

## A Multichannel Cross-Modal Fusion Framework for Electron Tomography

Guo, Yan; Aveyard, Richard; Rieger, Bernd

**DOI**

[10.1109/TIP.2019.2907461](https://doi.org/10.1109/TIP.2019.2907461)

**Publication date**

2019

**Document Version**

Accepted author manuscript

**Published in**

IEEE Transactions on Image Processing

**Citation (APA)**

Guo, Y., Aveyard, R., & Rieger, B. (2019). A Multichannel Cross-Modal Fusion Framework for Electron Tomography. *IEEE Transactions on Image Processing*, 28(9), 4206-4218. Article 8673880. <https://doi.org/10.1109/TIP.2019.2907461>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

# A Multichannel Cross-modal Fusion Framework for Electron Tomography

Yan Guo, *Student Member, IEEE*, Richard Aveyard, and Bernd Rieger

**Abstract**—In this paper, we present a multichannel cross-modal fusion algorithm to combine two complementary modalities in electron tomography: X-ray spectroscopy and scanning transmission electron microscopy (STEM). The former reveals compositions with high elemental specificity but low signal-to-noise ratio (SNR), while the latter characterizes structure with high SNR but little chemical information. We use multivariate regression to build a cross-modal fusion framework for these two modalities to simultaneously achieve high elemental specificity and high SNR for a target element chosen from the sample under study. Specifically, we first compute three-dimensional tomograms from tilt-series datasets of X-ray and STEM using different reconstruction algorithms. Then, we generate many feature images from each tomogram. Lastly, we adopt partial least squares regression to assess the connection between these feature images and the reconstruction of the target element. Based on simulated and experimental datasets of semiconductor devices, we demonstrate that our algorithm cannot only produce continuous edges, homogeneous foreground and clean background in its element-specific reconstructions, but also can more accurately preserve fine structures than state-of-the-art tomography techniques. Moreover, we show that it can deliver results with high fidelity even for X-ray datasets with limited tilts or low counts. This property is highly desired in the semiconductor industry where acquisition time and sample damage are essential.

**Index Terms**—Multimodal image fusion, electron tomography, HAADF-STEM, X-ray spectroscopy, EDS, nanomaterials

## I. INTRODUCTION

**E**LECTRON tomography is a powerful tool in materials science to characterize the complex three-dimensional (3D) structure of inorganic specimens on the nanoscale [1]. In transmission electron microscopy (TEM), the sample under study is exposed to an electron beam and tilted to obtain two-dimensional (2D) projection images at different angles. Several imaging modalities exist, e.g., bright-field TEM [2] and high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) [3]. In tomography, these projections are called a tilt-series, from which we can reconstruct a volume representing the sample [1]. Since the intensity of HAADF-STEM scales with the atomic number  $Z$  of the

element ( $\sim Z^n$ ,  $n = 1.6 - 1.9$ , depending on the inner and outer detector angles [4]), it can also indirectly reveal the compositional information of the sample. However, when the sample has elements with close atomic numbers, HAADF-STEM images may no longer be distinctive for these elements. To better understand more complex compositions, spectral imaging techniques like energy dispersive X-ray spectroscopy (EDS) [4] must be pursued.

Tomographic reconstruction is an ill-posed inverse problem because of the inevitable noise in the measurements such as Poisson noise and readout noise [5]; reconstruction becomes even more problematic for electron tomography where the number of projections is much smaller than the 3D volume [6]. Consequently, its solution might not be stable and unique. So far, dozens of reconstruction techniques have been proposed, and the classical filtered backprojection (FBP) is still frequently applied in practice thanks to its simplicity and speed [7]. Alternatively, iterative algorithms (e.g., simultaneous iterative reconstruction technique (SIRT) and its variants [8]) have also attracted large attention as they produce less artifacts for noisy datasets [1]. Moreover, prior knowledge has been incorporated to further enhance the reconstruction quality [6], [9]–[12]. For instance, assuming that the sample of interest has piecewise constant structures, Goris *et al.* incorporated total variation regularization into SIRT and effectively reduced missing wedge artifacts in the reconstruction [6]. While conventionally only a single modality and/or tilt-series is used in a reconstruction algorithm, advanced approaches tend to combine two (or more) datasets from multiple modalities for integrating complementary information [13]–[15]. Bimodal tomography [14], for example, links HAADF-STEM and EDS projections into a joint reconstruction scheme. The former modality is at atomic resolution with high signal-to-noise ratio (SNR) but not intrinsically element-specific, and the latter, conversely, is rich in chemical information but suffers from low SNR. Although great efforts have been dedicated to improving reconstruction techniques, directly combining reconstruction volumes at hand has still not been widely considered, to the best of our knowledge.

To generate a composite image benefiting from different modalities and/or reconstruction algorithms, one can employ various statistical approaches that project high-dimensional inputs onto low-dimensional outputs. With the capability of reducing redundancies yet highlighting similarities and differences, statistical methods have been broadly investigated and applied in multimodal image fusion, that is, multiple input images of different modalities are fused into a single output. For example, many infrared and visible image fusion frame-

This research is partially supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organization for Scientific Research, and partially by the Ministry of Economic Affairs, Agriculture and Innovation under project number 13314.

Y. Guo and B. Rieger are with the Department of Imaging Physics, Delft University of Technology, 2628CJ Delft, The Netherlands (email: y.guo-3@tudelft.nl; b.rieger@tudelft.nl).

R. Aveyard was with the Department of Imaging Physics, Delft University of Technology, 2628CJ Delft, The Netherlands. He is now with the York Neuroimaging Centre, University of York, YO10 5NY York, United Kingdom (email: richard.aveyard@york.ac.uk)

works have incorporated principal component analysis (PCA) for decorrelation [16]. As part of these methods, regression techniques cannot only decompose two sets of input images into lower dimensions but also assess their connections. For instance, with partial least squares (PLS) regression, van de Plas *et al.* built a linear model to fuse data obtained from mass spectrometry and optical microscopy for studying protein, peptide, lipid, and drug distributions in tissues [17].

Recently, we extended the concept of multimodal fusion to electron tomography [18]. Specifically, we adopted and modified the regression-based cross-modality modeling of van de Plas [17] to fuse X-ray and electron tomograms for reconstructing bi-elemental nanomaterials. We demonstrated our method on an experimental dataset of a core-shell nanoparticle (consisting of gold and silver with distinct atomic numbers of 79 and 47, respectively), and showed that it enabled reconstructions with sharper edges and smoother foreground and background than bimodal tomography [14]. However, the extension to more complex multi-elemental nanostructures has still been lacking thus far. Therefore, in this paper, we first present a 3D multichannel cross-modal fusion algorithm based on our preliminary work in 2D [18]. Then, we validate it on simulated and experimental datasets of semiconductor devices. In particular, we investigate the fusion quality for a small number of EDS acquisition angles and low-dose EDS maps. EDS maps with high SNR are currently hampered by the slow acquisition time in practice, and hence algorithms that can deal with lower SNR are desired.

The rest of this paper is organized as follows. In Section II, we briefly review the classical and state-of-the-art electron tomography along with statistical methods that have been employed in multimodal image fusion. Section III introduces our six-step cross-modal fusion framework and Section IV the experimental setup, including simulation and experimental datasets, and quantitative assessment procedures. Our results are presented in Section V and discussed in Section VI. Finally, we draw the conclusion in Section VII.

## II. CONTEXT AND PRIOR ART

### A. Electron Tomography

In this section, we refer to algorithms that jointly reconstruct a volume from multiple modalities as “multichannel”, and “single-channel” otherwise [19].

1) *Single-channel*: Single-channel approaches are either analytical or iterative. Analytical algorithms are based on the Fourier slice theorem, and directly calculate the reconstruction in a single step. While FBP is the most commonly employed, it aggravates thin streaks in the reconstruction if the number of projections is low. To alleviate such artifacts yet preserve desirable fine structures, Jin *et al.* trained a convolutional neural network to regress the FBP results towards a ground truth image [20]. Although the authors demonstrated the proposed FBPCNN on sparse-view X-ray projections, they stressed that it could be generalized to other modalities. Iterative algorithms, however, formulate the reconstruction problem as a large under-determined linear system and solve it iteratively [5]. Common methods include algebraic reconstruction

technique (ART), simultaneous iterative reconstruction technique (SIRT), simultaneous algebraic reconstruction technique (SART), and maximum likelihood expectation-maximization (ML-EM) [8]. With the capability of incorporating various types of prior knowledge, iterative methods are more robust to deal with ill-posed inverse problems. For instance, assuming that the sample under study only consists of a few elements, that is, the reconstruction only has a few discrete intensity levels, Batenburg and Sijbers developed discrete algebraic reconstruction technique (DART) to achieve a more accurate reconstruction from limited and/or noisy projections [9]. Furthermore, inspired by  $l_1$  regularization in compressive sensing, Goris *et al.* combined the popular total variation regularization (TVR) with SIRT and proposed total variation minimization (TVM) reconstruction to compensate for the missing wedge in electron tomography [6]. TVR was also combined with DART, such that the TVR-DART would require less tuning parameters [10]. Alternative  $l_1$  regularization approaches, such as higher order total variation (HOTV) [11] [12], have also been investigated. Since the solution of HOTV-based reconstruction algorithm is not limited to a piecewise constant function, it could more effectively recover fine features than the common TV [11].

