to consolidate the variance between observations into as few components as possible by orthogonalizing them. Thus, data fusion can be seen as pursuant to the goal of maximizing the explanatory power of the joint experiment, at the cost of maximizing the explained variance in an individual data block. We explore this distinction in detail below.

The coupling of data blocks in data fusion can roughly be divided into three levels.\textsuperscript{30} In this paper, we focus on a shared observation mode “low-level” data fusion approach via weighted matrix concatenation, but also show that, via appropriate data block weighting, this yields a simple mid-level approach via shared spatial factors. Low-level data fusion is generally defined as performing the data linking step prior to dimensionality reduction.\textsuperscript{29,58} The motivation is that the results can be directly understood in terms of the original collected variables. However, we have observed that one can also use the low-level fusion approach to tailor the analysis to either only study shared factors or to alternatively include distinct factors unique to only one of the data blocks. This provides the analyst with considerable flexibility when constructing a data model.

The framework we propose here for linking EDX and EELS datablocks follows that of Smilde \textit{et al.}\textsuperscript{58} It consists of an asymmetrical block-weighted matrix concatenation that extends the spectral (or energy) dimension of the fused dataset to include both the electron energy loss axis from EELS as well as the photon emission energy of EDX. Even though these are two physically distinct phenomena, since their spatial dimensions are shared, and since we are interested only in describing the variance of the observations, this concatenation has physical meaning. The fused datasets can be considered to extend each other. In the specific case of EELS and EDX fusion, the ELNES of EELS is now complimented by the full compositional

---

Figure 3: Schematic of the low-level data fusion approach proposed in this manuscript. The EELS and EDX datacubes are coupled via their shared spatial dimensions prior to the desired processing step. The weighting scheme can be used to either permit or preclude the discovery of distinct factors.
The weighted concatenation procedure is briefly summarized here (see the supplementary information for a mathematical description). First, each data block is centered by subtracting the spectral (or observational) mean. Then, each data block is individually normalized by its total variance. These two steps are key to achieving a physically-meaningful link by normalizing the observational variance between data blocks. Note that this is not the same as standardizing the observational variance within an individual data block (also known as z-scoring). The approach we describe here retains the dynamical range within an individual data block while simultaneously allowing for correlations between the different data blocks to be drawn.

Finally, immediately prior to concatenation, we scale each data block by an arbitrary weight $w_m$, where the subscript $m$ refers to the data block modality (in this paper, $m = 1$ for core-loss EELS and $m = 2$ for EDX). This weighting scheme establishes the significance each individual data blocks will have in the subsequent analysis. Large values of $w_m$ will increase the significance of its corresponding data block and, hence, its explanatory power. Such over-weighting yields a dominant data block that preferentially imposes its descriptive statistics onto the other fused data blocks. Sufficiently large values of $w_m$ will even force the entire multivariate structure of the overweighted data block onto the ones with smaller weights. This scenario essentially describes a mid-level fusion approach, where factors from one individual data block are used to explain variance in the others. On the other hand, when $w_m = 1$ for all data modes, each will contribute equally to the subsequent decomposition. In this case, any subsequent analysis (such as PCA decomposition) will consolidate joint observational variance into shared factors that explain features common to both data blocks (see figure 3). An interesting case arises for factors that are not necessarily shared between data sets. As we will show, these so-called "distinct factors" can, if desired, still be included into the analysis model by appropriately weighting the data block from which this factor originates.

Application to the studied material

In this section, we demonstrate how data fusion applied to simultaneously-acquired EELS and EDX data blocks can be employed as an
alternative to applying spatio-spectral smoothness constraints on the sparse EDX dataset, thereby addressing the first materials challenge highlighted above. The general concept is to inpaint the missing EDX values by using a PCA model for the EDX data that is based off of the multivariate structure extracted from the dense core-loss EELS data. This model reduces the reliance on the assumption of local smoothness, instead replacing it with an assumption that variations in compositional information from EDX can be related to both the inelastic scattering cross-section of core-loss EELS as well as variations in the electronic structure via ELNES. In this case study, this assumption appears to hold up sufficiently well to allow for a more reliable estimation of the P distribution from the EDX data. Moreover, by properly weighting the EDX block, we can additionally extract spectral features that are distinct to the EDX data but do not show up clearly in the PCA model for the EELS data. This produces a more comprehensive model of the entire system than is possible with PCA on individual blocks alone.

The effect of weighting the two fused EELS \((m = 1)\) and EDX \((m = 2)\) data blocks acquired for the investigation of this heterogeneous catalyst is visualized in figure 4. In this case, the EDX data block was subjected to a Gaussian blur to allow for any components unique to the EDX data to be extracted, but we emphasize that this is not necessary if one is only interested in using the EELS data to explain the EDX results. Prior to concatenation, \(w_1\) was fixed at unity whereas \(w_2\) was allowed to vary on a logarithmic scale from \(10^2 - 10^{-2}\). The fused data block was then unfolded along the spatial dimension, centered by subtracting the spectral mean, and subsequently decomposed using PCA. The resulting components were then split into the two original modalities and a low-rank data model for each mode was created. The fractional variance explained for each modality is plotted against \(w_2\) in figure 4 for the first 1 - 6 components of the EDX (figure 4a), core-loss EELS (figure 4b), and the total variance of the fused dataset (figure 4c). See the supple-
mentary information for details on how this was calculated. The variance explained by the joint PCA model for the EDX and EELS datasets without data fusion are labeled "pure" and are displayed in gray. $w_2 > 1$ serves to bias the PCA model towards the data variance captured by EDX, while $w_2 < 1$ overweights the core-loss EELS data. When $w = 1$, the EDX and EELS datasets both contribute equally to the model.

Figures 4a-b demonstrate how $w_2$ impacts the ability of the joint PCA model to explain the variances of the individual datasets. When $w_2$ is very large, the PCA model is dominated by the variance of the EDX data block. Consequently, the explanatory power of the EELS data block is reduced. The opposite is true when $w_2$ is very small. When $w_2$ is closer to unity, both data blocks contribute to the PCA model, as observed by the gradual change to the fractional variance explained. We denote this region the "mixed" region in figure 4.

