

This item is the archived peer-reviewed author-version of:

Automatic correction of nonlinear damping effects in HAADF-STEM tomography for nanomaterials of discrete compositions

Reference:

Zhong Zhichao, Aveyard Richard, Rieger Bernd, Bals Sara, Palenstijn Willem Jan, Batenburg K. Joost.- Automatic correction of nonlinear damping effects in HAADF-STEM tomography for nanomaterials of discrete compositions
Ultramicroscopy - ISSN 0304-3991 - 184:B(2018), p. 57-65
Full text (Publisher's DOI): <https://doi.org/10.1016/J.ULTRAMIC.2017.10.013>
To cite this reference: <https://hdl.handle.net/10067/1485010151162165141>

Automatic Correction of Nonlinear Damping Effects in HAADF-STEM Tomography for Nanomaterials of Discrete Compositions

Zhichao Zhong^a, Richard Aveyard^b, Bernd Rieger^b, Sara Bals^c, Willem Jan Palenstijn^a, K. Joost Batenburg^{a,d}

^aCentrum Wiskunde & Informatica, Amsterdam, The Netherlands

^bDepartment of Imaging Physics, Delft University of Technology, The Netherlands

^cEMAT, University of Antwerp, Antwerp, Belgium

^dMathematical Institute, Universiteit Leiden, Leiden, The Netherlands

Abstract

HAADF-STEM tomography is a common technique for characterizing the three-dimensional morphology of nanomaterials. In conventional tomographic reconstruction algorithms, the image intensity is assumed to be a linear projection of a physical property of the specimen. However, this assumption of linearity is not completely valid due to the nonlinear damping of signal intensities. The nonlinear damping effects increase w.r.t the specimen thickness and lead to so-called “cupping artifacts”, due to a mismatch with the linear model used in the reconstruction algorithm. Moreover, nonlinear damping effects can strongly limit the applicability of advanced reconstruction approaches such as Total Variation Minimization and discrete tomography.

In this paper, we propose an algorithm for automatically correcting the nonlinear effects and the subsequent cupping artifacts. It is applicable to samples in which chemical compositions can be segmented based on image gray levels. The correction is realized by iteratively estimating the nonlinear relationship between projection intensity and sample thickness, based on which the projections are linearized. The correction and reconstruction algorithms are tested on simulated and experimental data.

1. Introduction

In materials science, electron tomography (ET) is commonly used to characterize the three-dimensional (3D) structural and compositional information of nanomaterials. The 3D image is usually reconstructed from a tilt series of two-dimensional (2D) projections (projection images). The projection images should have a monotonic relationship between the measurement intensity and the integrated physical property of the specimen, which is referred to as the *projection requirement* in ET [1, 2]. Strictly speaking, the relationship should be linear, as most tomographic reconstruction algorithms are based on a linear mathematical model – the line integral model. It assumes that the projection is a measurement of a physical property integrated along the projection orientation (see

Chapter 3 in [3]).

High angle annular dark field (HAADF) scanning transmission electron microscopy (STEM) is commonly used for ET [1, 4] under the implicit assumption that the projection requirement can be approximately satisfied. The image intensity approximates to be proportional to the mass-thickness weighted by $Z^{\sim 2}$, where Z is the atomic number [4]. However, this approximation is not always valid. One example is that when projections of a crystalline material are acquired at zone-axis orientations, fringes and large overall intensity differences can be observed. Thus the tilts at zone-axis are usually excluded from the tomographic reconstruction step [5]. Another example is that the image intensity damps at high sample thickness due to the multiple scattering events redirecting electrons

1 outside the annular detector, which can occur in all
2 projection orientations. While the zone-axis effects
3 can be easily identified, intensity damping is not
4 easily seen in individual projections. In this pa-
5 per, we aim at addressing the nonlinear effects of
6 intensity damping for tomographic reconstruction.

7 The consequence of intensity damping appears as
8 the cupping artifact in tomographic reconstruction:
9 the gray levels in the center of the reconstructed
10 sample are underestimated while overestimated on
11 the exterior [6]. In Fig. 1(a), an example of the
12 cupping artifact is given. It is a 2D cross section of
13 an Au-Ag core-shell nanoparticle [7], reconstructed
14 using the SIRT algorithm [8]. If we look at the
15 line-profile of the 2D image (Fig. 1(b)), the curve
16 appears in a concave “cup” shape, while ideally it
17 should be flat. The cupping artifacts are caused by
18 the strong damping effects of Au at large thickness,
19 which is illustrated by the simulated relationships
20 between measurement intensity and sample thick-
21 nesses using the multislice simulation method [5] in
22 Fig. 2. In this example, the linear approximation
23 is only valid for thickness smaller than 8 nm due to
24 the clear damping effect for larger thickness.

25 It is important to correct the nonlinear effects
26 and the subsequent cupping artifacts for three rea-
27 sons. First of all, compositional analysis based on
28 gray levels becomes difficult when the cupping arti-
29 facts occur, as gray levels are not proportional any-
30 more to density and atomic numbers. Second, mor-
31 phological analysis based on segmentation of recon-
32 struction images is hindered by the cupping arti-
33 facts. Some straightforward segmentation meth-
34 ods, e.g. Otsu’s method [9], require that for each
35 chemical composition there should be one constant
36 gray level. Third, the nonlinear effects limit apply-
37 ing advanced reconstruction algorithms to address
38 a critical issue of ET – the missing wedge artifacts
39 caused by the limited tilt range of the sample. Al-
40 gorithms such as total variation minimization [10]
41 reduce the missing wedge artifacts by incorporating
42 prior knowledge i.e. sparsity of the unknown sam-
43 ple. Nevertheless, these algorithms have an even
44 stronger requirement for the linear forward model
45 which is inaccurate due to the nonlinear effects.

46 Despite these shortcomings of using uncorrected
47 data, there are few publications addressing the non-
48 linearity issue in ET [5, 6]. Nonlinear effects are

1 usually ignored or mitigated during image acquisi-
2 tion by increasing the inner angle of the HAADF
3 detector but at the cost of losing signal strength
4 [5]. An alternative to adjusting the acquisition pa-
5 rameters is to correct the measured data in a post-
6 processing step by linearizing the projection data,
7 provided that the incident beam intensity is known
8 [6]. The method described here requires only the
9 HAADF signal, consequently, it can be applied to
10 correct cupping artifacts in many existing datasets
11 acquired in a conventional manner. The mathe-
12 matical model of nonlinearity and the concept of
13 linearization in [6] are also used in this paper, which
14 will be explained in Section 2.1.

15 Here, we propose an iterative algorithm to au-
16 tomatically correct the nonlinear effects and the
17 cupping artifacts. It does not require the extra
18 measurement of the incident beam intensity as in
19 [6]. Instead, it automatically models the nonlinear
20 effects given the projection data. The algorithm
21 iteratively searches for the minimal distance be-
22 tween the acquired projections and the nonlinear
23 re-projections of chemical compositions by varying
24 the nonlinear model and the reconstruction image,
25 so as to estimate a nonlinear relationship between
26 the measured HAADF-STEM intensities and sam-
27 ple thickness for all chemical compositions. The
28 algorithm contains the following steps in every it-
29 eration: first a reconstruction image with contin-
30 uous gray levels is made; then the image is seg-
31 mented into several binary images, each of which
32 corresponds to a chemical composition; after that,
33 the nonlinear effects are modeled by minimizing the
34 projection distance; based on the model, the pro-
35 jection data is linearized at last. The concept of
36 iterative correction has been used to correct beam
37 hardening artifacts for X-ray computed tomogra-
38 phy, which is similarly caused by nonlinear intensi-
39 ties [11, 12, 13].

40 Our approach is only applicable to samples con-
41 sisting of several chemical compositions with uni-
42 form densities, such as homogeneous or core-shell
43 particles. It is assumed that for these samples the
44 volumetric distributions of the compositions can be
45 approximated well by segmenting the reconstructed
46 image based on gray levels and that this segmen-
47 tation improves as the correction model applied to
48 the measured data becomes more accurate. In fact,
49 these kinds of samples are commonly studied in

1 materials science. For example, the samples typi-
 2 cally studied in the context of discrete tomography
 3 [14, 15] match the requirements.

4 In Section 2, the correction algorithm is ex-
 5 plained in detail. In section 3, we demonstrate
 6 how the nonlinear effects are corrected using this
 7 algorithm for real experimental data and phantom
 8 simulations.

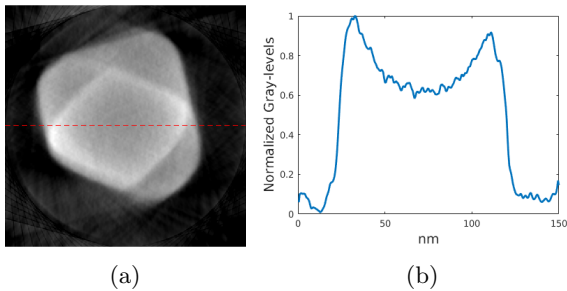


Figure 1: (a): 2D slice of the SIRT reconstruction of an Au-Ag nanoparticle. (b): Gray levels of the line-profile located at the dash line of the 2D slice.

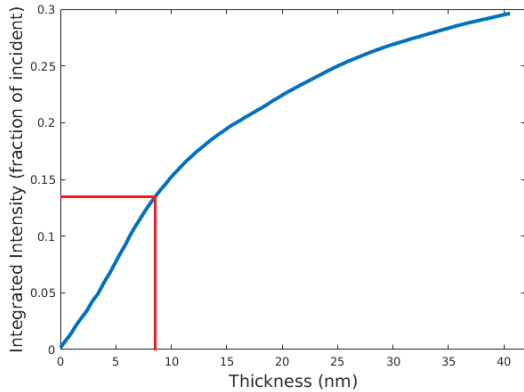


Figure 2: Normalized HAADF signal intensity w.r.t the thickness of Au slabs mistilted 10 degrees from the [100] zone axis about the $\langle 100 \rangle$ axis, simulated using the multislice method [5]. The accelerating voltage is 200 kV, the convergence angle is 10 mrad and the detector angular range is 50 - 250 mrad. The intensities are scaled by the incident beam intensity. The red lines indicate the region where intensity is approximately linear to thickness.

2. The Nonlinear Model and the Correction Algorithm

2.1. The Nonlinear Model

To linearize the projections, we first need to define a model that describes the nonlinear relationship mathematically. A precise mathematical model is possible but does not fit as a subroutine of the correction algorithm. The computation of a sophisticated model, such as the one used in multi-slice simulations which take into consideration the multiple scattering of electrons [5], is extremely time-consuming and costly. Therefore, a simple model is preferred here.

Here, we choose a model that has already been used for describing the nonlinear relationship. In [6, 16], it is illustrated we can assume that the HAADF detector collects electrons complementary to the electrons scattered to angles smaller than its inner detector angle. The electrons can also be scattered to angles beyond the outer detector angle, but the proportion is negligibly small. By pragmatically applying a simple Beer-Lambert description of electron scattering we can state that the number of electrons scattered to small angles p_t decreases exponentially to the sample thickness t along the beam direction. The p_t - t relationship is

$$p_t = I_0 \exp\left(-\sum_e^K \mu_e t\right), \quad (1)$$

where I_0 is the incident beam intensity, e is the index of chemical composition, K is the total number of chemical compositions, μ_e is the attenuation coefficient of chemical composition e . Therefore, the complementary HAADF signal intensity p at sample thickness t is:

$$p = I_0(1 - \exp(-\sum_e^K \mu_e t)) + p_b, \quad (2)$$

where p_b is the bias signal, which is influenced by the dark current, carbon grid, and possibly other factors.

This mathematical model has been used to correct the cupping artifacts successfully in [6], which is applicable only if the incident beam intensities can be measured. An advantage of this simple

1 model is that it can easily be transformed into a
 2 linear relationship by taking logarithms so that we
 3 can avoid solving nonlinear least-squared problems
 4 for tomographic reconstruction.

5 In the practice of ET, a series of projections are
 6 taken at different angles. The image intensity of
 7 each pixel corresponds to the electrons scattered for
 8 an electron beam transmitting through the sample,
 9 which is called a line projection here. In total, there
 10 are M pixels for all the images. The image intensity
 11 of the i_{th} pixel is now written as an entry p_i in
 12 $\mathbf{p} \in \mathbf{R}^M$. In addition, the space of reconstruction
 13 is a cubic volume partitioned into N voxels.

14 We also assume the chemical compositions are
 15 not mixed and voxels are small enough to resolve
 16 every chemical composition, which means that in
 17 each voxel only one element is present. As stated in
 18 the introduction, this algorithm is applied to sam-
 19 ples with uniform density. Thus we assume that
 20 each chemical composition is either present (1) or
 21 absent (0) in each voxel. The distribution of chem-
 22 ical composition e is described by binary variables
 23 s_{ej} , where $j = 1, \dots, N$ is the index of voxel.