2) *Multichannel*: In general, multichannel algorithms that simultaneously couple datasets from multiple sources are applied either in multispectral or multimodality reconstruction. In electron tomography, most methods combine the complementary information of HAADF-STEM and EDS projections. For example, Zanaga *et al.* used HAADF-STEM to supplement EDS for improving its shadowing effects and lower spatial resolution, thereby enabling a more reliable EDS quantification [13]. Zhong *et al.* introduced a manual parameter to weigh the HAADF-STEM and EDS channels in bimodal tomography for effectively suppressing noise and enhancing contrast [14]. Regularizations, such as total nuclear variation (TNV) derived from TV, have also been considered [15]. Different from TV that only promotes sparse gradients in the EDS reconstructions, TNV further incorporates HAADF-STEM to encourage anti-/parallel gradients for enforcing common edges in the joint reconstructions [15].

Compared with zero- and single-parameter algorithms (e.g., FBP and SIRT), advanced techniques may yield a more accurate result using a limited set of noisy projections. Many of them, however, involve extra parameters that need to be carefully tuned. For instance, the weighting factor  $\alpha$  in bimodal tomography [14] is currently chosen by comparing reconstructions over the whole range of  $\alpha \in (0, 1)$  to a hand-segmented ground truth, which is very impractical.

### B. Statistical Methods Applied to Multimodal Image Fusion

According to Sui *et al.*, statistical approaches applied to multimodal image fusion are either driven by data or by hypotheses [21].

1) *Data-driven Methods*: Data-driven statistical methods include, but are not limited to, principal component analysis (PCA), independent component analysis (ICA), and nonnegative matrix factorization (NMF). PCA has been a common

initializer for fusion algorithms in brain imaging to effectively remove redundancies [21]. It was also improved to robust PCA [22], and adopted to suppress noise yet retain useful information for infrared and visible image fusion [16]. As an extension of PCA, ICA can separate correlated input images into independent components. ICA-aided fusion algorithms usually involve other techniques like multiscale transforms. For instance, Ghahremani and Ghassemian chose ICA and curvelet transform to reduce the spectral distortion of pansharpened multispectral bands [23]. Moreover, incorporating ICA into training-based algorithms is also popular, in which a set of ICA bases are trained from patches with similar contents as the source images [16]. Different from PCA, NMF only allows additive, not subtractive, combinations due to its nonnegativity constraints [24]. As a result, it represents parts of the objects and corresponds better to the human perception mechanism. Wang *et al.* further extended the traditional NMF to a nonnegative sparse representation (NNSR) model for fusing infrared and visible images [25]. As the NNSR not only emphasizes the nonnegativity but also sparsity of the coefficients, it can achieve a rational (only with nonnegative intensities) and convenient (with just a few sparse components) image interpretation.

2) *Hypotheses-driven Methods*: Hypotheses-driven statistical methods, such as regression, can characterize the relationship among source images. In remote sensing, for instance, multivariate regression has been a powerful tool to merge multispectral (MS) and panchromatic (PAN) images. The former captures visible light in a small number of spectral bands at low resolution, and the latter is sensitive to all wavelengths of the light at high resolution. In this case, regression is adopted to estimate the weights between the MS channels and PAN image at both the reduced [26] and full scale [27] for pansharpening. As a well-studied member in the regression family, partial least squares (PLS) establishes a linear multivariate model to relate the inputs [28]. It was first applied to multimodal fusion in neuroimaging by Martínez-Montes *et al.* to concurrently analyze electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) data [29]. Since then, PLS has been broadly employed for multimodal fusion not only in brain imaging [21] but also in biomedical [17] and chemical imaging [30]. For instance, van de Plas *et al.* [17] chose this multivariate regression technique to fuse mass spectrometric and microscopic images. With a series of case studies, they showed that the proposed modeling could maintain both the high chemical specificity and high spatial resolution.

### III. METHOD

Let us assume that a sample has a number of  $E$  different chemical elements that make up the element set  $\mathcal{E}$ , and that its HAADF-STEM projections are acquired at a number of  $P_H$  angles. Meanwhile, X-ray spectral projections are recorded at a number of  $P_E$  angles, and deconvolved into a number of  $E$  EDS maps; each corresponds to one chemical element  $e \in \mathcal{E}$ . Note that the spectral images usually have fewer tilts than the HAADF projections ( $P_E < P_H$ ) and suffer from much lower

SNR. Binning may effectively increase the number of X-ray counts collected in EDS maps but also degrade their spatial resolution [31].

The proposed fusion algorithm has six steps, where the first five are illustrated in Fig. 1:

1. compute multiple tomograms from HAADF and each EDS tilt-series using a number of  $N$  different reconstruction algorithms;
2. i) check visibility of all elements and select a target element  $e^*$  for fusion;  
ii) denoise the EDS reconstructions of  $e^*$ ;
3. generate a number of  $M$  feature images for each HAADF reconstruction, and each EDS reconstruction of  $\check{e} : e \in \mathcal{E} \setminus e^*$ ;
4. (optional) upsample all EDS-related images if the X-ray spectral images have been binned;
5. build a cross-modality model between the feature images and the denoised EDS reconstruction of  $e^*$ , and apply it for fusion;
6. evaluate the reliability of the fusion result.

#### A. Step 1: Computing Tomograms

Our fusion framework starts with tomographic reconstruction. To date, a range of software packages have been issued for electron tomography, such as the open source ASTRA toolbox [32]. Given a HAADF tilt-series and a number of  $N$  available reconstruction algorithms, we can compute  $N$  volumetric images  $\mathbf{x}_n^H \in \mathbb{R}^{V_H \times 1}$ ,  $n = 1, \dots, N$  where  $V_H$  is the total number of voxels being reconstructed. Similarly, for each element  $e \in \mathcal{E}$ , we can also reconstruct  $\mathbf{x}_n^{(e)} \in \mathbb{R}^{V_E \times 1}$  with  $n = 1, \dots, N$ . Note that  $V_E \neq V_H$  if the original X-ray spectral images have been binned.

#### B. Step 2: Checking Visibility and Denoising

In principle, one can choose any element  $e \in \mathcal{E}$  as the fusion target  $e^*$ ; in practice, however,  $e^*$  should be visible in the HAADF reconstructions for building a representative cross-modality model. Our approach to measuring the visibility of  $\mathbf{x}_n^{(e)}$  in  $\mathbf{x}_n^H$  is taken from [26]. First, we upsample  $\mathbf{x}_n^{(e)}$  to  $\tilde{\mathbf{x}}_n^{(e)} \in \mathbb{R}^{V_H \times 1}$  if  $V_E \neq V_H$  using trilinear interpolation. Then, we calculate visibility weights  $w_e$ ,  $\forall e \in \mathcal{E}$  as

$$\mathbf{x}_n^H = w_0 + w_1 \tilde{\mathbf{x}}_n^{(1)} + \dots + w_E \tilde{\mathbf{x}}_n^{(E)} \quad (1)$$

with ordinary least squares regression. Note that this only needs to be done once for any reconstruction algorithm. Since the intensity of HAADF-STEM scales with the atomic number of the element, we suggest that the weight of the chosen  $e^*$  should be at least three times as high as the light ones (e.g., N, O, etc.)

Once we select  $e^*$ , we have  $N$  volumes  $\mathbf{x}_n^{(e^*)}$  at hand. Although they are computed by  $N$  different reconstruction algorithms, they share common patterns. To capture the most dominant structure among these reconstructions and reduce their pixel-specific variations, one can choose any dimension reduction technique mentioned in Section II-B1. Since all intensity values in the output image are inherently nonnegative, we perform NMF over all the reconstructions of  $e^*$  to obtain one denoised image  $\mathbf{x}_d^{(e^*)}$ .