In addition to the changes to fractional variance explained for the individual data blocks, it is instructive to inspect how much variance the PCA model describes from the joint data set. This is presented in figure 4c. We observe that, for large values of $w_2$, the explanatory power of the joint PCA model is quite low. This is a consequence of the sparsity of the EDX data block, which struggles to describe its own variance, much less the variance of the coupled EELS data block. In contrast, the explanatory power of the joint PCA model is considerably larger for very low values of $w_2$. Being both dense and also containing overlapping compositional information for Co and O distribution in the EDX data, the EELS block alone can describe a larger percentage of the total variance. In fact, we even observe that some of the EELS components describing modifications to the O and Co ELNES explains compositional changes that are observed in EDX.

Of particular note, however, is when $w_2$ is closer to unity. Here, we observe that the PCA model is actually capable of describing even more of the data variance in the fused dataset than the individual data blocks. This increase arises because not all of the factors are shared between the two datasets. Specifically, the impurity factor arising from EDX (see component 3 for the PCA model generated on the "blurred" EDX datablock in the supplementary information) does not appear to result in a modification of the EELS edges that is sufficiently significant for the PCA model on the EELS datablock alone to detect. We also observe a plateau-like structure in the "mixed" region. This describes the gradual transition of shared factors from being mostly described by the EDX data (larger $w_2$) to being mostly described by EELS (smaller $w_2$). The impact of this is that, since the EDX data were spatially blurred prior to concatenation in order to allow for the detection of the distinct impurity component, smaller values of $w_2$ draw the score maps back towards the non-blurred resolution of the EELS data. Taking all of these factors into account, we choose the lowest weighting factor that still reveals the EDX impurity phase. This is $w_2 = 0.25$, representing a preference for the EELS data but still retaining the variance of the EDX. In this way, we demonstrate how the weighting procedure can be tailored to suit the scientific inquiry under investigation to either return the most comprehensive model of the joint system or to favor one data modality over the others.

We now turn our focus to how the joint PCA model can be used as a powerful justification for the inpainting of missing EDX values. To demonstrate this, we reconstruct the full EDX data from a rank 6 joint PCA model for the weights $w_2 = 100$, $w_2 = 0.01$, and $w_2 = 0.25$. Validation of this model is performed by comparing the integrated line intensities for all detected elements, and this is presented in figure 5. Further details including an examination of the spectra models are presented in the supplementary information.

The top two rows in figure 5 present the integrated line intensities resulting from the raw data and using a standard spatio-spectral blur as applied in the commercial Velox software package. PCA was not used to generate these maps, so they are used as references to assess the subsequent decomposition. Of particular note is the S-$K_\alpha$ map for the "blurred" data. Here, we see evidence for a S-rich impurity phase that was not detected in the EELS ex-
periments. We also observe an enrichment of Si at the edges of this material. The Co-$K_{\alpha}$ and O-$K_{\alpha}$ maps appear much noisier than their EELS counterparts in figure 2, while the P-$K_{\alpha}$ map is far too noisy to draw any meaningful conclusions.

The third row with $w_2 = 100$ represents the scenario when the EDX data completely describe the joint PCA model, essentially replicating the established procedure of performing PCA directly on the blurred EDX data without any EELS contribution. This approach works by asserting a strong spatio-spectral smoothness constraint to fill in the missing values from the raw EDX data. We now observe a closer feature match between the EDX and EELS maps of Co and O, as presented in figure 2. Additionally, the EDX data provide complimentary information pertaining to the spatial distributions of the impurity elements Si, S, and Cl that were not easily interpreted in the EELS experiment. However, the Si-rich boundary of the catalyst is not fully retrieved, despite being visible in the raw data above. Finally, we observe that the P distribution closely follows that of Co.

We now examine the effect of $w_2 = 0.01$. This essentially represents the case when the EDX data are not allowed to make any contribution to the PCA model. Instead, the descriptive statistics extracted from the core-loss EELS PCA model are directly applied to the EDX data. In other words, the EELS data are used to explain the EDX data. The first observation is that the spatial resolution of the maps now matches the spatial resolution of the EELS dataset, which often are sufficiently dense to be viable for PCA without spatial blurring. The EDX maps generated in this manner can, thus, retain the spatial resolution of the original experiment, which is not straight-forward when employing PCA directly to the sparse EDX data. The second observation is that the maps show a much more clear distribution of elements compared to the raw data. It is especially remarkable all but two of the elements were not captured within the EELS data. Thus, we find that the fine structure variations on O and Co that dominate the EELS PCA model explain variations in the presence of impurities such as Si and Cl in the EDX data. In this case, this is sufficient to explain the variation of P, as this happens to be linked to changes in the white line ratio of the Co edge in the EELS data block (see figure 7c for the PCA model). However, we also observe that the S-rich impurities are lost when compared to the $w_2 = 100$ case or the raw data. This highlights the limitations of this approach: if the changes to the EELS fine structure due to the presence of these impurities is not sufficiently large to be detected in the PCA decomposition, then these features will fail to explain the variation present in the raw EDX data and these features will be excluded from the model.

The final case we examine is when the EDX and EELS data blocks are given nearly equal weighting by setting $w_2 = 0.25$. Here we observe that PCA finds a balance between the two datasets with which to maximize the variance of the fused data block. This allows for the extraction of both shared and distinct factors in a single data set. The bottom row in figure 5 presents the corresponding EDX maps. Both the S and Si maps are now recovered and are consistent with the raw ”blurred” data, while the noise level of the O map is greatly reduced when compared to $w_2 = 100$. We also now observe evidence for C inclusion in the S-rich impurity phase. This captures all the important features in the fused dataset and thus allows for the most comprehensive interpretation of the data, while still slightly favoring the higher spatial resolution and superior signal to noise properties of the EELS data cube.

**Hyperspectral unmixing fused data**

While PCA analysis of the fused data yields an efficient model to describe the signal subspace that can be used for an improved estimate of elemental distribution maps for both EELS and EDX, it cannot retrieve the source spectra from highly mixed datasets resulting from a two-dimensional projection of a three-dimensional object. In the system studied here, this corresponds to extracting the pure spectra
Figure 6: Abundance maps and endmember spectra resulting from the application of NMF-QMV to the fused data weighted with $w_2 = 0.25$. Endmembers 1, 3, 4, and 5 largely derive from the EELS datablock, while endmember 2 is distinct to the EDX datablock. Colors used for the names are also applied in figure 7.