24 Now we define the nonlinear relationship in the
 25 discrete form. For pixel i , the corresponding sample
 26 thickness of chemical composition e is now written
 27 as the ray-sum $\sum_{j=1}^N w_{ij}s_{ej}$, where the factor w_{ij}
 28 is determined by the area of intersection between
 29 the i_{th} line projection and the j_{th} voxel. The rela-
 30 tionship between projection intensities and binary
 31 volumes are:

$$p_i = I_0(1 - \exp(-\sum_{e=1}^K \mu_e \sum_{j=1}^N w_{ij}s_{ej})) + p_b, \quad (3)$$

32 where $i = 1, \dots, M$.

33 2.2. The Correction Algorithm

34 The basis of the correction algorithm is to esti-
 35 mate the nonlinear relationship of Eq. 3 based on
 36 the reconstructed distributions of chemical compo-
 37 sitions. The procedures of the automatic correc-
 38 tion algorithm are given in the flowchart (Fig. 3).
 39 The correction is realized iteratively through the
 40 following steps: (1) make a reconstruction image
 41 based on the linear model from the projections; (2)
 42 segment the reconstruction into a series of binary

1 images, one for each chemical composition; (3) esti-
 2 mate the parameters of the nonlinear model in Eq.
 3 3 given the projections and the binary images; (4)
 4 reduce the nonlinearities in the projections through
 5 a rescaling of the intensities based on the nonlinear
 6 model.

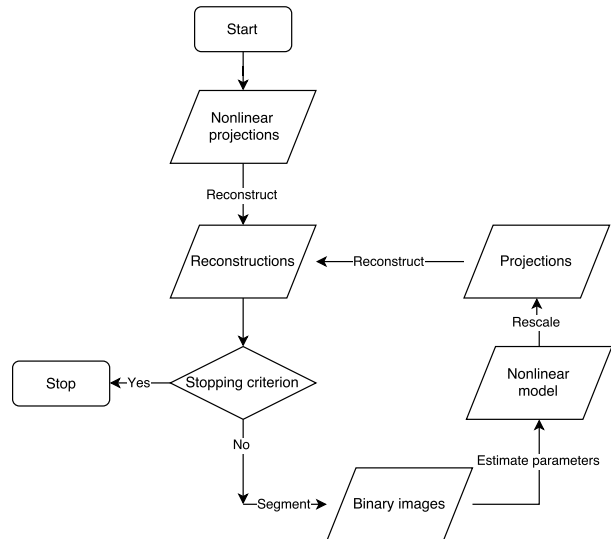


Figure 3: Flowchart of the correction algorithm

Before we explain the steps explicitly, we estab-
 lish an objective function which will be used to
 guide the optimization in the correction algorithm.
 We define it as the l_2 norm of the distance between
 the acquired projections and the re-projection of
 binary images based on our nonlinear model:

$$\mathcal{C}(I_0, p_b, \boldsymbol{\mu}, \mathbf{S}) = \|\mathbf{p} - I_0(1 - \exp(-\mathbf{W} \sum_{e=1}^K \mu_e \mathbf{s}_e)) - p_b\|_2^2, \quad (4)$$

where $\mathbf{W} = \{w_{ij}\}$, $\boldsymbol{\mu} = \{\mu_e\}$ and $\mathbf{S} = \{s_{ej}\}$.

We also define a stopping criterion. The cost
 value at the r_{th} iteration is denoted as the c^r . The
 loop is terminated if the cost is stable, which is
 when the following criterion is met:

$$\frac{c^r + c^{r-1}}{c^{r-2} + c^{r-3}} > t, \quad (5)$$

where $0 < t < 1$ is a thresholding value. Note that
 although we minimize the cost function in some
 steps of the algorithm, the cost is not guaranteed
 to reach a global minimum in the end.

1 *Step 1: Reconstruction*

2 As the first step, a reconstruction with continu-
 3 ous gray levels is made for determining the binary
 4 images in the next step. Though it is possible to
 5 reconstruct binary images directly using some dis-
 6 crete tomography algorithms (e.g. [14]), these algo-
 7 rithms will possibly not give better results than ba-
 8 sic algorithms given an inaccurate forward model.
 9 Thus, we choose to first make a reconstruction \mathbf{x}
 10 with continuous gray levels based on a linear model
 11 and then segment the reconstruction into binary
 12 images \mathbf{S} .

13 The reconstruction is computed using the simul-
 14 taneous iterative reconstruction technique (SIRT)
 15 [8] which solves the following least-squares prob-
 16 lem:

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}} \|\mathbf{p}_{lin} - \mathbf{W}\mathbf{x}\|_2^2. \quad (6)$$

17 The widely used SIRT algorithm is chosen for its
 18 robustness to noise and its easy implementation.

19 The input for this step is a set of “linearized”
 20 projections p_{lin} . For the first iteration, they are
 21 just the acquired projections. For the other itera-
 22 tions, they are adopted as the projections that have
 23 been rescaled in the previous iteration, which will
 24 be explained in Step 4.

25 *Step 2: Segmentation*

26 The binary images are then determined by seg-
 27 menting the reconstruction image \mathbf{x} . As gray lev-
 28 els are related to atomic numbers, we segment the
 29 SIRT reconstruction by global thresholding. The
 30 thresholds for the segmentation are determined by
 31 solving the following optimization problem:

$$\mathbf{S}^* = \operatorname{argmin}_{\mathbf{S} \in \mathcal{S}} \mathcal{C}(I_0, p_b, \boldsymbol{\mu}, \mathbf{S}). \quad (7)$$

32 The solution of this problem is found by straightfor-
 33 ward (brute-force) sampling of the space of thresh-
 34 olds, each time evaluating the cost function. In
 35 practice, the thresholds are sampled from the min-
 36 imum to the maximum of gray levels of the SIRT
 37 reconstruction in Step 1.

38 The first iteration is again an exception since pa-
 39 rameters have not yet been estimated and the ob-
 40 jective function cannot be computed. Thus, the
 41 above segmentation method is not applicable. In-
 42 stead, the thresholds are determined using Otsu’s

method which finds optimal thresholds based on
 the gray level histograms [9].

3 *Step 3: Nonlinear parameters estimation*

4 Given the binary images, we can update the free
 5 parameters of the nonlinear model $I_0, p_b, \boldsymbol{\mu}$ by min-
 6 imizing the objective function, which is a nonlinear
 7 regression problem. This nonlinear regression prob-
 8 lem is solved using the Nelder–Mead method [17].
 9 To improve the stability of the regression, the three
 10 parameters are estimated separately and iteratively
 11 in an inner loop:

For $l = 1 : L$

$$\begin{aligned} p_b^{l+1} &= \operatorname{argmin}_{p_b} \mathcal{C}(I_0^l, p_b, \boldsymbol{\mu}^l, \mathbf{S}^*); \\ \boldsymbol{\mu}^{l+1} &= \operatorname{argmin}_{\boldsymbol{\mu} > 0} \mathcal{C}(I_0^l, p_b^{l+1}, \boldsymbol{\mu}, \mathbf{S}^*); \\ I_0^{l+1} &= \operatorname{argmin}_{I_0 > \max(\mathbf{p})} \mathcal{C}(I_0, p_b^{l+1}, \boldsymbol{\mu}^{l+1}, \mathbf{S}^*). \end{aligned} \quad (8)$$

12 here l is the iteration number of the inner loop.
 13 The estimation algorithm requires initial parame-
 14 ter values. In the experiments, we found that the
 15 initial values have little influence on the conver-
 16 gence result but proper initial values help to con-
 17 verge faster. Since we know that the beam inten-
 18 sity I_0 should be at least the maximal image inten-
 19 sity and that the attenuation coefficients μ and the
 20 bias intensity p_b are very small, we can start from
 21 $I_0^1 = \max(\mathbf{p})$, $p_b^1 = 0$ and $\boldsymbol{\mu}^1 = \mathbf{0}$, which were used
 22 in all the experiments in the paper.

23 *Step 4: Projection intensities rescaling*

24 Given the parameters, we rescale the measured
 25 projections \mathbf{p} to reduce nonlinear damping effects
 26 using:

$$\mathbf{p}_{lin} = \log \frac{I_0 + p_b - \mathbf{p}}{I_0}, \quad (9)$$

27 where \mathbf{p}_{lin} is the rescaled projections and is used
 28 as the input data for Step 1. At the last iteration,
 29 the rescaled projections are returned as the output
 30 \mathbf{p}_{lin} . These correspond to the linearly projected
 31 sum of the attenuation coefficients.

32 **3. Experiments and Simulations**

33 We report the correction of cupping artifacts for
 34 two sets of experimental data and three phantom

1 simulations. The experimental data show strong
 2 nonlinear effects because the samples consist of
 3 thick metallic materials. Two phantom simulations
 4 resembling the experimental data were performed,
 5 as ground-truth is missing for quality assessment of
 6 the reconstruction image due to the lack of other
 7 measurement methods. In addition, a phantom of
 8 four chemical compositions was simulated to investigate
 9 the robustness of the algorithm when more
 10 chemical compositions are present, as the experimental
 11 samples consist of only one or two chemical
 12 compositions.

13 3.1. Experiments

14 The first experimental sample is an assembly
 15 consisting of 16 Pt nanoparticles, each of which
 16 has a diameter of about 60 nm (Fig. 4(a)) [18]. It
 17 has only one chemical composition and a relatively
 18 more complex structure than the second sample.

19 The second sample is a hetero-nanoparticle,
 20 which is an Ag nanoparticle with a diameter of approximately
 21 110 nm with an embedded Au octahedron [7]. It is studied
 22 as a case where the cupping artifacts reduce the image contrast
 23 between different chemical compositions. The specifications of
 24 data acquisition are listed in Table 1.
 25

26 This dataset has been used to investigate
 27 HAADF-EDS bimodal tomography (HEBT) in [7].
 28 In that study, the authors have noticed that the
 29 raw data had strong intensity damping which not
 30 only limited straightforward segmentation of the
 31 HAADF reconstructions but also undermined the
 32 validity of HEBT based on linear models. Therefore,
 33 in [7] the data has been linearized in the data
 34 preprocessing as mentioned in section 3.2 of [7].

35 3.1.1. Results: Nanoparticle Assembly

36 Fig. 5 (a) is the initial SIRT reconstruction,
 37 based on which a binary image (Fig. 5(c)) was segmented
 38 using Otsu’s method. Fig. 5 (b) and (d) are the
 39 reconstruction and the binary image acquired after applying
 40 the correction algorithm. To obtain morphological information
 41 which is difficult to observe in the reconstruction images,
 42 we plotted their edges (Fig. 5 (e)) which are detected using
 43 a Sobel filter that depends on the derivatives of gray levels.
 44

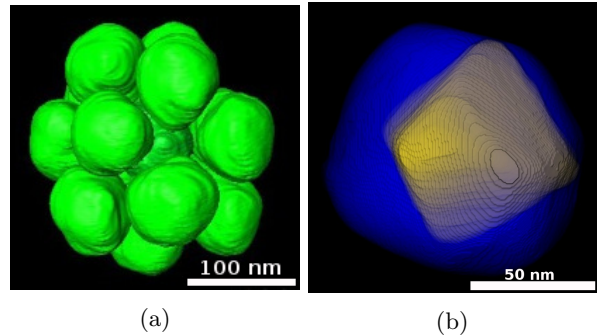


Figure 4: (a): 3D volume rendering of the Pt nanoparticle assembly. (b): 3D volume rendering of the Au-Ag nanoparticle.

In addition, the fidelity of the nonlinear regression for the nonlinear model was investigated. The fitted nonlinear model w.r.t thickness is plotted in Fig 6, where the thickness was computed as the forward projection of the binary image after correction. The error bars indicate the mean intensities and the standard deviations of the projection intensity.

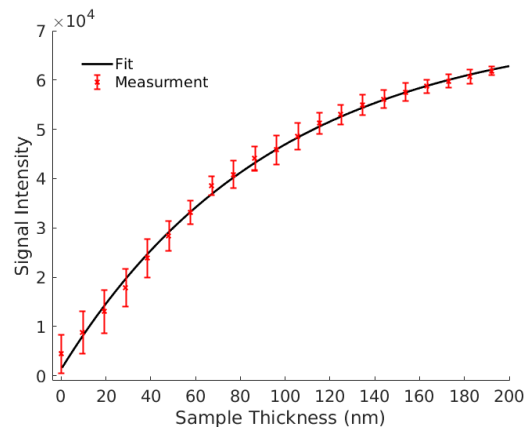


Figure 6: The nonlinear damping model fitted for projection signal intensity w.r.t. sample thickness of the nanoparticle assembly. The error bars indicate mean intensities and the standard deviations of the projection data.

3.1.2. Results: Au-Ag Core-shell Nanoparticle

For this experimental data, the SIRT reconstructions and segmented binary images before and after correction are shown in Fig. 7. In addition, the line profiles across the reconstruction images for some

Table 1: Data acquisition specifications.

specimen	nanoparticle assembly	core-shell nanoparticle
electron microscope	Tecnai G2, FEI company	Tecnai Osiris, FEI company
accelerating voltage	200 kV	120 kV
convergence angle	16 mrad	18 mrad
HAADF detector range	82-180 mrad	54-230 mrad
projection angles range	-74° to 74°	-75° to 75°
projection angle increment	2°	5°

1 iterations are plotted in Fig. 8 to demonstrate how
 2 gray levels evolve during a run of the correction
 3 algorithm.

