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

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, 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

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

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

It is important to correct the nonlinear effects and the subsequent cupping artifacts for three reasons. First of all, compositional analysis based on gray levels becomes difficult when the cupping artifacts occur, as gray levels are not proportional anymore to density and atomic numbers. Second, morphological analysis based on segmentation of reconstruction images is hindered by the cupping artifacts. Some straightforward segmentation methods, e.g. Otsu's method [9], require that for each chemical composition there should be one constant gray level. Third, the nonlinear effects limit applying advanced reconstruction algorithms to address a critical issue of ET – the missing wedge artifacts caused by the limited tilt range of the sample. Algorithms such as total variation minimization [10] reduce the missing wedge artifacts by incorporating prior knowledge i.e. sparsity of the unknown sample. Nevertheless, these algorithms have an even stronger requirement for the linear forward model which is inaccurate due to the nonlinear effects.

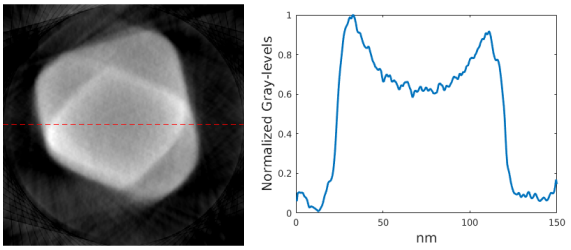
Despite these shortcomings of using uncorrected data, there are few publications addressing the nonlinearity issue in ET [5, 6]. Nonlinear effects are usually ignored or mitigated during image acquisition by increasing the inner angle of the HAADF detector but at the cost of losing signal strength [5]. An alternative to adjusting the acquisition parameters is to correct the measured data in a post-processing step by linearizing the projection data, provided that the incident beam intensity is known [6]. The method described here requires only the HAADF signal, consequently, it can be applied to correct cupping artifacts in many existing datasets acquired in a conventional manner. The mathematical model of nonlinearity and the concept of linearization in [6] are also used in this paper, which will be explained in Section 2.1.

Here, we propose an iterative algorithm to automatically correct the nonlinear effects and the cupping artifacts. It does not require the extra measurement of the incident beam intensity as in [6]. Instead, it automatically models the nonlinear effects given the projection data. The algorithm iteratively searches for the minimal distance between the acquired projections and the nonlinear re-projections of chemical compositions by varying the nonlinear model and the reconstruction image, so as to estimate a nonlinear relationship between the measured HAADF-STEM intensities and sample thickness for all chemical compositions. The algorithm contains the following steps in every iteration: first a reconstruction image with continuous gray levels is made; then the image is segmented into several binary images, each of which corresponds to a chemical composition; after that, the nonlinear effects are modeled by minimizing the projection distance; based on the model, the projection data is linearized at last. The concept of iterative correction has been used to correct beam hardening artifacts for X-ray computed tomography, which is similarly caused by nonlinear intensities [11, 12, 13].

Our approach is only applicable to samples consisting of several chemical compositions with uniform densities, such as homogeneous or core-shell particles. It is assumed that for these samples the volumetric distributions of the compositions can be approximated well by segmenting the reconstructed image based on gray levels and that this segmentation improves as the correction model applied to the measured data becomes more accurate. In fact, these kinds of samples are commonly studied in

materials science. For example, the samples typically studied in the context of discrete tomography [14, 15] match the requirements.

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



(a) (b)

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.