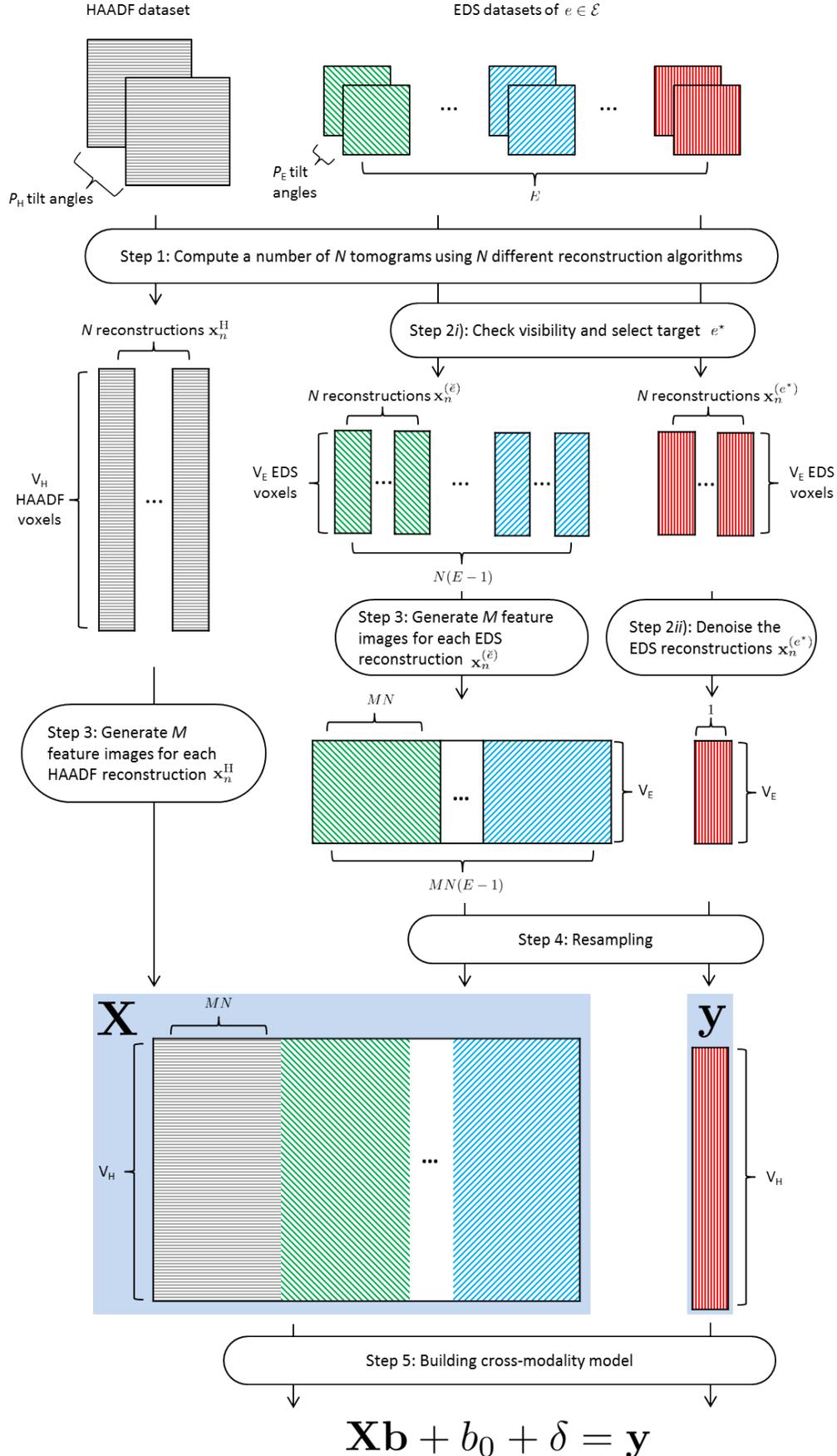


Fig. 1: The first five steps of fusion algorithm. Assume the sample of interest is composed of a set of chemical elements  $\mathcal{E}$ . Denote the target element chosen for fusion as  $e^*$ , and the rest  $\tilde{e} : e \in \mathcal{E} \setminus e^*$ . H symbolizes HAADF and E EDS. Details in Section III.

### C. Step 3: Generating Feature Images

The third step aims to generate more information for building the model. In particular, feature images of the  $N$  HAADF and  $N(E - 1)$  EDS reconstructions (with  $e^*$  being excluded) are generated with several texture filters. In our case, we use the entropy, range, standard deviation and Gaussian gradient magnitude filters; Table I lists the filter parameters. More filters could be included and provide the model with more feature images, namely higher degrees of freedom, to describe the denoised EDS reconstruction; however, they may simultaneously introduce unwanted patterns. For example, an orientation filter may reproduce undesirable star-shaped streaks in the fusion result. One can also extend the filtering operation to various scale spaces, and we choose a two-level Gaussian scale space. Note that more filters and deeper scale spaces would lead to higher computational costs both in time and memory.

TABLE I: Filters and related parameters in step (3)

Name of filter	Parameter
Local entropy	$3 \times 3 \times 3$ neighborhood
Local range	$3 \times 3 \times 3$ neighborhood
Local standard deviation	$3 \times 3 \times 3$ neighborhood
Gaussian gradient magnitude	Standard deviation $\sigma = 1$

### D. Step 4: Resampling

EDS reconstruction of  $e^*$  denoised in step (2) and feature images generated in step (3) should be of the same size for establishing the final model. Similarly, if the spectral images have been binned, we use the trilinear interpolation to upsample all EDS-related images at this stage.

### E. Step 5: Building Cross-modality Model

We formulate the cross-modality modeling as a linear regression task

$$\mathbf{y} = \mathbf{X}\mathbf{b} + b_0 + \delta \quad (2)$$

where  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_P]$  and  $\mathbf{b} = (b_1, \dots, b_P)^T$  with  $P = M \cdot N \cdot E$ . Each predictor variable  $\mathbf{x}_p$  denotes one (vectorized) feature image, and the response  $\mathbf{y}$  is the denoised EDS reconstruction of  $e^*$ .  $b_p$ ,  $p = 0, \dots, P$  are the scalar regression coefficients to be found, and  $\delta$  is the mismatch term. We employ partial least squares (PLS) regression, which is different from the ordinary least squares that directly establishes a linear model in the original data space. By first performing PCA to project both the predictor and response variables to  $N_{\text{comp}}$  components in another space [28], PLS can produce stable results with low variability even if the correlation among predictor variables is high. Since the variance explained in the response variable  $\mathbf{y}$  increases with the number of PLS components  $N_{\text{comp}}$ , we set  $N_{\text{comp}}$  to its maximum value  $N_{\text{comp}} = P - 1$  and solve this regression problem by `plsregress()` in MATLAB. Once we find all coefficients  $b_p$ , we fuse the image for  $e^*$  as  $\mathbf{x}_f^{(e^*)} = b_0 + b_1\mathbf{x}_1 + \dots + b_P\mathbf{x}_P$ , which is guaranteed to be the closest to the denoised EDS reconstruction  $\mathbf{x}_d^{(e^*)}$ .

### F. Step 6: Evaluating Reliability

Lastly, we evaluate the reliability of the fusion result, considering that PLS will always build a model regardless of whether there is a linear relationship, and that the PLS output is actually a prediction only. Specifically, we quantify the correspondence between the fused image  $\mathbf{x}_f^{(e^*)}$  and the denoised  $\mathbf{x}_d^{(e^*)}$  using the proportion of variance explained and the (Pearson) correlation coefficient. The former is returned by `plsregress()`, which is the ratio between the sum of squares of  $\mathbf{y}$  explained by the  $N_{\text{comp}}$  PLS components and the total sum of squares of  $\mathbf{y}$  [33]. The latter is calculated as

$$\text{CC} = \frac{\sum_i (x_{f,i} - \bar{x}_f)(x_{d,i} - \bar{x}_d)}{\sqrt{\sum_i (x_{f,i} - \bar{x}_f)^2 \sum_i (x_{d,i} - \bar{x}_d)^2}}, \quad (3)$$

which measures how well the relative intensity distribution of  $\mathbf{x}_f$  matches  $\mathbf{x}_d$  [17].  $x_{f,i}$  and  $x_{d,i}$  are the intensity values of the  $i$ -th voxel,  $\bar{x}_f$  and  $\bar{x}_d$  the average intensities over all voxels in  $\mathbf{x}_f$  and  $\mathbf{x}_d$ , respectively.

## IV. EXPERIMENTS

Based on simulated and experimental datasets, we investigated the performance of our method, and compared it to other state-of-the-art reconstruction techniques.

### A. Simulation and Experimental Datasets

We start with two noise-free multislice simulation datasets that were generated from two semiconductor models. The first

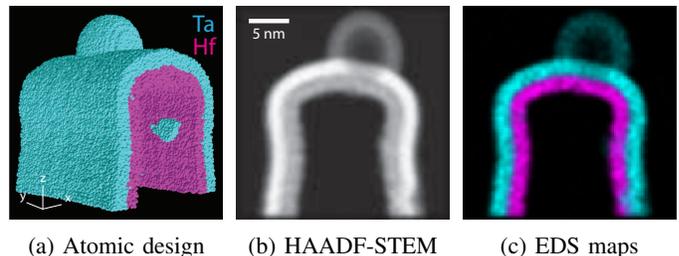


Fig. 2: (a) Atomic design of a defective finFET structure. Tantalum (Ta, cyan) layer and hafnium (Hf, magenta) in  $\text{HfO}_2$  layer are shown for clarity. (b) HAADF-STEM projection and (c) superposed EDS maps at  $2^\circ$ . Since the intensity of HAADF-STEM scales with the atomic number  $Z$ , only Ta ( $Z_{\text{Ta}} = 73$ ) and Hf ( $Z_{\text{Hf}} = 72$ ) are visible but O ( $Z_{\text{O}} = 8$ ) is not. Details in Section IV-A.

is a defective finFET structure with a size of  $25 \text{ nm} \times 25 \text{ nm} \times 25 \text{ nm}$  [34]. It has crystalline silicon (Si) as its source-drain fin, on top of which are oxygen (O), hafnium dioxide ( $\text{HfO}_2$ ), tantalum (Ta), and titanium aluminum nitride ( $\text{TiAlN}_2$ ). A few defects have been introduced, such as three pinholes in the  $\text{HfO}_2$  layer with diameters of 1, 2 and 3 nm allowing Ta to contact the fin. Moreover, a 7 nm ellipsoid carbon contaminant is trapped between the  $\text{HfO}_2$  and Ta layers. The atomic design of Ta and Hf is shown in Fig. 2(a). HAADF-STEM images were simulated with an accelerating voltage of 200 kV, a focused electron probe normalized to a total intensity of 1, a convergence angle of 10 mrad, and a detector with an inner

angle of 90 mrad and outer angle 230 mrad. For details of the simulation see [34]. X-ray maps were generated by summing up the probability of characteristic emission. The raw tilt-series of this finFET structure consists of 180 projections, ranging from  $0^\circ$  to  $358^\circ$  with an increment of  $2^\circ$  between consecutive projections. We only used the first 90 (unique) projections for reconstruction; each contains one HAADF-STEM image with a size of  $128 \text{ pixel} \times 128 \text{ pixel}$  ( $\approx 2\text{\AA}/\text{pixel}$ ) and eight elemental EDS maps of the same size and resolution. The second model is a larger region of PMOS with a size of  $70 \text{ nm} \times 70 \text{ nm} \times 70 \text{ nm}$ . Besides two Si fins, Ta metal gate and  $\text{HfO}_2$  layer, a tungsten (W) contact was also added, see Fig. 3(a). Projections of this PMOS device were simulated over  $[0^\circ, 180^\circ)$  in every  $2.5^\circ$  with the same parameters set for the previous case. At each angle, one HAADF-STEM image with a size of  $256 \text{ pixel} \times 256 \text{ pixel}$  ( $\approx 3.8\text{\AA}/\text{pixel}$ ) and the equi-sized EDS maps of all chemical elements were recorded.