4 As discussed in the introduction, the nonlinear
 5 effects also hinder adopting prior knowledge to re-
 6 duce missing wedge artifacts. In this data, the pro-
 7 jections were only acquired from -75° to 75° . We
 8 thus compared reconstructions using advanced re-
 9 construction algorithms: total-variation minimiza-
 10 tion (TV-min) [10], discrete algebraic reconstruc-
 11 tion technique (DART) [14] and total variation regu-
 12 larized DART (TVR-DART) [15], which incorpo-
 13 rate the prior knowledge of image sparsity, discrete
 14 gray levels and image sparsity combined with dis-
 15 crete gray levels respectively. The images recon-
 16 structed from the nonlinear projections and the cor-
 17 rected projections are given in In Fig. 9.

18 Finally, we plotted the normalized residuals of
 19 the cost function w.r.t. iterations for the two ex-
 20 perimental data (Fig. 10). For the first and sec-
 21 ond experiments, the cost values converge to stable
 22 minimums after 16 and 12 iterations respectively.

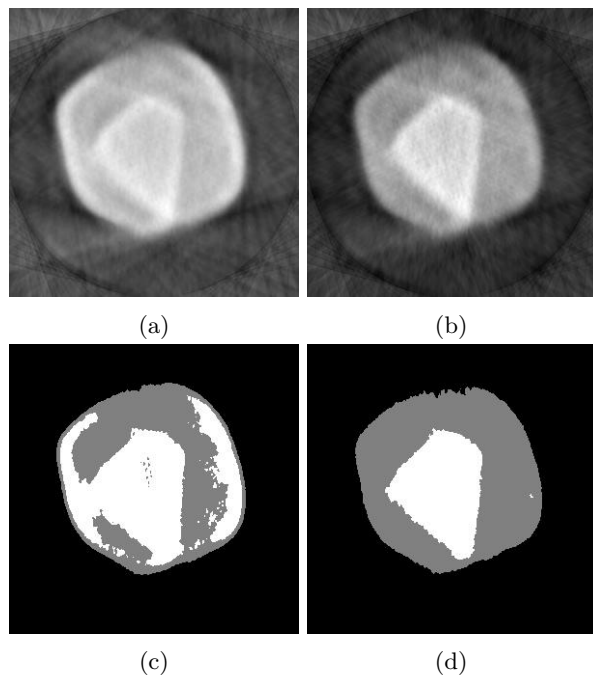


Figure 7: (a) and (b): SIRT reconstructions of the Au-Ag nanoparticle from the nonlinear projections and corrected projections. (c) and (d): Binary images segmented based on the reconstruction images (a) and (b) respectively.

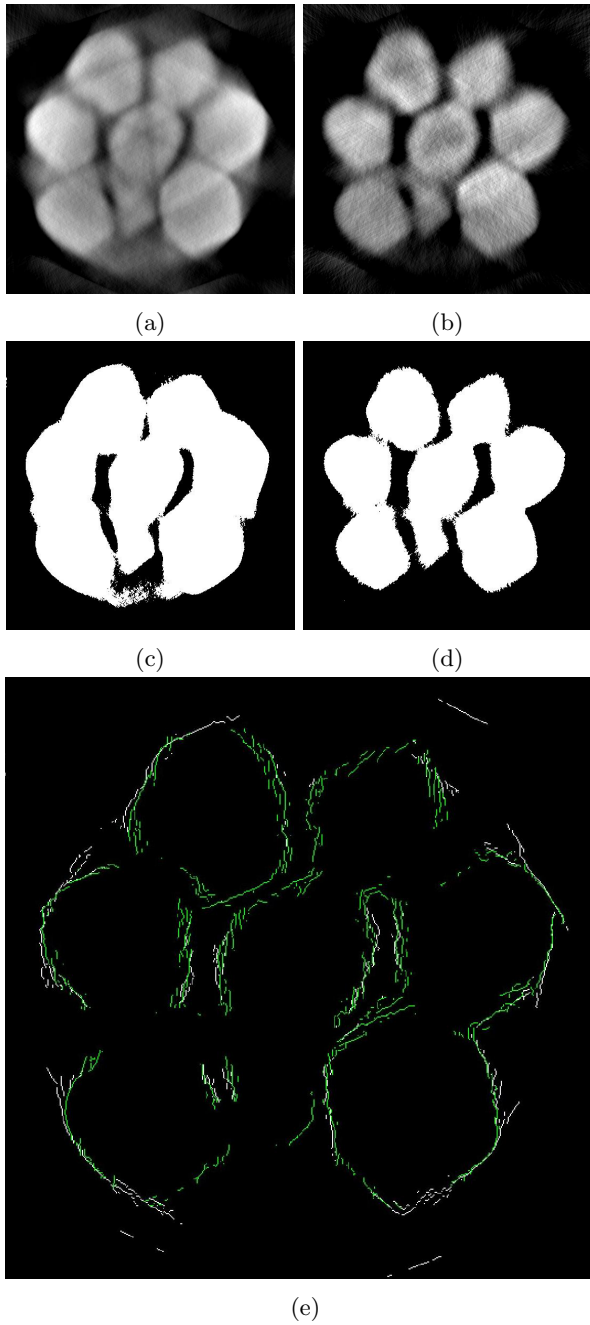


Figure 5: (a) and (b): SIRT reconstructions of the Pt nanoparticle assembly from the nonlinear projections and corrected projections respectively. (c) and (d): Binary images obtained by segmenting (a) and (b) respectively. (e) Edges of reconstructions before (white) and after correction (green).

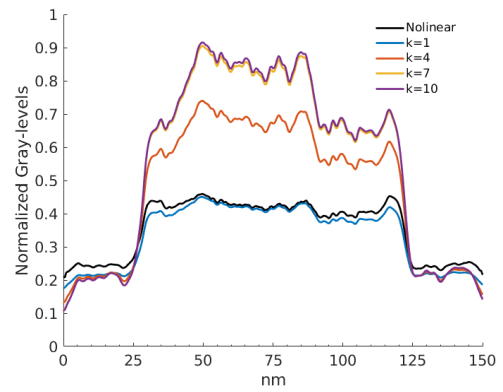


Figure 8: Cross-section line profiles of the SIRT reconstructions of the Au-Ag nanoparticle at different iterations.

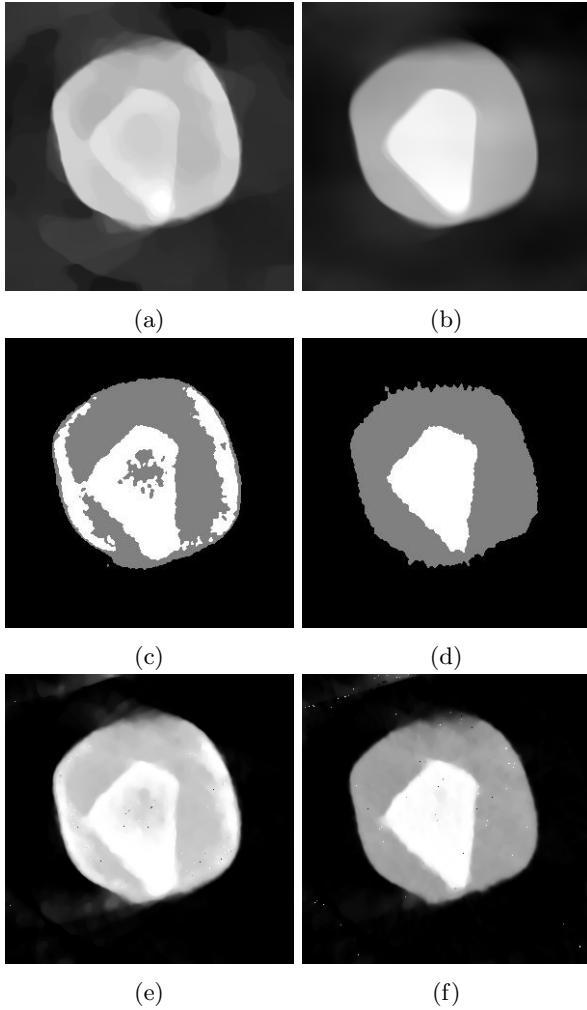


Figure 9: (a)/(b), (c)/(d) and (e)/(f) are the TV-min, DART and TVR-DART reconstructions of the Au-Ag nanoparticle from projections before/after the correction respectively.

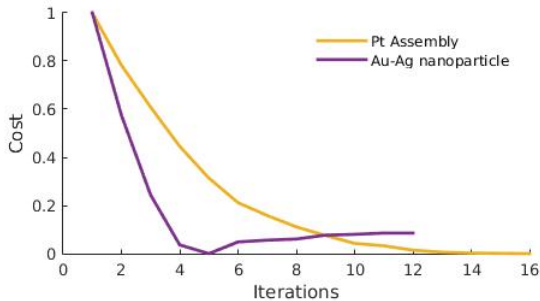


Figure 10: The residuals of cost function (Eq. 10) w.r.t. iterations for the two experimental datasets.

3.2. Phantom Simulations

First of all, two phantom simulations were made resembling the two experimental datasets. Note that the purpose of the simulation is not to validate the nonlinear model, but to assess the quality of nonlinear correction assuming the nonlinear forward model is accurate once all model parameters have been accurately obtained. For each sample, we first applied the correction algorithm to the experimental data to obtain binary images and nonlinear forward models. Afterwards, projections were simulated by projecting the binary images based on the nonlinear model. In addition, Gaussian noise was added to the projections to make the simulation more realistic.

The simulations provide ground-truth to quantify the quality of reconstructions. Here, the error metric is defined as the mean difference between the reconstructed and the ground-truth binary images:

$$err = \frac{1}{K} \sum_e \sum_j^N \| s_{ej} - g_{ej} \| / \sum_j^N g_{ej}, \quad (10)$$

where $\{g_{ej}\}$ are the ground-truth binary images.

The third phantom simulation, focused on the correction for more than two chemical compositions, was made using the same shapes as the nanoparticle assembly phantom. What is different is that instead of having one composition for all particles, there are particles of four different compositions, each having a different atomic number. Then projections were made by projecting the particles based on the nonlinear model.

3.2.1. Results of Simulations

The first phantom resembles the nanoparticle assembly, whose contours are plotted in Fig. 11 (c) and (d). Fig. 11 (a) is the initial SIRT reconstruction before correction, based on which a binary image (Fig. 1(c)) was segmented. Fig. 11 (b) and (d) show the SIRT reconstruction and the binary image after applying the correction algorithm. The error metrics of the binary images are respectively 5% and 2% before and after correction.

The results of the second phantom simulation are shown in Fig. 12, where (a) and (b) are the SIRT

1 reconstructions before and after correction respec-
 2 tively. The binary images in Fig. 12 (c) and (d)
 3 were segmented from the SIRT reconstruction im-
 4 ages. The ground-truth phantom is plotted using
 5 red and green contours for Au and Ag respectively.
 6 The error metrics of the binary images are respec-
 7 tively 56% and 1% before and after correction.

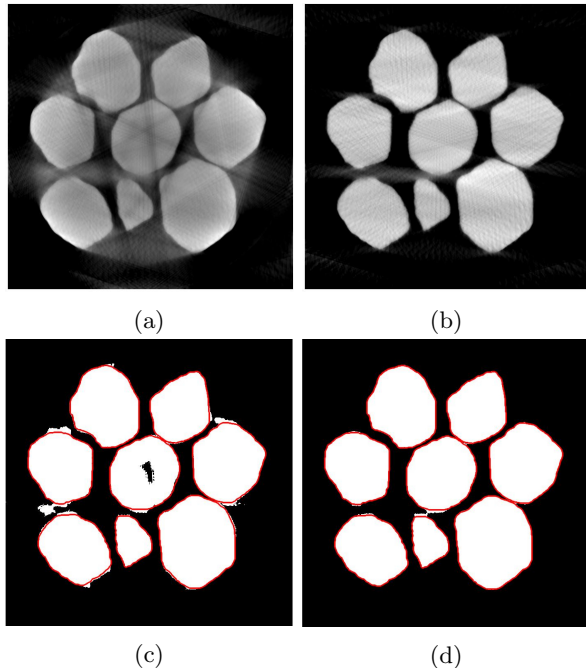


Figure 11: (a) and (b): SIRT Reconstruction images of the nanoparticle assembly phantom simulation before and after the nonlinearity correction. (c) and (d): Binary images segmented based on (a) and (b) respectively. The red contour shows the shape of the phantom.

8 The third phantom simulation presents the case
 9 when four chemical compositions exist in the same
 10 phantom. The SIRT reconstruction images before
 11 and after correcting the nonlinearity are shown in
 12 Fig. 13 (a) and (b) respectively, while the corre-
 13 sponding binary images are given in Fig. 13 (c)
 14 and (d). The error metrics of the binary images
 15 are respectively 69% and 20% before and after cor-
 16 rection.