Estimating pure spectral signatures (also known as endmembers) in the presence of strong mixing is challenge that has been extensively explored in the remote sensing literature. Geometric unmixing techniques have been applied to hyperspectral EELS datasets in the past and are useful for datasets where it can be assumed that each individual spectrum can be described as a linear combination of $p$ source spectra whose pure signature is not isolated within the experimental data. These pure spectra are further assumed to constitute the vertices of a simplex object (the convex hull of the data) projected onto $p-1$-dimensional affine space. While many approaches exist to estimate the vertices of this simplex, the QMV-NMF algorithm explored in this paper achieves this by minimizing an objective function balancing a reduction of the simplex volume against data reconstruction errors, subject to constraints such as non-negativity that can be physically justified on both the EELS and EDX data blocks. Application of the QMV-NMF function to the data presented here is described in the supplementary information, while a much more detailed explanation is provided in the work of Zhuang et al. and references.
The core-loss EELS and EDX data blocks were processed as described in the methods section and subsequently fused with $w_2 = 0.25$. PCA decomposition suggested $p = 5$ and this was used as justification for the number of endmembers here. Increasing $p$ was not observed to improve the explanatory power of the decomposition (see supplementary information for details and additional discussion for this justification).

The endmembers and abundance maps from this analysis are presented in figure 6. The five endmembers appear to describe five chemically and morphologically distinct subsets of the original dataset. Endmember 1 reveals a Co-$L_{2,3}$ edge that is strongly reminiscent of metallic Co (see Zhao et al.\textsuperscript{49}). The white-line ratio of this edge is $3.0 \pm 0.2$ and the O-$K$ edge is nearly completely absent. Moreover, its distribution closely follows the regions of high Co concentration observed in figure 2g. The EDX dataset moreover clearly shows Co-$K_{\alpha,\beta}$ and $L_\alpha$ edges while the O-$K_\alpha$ edge is strongly reduced in intensity. Of great significance, however, is the presence of a strong P-$K_\alpha$ edge. Moreover, other impurity elements such as Si, S, and Cl are greatly reduced.

The abundance map for endmember 2 closely follows the S-rich impurity map in figure 5 that was distinct to the EDX dataset. Accordingly, we observe that the EDX spectrum contains strong contributions from Si and S impurities but lacks P and Cl. The EELS spectrum shows a fine structure reminiscent of CoO with a white-line ratio of $4.7 \pm 0.2$, although the corresponding EDX peaks appear to be reduced in intensity. Additionally, a very weak additional peak for F is possibly observed. F is difficult to see in the EDX spectrum, as it overlaps the Co $L_\alpha$ edge and the EDX datacube was subject to spatio-spectral blurring. However, this may also be related to a camera artifact in the EELS spectrometer. It is curious that the CoO EELS spectrum appears so strongly in this component if it only corresponds to an impurity phase. The component this abundance map most closely resembles was distinct to the EDX datablock and was absent in the EELS PCA model. In spite of this, there are some small variations to the ELNES compared to the other components that suggest genuine albeit very weak modifications to the local Co and O electronic environment due to the impurity phase. At this stage, however, we cannot rule out that these variations aren’t caused by errors in the placement of the endmember vertex. While this is not important for addressing the two materials questions posed at the beginning of this manuscript, we see this as a potentially useful way to explore very weak ELNES variations that are not detected in PCA though use of additional EDX information.

Endmembers 3 and 4 both represent Co oxide phases. However, in contrast to the PCA analysis, the unmixing approach strongly suggests that there are two chemically distinct Co phases present in this material. Between both endmembers, we observe very small changes to the O-$K$ ELNES at 550 eV as well as a slight modification to the Co-$L_{2,3}$ white line ratio. Also, the extracted EDX spectra point to a distinctly different composition between these two components. Similar to endmember 1, endmember 3 also contains a significant amount of P. Endmember 4, on the other hand, also contains a significant contribution from Si and Cl impurities. Moreover, the Co white-line ratio between these two is slightly different, with endmember 3 being very consistent with Co in the 2+ oxidation state.\textsuperscript{49}

Finally, endmember 5 reveals the background. As this region has very low counts in the raw data, it is difficult for the QMV-NMF algorithm to accurately describe it. Consequently, we instead use the vertex describing the local extrema of the data simplex as estimated with Vertex Component Analysis (VCA).\textsuperscript{63} The EELS spectrum shows a slight O-$K$ signature while a negative spectrum representing the residual Co-$L_{2,3}$ edge is visible. This is likely due to the difficulties in accurately estimating this vertex position with such low counts in the background. The EDX spectrum mostly reveals a significant contribution of C-$K_\alpha$, which comes from the suspension film. Si is also detected, and this likely arises from an incomplete unmixing on the pixels where the catalyst is present.
Figure 7: (a) Z-Contrast image of the investigated region. (b) Color map of pixel labels resulting from application of k-means clustering to the abundance maps in figure 6. (c) Principal Component (PCs) describing the orthogonal axes of the scatter plots. (d-f) 2D scatter plots for the fused dataset projected along PC04 (d), PC03 (e), and PC02 (f). Note that these projections exclude PC01, as this mostly describes the background contribution. The colors of the datapoints correspond to the labels in (b). The simplex best describing this dataset is overlaid in light gray. Vertices of this simplex giving rise to the endmember spectra in figure 6 are represented as stars and are colored to match the labels in b and figure 6. (g) Projection of the 5D data cloud onto 3D affine space. Again, PC01 is excluded.

The other elements are largely absent.

To improve our understanding of the physical justification of the endmember extraction, it is crucial to refer to the data geometry. This is presented in figure 7. Figure 7a again shows the HAADF survey overview, while next to it in figure 7b, the labels for the endmembers are mapped. The labels were created by applying k-means clustering to the abundance maps in figure 6. The colors for these labels are used to distinguish the data geometry at right, and these same colors are used for the endmember labels in figure 6. Note that endmember 5 is referred to as "BG" in figure 7 to emphasize that it does not contribute significantly to the description of the target material but rather only describes the "background" substrate material.

In figure 7c, the first four principal components representing the affine projection are plotted. The components define the orthogonal axes of the data geometry in figure 7d - g. These scatter plots consist of the score maps plotted against each other in \((p - 1) = 4\)-dimensional affine space. This is difficult to represent as a 2-dimensional object for a figure. To this end, we observe that PC01 largely describes variations from the spectral mean and, hence, is most significant in describing the sample thickness rather than its chemical heterogeneity. Accordingly, we omit this axis in the visualization here,
instead plotting the data against PC02, PC03, and PC04. This results in a 3-dimensional object, which is presented in figure 7g. The 3-dimensional visualization greatly aids with the assessment of the endmember columns. The vertices of the mixing matrix are presented as stars whose color represents the endmember they construct. A 3D video of this data simplex is presented in the supplementary information.