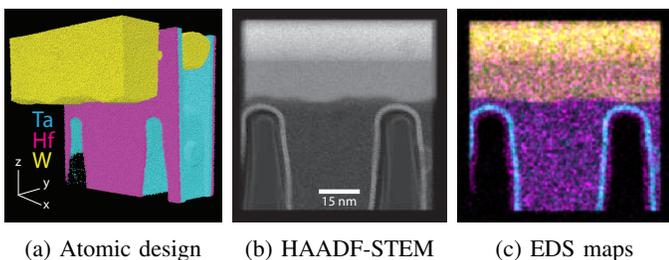


Fig. 3: (a) Atomic design of a PMOS model where tantalum (Ta, cyan), hafnium (Hf, magenta) and tungsten (W, yellow) are shown. (b) HAADF-STEM projection and (c) superposed EDS maps at  $2.5^\circ$ . Ta ( $Z_{\text{Ta}} = 73$ ), Hf ( $Z_{\text{Hf}} = 72$ ) and W ( $Z_{\text{W}} = 74$ ) are clearly visible in the HAADF-STEM image. Details in Section IV-A.

To produce projections that are more comparable to real experimental data, we performed some post-processing steps on the simulated tilt-series. For HAADF-STEM images, we first applied Gaussian smoothing ( $\sigma = 1.0 \text{ pixel}$ ) to simulate a less focused lens system. Then, we added Poisson noise with a mean value of the HAADF intensity, and Gaussian noise with a standard deviation of 0.2 to corrupt the noiseless dataset. Finally, we excluded zone-axis projections ( $0^\circ$  and  $90^\circ$ ) because of the significant channeling effect present, as is custom in practice [35]. Compared to HAADF-STEM images, X-ray counts collected in the spectral images are usually fewer (max. 30 per pixel [4]), resulting in much noisier EDS maps. Therefore, we employed a Gaussian filter ( $\sigma = 1.0 \text{ pixel}$ ) for denoising. Moreover, since the number of EDS projections is always smaller than the STEM projections (due to time) in real experiments [31], we subsampled the two EDS tilt-series by factors of 3 and 2, respectively, assuming that they were recorded in every  $6^\circ$  and  $5^\circ$  over  $[0^\circ, 180^\circ)$ . Post-processed projections of the defective finFET structure and PMOS model are shown in Fig. 2 and Fig. 3.

Our experimental dataset is a pillar-shaped semiconductor device comprised of eight chemical elements: N, O, Al, Si, Ti, Hf, Ta and Co [36]. It was placed on a Fischione on-

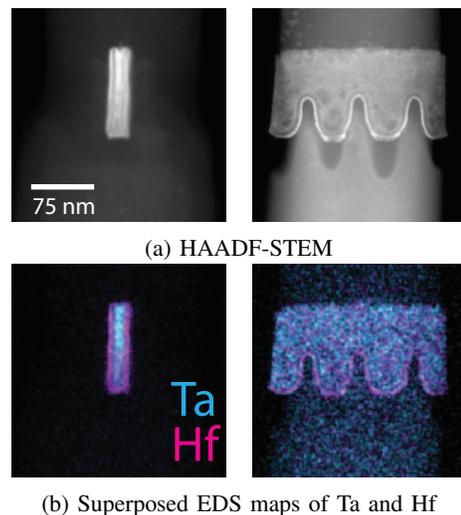


Fig. 4: Experimental (a) HAADF-STEM projections and (b) EDS maps of a pillar-shaped semiconductor device at  $0^\circ$  (left) and  $90^\circ$  (right). Titanium with an atomic number of  $Z_{\text{Ti}} = 22$  is surrounded by tantalum (cyan,  $Z_{\text{Ta}} = 73$ ) and hafnium (magenta,  $Z_{\text{Hf}} = 72$ ). Details in Section IV-A.

axis rotation tomography holder allowing a  $360^\circ$  projection acquisition, and scanned in a FEI Titan electron microscope equipped with four Super-X energy dispersive silicon drift detectors. The microscope was operated at an accelerating voltage of 120 kV with a beam current of 280 pA and a convergence angle of 10 mrad. HAADF-STEM projection images were acquired at 221 angles uniformly distributed between  $0^\circ$  and  $220^\circ$ . In addition, 47 full spectral images were recorded from  $0^\circ$  to  $216^\circ$  in approximately every  $5^\circ$ ; each had a constant acquisition time of 270 s. They were later deconvolved into eight EDS maps corresponding to the eight aforementioned components in the sample. An example of the experimental tilt-series is given in Fig. 4.

### B. Fusion Framework

Besides the widely used reconstructions via FBP and SIRT, we also added SART and ML-EM to our fusion framework. The former combines the best of ray-by-ray ART and all-inclusive SIRT, and can yield reconstructions of good quality along with high numerical accuracy in only a few iterations; the latter assumes the Poisson nature of electron/photon noise (rather than Gaussian) and inherently includes the nonnegativity constraint. For the EDS datasets, we chose Hanning window (rather than Ram-Lak filter) for FBP to deemphasize high frequencies, and 50 iterations for SIRT to avoid reconstructions overfitting to the noise. Moreover, we set 5 and 50 iterations for SART and ML-EM, respectively. For the less-noisy HAADF datasets, the number of iterations used for all iterative algorithms were doubled. Although we only incorporated four reconstruction techniques, we stress that our framework is general, and algorithms can be easily added (or removed) if necessary. Throughout this paper, visibility weights were calculated using the SIRT reconstructions of EDS and HAADF.

### C. Tilt and X-ray Count Settings

We used the simulated datasets to study the fusion accuracy of our method for two cases: “limited-tilt” and “low-count”. Compared to HAADF-STEM images, EDS spectra usually require much longer acquisition time at each tilt angle to achieve an acceptable SNR. Consequently, either case can effectively reduce the amount of time for nanomaterials being exposed to the electron beam and hence limit the sample damage. For the “limited-tilt” case, we reduced the number of EDS elemental maps in the two datasets from 30 and 36 to 8 and 9, respectively, guaranteeing that the remaining angles were uniformly distributed between  $[0^\circ, 180^\circ)$ . For the “low-count” case, we fixed the number of tilt angles for both EDS datasets to 30 and 36, and lowered their elemental X-ray counts by decreasing the maximal counts per pixel by three quarters. For the experimental dataset, we only considered the “limited-tilt” case by subsampling the original EDS tilt-series with 47 projections to 25 and 14, respectively.

### D. Benchmark Algorithms and Evaluation Metrics

We compared our proposed scheme to the classical FBP (with Hanning window) and SIRT (50 iterations), and two more advanced regularized approaches: TV and TNV, for which we set 200 and 400 iterations to guarantee convergence [15]. Moreover, since reconstruction qualities of TV and TNV highly depend on the regularization coefficient, we manually tuned their inputs for reasonable outputs, following the guidelines in [15].

We chose the structural similarity index (SSIM) and correlation coefficient to evaluate the reconstruction quality. SSIM is defined as

$$\text{SSIM}(f, g) = l(f, g) \times c(f, g) \times s(f, g) \quad (4)$$

where

$$\begin{aligned} l(f, g) &= \frac{2\mu_f\mu_g + C_1}{\mu_f^2 + \mu_g^2 + C_1} \\ c(f, g) &= \frac{2\sigma_f\sigma_g + C_2}{\sigma_f^2 + \sigma_g^2 + C_2} \\ s(f, g) &= \frac{\sigma_{fg} + C_3}{\sigma_f\sigma_g + C_3} \end{aligned}$$

which measures the similarity between the reconstructed image  $f$  and ground truth  $g$  in three aspects: luminance ( $l$ ), contrast ( $c$ ) and structure ( $s$ ) [37]. A higher SSIM value corresponds to a better reconstruction.  $\mu_f$  and  $\mu_g$  are the average intensity of  $f$  and  $g$ ;  $\sigma_f$  and  $\sigma_g$  are the standard deviation;  $\sigma_{fg}$  is the covariance between  $f$  and  $g$ . Moreover,  $C_1$ ,  $C_2$  and  $C_3$  are the constants introduced to avoid denominators being close to zero. By default,  $C_1 = (0.01L)^2$ ,  $C_2 = (0.03L)^2$  and  $C_3 = C_2/2$  with  $L$  denoting the dynamic range of  $f$  and  $g$  [37]. Note that  $s(f, g)$  reduces to the correlation coefficient in Eq. (3) if  $C_3 = 0$ . For the simulation datasets, the ground truth  $g$  was computed by SIRT with 100 iterations given the full-view noiseless (element-wise) EDS maps. We did not use the mask generated from the atoms’ coordinates as reference image because it does not involve the interaction between the atoms and the incident electron beam.