17 3.3. Discussion

18 In the initial reconstruction of the nanoparticle
 19 assembly (Fig. 5(a)), the artifacts appear, on one

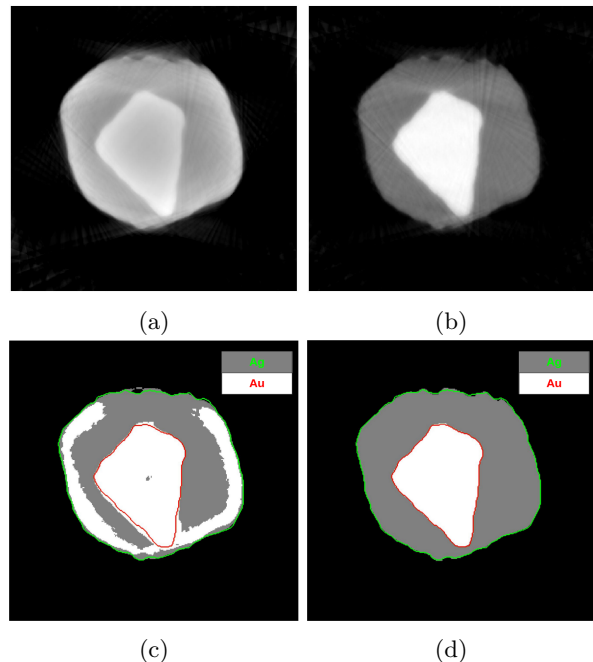


Figure 12: (a) and (b): SIRT reconstructions of the Au-Ag nanoparticle phantom simulation before and after the nonlinearity correction. (c) and (d): Binary images segmented based on (a) and (b) respectively. The red and green contours show the shape of the phantoms of Au and Ag respectively.

hand, as dark streaks elongated from the gaps be-
 2 tween particles. On the other, they appear as un-
 3 derestimated gray levels in the interior, for which
 4 we see missing pixels in the binary image (Fig.
 5 5(c)).

6 The correction algorithm successfully reduced
 7 these artifacts and produced images easier to in-
 8 terpret. The correction algorithm also changed
 9 the morphology of the reconstruction image (Fig.
 10 5(b)), as can be seen from the plot of edges. The
 11 change may be due to the removal of the overes-
 12 timated gray levels on the background. The plot
 13 of fitting (Fig. 6) shows that the experimental
 14 data matches our nonlinear model, demonstrating
 15 a damping effect following the exponential rule. It
 16 is also noticeable that the standard deviations de-
 17 crease at large thickness, which can be explained by
 18 noting that the errors introduced by segmentation
 19 are relatively smaller at larger thickness.

20 In the initial SIRT reconstruction image of the
 21 Au-Ag nanoparticle (Fig. 7(a)), the cupping arti-

Table 2: Errors Metrics of Binary Images.

	before correction	after correction
nanoparticle assembly phantom	5%	2%
Au-Ag nanoparticle phantom	56%	1%
phantom of four chemical compositions	69%	20%

1 facts caused the loss of contrast between Au and
 2 Ag, even though Au and Ag have a large difference
 3 in atomic number. As a result, many pixels were
 4 misclassified in the binary images (Fig. 7(b)). The
 5 algorithm corrected the experimental data and en-
 6 hanced the contrast between Au and Ag. Demon-
 7 strated in Fig. 8, the contrast between Au(center)
 8 and Ag(outskirts) was enhanced step by step. At
 9 last, the Au and Ag particles were segmented cor-
 10 rectly based on gray levels.

11 The Au-Ag nanoparticle should be suitable for
 12 incorporating prior knowledge to correct missing
 13 wedge artifacts. It contains two distinct compo-
 14 sitions with uniform densities, and thus the recon-
 15 struction image should be sparse and have constant
 16 gray levels. However, before the correction, incor-
 17 porating different variants of prior knowledge in the
 18 reconstruction actually appears to be detrimental
 19 to the image quality, as can be seen in Fig. 9. Espe-
 20 cially the tip of the Au particle was expanded. The
 21 expanded tip probably is a mixture of cupping arti-
 22 facts and missing wedge artifacts. After correcting
 23 the nonlinear effects, the linearized projection data
 24 was suitable for using the advanced algorithms as
 25 the reconstructions show.

26 The first two phantom simulations show artifacts
 27 (in Fig. 11(a) and Fig. 12(a)) very similar to those
 28 from the experimental data, which indicates that
 29 the modeling of nonlinear effects is accurate. Both
 30 reconstructions after correction are free of these arti-
 31 facts, and are in good agreement with the ground-
 32 truth phantom, as the error metrics were reduced
 33 (Table 2).

34 For the third simulation, we see cupping artifacts
 35 (Fig. 13(a)) with features observed in the previ-
 36 ous two cases. First, there are dark streaks and
 37 underestimated gray levels. Second, the contrast

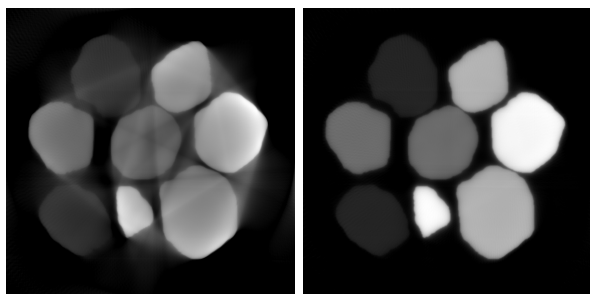
between different chemical compositions is blurred.
 These artifacts were corrected after applying the
 correction algorithm (Fig. 13(b)).

The segmented binary images after correction
 (Fig. 13(d)) show a stack of different chemical com-
 positions at the borders of some particles. How-
 ever, these misclassified pixels are not caused by
 the cupping artifacts, but due to the limitation of
 the global thresholding [19]. The gray levels in the
 reconstruction image are continuously dropping at
 the borders. These pixels were classified into par-
 ticles of smaller gray levels. Despite the imperfect
 segmentation, the correction algorithm converged
 to a result free from cupping artifacts, which also
 indicates the good robustness of the algorithm.

4. Conclusion

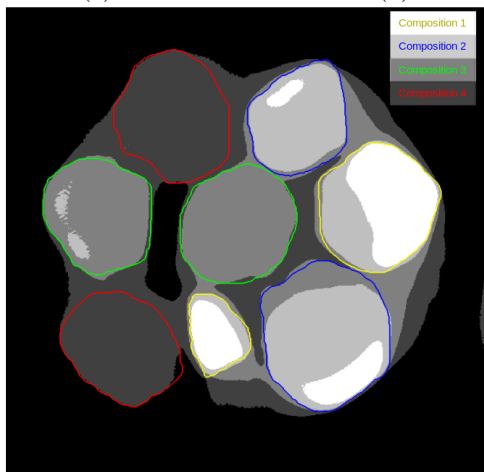
In this paper, we proposed an iterative algorithm
 to automatically correct the cupping artifacts in
 tomographic reconstructions from HAADF-STEM
 projections with nonlinearly damping intensities
 using only the projection data. The correction is
 based on modeling the nonlinear relationship be-
 tween projection intensities and sample thickness
 as an exponential function.

We showed that the algorithm is an effective tool
 in achieving better tomographic reconstructions. It
 successfully corrected the nonlinear damping effects
 and the subsequent cupping artifacts in three cases
 where one, two and four chemical compositions are
 present respectively. The correction is useful for
 improving the accuracy of morphological analysis
 and compositional analysis for 3D nanostructures
 and nanomaterials. In addition, users can benefit
 from it in enhancing the Z-contrast between chem-
 ical compositions as well as in facilitating incorpo-
 rating prior knowledge to correct the missing wedge
 artifacts.

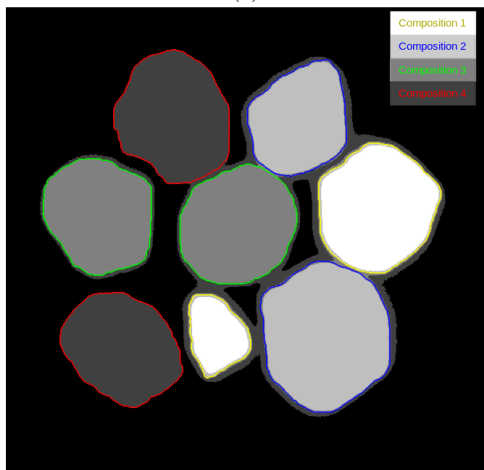


(a)

(b)



(c)



(d)

Figure 13: (a) and (b): SIRT reconstructions of the phantom simulation with four chemical compositions before and after correcting the nonlinear effects. (c) and (d): Binary images segmented based on (a) and (b) respectively. The colorful contours show the shape of the phantom particles of four different chemical compositions.

For limited data (e.g. with only a small range of tilts), the correction algorithms may fail due to the inaccurate segmentation caused by the dominant missing wedge artifacts. Potentially, this issue may be addressed by replacing SIRT and possibly the segmentation step by an advanced reconstruction algorithm (e.g. TVR-DART). However, it is still an unsolved question how to automatically set the parameters of the reconstruction algorithms, which has to be done in each iteration of the correction algorithm.

Note that the algorithm is only applicable to samples consist of several chemical compositions with homogeneous densities that can be segmented based on images gray levels. This is because the graylevel-based segmentation method fails easily when the chemical compositions are mixed or have similar atomic numbers. Moreover, this segmentation method is a global thresholding method. It may lead to poor initial segmentation results and consequently failed corrections when the cupping artifacts are very strong. Consequently, the next step of improving the algorithm is to incorporate advanced segmentation methods or spectroscopic techniques to determine the distributions of chemical compositions.

5. Acknowledgement

This research is supported by the Dutch Technology Foundation STW (<http://www.stw.nl/>), which is part of the Netherlands Organization for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs, Agriculture and Innovation under project number 13314. Funding from the European Research Council (Starting Grant No. COLOURATOMS 335078) is acknowledged by S. Bals. The authors would like to thank Dr. Thomas Altantzis and Dr. Bart Goris for providing the experimental data, and Prof. Dr. Luis M. Liz-Marzan for providing the investigated samples.

- 1 [1] P. A. Midgley, M. Weyland, J. M. Thomas, B. F. G. Johnson, Z-Contrast tomography: a technique in three-dimensional nanostructural analysis based on Rutherford scattering, *Chem. Commun.* (2001) 907–908.
- 2 [2] C. Kübel, A. Voigt, R. Schoenmakers, M. Otten, D. Su, T.-C. Lee, A. Carlsson, J. Bradley, Recent advances in electron tomography: TEM and HAADF-STEM tomography for materials science and semiconductor applications, *Microscopy and Microanalysis* 11 (05) (2005) 378–400.
- 3 [3] A. C. Kak, M. Slaney, *Principles of computerized tomographic imaging*, IEEE press, 1988.
- 4 [4] P. A. Midgley, M. Weyland, 3D electron microscopy in the physical sciences: the development of Z-contrast and EFTEM tomography, *Ultramicroscopy* 96 (3) (2003) 413–431.
- 5 [5] R. Aveyard, Z. Zhong, K. J. Batenburg, B. Rieger, Optimizing experimental parameters for the projection requirement in HAADF-STEM tomography, *Ultramicroscopy* 177 (2017) 84 – 90.
- 6 [6] W. Van den Broek, A. Rosenauer, B. Goris, G. T. Martinez, S. Bals, S. Van Aert, D. Van Dyck, Correction of non-linear thickness effects in HAADF STEM electron tomography, *Ultramicroscopy* 116 (2012) 8–12.
- 7 [7] Z. Zhong, B. Goris, R. Schoenmakers, S. Bals, K. J. Batenburg, A bimodal tomographic reconstruction technique combining EDS-STEM and HAADF-STEM, *Ultramicroscopy* 174 (2017) 35 – 45.
- 8 [8] J. Gregor, T. Benson, Computational Analysis and Improvement of SIRT, *IEEE Transactions on Medical Imaging* 27 (7) (2008) 918–924.
- 9 [9] N. Otsu, A threshold selection method from gray-level histograms, *IEEE Transactions on Systems, Man, and Cybernetics* 9 (1) (1979) 62–66.
- 10 [10] B. Goris, W. Van den Broek, K. J. Batenburg, H. Heidari Mezerji, S. Bals, Electron tomography based on a total variation minimization reconstruction technique, *Ultramicroscopy* 113 (2012) 120–130.
- 11 [11] G. Van Gompel, K. Van Slambrouck, M. De-frise, K. J. Batenburg, J. de Mey, J. Sijbers, J. Nuyts, Iterative correction of beam hardening artifacts in CT, *Medical Physics* 38 (S1) (2011) S36.
- 12 [12] J. Hsieh, R. C. Molthen, C. a. Dawson, R. H. Johnson, An iterative approach to the beam hardening correction in cone beam CT, *Medical physics* 27 (1) (2000) 23–29.
- 13 [13] E. V. de Castele, D. V. Dyck, J. Sijbers, E. Raman, An energy-based beam hardening model in tomography, *Physics in Medicine & Biology* 47 (23) (2002) 4181.
- 14 [14] K. J. Batenburg, S. Bals, J. Sijbers, C. Kübel, P. A. Midgley, J. C. Hernandez, U. Kaiser, E. R. Encina, E. a. Coronado, G. Van Tendeloo, 3D imaging of nanomaterials by discrete tomography, *Ultramicroscopy* 109 (6) (2009) 730–740.
- 15 [15] X. Zhuge, W. J. Palenstijn, K. J. Batenburg, TVR-DART: a more robust algorithm for discrete tomography from limited projection data with automated gray value estimation, *IEEE Transactions on Image Processing* 25 (1) (2016) 455–468.
- 16 [16] D. B. Williams, C. B. Carter, *The transmission electron microscope*, Springer, 1996.
- 17 [17] J. Lagarias, J. Reeds, H. Wright, P. Wright, Convergence properties of the Nelder–Mead simplex method in low dimensions, *Journal of Optimization* 9 (1) (1998) 112–117.
- 18 [18] A. Sánchez-Iglesias, M. Grzelczak, T. Altantzis, B. Goris, J. Pérez-Juste, S. Bals, G. Van Tendeloo, S. H. Donaldson, B. F. Chmelka, J. N. Israelachvili, L. M. Liz-Marzán, Hydrophobic interactions modulate self-assembly of nanoparticles, *ACS Nano* 6 (12) (2012) 11059–11065.
- 19 [19] K. Batenburg, J. Sijbers, Adaptive thresholding of tomograms by projection distance minimization, *Pattern Recognition* 42 (10) (2009) 2297 – 2305.