Figure 7 provides a strong justification for the construction of the simplex defining the convex hull of the fused data. First, we observe that the facets of this simplex lie along clearly exposed facets within the data itself. The volume regularization parameter $\beta$ used in the QMV-NMF optimization determines the ability of the simplex to resist outliers. Figure 7 also allows us to better understand the origins of the endmembers observed in figure 6. We observe that the PCA model alone is very difficult to interpret in a physically-meaningful manner. For example, we observe that, at first glance, endmember 1 (interpreted as metallic Co) and endmember 3 (interpreted as CoO with P) appear to be primarily separated along PC02 (see figure 7d). This component describes a positive correlation between the EELS O-K edge and the EELS Co-$L_{2,3}$ edge that is negatively correlated to the EDX Co, O, and P peaks. This suggests that larger values of PC02 describe a more CoO-like structure, which would seem to contradict the metallic appearance of endmember 1. However, we also need to consider the contribution from PC03, which describes an anticorrelation between the EELS O-K edge and the EELS Co-$L_{2,3}$ with a clear white-line ratio modification. Between EELS and EDX, O is positively associated with an increase in Si, while Co is positively associated to an increase in P (since they are both negative). Thus, the placement of the vertex corresponding to endmember 1 at a strongly negative value of PC03 cancels out the CoO-like effect from PC02 and constructs the metallic Co signature, as well as increasing the P concentration. This clustering approach nicely highlights the complications involved when directly interpreting PCA results, many of which can be better understood by applying physically-constrained unmixing routines.

**Discussion**

The methodology presented here is well suited to handle the high complexity of this heterogeneous nanostructured catalyst, providing a much more complete picture of its nanoscale chemical and morphological landscape. The PCA analysis works well on the core-loss EELS data to reduce the dimensionality of the system and can be employed as a noise reduction approach to assist with the interpretation of the chemical composition. However, this approach fails to address the two most important questions of this investigation. First, what is the distribution of P? And second, does the Co segregate out as a metallic phase?

The weighted data fusion method we propose allows us to directly address the first of these two questions. By imposing the features extracted with PCA from the core-loss EELS datacube onto the EDX datacube, we are able to overcome many of the problems associated with its high sparsity, allowing us to map out the P distribution, as performed in figure 5. Moreover, by normalizing and arbitrarily weighting the variance of the EDX and EELS datasets, we can even extract very weak impurity features that were not captured by the EELS experiment. We see this as an exceptionally powerful tool that can provide researchers with considerable flexibility in developing a detailed model of their investigated system. In our case, we are able to surmise that the impurities of Si, S, and Cl are likely to originate from the synthesis and TEM sample preparation in a glass vial, painting a broader picture of the system that can guide future synthesis and sample preparation.

QMV-NMF allows us to address the second question by constructing a rich descriptive model for the elemental distributions within the catalyst. Figure 6 reveals that there is a high probability that the Co segregates out as a metal in distinct nano-sized features, and that these particles contain P. These are likely to be encapsulated within (or covered by) a complex Co oxide shell structure. In contrast
to the findings of Oh et al.,\textsuperscript{28} our shell structure exhibits chemical features that are best described as a linear combination of two Co oxide / phosphide phases. These exhibit slightly different white-line ratios and chemical compositions from each other as well as a modulation of the O-K ELNES. Finally, there appears to be a S-rich impurity phase that likely sits on the surface region of this material that we suspect is related to the reagents.

![Figure 8: Schematic cross-section of the proposed heterogeneous catalyst derived from the experimental methods.](image)

Taken all together, we are able construct a more reasonable materials model for this system, which is schematically presented in figure 8. We feel that it is best described as a ternary core-shell morphology consisting of Co(0):P@CoO(II):P@Co(II or higher)PO\textsubscript{x}. The core is best described by endmember 1, which we argue is metallic based on the Co white-line ratio as well as ELNES fingerprinting.\textsuperscript{49} Also, EDX reveals a significant amount of P present in this endmember, suggesting that it may be alloyed with the metallic Co. The inner shell is likely to be best described by endmember 3. Here, the O and Co ELNES features nearly exactly match those from CoO\textsuperscript{49,64} and the white-line ratio is 4.6 ± 0.2, which is fully consistent with Co in a 2+ oxidation state. We also observe additional P from EDX here. Since the ELNES so closely resembles CoO, we believe that P may be alloyed in a manner similar to the metallic Co core. However, we cannot rule out the formation of some PO\textsubscript{x} phase. The outer shell is likely best described by endmember 4. The ELNES for this compound does not closely resemble any published Co compound spectra. It exhibits a white-line ratio of 4.9±0.3 and an unusual dip between the L\textsubscript{3} and L\textsubscript{2} edges is observed. Additionally, a second peak on the O-K edge at 560 eV is visible that is not observed in CoO. P is observed here in lower concentrations, but so is Si and trace amounts of S. In the thinnest regions of the sample, this phase seems to mix exclusively with the metallic Co, whereas in the thicker regions, both endmember 3 and 4 are present. For these reasons, we feel that this best describes an outer reaction layer that may have been contaminated by the glass during sample preparation.

In conclusion, we have presented a detailed description of a technique that can be used to derive considerably deeper insight into the true morphological, chemical, and structural nature of highly complex nanoscale heter structured systems. By fusing EDX and EELS data, we are able to effectively extend the informational range of each technique to include both electronic features extracted from analysis of the ELNES but also the full compositional range accessed via EDX. Our proposed weighting scheme allows for both shared and distinct features to be flexibly factored into any subsequent analysis. We then demonstrate how this can be used to provide a much more insightful model of a highly-complex Co-P heterocatalyst, revealing strong evidence for a ternary core-shell morphology. It also provides a compelling argument to be made in favor of a metallic Co core that is alloyed with P. As a final note, while this study employed a relatively large field of view, the only restrictions on the spatial resolution of these techniques are those of the instrumentation themselves. In fact, we expect that application of data fusion and unmixing on atomic-resolution maps could provide unprecedented insight into the atomic-level compositional and electronic fluctuation in crystalline materials.
Experimental