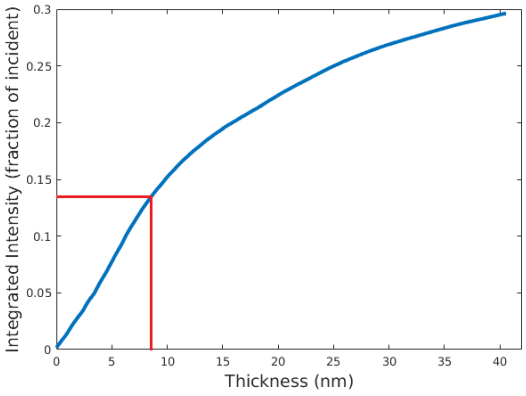


Figure 2: Normalized HAADF signal intensity w.r.t the thickness of Au slabs mistilted 10 degrees from the [100] zone axis about the <100> axis, simulated using the multislice method [5]. The accelerating voltage is 200 kV, the convergence angle is 10 mrad and the detector angular range is 50 - 250 mrad.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

$$p_t = I_0 \exp(-\sum_e^K \mu_e t), \tag{1}$$

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:

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

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

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

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

We also assume the chemical compositions are not mixed and voxels are small enough to resolve every chemical composition, which means that in each voxel only one element is present. As stated in the introduction, this algorithm is applied to samples with uniform density. Thus we assume that each chemical composition is either present (1) or absent (0) in each voxel. The distribution of chemical composition e is described by binary variables s_{ej}, where $j = 1, \ldots, N$ is the index of voxel.

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

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

where $i = 1, \ldots, M$.

2.2. The Correction Algorithm

The basis of the correction algorithm is to estimate the nonlinear relationship of Eq. 3 based on the reconstructed distributions of chemical compositions. The procedures of the automatic correction algorithm are given in the flowchart (Fig. 3). The correction is realized iteratively through the following steps: (1) make a reconstruction image based on the linear model from the projections; (2) segment the reconstruction into a series of binary images, one for each chemical composition; (3) estimate the parameters of the nonlinear model in Eq. 3 given the projections and the binary images; (4) reduce the nonlinearities in the projections through a rescaling of the intensities based on the nonlinear model.

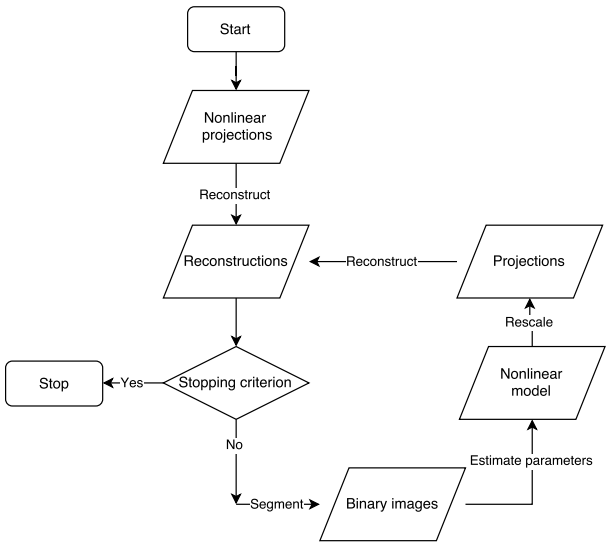


Figure 3: Flowchart of the correction algorithm

Before we explain the steps explicitly, we establish 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 $\boldsymbol{W} = \{w_{ij}\}$, $\boldsymbol{\mu} = \{\mu_e\}$ 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.

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. [14]), 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 and then segment the reconstruction into binary images $\mathbf{S}$.

The reconstruction is computed using the simultaneous iterative reconstruction technique (SIRT) [8] which solves the following least-squares problem:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \parallel \mathbf{p}_{lin} - \mathbf{W}\mathbf{x} \parallel_2^2 . \tag{6}$$

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

The input for this step is a set of "linearized" projections p_{lin}. For the first iteration, they are just the acquired projections. For the other iterations, they are adopted as the projections that have been rescaled in the previous iteration, which will be explained in Step 4.

Step 2: Segmentation

The binary images are then determined by segmenting the reconstruction image $\mathbf{x}$. As gray levels are related to atomic numbers, we segment the SIRT reconstruction by global thresholding. The thresholds for the segmentation are determined by solving the following optimization problem:

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

The solution of this problem is found by straightforward (brute-force) sampling of the space of thresholds, each time evaluating the cost function. In practice, the thresholds are sampled from the minimum to the maximum of gray levels of the SIRT reconstruction in Step 1.