- An algorithm is proposed to correct cupping artifacts in HAADF-STEM tomography.
- It is an automated algorithm that requires no extra measurement.
- The algorithm is applied on different hetero nanostructures.

Automatic Correction of Nonlinear Damping Effects in HAADF-STEM Tomography for Nanomaterials of Discrete Compositions

Zhichao Zhong^a, Richard Aveyard^b, Bernd Rieger^b, Sara Bals^c, Willem Jan Palenstijn^a, K. Joost Batenburg^{a,d}

^a*Centrum Wiskunde & Informatica, Amsterdam, The Netherlands*

^b*Department of Imaging Physics, Delft University of Technology, The Netherlands*

^c*EMAT, University of Antwerp, Antwerp, Belgium*

^d*Mathematical Institute, Universiteit Leiden, Leiden, The Netherlands*

Abstract

HAADF-STEM tomography is a common technique for characterizing the three-dimensional morphology of nanomaterials. In conventional tomographic reconstruction algorithms, the image intensity is assumed to be a linear projection of a physical property of the specimen. However, in HAADF-STEM imaging, this linearity assumption is not completely valid due to nonlinear effects caused by the multiple scattering of electrons. Nonlinear imaging effects increase with the specimen thickness and lead to so-called “cupping artifacts”, due to a mismatch with the linear model used in the reconstruction algorithm. Moreover, nonlinear imaging effects can strongly limit the applicability of advanced reconstruction approaches such as Total Variation Minimization and discrete tomography.

In this paper, we propose an algorithm for automatically correcting the nonlinear effects and the subsequent cupping artifacts. It is applicable to samples in which chemical compositions can be segmented based on image gray levels. The correction is realized by iteratively estimating the nonlinear relationship between projection intensity and sample thickness, based on which the projections are linearized. The correction and reconstruction algorithms are tested on simulated and experimental data.

1. Introduction

In materials science, electron tomography (ET) is commonly used to characterize the three-dimensional (3D) structural and compositional information of nanomaterials. The 3D image is usually reconstructed from a tilt series of two-dimensional (2D) projections (projection images). The projection images should have a monotonic relationship between the measurement intensity and the integrated physical property of the specimen, which is referred to as the *projection requirement* in ET [1, 2]. Strictly speaking, the relationship should be linear, as most tomographic reconstruction algorithms are based on a linear mathematical model – the line integral model. It assumes that the projection is a measurement of a physical property integrated along the projection orientation (see

Chapter 3 in [3]).

High angle annular dark field (HAADF) scanning transmission electron microscopy (STEM) is commonly used for ET [1, 4] under the implicit assumption that the projection requirement can be approximately satisfied. The image intensity approximates to be proportional to the mass-thickness weighted by Z^2 , where Z is the atomic number [4]. However, this approximation is not always valid. One example is that channeling effects cause fringes on the projections at zone-axis orientations, which therefore are usually excluded from the input data [5]. Another example is that the image intensity damps at high sample thickness due to multiple scattering of electrons, which can occur in all projection orientations. While the zone-axis effects can be easily identified,

1 intensity damping is not easily seen in individual
2 projections. In this paper, we aim at addressing
3 the nonlinear effects of intensity damping for
4 tomographic reconstruction.

5
6 The consequence of intensity damping appears as
7 the cupping artifact in tomographic reconstruction:
8 the gray levels in the center of the reconstructed
9 sample are underestimated while overestimated on
10 the exterior [6]. In Fig. 1(a), an example of the
11 cupping artifact is given. It is a 2D cross section of
12 an Au-Ag core-shell nanoparticle [7], reconstructed
13 using the SIRT algorithm [8]. If we look at the
14 line-profile of the 2D image (Fig. 1(b)), the curve
15 appears in a concave “cup” shape, while ideally it
16 should be flat. The cupping artifacts are caused by
17 the strong damping effects of Au at large thickness,
18 which is illustrated by the simulated relationships
19 between measurement intensity and sample thick-
20 nesses using the multislice simulation method [5] in
21 Fig. 2.

22 It is important to correct the nonlinear effects
23 and the subsequent cupping artifacts for three rea-
24 sons. First of all, compositional analysis based on
25 gray levels becomes difficult when the cupping arti-
26 facts occur, as gray levels are not proportional any-
27 more to density and atomic numbers. Second, mor-
28 phological analysis based on segmentation of recon-
29 struction images is hindered by the cupping arti-
30 facts. Some straightforward segmentation meth-
31 ods, e.g. Otsu’s method [9], require that for each
32 chemical composition there should be one constant
33 gray level. Third, the nonlinear effects limit apply-
34 ing advanced reconstruction algorithms to address
35 a critical issue of ET – the missing wedge artifacts
36 caused by the limited tilt range of the sample. Al-
37 gorithms such as total variation minimization [10]
38 reduce the missing wedge artifacts by incorporating
39 prior knowledge i.e. sparsity of the unknown sam-
40 ple. Nevertheless, these algorithms have an even
41 stronger requirement for the linear forward model
42 which is inaccurate due to the nonlinear effects.

43 Despite these shortcomings of using uncorrected
44 data, there are few publications addressing the non-
45 linearity issue in ET [5, 6]. Nonlinear effects are
46 usually ignored or mitigated during image acquisi-
47 tion by increasing the inner angle of the HAADF
48 detector but at the cost of losing signal strength

1 [5]. An alternative to adjusting the acquisition pa- 1
2 rameters is to correct the measured data in a post- 2
3 processing step by linearizing the projection data, 3
4 provided that the incident beam intensity is known 4
5 [6]. The method described here requires only the 5
6 HAADF signal, consequently, it can be applied to 6
7 correct cupping artifacts in many existing datasets 7
8 acquired in a conventional manner. The mathe- 8
9 matical model of nonlinearity and the concept of 9
10 linearization in [6] are also used in this paper, which 10
11 will be explained in Section 2.1. 11

12 Here, we propose an iterative algorithm to au- 12
13 tomatically correct the nonlinear effects and the 13
14 cupping artifacts. It does not require the extra 14
15 measurement of the incident beam intensity as in 15
16 [6]. Instead, it automatically models the nonlinear 16
17 effects given the projection data. The algorithm 17
18 iteratively searches for the minimal distance be- 18
19 tween the acquired projections and the nonlinear 19
20 re-projections of chemical compositions by varying 20
21 the nonlinear model and the reconstruction image, 21
22 so as to estimate a nonlinear relationship between 22
23 the measured HAADF-STEM intensities and sam- 23
24 ple thickness for all chemical compositions. The 24
25 algorithm contains the following steps in every it- 25
26 eration: first a reconstruction image with contin- 26
27 uous gray levels is made; then the image is seg- 27
28 mented into several binary images, each of which 28
29 corresponds to a chemical composition; after that, 29
30 the nonlinear effects are modeled by minimizing the 30
31 projection distance; based on the model, the pro- 31
32 jection data is linearized at last. A similar concept 32
33 of iterative correction has been introduced to cor- 33
34 rect beam hardening artifacts for X-ray computed 34
35 tomography [11]. 35

36 Our approach is only applicable to samples con- 36
37 sisting of several chemical compositions with uni- 37
38 form densities, such as homogeneous or core-shell 38
39 particles. It is assumed that for these samples the 39
40 volumetric distributions of the compositions can be 40
41 approximated well by segmenting the reconstructed 41
42 image based on gray levels and that this segmenta- 42
43 tion improves as the correction model applied to the 43
44 measured data becomes more accurate. In fact, this 44
45 kind of samples are commonly studied in materials 45
46 science. For example, the samples typically stud- 46
47 ied in the context of discrete tomography [12, 13] 47
48 match the requirements. 48

1 In Section 2, the correction algorithm is ex- 1
 2 plained in detail. In section 3, we demonstrate 2
 3 how the nonlinear effects are corrected using this 3
 4 algorithm for real experimental data and phantom 4
 5 simulations. 5

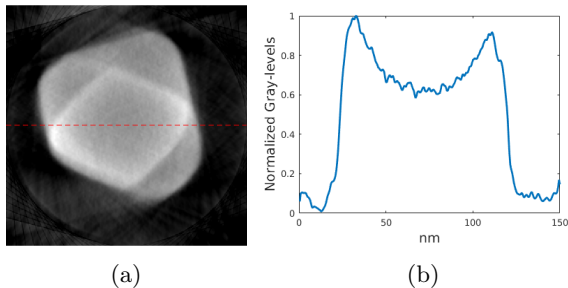


Figure 1: (a): 2D slice of the SIRT reconstruction of an Au-Ag nanoparticle. (b): Gray levels of the line-profile located at the dash line of the 2D slice.

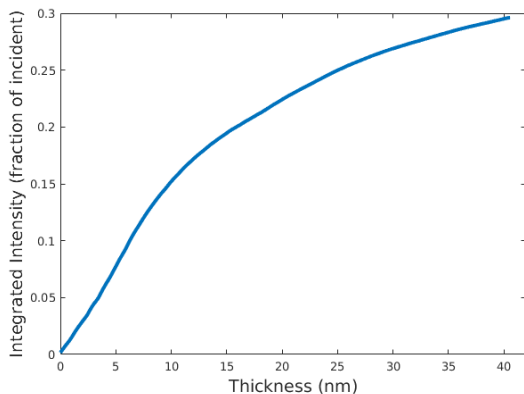


Figure 2: Normalized HAADF signal intensity w.r.t the thickness of Au slabs mistilted 10 degrees from the [100] zone axis about the $\langle 100 \rangle$ axis, simulated using the multislice method [5]. The accelerating voltage is 200 kV, the convergence angle is 10 mrad and the detector angular range is 50 - 250 mrad. The intensities are scaled by the incident beam intensity.

6 2. The Nonlinear Model and the Correction 7 7 Algorithm

8 2.1. The Nonlinear Model

9 To linearize the projections, we first need to 9
 10 define a model that describes the nonlinear rela- 10
 11 tionship mathematically. A precise mathematical 11

model is possible but does not fit as a subrou- 1
 tine of the correction algorithm. The computa- 2
 tion of a sophisticated model, such as the one used 3
 in multi-slice simulations which take into considera- 4
 tion the multiple scattering of electrons [5], is ex- 5
 tremely time-consuming and costly. Therefore, a 6
 simple model is preferred here. 7

Here, we choose a model that has already been 8
 used for describing the nonlinear relationship. In 9
 [6, 14], it is illustrated that the HAADF detec- 10
 tor collects electrons complementary to the elec- 11
 trons scattered to angles smaller than its inner de- 12
 tector angle. By pragmatically applying a simple 13
 Beer-Lambert description of electron scattering we 14
 can state that the number of electrons scattered to 15
 small angles p_t decreases exponentially to the sam- 16
 ple thickness t along the beam direction. The p_t - t 17
 relationship is 18

$$p_t = I_0 \exp\left(-\sum_e^K \mu_e t\right), \quad (1)$$

where I_0 is the incident beam intensity, e is the 19
 index of chemical composition, K is the total num- 20
 ber of chemical compositions, μ_e is the attenua- 21
 tion coefficient of chemical composition e . There- 22
 fore, the complementary HAADF signal intensity p at 23
 sample thickness t is: 24

$$p = I_0(1 - \exp(-\sum_e^K \mu_e t)) + p_b, \quad (2)$$

where p_b is the bias signal, which is influenced by 25
 the dark current, carbon grid, and possibly other 26
 factors. This mathematical model has been used 27
 to correct the cupping artifacts successfully in [6], 28
 which is applicable only if the incident beam inten- 29
 sities can be measured. An advantage of this 30
 simple model is that it can easily be transformed 31
 into a linear relationship by taking logarithms so 32
 that we can avoid solving nonlinear least-squared 33
 problems for tomographic reconstruction. 34

In the practice of ET, a series of projections are 35
 taken at different angles. The image intensity of 36
 each pixel corresponds to the electrons scattered for 37
 an electron beam transmitting through the sample, 38
 which is called a line projection here. In total, there 39
 are M pixels for all the images. The image intensity 40
 of the i_{th} pixel is now written as an entry p_i in 41

1 $\mathbf{p} \in \mathbf{R}^M$. In addition, the space of reconstruction
 2 is a cubic volume partitioned into N voxels.