## V. RESULTS

### A. Simulated FinFET Dataset

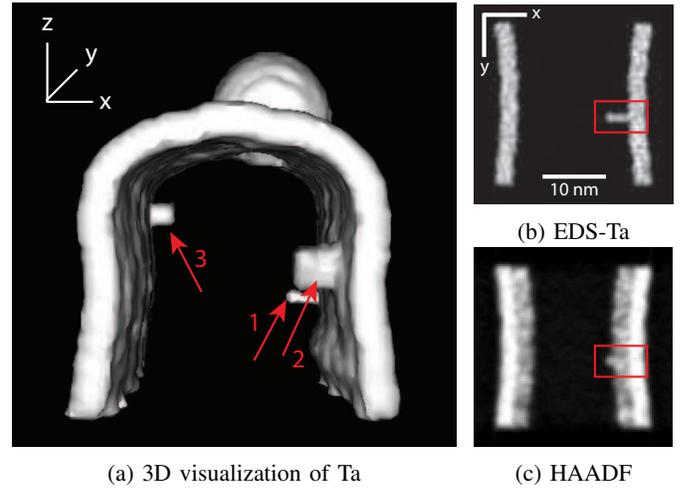


Fig. 5: (a) Volume rendering of the noiseless Ta (tantalum) reconstruction for the simulated finFET dataset; SIRT with 100 iterations using 90 elemental maps between  $[0^\circ, 180^\circ)$ . Three arrows point to the Ta penetration caused by the pinhole defects in the inner  $\text{HfO}_2$  layer. (b) and (c) are the Ta and HAADF reconstruction  $xy$ -slices at  $z = 37$ , corresponding to the first arrow in (a). Red rectangles indicate the defect.

For the first simulated dataset, we select Ta as the target  $e^*$ , which has the highest visibility weight  $w_{\text{Ta}} = 0.84$ . Fig. 5(a) shows a volume rendering of its noiseless reconstruction, in which penetrations (indicated by red arrows) result from pinhole defects in the inner  $\text{HfO}_2$  layer. We consider this 3D volume as the ground truth. Fig. 5(b) shows an orthoslice at location 1 and the red rectangle highlights a 1 nm defect. Since the atomic numbers of Ta and Hf are close ( $Z_{\text{Ta}} = 73$ ,  $Z_{\text{Hf}} = 72$ ), they yield similar  $Z$ -contrast in the HAADF reconstruction and make discrimination difficult, see Fig. 5(c).

Fig. 6 depicts the  $xy$ -slices of Ta reconstructions at  $z = 37$ , which are generated by FBP, SIRT, TV, TNV and our fusion algorithm. Images in the last column are the (NMF denoised) response variable  $\mathbf{y}$  in Eq. (2). In the first row, the number of used projections and the maximal X-ray counts per pixel in the EDS datasets are 30 and 20, respectively. We refer to this as the “normal” case for brevity because it is comparable to a typical experiment. Moreover, the second and third rows show the “limited-tilt” and “low-count” cases, in which either the number of projections or X-ray counts is reduced by three quarters.

Although FBP successfully reproduces the penetration defect in the normal case (Fig. 6(a), top row), it also introduces visible line artifacts in the background. This can be largely suppressed by SIRT (Fig. 6(b)). Tomograms from regularized TV and TNV are visually indistinguishable regarding the continuity of edges and smoothness of foreground. Our algorithm cannot only get rid of background noise as TV and TNV, but also keep the structural details of the Ta layer (e.g., rectangular ends) that are somewhat smeared by the other two techniques.

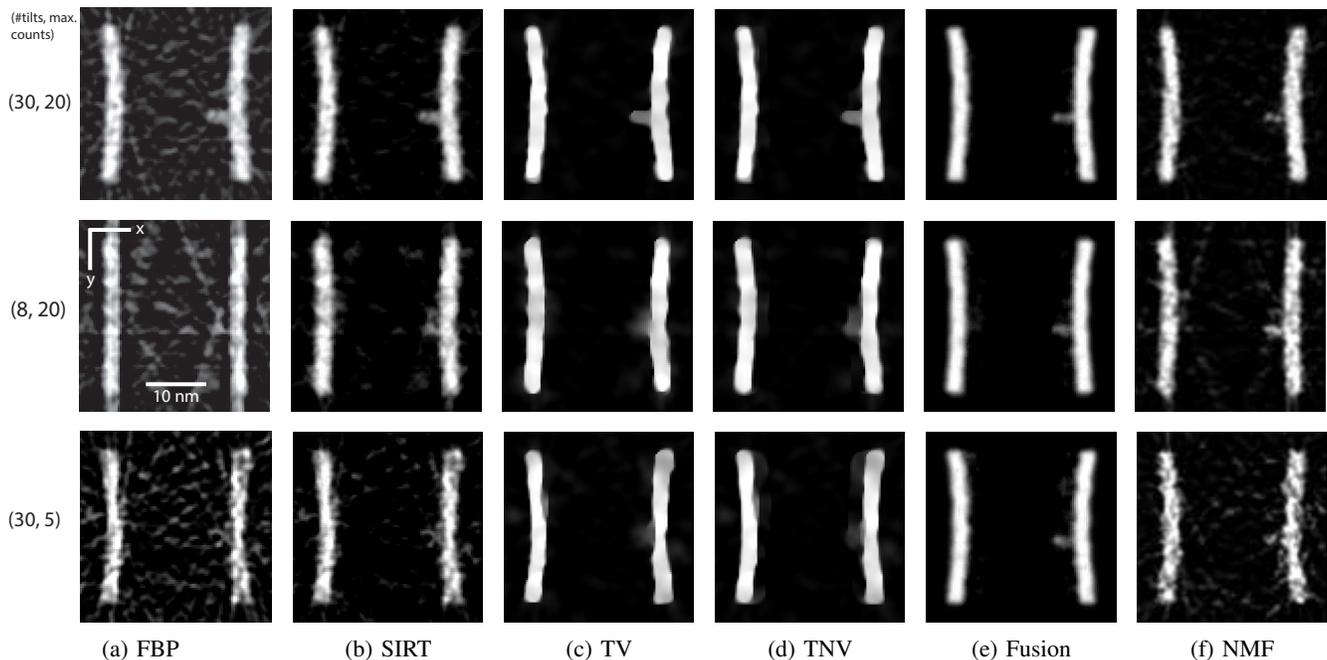


Fig. 6: Ta  $xy$ -slices of (a) FBP, (b) SIRT, (c) TV, (d) TNV, (e) fusion and (f) NMF denoising for the simulated finFET at  $z = 37$  (the height of the first defect in Fig. 5(a)). The top row is the “normal” case where the number of EDS projections and the maximal X-ray counts per pixel are 30 and 20. Middle and bottom rows are the limited-tilt case (with 8 projections) and low-count case (max. 5 X-ray counts). For better visualization, we perform percentile contrast stretching from 0 to 99%.

When the number of projections is reduced from 30 to 8 (Fig. 6, middle row), FBP, SIRT and NMF deform the curved Ta layer because of the limited tilts. Regularized TV and TNV manage to retain its shape but simultaneously introduce jaggy edges and paintbrush artifacts as both encourage piecewise constant structures [15]. However, TNV better preserves the rectangular ends of the Ta layer thanks to the augmentation from HAADF-STEM projections. The image computed by our algorithm is visually pleasing due to its continuous edges and homogeneous foreground. Additionally, it maintains the curved Ta layer to some extent even though it highly relies on the NMF denoised reconstruction. The tiny defect can be seen in TV, TNV, NMF and our fusion results, whereas it is distorted from a tip to a bulb in the former two reconstructions and contaminated by streak artifacts in the third. For the low-count case (Fig. 6, bottom row), classical FBP and SIRT, and the resulting NMF create a lot of undesirable lines in their fore- and background due to the low signal. To significantly restrain such artifacts, we set high regularization coefficients  $\lambda$  for TV ( $\lambda_{TV} = 0.045$ ) and TNV ( $\lambda_{TNV} = 0.1$ ), but inevitably distort the shape of the Ta layer and fail to satisfactorily reconstruct the defect. Although the penetration in our fused image is mixed with noise along the edges, one can still see it thanks to its relatively high intensity (at least three times higher than the noise). From top to bottom conditions in Fig. 6, the corresponding proportion of variance explained and correlation of the fusion to the denoised image are (0.95, 0.98), (0.91, 0.96) and (0.83, 0.91), respectively.

Table II summarizes the SSIM and correlation coefficient (CC) values of the five aforementioned algorithms on the three

settings in Fig. 6. Different noise realizations do not change the values in Table II to the shown digits. Our scheme ranks the best in all three cases. FBP falls far behind in SSIM due to the line artifacts visible in Fig. 6. The two regularized techniques achieve similar quantitative performance, but are surpassed by SIRT in terms of CC because regularizations may oversmooth the underlying structure and make reconstructions less accurate. Note that our fusion algorithm can produce stable results regardless of limited or noisy datasets, as demonstrated in Fig. 6 and Table II.