The synthesis of cobalt-phosphorous thin films was carried out according to a slightly-modified procedure reported by Cho et al.\textsuperscript{27} A nickel-coated steel plate (3 cm × 1 cm) was used as substrate for the electrochemical deposition of cobalt and cobalt-phosphorus thin films. The substrate was degreased with ethanol and the electrochemical deposition was performed in a two-electrode setup with a stirring rate of 100 rpm at room temperature. For the deposition of cobalt, a solution containing 0.05 M CoSO\textsubscript{4}-7H\textsubscript{2}O (Sigma Aldrich, ≥99 wt.%), 0.2 M C\textsubscript{6}H\textsubscript{5}Na\textsubscript{3}O\textsubscript{7}-2H\textsubscript{2}O (Sigma Aldrich, ≥99 wt. %) and 0.5 (NH\textsubscript{4})\textsubscript{2}SO\textsubscript{4} (Sigma Aldrich, ≥99 wt. %) was prepared. The cobalt was deposited electrochemically for 12 min at a current density of -0.01 A cm\textsuperscript{-2}. For the deposition of cobalt-phosphorus layer, a solution containing 0.05 M CoSO\textsubscript{4}-7H\textsubscript{2}O (Sigma Aldrich, ≥99 wt. %), 0.2 M C\textsubscript{6}H\textsubscript{5}Na\textsubscript{3}O\textsubscript{7}-2H\textsubscript{2}O (Sigma Aldrich, ≥99 wt. %) and 0.2 M NaH\textsubscript{2}PO\textsubscript{2}-H\textsubscript{2}O (Sigma Aldrich, ≥99 wt. %) was used. The cobalt-phosphorus layer was deposited at a constant current density of 0.01 A cm\textsuperscript{-2} for 30 minutes. After electrochemical deposition, the thin film was washed with de-ionized water and dried under vacuum for 24 hours at room temperature.

The TEM sample was prepared by mechanically scraping the catalyst onto a glass vial and diluting it with ethanol. This was ultrasonicated and a drop was placed onto a thin carbon layer suspended in a copper grid. STEM experiments were performed on an aberration-corrected Themis Z electron microscope from Thermo Fischer. The instrument was operated in STEM mode at 300 kV with probe aberrations corrected up to the fifth order. The electron probe was formed using a convergence angle of 30 mrad and a probe current of 250 pA. The instrument is equipped with a SuperX EDX system from Thermo Fischer as well as a Quantum Gatan Image Filter (GIF) from Gatan, Inc. EELS and EDX data were acquired simultaneously using the Velox software. The probe was scanned in increments of 8.5 Å and a dwell time of 2 ms was used. The spectrometer was set to a dispersion of 0.25 eV/channel and was operated in dual-EELS mode, allowing for the simultaneous collection of the low-loss and core-loss energy ranges. The low-loss spectrum contains the zero-loss peak and contributions from plasmon interactions. This was used primarily to correct for energy shifts and to remove the contribution from plural scattering via fourier ratio deconvolution.\textsuperscript{16,17} The core-loss energy range extended from 482 – 994 eV and captured the O-K and Co-L\textsubscript{2,3} ionization edges as well as their Energy-Loss Near Edge Structure (ELNES).

Data processing was performed using custom written code in the Matlab programming environment and utilizes some functions from the tensorlab package.\textsuperscript{65} Both EELS and EDX data were pre-treated prior to decomposition in the following manner. The EELS datasets were first corrected for spectral and instrumental artifacts using robust multivariate methods.\textsuperscript{66,67} They were both assumed to have a mixed Gaussian / Poissonian noise model, while the EDX dataset was assumed to consist primarily of Poissonian noise. Accordingly, noise homoscedasticity in the EELS data was enforced by application of the Amscombe transformation\textsuperscript{68,69} while the EDX datablock was weighted by the inverse root of the spectral mean.\textsuperscript{70} PCA was performed using the singular value decomposition algorithm in Matlab. The elemental EELS maps in figure 2 were generated on the PCA model using the EELS Analysis plugin for the Digital Micrograph software package by Gatan, Inc.

Prior to QMV-NMF, plural scattering was removed from the core-loss EELS data block using Fourier ratio deconvolution method\textsuperscript{17,71} implemented in Digital Micrograph. For this, the pre-edge background was removed using the data compression approach detailed in Spiegelberg et al.\textsuperscript{72} After deconvolution, the background was added back into the EELS data. Deconvolution ensures that the linear mixing model holds while the inclusion of the pre-edge background has been noted to be beneficial for non-negative matrix factorizations.\textsuperscript{57} QMV-NMF results for the background-subtracted
data are included in the supplementary information. Removal of plural scattering reduces the estimated rank of the PCA model to 5, which is our initial justification for the number of endmembers. We assume that this reflects the number of endmembers in the scene and set $p = 5$ for the unmixing algorithm. This choice of endmembers is further justified by exploring what happens if more endmembers are included. We observe that the additional endmembers return largely sparse abundance maps, suggesting that these additional features are overfitting the original data. This observation has been previously used to automate the estimation of $p$ using sparse regression techniques. However, we did not adopt this automated procedure to our analytical workflow at this time.

QMV-NMF was performed by first projecting the data onto its 4-dimensional signal subspace. Starting parameters were procured by estimating the vertices of the dataset using Vertex Component Analysis (VCA) on this subspace. The final mixing matrix containing 5 endmembers was estimated by applying QMV-NMF using the total variance option and a volume minimization parameter of $\beta = 20$. Abundance maps were calculated by applying the SUNsAL algorithm to the mixing matrix subject to the non-negativity and sum-to-one constraints.

Acknowledgement T. T. acknowledges funding from the Swedish Research Council (project nr. 2016-05113). This work was additionally partially supported by the Austrian COMET Program (project K2 XTribology, no. 849109 and project K2 InTribology, no. 872176) and carried out in collaboration with “Excellence Centre of Tribology” AC2T research GmbH.

Supporting Information Available

Supporting information is available online.

References

1. Rao, P. M.; Cai, L.; Liu, C.; Cho, I. S.; Lee, C. H.; Weisse, J. M.; Yang, P.; Zheng, X. Simultaneously Efficient Light Absorption and Charge Separation in $\text{WO}_3/\text{BiVO}_4$ Core/Shell Nanowire Photoanode for Photoelectrochemical Water Oxidation. Nano Lett. 2014, 14, 1099–1105.