3 We also assume the chemical compositions are
 4 not mixed and voxels are small enough to resolve
 5 every chemical composition, which means that in
 6 each voxel only one element is present. As stated
 7 in Introduction, this algorithm is applied to sam-
 8 ples with uniform density. Thus we assume that
 9 each chemical composition is either present (1) or
 10 absent (0) in each voxel. The distribution of chem-
 11 ical composition e is described by binary variables
 12 s_{ej} , where $j = 1, \dots, N$ is the index of voxel.

13 Now we define the nonlinear relationship in the
 14 discrete form. For pixel i , the corresponding sample
 15 thickness of chemical composition e is now written
 16 as the ray-sum $\sum_{j=1}^N w_{ij}s_{ej}$, where the factor w_{ij}
 17 is determined by the area of intersection between
 18 the i_{th} line projection and the j_{th} voxel. The rela-
 19 tionship between projection intensities and binary
 20 volumes are:

$$p_i = I_0(1 - \exp(-\sum_{e=1}^K \mu_e \sum_{j=1}^N w_{ij}s_{ej})) + p_b, \quad (3)$$

21 where $i = 1, \dots, M$.

22 2.2. The Correction Algorithm

23 The basis of the correction algorithm is to esti-
 24 mate the nonlinear relationship of Eq. 3 based on
 25 the reconstructed distributions of chemical compo-
 26 sitions. The procedures of the automatic correc-
 27 tion algorithm are given in the flowchart (Fig. 3).
 28 The correction is realized iteratively through the
 29 following steps: (1) make a reconstruction image
 30 based on the linear model from the projections; (2)
 31 segment the reconstruction into a series of binary
 32 images, one for each chemical composition; (3) esti-
 33 mate the parameters of the nonlinear model in Eq.
 34 3 given the projections and the binary images; (4)
 35 reduce the nonlinearities in the projections through
 36 a rescaling of the intensities based on the nonlinear
 37 model.

Before we explain the steps explicitly, we estab-
 lish an objective function which will be used to
 guide the optimization in the correction algorithm.
 We define it as the l_2 norm of the distance between
 the acquired projections and the re-projection of

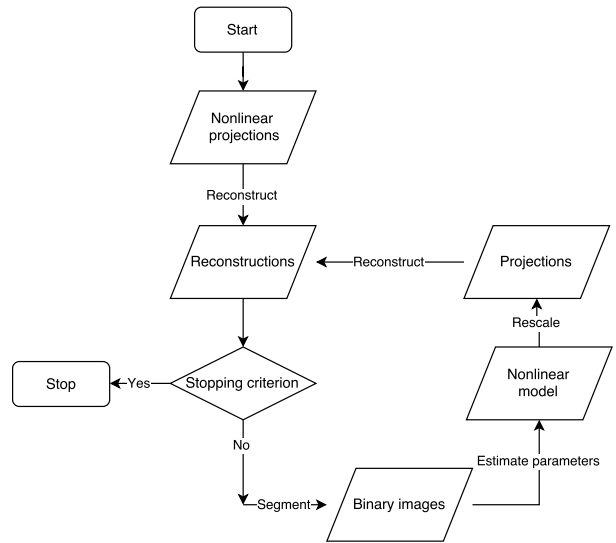


Figure 3: Flowchart of the correction algorithm

binary images based on our nonlinear model:

$$\mathcal{C}(I_0, p_b, \boldsymbol{\mu}, \mathbf{S}) = \|\mathbf{p} - I_0(1 - \exp(-\mathbf{W} \sum_{e=1}^K \mu_e \mathbf{s}_e)) - p_b\|_2^2, \quad (4)$$

where $\mathbf{W} = \{w_{ij}\}$, $\boldsymbol{\mu} = \{\mu_e\}$ and $\mathbf{S} = \{s_{ej}\}$.

We also define a stopping criterion. The cost value at the w_{th} iteration is denoted as the c^w . The loop is terminated if the cost is stable, which is when the following criterion is met:

$$\frac{c^w + c^{w-1}}{c^{w-2} + c^{w-3}} > t, \quad (5)$$

where $0 < t < 1$ is a thresholding value. Note that although we minimize the cost function in some steps of the algorithm, the cost is not guaranteed to reach a global minimum in the end.

Step 1: Reconstruction

As the first step, a reconstruction with continuous gray levels is made for determining the binary images in the next step. Though it is possible to reconstruct binary images directly using some discrete tomography algorithms (e.g. [12]), these algorithms will possibly not give better results than basic algorithms given an inaccurate forward model. Thus, we choose to first make a reconstruction \mathbf{x} with continuous gray levels based on a linear model

1 and then segment the reconstruction into binary
2 images \mathbf{S} .

3 The reconstruction is computed using the simul-
4 taneous iterative reconstruction technique (SIRT)
5 [8] which solves the following least-squares prob-
6 lem:

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}} \|\mathbf{p}_{lin} - \mathbf{W}\mathbf{x}\|_2^2. \quad (6)$$

7 The widely used SIRT algorithm is chosen for its
8 robustness to noise and its easy implementation.

9 The input for this step is a set of “linearized”
10 projections p_{lin} . For the first iteration, they are
11 just the acquired projections. For the other itera-
12 tions, they are adopted as the projections that have
13 been rescaled in the previous iteration, which will
14 be explained in Step 4.

15 *Step 2: Segmentation*

16 The binary images are then determined by seg-
17 menting the reconstruction image \mathbf{x} . As gray lev-
18 els are related to atomic numbers, we segment the
19 SIRT reconstruction by global thresholding. The
20 thresholds for the segmentation are determined by
21 solving the following optimization problem:

$$\mathbf{S}^* = \operatorname{argmin}_{\mathbf{S} \in \mathcal{S}} \mathcal{C}(I_0, p_b, \boldsymbol{\mu}, \mathbf{S}). \quad (7)$$

22 The solution of this problem is found by straightfor-
23 ward (brute-force) sampling of the space of thresh-
24 olds, each time evaluating the cost function. In
25 practice, the thresholds are sampled from the min-
26 imum to the maximum of gray levels of the SIRT
27 reconstruction in Step 1.

28 The first iteration is again an exception since pa-
29 rameters have not yet been estimated and the ob-
30 jective function cannot be computed. Thus, the
31 above segmentation method is not applicable. In-
32 stead, the thresholds are determined using Otsu’s
33 method which finds optimal thresholds based on
34 the gray level histograms [9].

35 *Step 3: Nonlinear parameters estimation*

36 Given the binary images, we can update the free
37 parameters of the nonlinear model $I_0, p_b, \boldsymbol{\mu}$ by min-
38 imizing the objective function, which is a nonlinear
39 regression problem. This nonlinear regression prob-
40 lem is solved using the Nelder–Mead method [15].
41 To improve the stability of the regression, the three
42 parameters are estimated separately and iteratively
43 in an inner loop:

For $l = 1 : L$

$$\begin{aligned} p_b^{l+1} &= \operatorname{argmin}_{p_b} \mathcal{C}(I_0, p_b, \boldsymbol{\mu}^l, \mathbf{S}^*); \\ \boldsymbol{\mu}^{l+1} &= \operatorname{argmin}_{\boldsymbol{\mu} > 0} \mathcal{C}(I_0, p_b^{l+1}, \boldsymbol{\mu}, \mathbf{S}^*); \\ I_0^{l+1} &= \operatorname{argmin}_{I_0 > \max(\mathbf{p})} \mathcal{C}(I_0, p_b^{l+1}, \boldsymbol{\mu}^{l+1}, \mathbf{S}^*). \end{aligned} \quad (8)$$

here l is the iteration number of the inner loop. The
estimation requires an initial guess for the param-
eters, which is $I_0^1 = 3 \cdot \max(\mathbf{p})$, $p_b^1 = 0$ and $\boldsymbol{\mu}^1 = \mathbf{0}$
in our experiments.

5 *Step 4: Projection intensities rescaling*

6 Given the parameters, we rescale the measured
7 projections \mathbf{p} to reduce nonlinear damping effects
8 using:

$$\mathbf{p}_{lin}^i = \log \frac{I_0 + p_b - \mathbf{p}}{I_0}, \quad (9)$$

9 where \mathbf{p}_{lin}^i is the rescaled projections and is used as
10 the input data for Step 1. At the last iteration, the
11 rescaled projections are return as the output \mathbf{p}_{lin} .

12 **3. Experiments and Simulations**

13 We report the correction of cupping artifacts for
14 two sets of experimental data and three phantom
15 simulations. The experimental data show strong
16 nonlinear effects because the samples consist of
17 thick metallic materials. Two phantom simulations
18 resembling the experimental data were performed,
19 as ground-truth is missing for quality assessment of
20 the reconstruction image due to the lack of other
21 measurement methods. In addition, a phantom of
22 four chemical compositions was simulated to inves-
23 tigate the robustness of the algorithm when more
24 chemical compositions are present, as the experi-
25 mental samples consist of only one or two chemical
26 compositions.

27 *3.1. Experiments*

28 The first experimental sample is an assembly
29 consisting of 16 Pt nanoparticles, each of which
30 has a diameter of about 60 nm (Fig. 4(a)) [16]. It
31 has only one chemical composition and a relatively
32 more complex structure than the second sample.

1 The second sample is a hetero-nanoparticle,
 2 which is an Ag nanoparticle with a diameter of ap-
 3 proximately 110 nm with an embedded Au octahe-
 4 dron [7]. It is studied as a case where the cupping
 5 artifacts reduce the image contrast between differ-
 6 ent chemical compositions. The specifications of
 7 data acquisition are listed in Table 1.

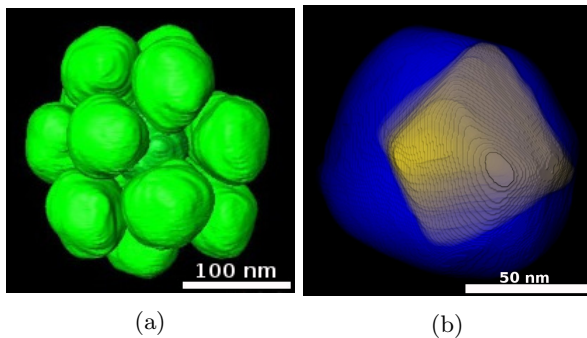


Figure 4: (a): 3D volume rendering of the Pt nanoparticle assembly. (b): 3D volume rendering of the Au-Ag nanoparticle.

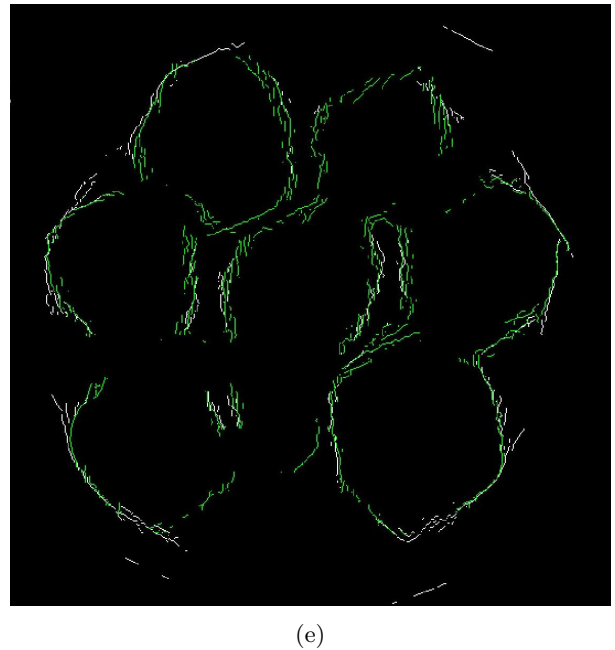
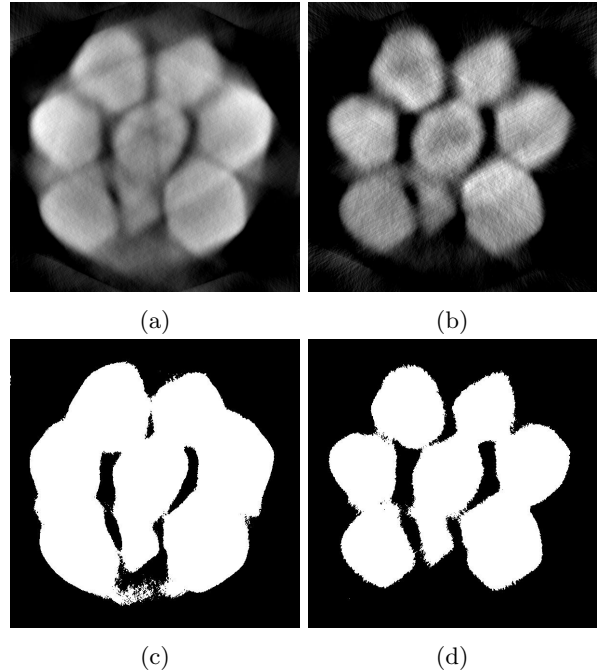


Figure 5: (a) and (b): SIRT reconstructions of the Pt nanoparticle assembly from the nonlinear projections and corrected projections respectively. (c) and (d): Binary images obtained by segmenting (a) and (b) respectively. (e) Edges of reconstructions before (white) and after correction (green).