TABLE II: Comparison of SSIM and correlation coefficient (CC) for Ta reconstruction of simulated finFET dataset

(#tilts, counts)	Metrics	FBP	SIRT	TV	TNV	Fusion
(30, 20)	SSIM	0.33	0.81	0.87	0.88	<b>0.96</b>
	CC	0.95	0.97	0.90	0.90	<b>0.98</b>
(8, 20)	SSIM	0.29	0.79	0.85	0.86	<b>0.95</b>
	CC	0.80	0.95	0.88	0.89	<b>0.98</b>
(30, 5)	SSIM	0.22	0.70	0.80	0.80	<b>0.95</b>
	CC	0.86	0.93	0.90	0.91	<b>0.98</b>

### B. Simulated PMOS Dataset

The noiseless Ta reconstruction of the simulated PMOS dataset is rendered in 3D in Fig. 7(a), and its  $xy$ -slice at  $z = 94$  depicted in Fig. 7(b). This uniformly distributed element ranks second in the visibility check, surpassed by W ( $w_{Ta} = 0.42$  and  $w_W = 0.85$ ). Fig. 7(c) is the associated HAADF reconstruction where thin Ta/Hf layers, and two Si

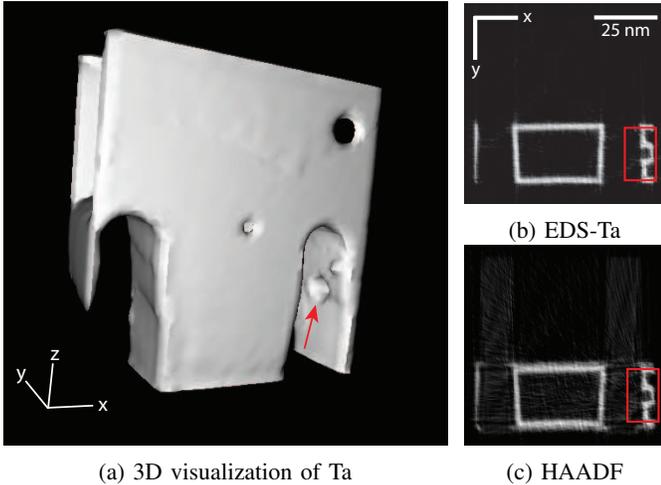


Fig. 7: (a) Volume rendering of the noiseless Ta reconstruction for the simulated PMOS dataset; SIRT with 100 iterations using 72 elemental maps between  $[0^\circ, 180^\circ]$ . (b) and (c) are the Ta and HAADF reconstruction  $xy$ -slices at  $z = 94$ , in which red rectangles encompass a fine structure indicated by the arrow in (a).

fins are visible. Red rectangles encompass a fine structure that is pointed out by the red arrow in Fig. 7(a).

Fig. 8 shows the corresponding  $xy$ -slices reconstructed by all techniques. As before, the top row is the normal case where the number of projections and the maximal X-ray counts per pixel in the EDS tilt-series are 36 and 20, respectively; the middle row is the limited-tilt case using only 9 projections, and the bottom the low-count case with no more than 5 X-ray counts per pixel. As illustrated in Fig. 8(a), background of FBP reconstructions are severely contaminated by lines if the sample of interest contains laminate. SIRT can suppress such artifacts but fails to deliver clear edges. Despite that TV and TNV achieve cleaner background than FBP and SIRT, they also noticeably smear the Ta layer especially for the limited-tilt and low-count cases: rectangular corners are rounded; laminate becomes thicker; line segments on the left are much shorter than they are supposed to be (see Fig. 7(b)). In general, such smearing effects are more pronounced in TV than TNV because the latter can incorporate the edge information from HAADF-STEM. Our fusion algorithm can generate a volume that is qualitatively comparable to the ground truth for the normal case (see Fig. 7(b) and Fig. 8(e)). In the other two cases, fine structures are still kept intact even though they are corrupted by noise in the regression targets; however, line artifacts in the background (with intensities no more than 15% of the foreground) show up. Quantitative results in Table III confirm this, in which our algorithm is in the top rank except for SSIM in the limited-tilt case. Outputs of the reliability validation for the aforementioned three cases are (variance explained, correlation): (0.86, 0.94), (0.74, 0.88) and (0.61, 0.8), respectively. The low variance explained for the low-count case is due to the noisy NMF, but our fusion remains relatively smooth and clear.

TABLE III: Comparison of SSIM and correlation coefficient (CC) for Ta reconstruction of simulated PMOS dataset

(#tilts, counts)	Metrics	FBP	SIRT	TV	TNV	Fusion
(36, 20)	SSIM	0.48	0.86	<b>0.91</b>	<b>0.91</b>	0.90
	CC	0.92	0.97	0.94	0.94	<b>0.98</b>
(9, 20)	SSIM	0.37	0.80	<b>0.90</b>	<b>0.90</b>	0.83
	CC	0.68	0.92	0.90	0.92	<b>0.93</b>
(36, 5)	SSIM	0.25	0.76	0.86	0.85	<b>0.88</b>
	CC	0.80	0.92	0.90	0.91	<b>0.96</b>

### C. Experimental Dataset

Fig. 9 depicts an orthoslice of the HAADF reconstruction for the experimental dataset, which clearly reveals the structural information of this pillar-shaped semiconductor device. For example, the dark contrast highlighted by the second red arrow possibly results from a defect (void inside the laminate). However, since the innermost Ta layer and the outermost Hf yield similar  $Z$ -contrast, directly discerning them from the HAADF reconstruction without any chemical information would be challenging.

Fig. 10 illustrates the Ta (cyan) and Hf (magenta) reconstructions generated by FBP, SIRT, TV, TNV and our fusion algorithm using 47, 25 and 14 EDS projections; the last column shows the response variable  $y$  in Eq. (2). Visibility weights of Ta and Hf are 0.2 and 0.25, respectively, ranking behind Si with  $w_{\text{Si}} = 0.41$ . From top to bottom, the variance explained and the correlation to  $y$  are (0.88, 0.92), (0.84, 0.91) and (0.82, 0.90) for Ta, and (0.89, 0.93), (0.84, 0.91) and (0.82, 0.90) for Hf. When the number of projections is decreasing, thin streaks in FBP drastically degrade its reconstruction quality, because the structure under study aligns with the projection direction, see Fig. 10(a); moreover, the background of SIRT reconstructions is also obviously getting noisier. Note that such noise is much lower in the resulting NMF. TV and TNV are also able to suppress the background noise; however, they simultaneously smear the Ta and Hf layers and make their boundary almost indistinguishable, especially for the 14-tilt case. Despite the spotty foreground in NMF, our fusion algorithm still achieves the best performance in noise suppression and structure preservation (e.g., the squeezing pattern indicated by the first arrow in Fig. 9) regardless of the number of projections. Moreover, it also maintains the void observed in the HAADF reconstruction to some extent, which is almost unrecognizable in the other five reconstructions. Therefore, our method can enable easier and more accurate fault analysis in the subsequent process. As no ground truth is available for this experimental dataset, we cannot compute the SSIM and correlation coefficient.

## VI. DISCUSSION

In the previous section, we demonstrated that our fusion algorithm is more robust for the limited and noisy datasets than other state-of-the-art tomography techniques. For the simulated datasets, we reduced either the number of EDS elemental maps or the maximal X-ray counts per pixel to

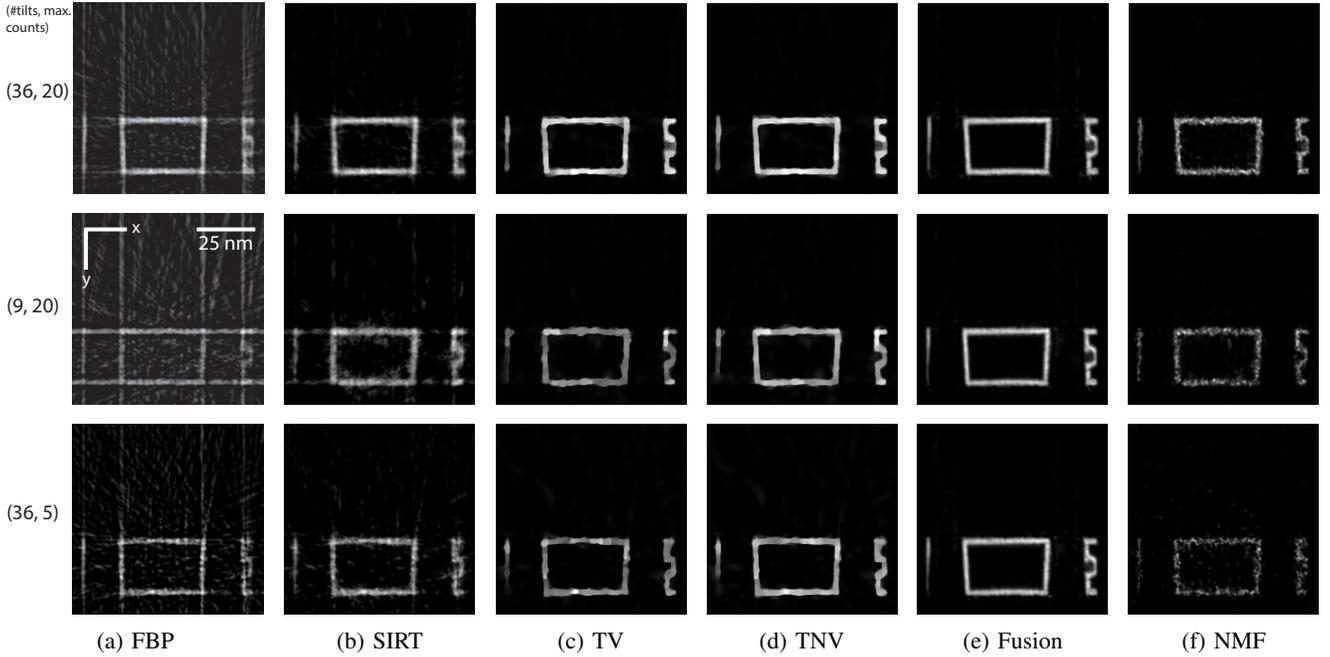


Fig. 8: Ta  $xy$ -slices of (a) FBP, (b) SIRT, (c) TV, (d) TNV, (e) fusion and (f) NMF denoising for the simulated PMOS at  $z = 94$  (the height of the arrow in Fig. 7(a)). The top row is the “normal” case where the number of EDS projections and the maximal X-ray counts per pixel are 36 and 20. Middle and bottom rows are the limited-tilt case (with 9 projections) and low-count case (max. 5 X-ray counts), respectively.