2. Liu, J.; Wang, S.; Kravchyk, K.; Ibáñez, M.; Krumeich, F.; Widmer, R.; Nasiou, D.; Meyns, M.; Llorca, J.; Arbiol, J.; Kovalenko, M. V.; Cabot, A. SnP nanocrystals as anode materials for Na-ion batteries. J. Mater. Chem. A 2018, 6, 10958–10966.

3. Schmidt, D. O.; Hoffmann-Eifert, S.; Zhang, H.; Torre, C. L.; Besmehn, A.; Noyong, M.; Waser, R.; Simon, U. Resistive Switching of Individual, Chemically Synthesized TiO$_2$ Nanoparticles. Small 2015, 11, 6444–6456.

4. Kim, D.; Resasco, J.; Yu, Y.; Asiri, A. M.; Yang, P. Synergistic geometric and electronic effects for electrochemical reduction of carbon dioxide using gold–copper bimetallic nanoparticles. Nature Communications 2014, 5, 1–8.

5. Liang, C.; Yu, Y. Understanding the formation of multiply twinned structure in decahedral intermetallic nanoparticles. IUCrJ 2019, 6, 447–453.

6. Xiong, Y.; Yang, Y.; DiSalvo, F. J.; Abrüna, H. D. Pt-Decorated Composition-Tunable Pd–Fe@Pd/C Core–Shell Nanoparticles with Enhanced Electrocatalytic Activity toward the Oxygen Reduction Reaction. J. Am. Chem. Soc. 2018, 140, 7248–7255.

7. Lord, R. W.; Holder, C. F.; Fenton, J. L.; Schaak, R. E. Seeded Growth of Metal Nitrides on Noble-Metal Nanoparticles To Form Complex Nanoscale Heterostructures. Chem. Mater. 2019, 31, 4605–4613.
8. Armbrüster, M.; Kovnir, K.; Friedrich, M.; Teschner, D.; Wowsnick, G.; Hahne, M.; Gille, P.; Szentmiklósi, L.; Feuerbacher, M.; Heggen, M.; Girgsdies, F.; Rosenthal, D.; Schlögl, R.; Grin, Y. Al_{13}Fe_{4} as a low-cost alternative for palladium in heterogeneous hydrogenation. *Nature Materials* 2012, 11, 690–693.

9. Boldt, K.; Kirkwood, N.; Beane, G. A.; Mulvaney, P. Synthesis of Highly Luminous and Photo-Stable, Graded Shell CdSe/Cd_{x}Zn_{1–x}S Nanoparticles by In Situ Alloying. *Chem. Mater.* 2013, 25, 4731–4738.

10. Brydson, R., Ed. *Aberration-corrected analytical transmission electron microscopy*; RMS-Wiley: Chichester, West Sussex, U.K, 2011.

11. Pennycook, S. J.; Nellist, P. D. *Scanning transmission electron microscopy: imaging and analysis*; Springer, 2011.

12. Krumeich, F.; Müller, E.; Wepf, R. A.; Nesper, R. Characterization of Catalysts in an Aberration-Corrected Scanning Transmission Electron Microscope. *J. Phys. Chem. C* 2011, 115, 1080–1083.

13. Krivanek, O. L.; Chisholm, M. F.; Nicolosi, V.; Pennycook, T. J.; Corbin, G. J.; Dellby, N.; Murfitt, M. F.; Own, C. S.; Szilagyi, Z. S.; Oxley, M. P.; Pantelides, S. T.; Pennycook, S. J. Atom-by-atom structural and chemical analysis by annular dark-field electron microscopy. *Nature* 2010, 464, 571–574.

14. Ertl, M.; Ma, Z.; Thersleff, T.; Lyu, P.; Huettner, S.; Nachtigall, P.; Breu, J.; Slabon, A. Mössbauerite as Iron-Only Layered Oxyhydroxide Catalyst for WO_{3} Photoanodes. *Inorg. Chem.* 2019, 58, 9655–9662.

15. Huang, X.; Teschner, D.; Dimitrakopoulou, M.; Fedorov, A.; Frank, B.; Kraehnert, R.; Rosowski, F.; Kaiser, H.; Schunk, S.; Kuretschka, C.; Schlögl, R.; Willinger, M.-G.; Trunschke, A. Atomic-Scale Observation of the Metal–Promoter Interaction in Rh-Based Syngas-Upgrading Catalysts. *Angewandte Chemie International Edition* 2019, 58, 8709–8713.

16. Brydson, R. *Electron energy loss spectroscopy*; Microscopy handbooks 48; Bios in association with the Royal Microscopical Society: Oxford, 2001.

17. Egerton, R. F. *Electron energy-loss spectroscopy in the electron microscope*; Springer, 2011.

18. Goldstein, J., Ed. *Scanning electron microscopy and x-ray microanalysis*, 3rd ed.; Kluwer Academic/Plenum Publishers: New York, 2003.

19. Ma, Z.; Thersleff, T.; Görne, A. L.; Cordes, N.; Liu, Y.; Jakobi, S.; Robicinska, A.; Schichtl, Z. G.; Coridan, R. H.; Kustrowski, P.; Schnick, W.; Dronskowski, R.; Slabon, A. Quaternary Core–Shell Oxynitride Nanowire Photoanode Containing a Hole-Extraction Gradient for Photoelectrochemical Water Oxidation. *ACS Appl. Mater. Interfaces* 2019, 11, 19077–19086.

20. D’Alfonso, A. J.; Freitag, B.; Klenov, D.; Allen, L. J. Atomic-resolution chemical mapping using energy-dispersive x-ray spectroscopy. *Phys. Rev. B* 2010, 81, 100101.

21. Muller, D. A.; Kourkoutis, L. F.; Murfitt, M.; Song, J. H.; Hwang, H. Y.; Silcox, J.; Dellby, N.; Krivanek, O. L. Atomic-Scale Chemical Imaging of Composition and Bonding by Aberration-Corrected Microscopy. *Science* 2008, 319, 1073–1076.

22. Cueva, P.; Hovden, R.; Mundy, J. A.; Xin, H. L.; Muller, D. A. Data Processing for Atomic Resolution Electron Energy Loss Spectroscopy. *Microsc. Microanal.* 2012, 18, 667–675.

23. Sun, H.; Meng, J.; Jiao, L.; Cheng, F.; Chen, J. A review of transition-metal...
boride/phosphide-based materials for catalytic hydrogen generation from hydrolysis of boron-hydrides. Inorganic Chemistry Frontiers 2018, 5, 760–772.