8 3.1.1. Results: Nanoparticle Assembly

9 Fig. 5 (a) is the initial SIRT reconstruction,
 10 based on which a binary image (Fig. 5(c)) was seg-
 11 mented using Otsu's method. Fig. 5 (b) and (d) are
 12 the reconstruction and the binary image acquired
 13 after applying the correction algorithm. To obtain
 14 morphological information which is difficult to ob-
 15 serve in the reconstruction images, we plotted their
 16 edges (Fig. 5 (e)) which are detected using Sobel
 17 filter that depends on the derivatives of gray levels.

18 In addition, the fidelity of the nonlinear regres-
 19 sion for the nonlinear model was investigated. The
 20 fitted nonlinear model w.r.t thickness is plotted in
 21 Fig 6, where the thickness was computed as the
 22 forward projection of the binary image after cor-
 23 rection. The error bars indicate the mean intensi-
 24 ties and the standard deviations of the projection
 25 intensity.

Table 1: Data acquisition specifications.

specimen	nanoparticle assembly	core-shell nanoparticle
electron microscope	Tecnai G2, FEI company	Tecnai Osiris, FEI company
accelerating voltage	200 kV	120 kV
convergence angle	16 mrad	18 mrad
HAADF detector range	82-180 mrad	54-230 mrad
projection angles range	-74° to 74°	-75° to 75°
projection angle increment	2°	5°

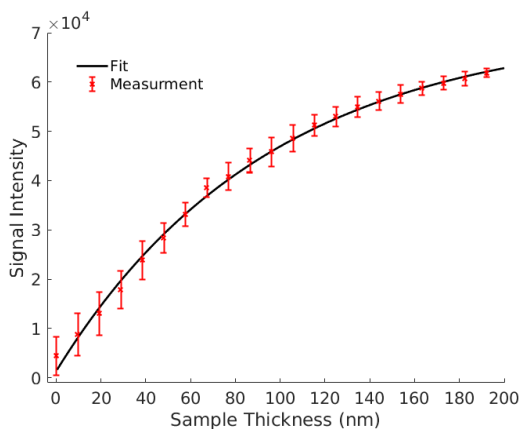


Figure 6: The nonlinear damping model fitted for projection signal intensity w.r.t. sample thickness of the nanoparticle assembly. The error bars indicate mean intensities and the standard deviations of the projection data.

3.1.2. Results: Au-Ag Core-shell Nanoparticle

For this experimental data, the SIRT reconstructions and segmented binary images before and after correction are shown in Fig. 7. In addition, the line profiles across the reconstruction images for some iterations are plotted in Fig. 8 to demonstrate how gray levels evolve during a run of the correction algorithm.

As discussed in the introduction, the nonlinear effects also hinder adopting prior knowledge to reduce missing wedge artifacts. In this data, the projections were only acquired from -75° to 75° . We thus compared reconstructions using advanced reconstruction algorithms: total-variation minimization (TV-min) [10], discrete algebraic reconstruction technique (DART) [12] and total variation regularized DART (TVR-DART) [13], which incorpo-

rate the prior knowledge of image sparsity, discrete gray levels and image sparsity combined with discrete gray levels respectively. The images reconstructed from the nonlinear projections and the corrected projections are given in In Fig. 9.

Finally, we plotted the normalized residuals of the cost function w.r.t. iterations for the two experimental data (Fig. 10). For the first and second experiments, the cost values converge to stable minimums after 16 and 12 iterations respectively.

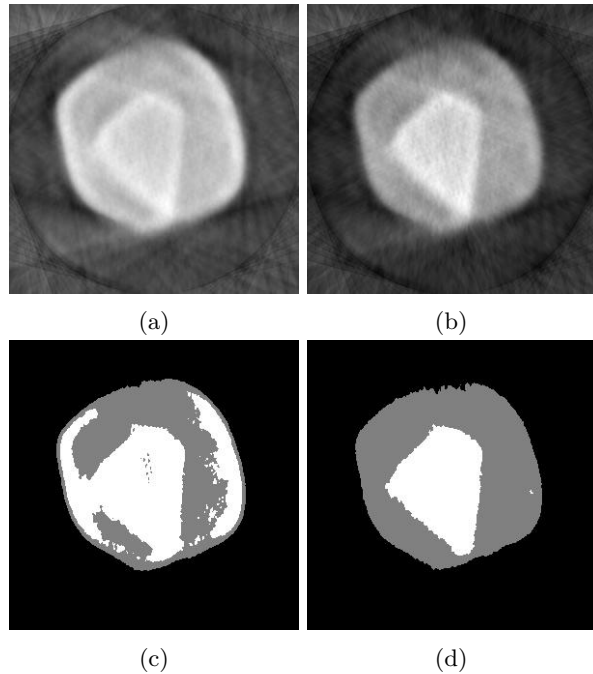


Figure 7: (a) and (b): SIRT reconstructions of the Au-Ag nanoparticle from the nonlinear projections and corrected projections. (c) and (d): Binary images segmented based on the reconstruction images (a) and (b) respectively.

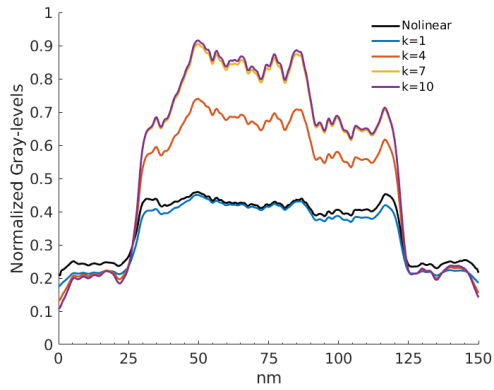


Figure 8: Cross-section line profiles of the SIRT reconstructions of the Au-Ag nanoparticle at different iterations.

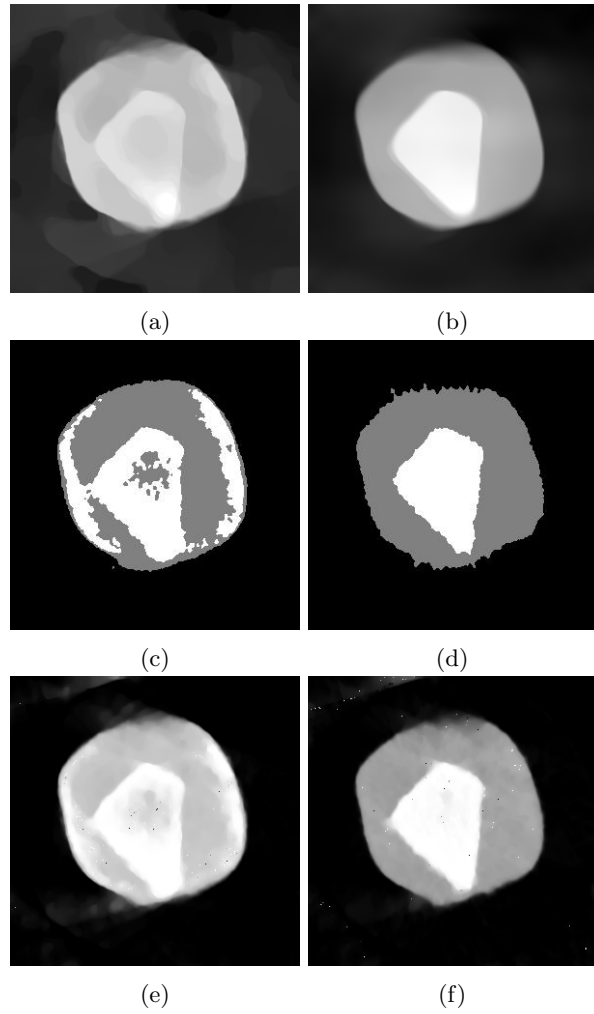


Figure 9: (a)/(b), (c)/(d) and (e)/(f) are the TV-min, DART and TVR-DART reconstructions of the Au-Ag nanoparticle from projections before/after the correction respectively.

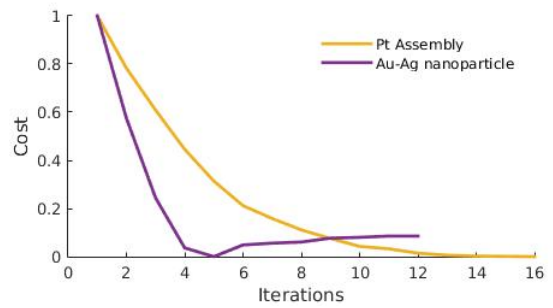


Figure 10: The residuals of cost function (Eq. 10) w.r.t. iterations for the two experimental datasets.

3.2. Phantom Simulations

First of all, two phantom simulations were made resembling the two experimental datasets. Note that the purpose of the simulation is not to validate the nonlinear model, but to assess the quality of nonlinear correction assuming the nonlinear forward model is accurate once all model parameters have been accurately obtained. For each sample, we first applied the correction algorithm to the experimental data to obtain binary images and nonlinear forward models. Afterwards, projections were simulated by projecting the binary images based on the nonlinear model. In addition, Gaussian noise was added to the projections to make the simulation more realistic.

The simulations provide ground-truth to quantify the quality of reconstructions. Here, the error metric is defined as the mean difference between the reconstructed and the ground-truth binary images:

$$err = \frac{1}{K} \sum_e \sum_j^N \| s_{ej} - g_{ej} \| / \sum_j^N g_{ej}, \quad (10)$$

where $\{g_{ej}\}$ are the ground-truth binary images.

The third phantom simulation, focused on the correction for more than two chemical compositions, was made using the same shapes as the nanoparticle assembly phantom. What is different is that instead of having one composition for all particles, there are particles of four different compositions, each having a different atomic number. Then projections were made by projecting the particles based on the nonlinear model.

3.2.1. Results of Simulations

The first phantom resembles the nanoparticle assembly, whose contours are plotted in Fig. 11 (c) and (d). Fig 11 (a) is the initial SIRT reconstruction before correction, based on which a binary image (Fig. 1(c)) was segmented. Fig. 11 (b) and (d) show the SIRT reconstruction and the binary image after applying the correction algorithm. The error metrics of the binary images are respectively 5% and 2% before and after correction.

The results of the second phantom simulation are shown in Fig. 12, where (a) and (b) are the SIRT

reconstructions before and after correction respectively. The binary images in Fig. 12 (c) and (d) were segmented from the SIRT reconstruction images. The ground-truth phantom is plotted using red and green contours for Au and Ag respectively. The error metrics of the binary images are respectively 56% and 1% before and after correction.

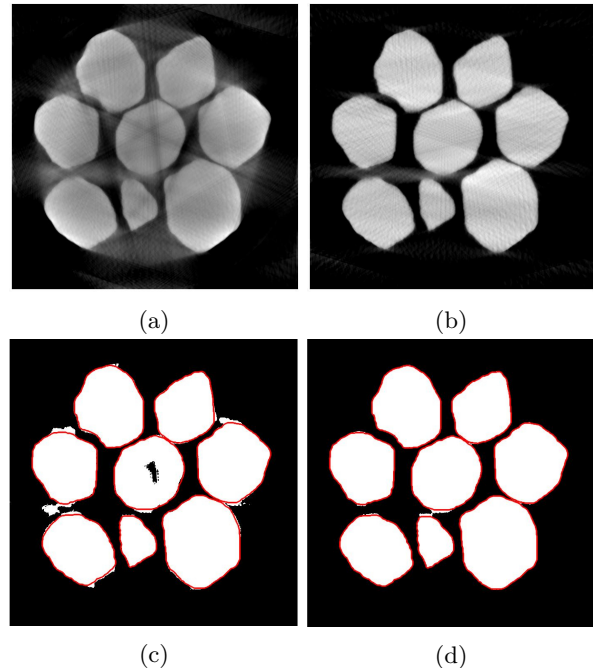


Figure 11: (a) and (b): SIRT Reconstruction images of the nanoparticle assembly phantom simulation before and after the nonlinearity correction. (c) and (d): Binary images segmented based on (a) and (b) respectively. The red contour shows the shape of the phantom.

The third phantom simulation presents the case when four chemical compositions exist in the same phantom. The SIRT reconstruction images before and after correcting the nonlinearity are shown in Fig. 13 (a) and (b) respectively, while the corresponding binary images are given in Fig. 13 (c) and (d). The error metrics of the binary images are respectively 69% and 20% before and after correction.