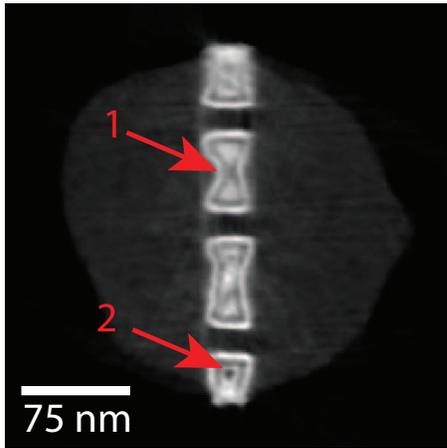


Fig. 9: An orthoslice of HAADF reconstruction for the experimental dataset. The first red arrow indicates a squeezing feature, and the second a dark contrast possibly results from some void inside the laminate.

one quarter of the initial values, and showed that our method could deliver stable outputs which were visually consistent with the ground truth of the Ta layer. We also fused Hf (for both finFET and PMOS) and W (only for PMOS), and found that the homogeneously distributed W resulted in visually better fusion. Because Hf was mixed with the light element O ( $Z_0 = 8$ ), its signals in the HAADF-STEM projections were weaker ( $w_{\text{Hf}} = 0.42$  for finFET and  $w_{\text{Hf}} = 0.18$  for PMOS). Consequently, fused Hf images were less smooth in the foreground and contaminated by the heavier Ta (and W) in the

background; such contamination could also be inferred from the low correlation to the denoised Hf (e.g., only 0.68 in the low-count case for PMOS) because this metric focuses on the relative pattern similarity [17]. For the experimental dataset, we subsampled the EDS tilt-series to simulate the “limited-tilt” scenario. Our method could reproduce fine structures and maintain clear boundaries between the Hf and Ta layers even if only 14 tilts were available.

Since both TV and TNV noticeably distort fine details in their reconstructions, we further tested HOTV [11] (online available [38]) on the simulated finFET dataset for reconstructing Ta. We found that the performance of HOTV (second- or third-order with regularization parameters as Eq. (15) in [11]) lies between SIRT and TV. That is, it could better recover the penetration defect in Fig. 5(b) but could not adequately suppress noise especially in the foreground. Consequently, desired fine features were mixed with (undesired) noise and hence did not stand out.

All datasets we used have no missing wedge (sample rotation was not limited within a certain range in the electron microscope due to mechanical constraints, e.g.,  $\pm 70^\circ$  [31]); these pillar-shaped samples and rotation holders will eventually replace the traditional tomography holders. However, we tested the robustness of the precedent 2D version [18] on the missing wedge artifact using an experimental dataset of a core-shell nanoparticle consisting of gold (Au) and silver (Ag), which only contained 31 tilts ranging from  $-75^\circ$  to  $+75^\circ$  with an increment of  $5^\circ$  between the consecutive projections. Qualitatively, our method outperformed HAADF-EDS bimodal tomography [14] in terms of the sharpness of edges and smoothness of fore- and background. Quantitatively,

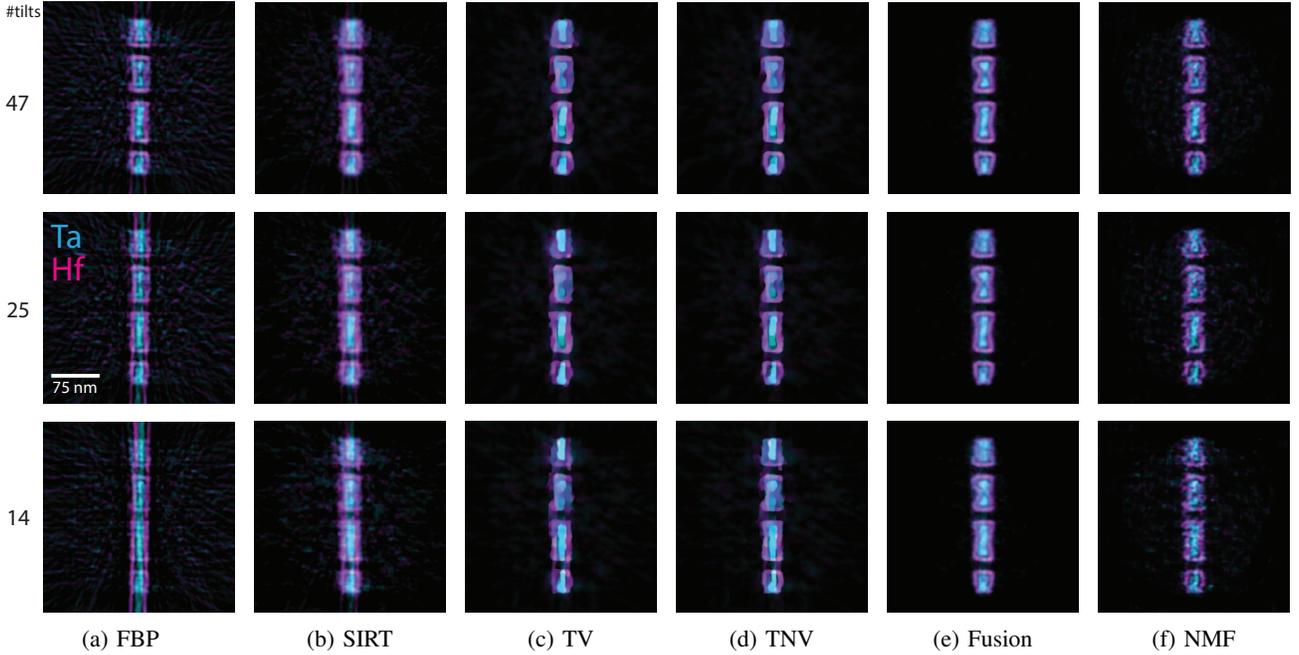


Fig. 10: Superposed  $xy$ -slices of Ta (cyan) and Hf (magenta) reconstructions computed by (a) FBP, (b) SIRT, (c) TV, (d) TNV, (e) fusion and (f) NMF denoising for the experimental dataset using 47 (top), 25 (middle) and 14 (bottom) EDS projections.

it achieved higher correlation coefficient between the reconstruction and the hand-segmented ground truth both for Au and Ag.

One prerequisite for our cross-modality framework to work is that chemical element chosen for fusion should be visible in the HAADF-STEM images; this is quantified by the visibility weight  $w_e$  in Eq. (2). Since light elements such as N and O never show up in the HAADF, we set the threshold for the target  $e^*$  three times as high as the weights of the light ones (e.g., O,  $3 \times w_0 = 0.15$  for finFET). This limitation may be overcome by incorporating other imaging modalities, such as bright-field or dark-field STEM with different detector geometries.

In addition, we also validated whether the resulting model is representative (i.e., whether fused images can be relied upon) through the proportion of variance explained and the correlation to the denoised image. For instance, if we try to fuse the carbon contaminant from Fig. 2 whose visibility weight is almost zero (even though its structure is clearly sketched by the surrounding  $\text{HfO}_2$  and Ta layers, see Fig. 2(b)), the corresponding variance explained is only 0.17 and correlation coefficient 0.5. As a rule of thumb, these two metrics should be above 0.6 and 0.8, respectively, to achieve a reliable fusion.

We further investigated how the choice of three popular 3D upsampling techniques: nearest-neighbor, trilinear and tricubic, influences the fusion quality. Since the resolution of EDS reconstructions decreases rapidly with an increased binning ratio  $r$ , we only considered  $r = 2$  with  $V_E = 1/8V_H$ . The nearest-neighbor interpolation produced the least favorable result; therefore, we left it out even though it has the highest speed. Moreover, we found that the trilinear and tricubic interpolation are comparable in terms of their visually indistinguishable

fusion results, and similar SSIM and CC values. We chose the trilinear because it is slightly faster than tricubic interpolation.