24. Carenco, S.; Portehault, D.; Boissière, C.; Mézailles, N.; Sanchez, C. Nanoscaled Metal Borides and Phosphides: Recent Developments and Perspectives. Chem. Rev. 2013, 113, 7981–8065.

25. Popczun, E. J.; McKone, J. R.; Read, C. G.; Biacchi, A. J.; Wiltrout, A. M.; Lewis, N. S.; Schaak, R. E. Nanostructured Nickel Phosphide as an Electrocatalyst for the Hydrogen Evolution Reaction. J. Am. Chem. Soc. 2013, 135, 9267–9270.

26. Popczun, E. J.; Read, C. G.; Roske, C. W.; Lewis, N. S.; Schaak, R. E. Highly Active Electrocatalysis of the Hydrogen Evolution Reaction by Cobalt Phosphide Nanoparticles. Angewandte Chemie International Edition 2014, 53, 5427–5430.

27. Cho, K. W.; Kwon, H. S. Effects of electrodeposited Co and Co–P catalysts on the hydrogen generation properties from hydrolysis of alkaline sodium borohydride solution. Catalysis Today 2007, 120, 298–304.

28. Oh, S.; Kim, H.; Kwon, Y.; Kim, M.; Cho, E.; Kwon, H. Porous Co–P foam as an efficient bifunctional electrocatalyst for hydrogen and oxygen evolution reactions. Journal of Materials Chemistry A 2016, 4, 18272–18277.

29. Van Mechelen, I.; Smilde, A. K. A generic linked-mode decomposition model for data fusion. Chemometrics and Intelligent Laboratory Systems 2010, 104, 83–94.

30. Cocchi, M. Data Fusion Methodology and Applications; Elsevier, 2019.

31. Kovarik, L.; Stevens, A.; Liyu, A.; Browning, N. D. Implementing an accurate and rapid sparse sampling approach for low-dose atomic resolution STEM imaging. Appl. Phys. Lett. 2016, 109, 164102.

32. Stevens, A.; Luzzi, L.; Yang, H.; Kovarik, L.; Mehti, B. L.; Liyu, A.; Gehm, M. E.; Browning, N. D. A sub-sampled approach to extremely low-dose STEM. Appl. Phys. Lett. 2018, 112, 043104.

33. Li, J.; Bioucas-Dias, J. M.; Plaza, A.; Liu, L. Robust Collaborative Nonnegative Matrix Factorization for Hyperspectral Unmixing. IEEE T. Geosci. Remote 2016, 54, 6076–6090.

34. Zhuang, L.; Lin, C.-H.; Figueiredo, M. A. T.; Bioucas-Dias, J. M. Regularization Parameter Selection in Minimum Volume Hyperspectral Unmixing. IEEE Transactions on Geoscience and Remote Sensing 2019, 1–20.

35. Jolliffe, I. T. Principal component analysis, 2nd ed.; Springer series in statistics; Springer: New York, 2002.

36. Lichtert, S.; Verbeeck, J. Statistical consequences of applying a PCA noise filter on EELS spectrum images. Ultramicroscopy 2013, 125, 35–42.

37. Spiegelberg, J.; Rusz, J. Can we use PCA to detect small signals in noisy data? Ultramicroscopy 2017, 172, 40–46.

38. Potapov, P. On the loss of information in PCA of spectrum-images. Ultramicroscopy 2017, 182, 191–194.

39. Wei, Y.; Akansel, S.; Thersleff, T.; Harward, I.; Brucas, R.; Ranjarb, M.; Jana, S.; Lansaker, P.; Pogoryelov, Y.; Dumas, R. K.; Leifer, K.; Karis, O.; Åkerman, J.; Celinski, Z.; Svedlindh, P. Exponentially decaying magnetic coupling in sputtered thin film FeNi/Cu/FeCo trilayers. Appl. Phys. Lett. 2015, 106, 042405.

40. Giraldo, S.; Neuschitzer, M.; Thersleff, T.; López-Marino, S.; Sánchez, Y.; Xie, H.; Colina, M.; Placidi, M.; Pistor, P.;
Izquierdo-Roca, V.; Leifer, K.; Pérez-Rodríguez, A.; Saucedo, E. Large Efficiency Improvement in Cu$_2$ZnSnSe$_4$ Solar Cells by Introducing a Superficial Ge Nanolayer. *Adv. Energy Mater.* 2015, 1501070.

41. Thersleff, T.; Giraldo, S.; Neuschitzer, M.; Pistor, P.; Saucedo, E.; Leifer, K. Chemically and morphologically distinct grain boundaries in Ge-doped Cu$_2$ZnSnSe$_4$ solar cells revealed with STEM-EELS. *Mater. Design* 2017, 122, 102–109.

42. Bosman, M.; Watanabe, M.; Alexander, D. T. L.; Keast, V. J. Mapping chemical and bonding information using multivariate analysis of electron energy-loss spectrum images. *Ultramicroscopy* 2006, 106, 1024–1032.

43. Thersleff, T.; Muto, S.; Werwiński, M.; Spiegelberg, J.; Kvashnin, Y.; Hjörvarsson, B.; Eriksson, O.; Rusz, J.; Leifer, K. Towards sub-nanometer real-space observation of spin and orbital magnetism at the Fe/MgO interface. *Sci. Rep.* 2017, 7, 48802.

44. Pearson, D. H.; Fultz, B.; Ahn, C. C. Measurements of 3d state occupancy in transition metals using electron energy loss spectrometry. *Appl. Phys. Lett.* 1988, 53, 1405–1407.

45. Wang, Z. L.; Yin, J. S.; Jiang, Y. D. EELS analysis of cation valence states and oxygen vacancies in magnetic oxides. *Micron* 2000, 31, 571–580.

46. Potapov, P.; Longo, P.; Okumishi, E. Enhancement of noisy EDX HRSTEM spectrum-images by combination of filtering and PCA. *Micron* 2017, 96, 29–37.

47. Potapov, P.; Lubk, A. Optimal principal component analysis of STEM XEDS spectrum images. *Advanced Structural and Chemical Imaging* 2019, 5, 4.

48. Pearson, D. H.; Ahn, C. C.; Fultz, B. White lines and d-electron occupancies for the 3d and 4d transition metals. *Phys. Rev. B* 1993, 47, 8471–8478.