3.3. Discussion

In the initial reconstruction of the nanoparticle assembly (Fig. 5(a)), the artifacts appear, on one

Table 2: Errors Metrics of Binary Images.

	before correction	after correction
nanoparticle assembly phantom	5%	2%
Au-Ag nanoparticle phantom	56%	1%
phantom of four chemical compositions	69%	20%

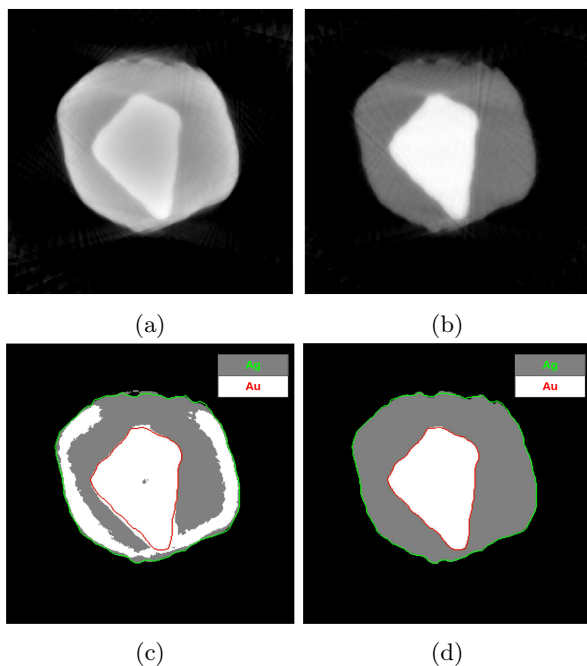


Figure 12: (a) and (b): SIRT reconstructions of the Au-Ag nanoparticle phantom simulation before and after the non-linearity correction. (c) and (d): Binary images segmented based on (a) and (b) respectively. The red and green contours show the shape of the phantoms of Au and Ag respectively.

hand, as dark streaks elongated from the gaps between particles. On the other, they appear as underestimated gray levels in the interior, for which we see missing pixels in the binary image (Fig. 5(c)).

The correction algorithm successfully reduced these artifacts and produced images easier to interpret. The correction algorithm also changed the morphology of the reconstruction image (Fig. 5(b)), as can be seen from the plot of edges. The change may be due to the removal of the over-

estimated gray levels on the background. The plot of fitting (Fig. 6) shows that the experimental data matches our nonlinear model, demonstrating a damping effect following the exponential rule. It is also noticeable that the standard deviations decrease at large thickness, which can be explained that the errors introduced by segmentation are relatively smaller at larger thickness.

In the initial SIRT reconstruction image of the Au-Ag nanoparticle (Fig. 7(a)), the cupping artifacts caused the loss of contrast between Au and Ag, even though Au and Ag have a large difference in atomic numbers. As a result, many pixels were misclassified in the binary images (Fig. 7(b)). The algorithm corrected the experimental data and enhance the contrast between Au and Ag. Demonstrated in Fig. 8, the contrast between Au(center) and Ag(outskirts) was enhanced step by step. At last, the Au and Ag particles were segmented correctly based on gray levels.

The Au-Ag nanoparticle should be suitable for incorporating prior knowledge to correct missing wedge artifacts. It contains two distinct compositions with uniform densities, and thus the reconstruction image should be sparse and have constant gray levels. However, before the correction, incorporating different variants of prior knowledge in the reconstruction actually appears to be detrimental to the image quality, as can be seen in Fig. 9. Especially the tip of the Au particle was expanded. The expanded tip probably is a mixture of cupping artifacts and missing wedge artifacts. After correcting the nonlinear effects, the linearized projection data was suitable for using the advanced algorithms as the reconstructions show.

The first two phantom simulations show artifacts (in Fig. 11(a) and Fig. 12(a)) very similar to those from the experimental data, which indicates that the modeling of nonlinear effects is accurate. Both

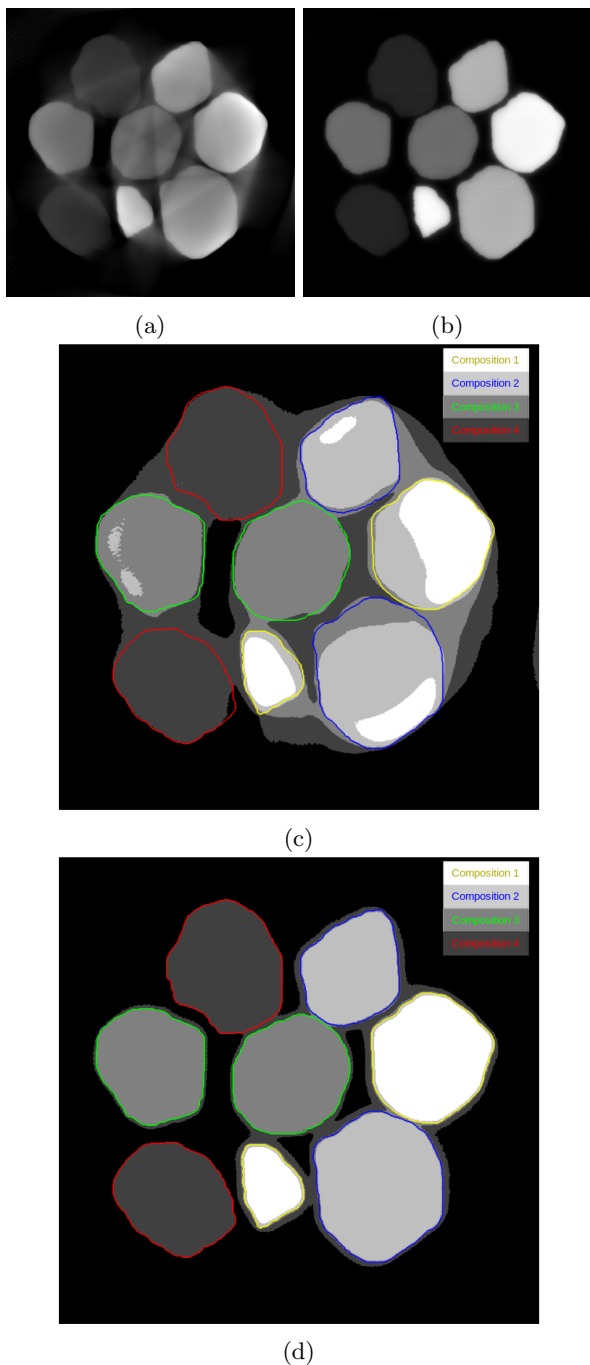


Figure 13: (a) and (b): SIRT reconstructions of the phantom simulation with four chemical compositions before and after correcting the nonlinear effects. (c) and (d): Binary images segmented based on (a) and (b) respectively. The colorful contours show the shape of the phantom particles of four different chemical compositions.

reconstructions after correction are free of these artifacts, and are in good agreement with the ground-truth phantom, as the error metrics were reduced (Table 2).

For the third simulation, we see cupping artifacts (Fig. 13(a)) with features observed in the previous two cases. First, there are dark streaks and underestimated gray levels. Second, the contrast between different chemical compositions is blurred. These artifacts were corrected after applying the correction algorithm (Fig. 13(b)).

The segmented binary images after correction (Fig. 13(d)) show a stack of different chemical compositions at the borders of some particles. However, these misclassified pixels are not caused by the cupping artifacts, but due to the limitation of the global thresholding [17]. The gray levels in the reconstruction image are continuously dropping at the borders. These pixels were classified into particles of smaller gray levels. Despite the imperfect segmentation, the correction algorithm converged to a result free from cupping artifacts, which also indicates the good robustness of the algorithm.

4. Conclusion

In this paper, we proposed an iterative algorithm to automatically correct the cupping artifacts in tomographic reconstructions from HAADF-STEM projections with nonlinearly damping intensities using only the projection data. The correction is based on modeling the nonlinear relationship between projection intensities and sample thickness as an exponential function.

We showed that the algorithm is an effective tool in achieving better tomographic reconstructions. It successfully corrected the nonlinear damping effects and the subsequent cupping artifacts in three cases where one, two and four chemical compositions are present respectively. The correction is useful for improving the accuracy of morphological analysis and compositional analysis for 3D nanostructures and nanomaterials. In addition, users can benefit from it in enhancing the Z-contrast between chemical compositions as well as in facilitating incorporating prior knowledge to correct the missing wedge artifacts.

1 Note that the algorithm is only applicable to
2 samples consist of several chemical compositions
3 with homogeneous densities that can be segmented
4 based on images gray levels. This is because the
5 graylevel-based segmentation method fails easily
6 when the chemical compositions are mixed or have
7 similar atomic numbers. Moreover, this segmenta-
8 tion method is a global thresholding method. It
9 may lead to poor initial segmentation results and
10 consequently failed corrections when the cupping
11 artifacts are very strong. Consequently, the next
12 step of improving the algorithm is to incorporate
13 advanced segmentation methods or spectroscopic
14 techniques to determine the distributions of chem-
15 ical compositions.

16 5. Acknowledgement

17 This research is supported by the Dutch Technol-
18 ogy Foundation STW (<http://www.stw.nl/>), which
19 is part of the Netherlands Organization for Scien-
20 tific Research (NWO), and which is partly funded
21 by the Ministry of Economic Affairs, Agriculture
22 and Innovation under project number 13314. Fund-
23 ing from the European Research Council (Starting
24 Grant No. COLOURATOMS 335078) is acknowl-
25 edged by S. Bals. The authors would like to thank
26 Dr. Thomas Altantzis and Dr. Bart Goris for pro-
27 viding the experimental data, and Prof. Dr. Luis
28 M. Liz-Marzan for providing the investigated sam-
29 ples.

- [1] P. A. Midgley, M. Weyland, J. M. Thomas, B. F. G. Johnson, Z-Contrast tomography: a technique in three-dimensional nanostructural analysis based on Rutherford scattering, *Chem. Commun.* (2001) 907–908.
- [2] C. Kübel, A. Voigt, R. Schoenmakers, M. Otten, D. Su, T.-C. Lee, A. Carlsson, J. Bradley, Recent advances in electron tomography: TEM and HAADF-STEM tomography for materials science and semiconductor applications, *Microscopy and Microanalysis* 11 (05) (2005) 378–400.
- [3] A. C. Kak, M. Slaney, Principles of computerized tomographic imaging, IEEE press, 1988.
- [4] P. A. Midgley, M. Weyland, 3D electron microscopy in the physical sciences: the development of Z-contrast and EFTEM tomography, *Ultramicroscopy* 96 (3) (2003) 413–431.
- [5] R. Aveyard, Z. Zhong, K. J. Batenburg, B. Rieger, Optimizing experimental parameters for the projection requirement in HAADF-STEM tomography, *Ultramicroscopy* 177 (2017) 84 – 90.
- [6] W. Van den Broek, A. Rosenauer, B. Goris, G. T. Martinez, S. Bals, S. Van Aert, D. Van Dyck, Correction of non-linear thickness effects in HAADF STEM electron tomography, *Ultramicroscopy* 116 (2012) 8–12.
- [7] Z. Zhong, B. Goris, R. Schoenmakers, S. Bals, K. J. Batenburg, A bimodal tomographic reconstruction technique combining EDS-STEM and HAADF-STEM, *Ultramicroscopy* 174 (2017) 35 – 45.
- [8] J. Gregor, T. Benson, Computational Analysis and Improvement of SIRT, *IEEE Transactions on Medical Imaging* 27 (7) (2008) 918–924.
- [9] N. Otsu, A threshold selection method from gray-level histograms, *IEEE Transactions on Systems, Man, and Cybernetics* 9 (1) (1979) 62–66.
- [10] B. Goris, W. Van den Broek, K. J. Batenburg, H. Heidari Mezerji, S. Bals, Electron tomography based on a total variation minimization reconstruction technique, *Ultramicroscopy* 113 (2012) 120–130.

- 1 [11] J. Hsieh, R. C. Molthen, C. a. Dawson, R. H.
2 Johnson, An iterative approach to the beam
3 hardening correction in cone beam CT, *Medi-
4 cal physics* 27 (1) (2000) 23–29.
- 5 [12] K. J. Batenburg, S. Bals, J. Sijbers, C. Kübel,
6 P. A. Midgley, J. C. Hernandez, U. Kaiser,
7 E. R. Encina, E. a. Coronado, G. Van Tende-
8 loo, 3D imaging of nanomaterials by discrete
9 tomography, *Ultramicroscopy* 109 (6) (2009)
10 730–740.
- 11 [13] X. Zhuge, W. J. Palenstijn, K. J. Baten-
12 burg, TVR-DART: a more robust algorithm
13 for discrete tomography from limited projec-
14 tion data with automated gray value estima-
15 tion, *IEEE Transactions on Image Processing*
16 25 (1) (2016) 455–468.
- 17 [14] D. B. Williams, C. B. Carter, *The transmis-
18 sion electron microscope*, Springer, 1996.
- 19 [15] J. Lagarias, J. Reeds, H. Wright, P. Wright,
20 Convergence properties of the Nelder–Mead
21 simplex method in low dimensions, *Journal of
22 Optimization* 9 (1) (1998) 112–117.
- 23 [16] A. Sánchez-Iglesias, M. Grzelczak, T. Al-
24 tantzis, B. Goris, J. Pérez-Juste, S. Bals,
25 G. Van Tendeloo, S. H. Donaldson, B. F.
26 Chmelka, J. N. Israelachvili, L. M. Liz-
27 Marzán, Hydrophobic interactions modulate
28 self-assembly of nanoparticles, *ACS Nano*
29 6 (12) (2012) 11059–11065.
- 30 [17] K. Batenburg, J. Sijbers, Adaptive threshold-
31 ing of tomograms by projection distance min-
32 imization, *Pattern Recognition* 42 (10) (2009)
33 2297 – 2305.