Computational time of our fusion algorithm is mainly spent on computing tomograms (step 1), generating feature images (step 2) and building cross-modality model (step 5). In our case, for instance, it took 300 s for tomographic reconstruction, 25 s for feature image generation, and another 300 s for cross-modal modeling for Ta in the simulated finFET structure using MATLAB 2017a on a desktop equipped with eight Intel Xeon X5550 CPU cores (24 GB memory) and NVIDIA GeForce GTX670 GPU (4 GB memory). Note that only the last 300 s (for building the model) are required to fuse any other element in this sample. Throughout this paper, FBP, SIRT, SART and ML-EM reconstructions were computed with the ASTRA toolbox [32]; regularized TV and TNV algorithms were realized by the Douglas-Rachford primal-dual splitting algorithm with the operator discretization library [39]. Although our method is slower than FBP (3 s), SIRT (12 s) and TV (273 s), it is still much faster than TNV ( $\sim 80$  mins); manually tuning regularization coefficients for TV and TNV to suppress noise yet avoiding over-regularization also consumes a lot of time. Moreover, considering that the acquisition time for each spectral image varies from 236 s to 895 s [31], our algorithm is still quite appealing.

## VII. CONCLUSION

In this paper, we have presented a regression-based cross-modal fusion framework for electron tomography, which does not require any fine-tuning parameter. We have adopted it to combine the EDS and HAADF reconstructions, and investigated its performance using simulated and experimental datasets of semiconductor devices that contain chemical elements with close atomic numbers. Results have shown that

our fusion algorithm can consistently yield more accurate reconstructions than the conventional yet widely employed FBP and SIRT, and the advanced regularization-based TV and TNV. Furthermore, it can still restore fine structures and achieve a high reconstruction quality even for limited and noisy EDS datasets. Such properties are highly desired in the semiconductor industry where the number of EDS maps is limited aiming for a shorter acquisition time, and the incident electron dose is usually low to minimize the sample damage.

#### ACKNOWLEDGMENT

The authors would like to thank Dr. Yang Qiu (IMEC, Leuven, Belgium—presently at South University of Science and Technology of China) and Dr. Hugo Bender (IMEC, Leuven, Belgium) for providing the experimental dataset of the semiconductor device, and Dr. Zhichao Zhong and Dr. Willem Jan Palenstijn (CWI, Amsterdam, The Netherlands) for the reconstruction codes of TV and TNV.

#### REFERENCES

- [1] M. Weyland and P. A. Midgley, "Electron tomography," *Materials Today*, vol. 7, no. 12, pp. 32–40, 2004.
- [2] A. J. Koster, U. Ziese, A. J. Verkleij, A. H. Janssen, and K. P. de Jong, "Three-dimensional transmission electron microscopy: a novel imaging and characterization technique with nanometer scale resolution for materials science," *Journal of Physical Chemistry B*, vol. 104, no. 40, pp. 9368–9370, 2000.
- [3] P. A. Midgley and M. Weyland, "3D electron microscopy in the physical sciences: the development of Z-contrast and EFTEM tomography," *Ultramicroscopy*, vol. 96, pp. 413–431, 2003.
- [4] A. Genc, L. Kovarik, M. Gu, H. Cheng, P. Plachinda, L. Pullan, B. Freitag, and C. Wang, "XEDS STEM tomography for 3D chemical characterization of nanoscale particles," *Ultramicroscopy*, vol. 131, pp. 24–32, 2013.
- [5] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. IEEE Press, 1988.
- [6] B. Goris, W. van den Broek, K. J. Batenburg, H. H. Mezerji, and S. Bals, "Electron tomography based on a total variation minimization reconstruction technique," *Ultramicroscopy*, vol. 113, pp. 120–130, 2012.
- [7] P. Ercius, O. Alaidi, M. J. Rames, and G. Ren, "Electron tomography: a three-dimensional analytic tool for hard and soft materials research," *Advanced Materials*, vol. 27, pp. 5638–5663, 2015.
- [8] M. Beister, D. Kolditz, and W. A. Kalender, "Iterative reconstruction methods in X-ray CT," *Physica Medica*, vol. 28, pp. 94–108, 2012.
- [9] K. J. Batenburg and J. Sijbers, "DART: a practical reconstruction algorithm for discrete tomography," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2542–2553, 2011.
- [10] X. Zhuge, W. J. Palenstijn, and K. J. Batenburg, "TVR-DART: a more robust algorithm for discrete tomography from limited projection data with automated gray value estimation," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 455–468, 2016.
- [11] T. Sanders, A. Gelb, R. B. Platte, I. Arslan, and K. Landskron, "Recovering fine details from under-resolved electron tomography data using higher order total variation  $l_1$  regularization," *Ultramicroscopy*, vol. 174, pp. 97–105, 2017.
- [12] T. Sanders and R. B. Platte, "Multiscale higher order TV operators for  $l_1$  regularization," *Advanced Structural and Chemical Imaging*, vol. 4, pp. 12–29, 2018.
- [13] D. Zanaga, T. Altantzis, L. Polavarapu, L. M. Liz-Marzán, B. Freitag, and S. Bals, "A new method for quantitative XEDS tomography of complex heteronanostructures," *Particle & Particle Systems Characterization*, vol. 33, 2016.
- [14] Z. Zhong, B. Goris, R. Schoenmakers, S. Bals, and K. J. Batenburg, "A bimodal tomographic reconstruction technique combining EDS-STEM and HAADF-STEM," *Ultramicroscopy*, vol. 174, pp. 35–45, 2017.
- [15] Z. Zhong, W. J. Palenstijn, J. Adler, and K. J. Batenburg, "EDS tomographic reconstruction regularized by total nuclear variation joined with HAADF-STEM tomography," *Ultramicroscopy*, vol. 191, pp. 34–43, 2018.
- [16] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: a survey," *Information Fusion*, vol. 45, pp. 153–178, 2019.
- [17] R. van de Plas, J. Yang, J. Spraggins, and R. M. Caprioli, "Image fusion of mass spectrometry and microscopy: a multimodality paradigm for molecular tissue mapping," *Nature Methods*, vol. 12, pp. 366–374, 2015.
- [18] Y. Guo and B. Rieger, "Image fusion of X-ray and electron tomograms," in *Proc. of IEEE International Conference on Image Processing*, October 2018, to be published.
- [19] D. S. Rigie and P. J. L. Riviere, "Joint reconstruction of multi-channel spectral CT data via constrained total nuclear variation minimization," *Physics in Medicine & Biology*, vol. 60, pp. 1741–1762, 2015.
- [20] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, 2017.
- [21] J. Sui, T. Adali, Q. Yu, J. Chen, and V. D. Calhoun, "A review of multivariate methods for multimodal fusion of brain imaging data," *Journal of Neuroscience Methods*, vol. 204, pp. 68–81, 2012.
- [22] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, 2011.
- [23] M. Ghahremani and H. Ghassemian, "Remote-sensing image fusion based on curvelets and ICA," *International Journal of Remote Sensing*, vol. 36, pp. 4131–4143, 2015.
- [24] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.
- [25] J. Wang, J. Peng, X. Feng, G. He, and J. Fan, "Fusion method for infrared and visible images by using non-negative sparse representation," *Infrared Physics & Technology*, vol. 67, pp. 477–489, 2014.
- [26] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS+Pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, 2007.
- [27] G. Vivone, R. Restaino, and J. Chanussot, "Full scale regression-based injection coefficients for panchromatic sharpening," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3418–3431, 2018.
- [28] R. D. Tobias, "An introduction to partial least squares regression," in *Proceedings of the Twentieth Annual SAS Users Group International Conference*, Cary, USA, 1995, pp. 1250–1257.
- [29] E. Martínez-Montes, P. A. Valdés-Sosa, F. Miwakeichi, R. I. Goldman, and M. S. Cohen, "Concurrent EEG/fMRI analysis by multiway partial least squares," *NeuroImage*, vol. 22, pp. 1023–1034, 2004.
- [30] A. A. Gowen and R. M. Dorrepaal, "Multivariate chemical image fusion of vibrational spectroscopic imaging modalities," *Molecules*, vol. 21, no. 7, pp. 870–888, 2016.
- [31] T. J. A. Slater, A. Janssen, P. H. C. Camargo, M. G. Burke, N. J. Zaluzec, and S. J. Haigh, "STEM-EDX tomography of bimetallic nanoparticles: a methodological investigation," *Ultramicroscopy*, vol. 162, pp. 61–73, 2016.
- [32] W. van Aarle, W. J. Palenstijn, J. D. Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers, "The ASTRA toolbox: a platform for advanced algorithm development in electron tomography," *Ultramicroscopy*, vol. 157, pp. 35–47, 2015.
- [33] S. de Jong, "SIMPLS: an alternative approach to partial least squares regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 18, pp. 251–263, 1993.
- [34] R. Aveyard and B. Rieger, "Tilt series STEM simulation of a 25x25x25 nm semiconductor with characteristic X-ray emission," *Ultramicroscopy*, vol. 171, pp. 96–103, 2016.
- [35] M. C. Scott, C. C. Chen, M. Mecklenburg, C. Zhu, R. Xu, P. Ercius, U. Dahmen, B. C. Regan, and J. Miao, "Electron tomography at 2.4-ångström resolution," *Nature*, vol. 483, pp. 444–447, 2012.
- [36] Y. Qiu, P. van Marcke, O. Richard, H. Bender, and V. Wilfried, "3D imaging of Si FinFETs by combined HAADF-STEM and EDS tomography," in *18th International Microscopy Congress*, Prague, Czech Republic, 2014, pp. 975–976.
- [37] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.
- [38] T. Sanders, "Open access codes for imaging problems," 2016. [Online]. Available: <http://www.toby-sanders.com/software>
- [39] J. Adler, H. Kohr, and O. Öktem, "Operator discretization library (ODL)," 2017. [Online]. Available: <https://github.com/odlgroup/odl>