49. Zhao, Y.; Feltes, T. E.; Regalbuto, J. R.; Meyer, R. J.; Klie, R. F. *In situ* electron energy loss spectroscopy study of metallic Co and Co oxides. *Journal of Applied Physics* 2010, 108, 063704.

50. Zheng, H. Electronic structure of CoO. *Physica B: Condensed Matter* 1995, 212, 125–138.

51. Khaleghi, B.; Khamis, A.; Karray, F. O.; Razavi, S. N. Multisensor data fusion: A review of the state-of-the-art. *Information Fusion* 2013, 14, 28–44.

52. Castanedo, F. A Review of Data Fusion Techniques. *The Scientific World Journal* 2013, 2013, 704504.

53. Camacho, J.; Magán-Carrión, R.; García-Teodoro, P.; Treinen, J. J. Networkmetrics: multivariate big data analysis in the context of the internet. *Journal of Chemometrics* 2016, 30, 488–505.

54. Alam, F.; Mehmood, R.; Katib, I.; Albogami, N. N.; Albeshti, A. Data Fusion and IoT for Smart Ubiquitous Environments: A Survey. *IEEE Access* 2017, 5, 9533–9554.

55. Lahat, D.; Adali, T.; Jutten, C. Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. *Proceedings of the IEEE* 2015, 103, 1449–1477.

56. Spiegelberg, J. Signal Processing Tools for Electron Microscopy. PhD Thesis, Uppsala University, Materials Theory, 2018.

57. Muto, S.; Shiga, M. Application of machine learning techniques to electron microscopic/spectroscopic image data analysis. *Microscopy (Oxf)* 2019.

58. Smilde, A. K.; Westerhuis, J. A.; Jong, S. d. A framework for sequential multiblock component methods. *Journal of Chemometrics* 2003, 17, 323–337.
59. Bioucas-Dias, J. M.; Plaza, A.; Dobigeon, N.; Parente, M.; Du, Q.; Gader, P.; Chanussot, J. Hyperspectral Unmixing Overview: Geometrical, Statistical, and Sparse Regression-Based Approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 2012, 5, 354–379.

60. Spiegelberg, J.; Rusz, J.; Thersleff, T.; Pelckmans, K. Analysis of electron energy loss spectroscopy data using geometric extraction methods. *Ultramicroscopy* 2017, 174, 14–26.

61. Tsung-Han Chan,; Wing-Kin Ma,; Chong-Yung Chi,; Yue Wang, A Convex Analysis Framework for Blind Separation of Non-Negative Sources. *IEEE Trans. Signal Process.* 2008, 56, 5120–5134.

62. Tsung-Han Chan,; Chong-Yung Chi,; Yu-Min Huang,; Wing-Kin Ma, A Convex Analysis-Based Minimum-Volume Enclosing Simplex Algorithm for Hyperspectral Unmixing. *IEEE Trans. Signal Process.* 2009, 57, 4418–4432.

63. Nascimento, J.; Dias, J. Vertex component analysis: a fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sensing* 2005, 43, 898–910.

64. Lu, C.; R. Jothi, P.; Thersleff, T.; M. Budnyak, T.; Rokicinska, A.; Yubuta, K.; Dronskowsi, R.; Kuśtrowski, P.; T. Fokwa, B. P.; Slabon, A. Nanostructured core–shell metal borides–oxides as highly efficient electrocatalysts for photoelectrochemical water oxidation. *Nanoscale* 2020, 12, 3121–3128.

65. Vervliet, N.; Debals, O.; Sorber, L.; Van Barel, M.; De Lathauwer, L.

66. Hubert, M.; Rousseeuw, P. J.; Vanden Branden, K. ROBPCA: A New Approach to Robust Principal Component Analysis. *Technometrics* 2005, 47, 64–79.

67. Verboven, S.; Hubert, M. LIBRA: a MATLAB library for robust analysis. *Chemometrics and Intelligent Laboratory Systems* 2005, 75, 127–136.

68. Starck, J.-L.; Murtagh, F. D.; Bijaoui, A. *Image Processing and Data Analysis: The Multiscale Approach*; Cambridge University Press: Cambridge, 1998.

69. Makitalo, M.; Foi, A. Optimal Inversion of the Generalized Anscombe Transformation for Poisson-Gaussian Noise. *IEEE Transactions on Image Processing* 2013, 22, 91–103.

70. Keenan, M. R.; Kotula, P. G. Accounting for Poisson noise in the multivariate analysis of ToF-SIMS spectrum images. *Surf. Interface Anal.* 2004, 36, 203–212.

71. Egerton, R. F.; Wang, F.; Malac, M.; Moreno, M. S.; Hofer, F. Fourier-ratio deconvolution and its Bayesian equivalent. *Micron* 2008, 39, 642–647.

72. Spiegelberg, J.; Rusz, J.; Leifer, K.; Thersleff, T. The usage of data compression for the background estimation of electron energy loss spectra. *Ultramicroscopy* 2017, 181, 117–122.

73. Shiga, M.; Tatsumi, K.; Muto, S.; Tsuda, K.; Yamamoto, Y.; Mori, T.; Tanji, T. Sparse modeling of EELS and EDX spectral imaging data by nonnegative matrix factorization. *Ultramicroscopy* 2016, 170, 43–59.

74. Bioucas-Dias, J. M.; Figueiredo, M. A. T. Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing. 2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing. 2010; pp 1–4.
Graphical TOC Entry

Data fusion + Inpainting + Core / shell dissection

Endmember 1:
Co(0):P
Endmember 4:
Co(II+)POx
Endmember 2:
Impurities
Endmember 3:
CoO(II):P

BG (carbon film)
Co(0):P
CoO(II):P
Co(II+)POx
P
K

Multimodal STEM spectroscopy
EELS
EDX
Block coupling

Raw
Fused
Encapsulated metal core

Energy [eV]
500 600 700 800 900
Abundance
Endmember 01
Endmember 02
Endmember 03
Endmember 04
Endmember 05

Energy [keV]

Energy [eV]
0 2 4 6 8 10
Abundance
Endmember 01
Endmember 02
Endmember 03
Endmember 04
Endmember 05

Energy [eV]
500 600 700 800 900
Abundance
Endmember 01
Endmember 02
Endmember 03
Endmember 04
Endmember 05

Energy [keV]

Raw
Fused
EELS
EDX
P Co
O