The first iteration is again an exception since parameters have not yet been estimated and the objective function cannot be computed. Thus, the above segmentation method is not applicable. Instead, the thresholds are determined using Otsu's method which finds optimal thresholds based on the gray level histograms [9].

Step 3: Nonlinear parameters estimation

Given the binary images, we can update the free parameters of the nonlinear model $I_0, p_b, \boldsymbol{\mu}$ by minimizing the objective function, which is a nonlinear regression problem. This nonlinear regression problem is solved using the Nelder–Mead method [17]. To improve the stability of the regression, the three parameters are estimated separately and iteratively in an inner loop:

For $l = 1 : L$

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

here l is the iteration number of the inner loop. The estimation algorithm requires initial parameter values. In the experiments, we found that the initial values have little influence on the convergence result but proper initial values help to converge faster. Since we know that the beam intensity I_0 should be at least the maximal image intensity and that the attenuation coefficients μ and the bias intensity p_b are very small, we can start from $I_0^1 = max(\mathbf{p})$, $p_b^1 = 0$ and $\boldsymbol{\mu}^1 = \mathbf{0}$, which were used in all the experiments in the paper.

Step 4: Projection intensities rescaling

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

$$\mathbf{p}_{lin}^{'} = \log \frac{I_0 + p_b - \mathbf{p}}{I_0}, \tag{9}$$

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

3. Experiments and Simulations

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

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

3.1. Experiments

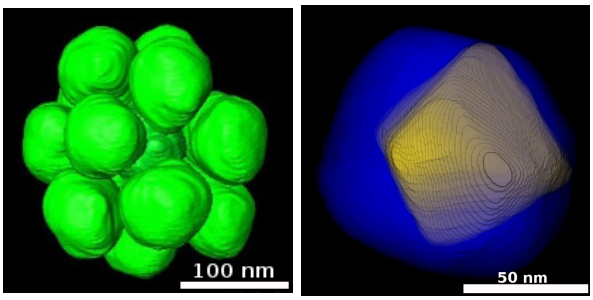
The first experimental sample is an assembly consisting of 16 Pt nanoparticles, each of which has a diameter of about 60 nm (Fig. 4(a)) [18]. It has only one chemical composition and a relatively more complex structure than the second sample.

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

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

3.1.1. Results: Nanoparticle Assembly

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



(a) (b)

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.

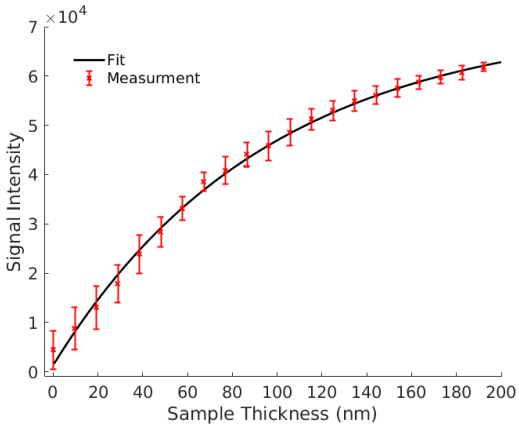


Figure 6: The nonlinear damping model fitted for projection signal intensity w.r.t. sample thickness of the nanoparticle assembly. The error bars indicate mean intensities and the standard deviations of the projection data.

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^o to 74^o	-75^o to 75^o
projection angle increment	2^o	5^o

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^o to 75^o. We thus compared reconstructions using advanced reconstruction algorithms: total-variation minimization (TV-min) [10], discrete algebraic reconstruction technique (DART) [14] and total variation regularized DART (TVR-DART) [15], which incorporate 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.

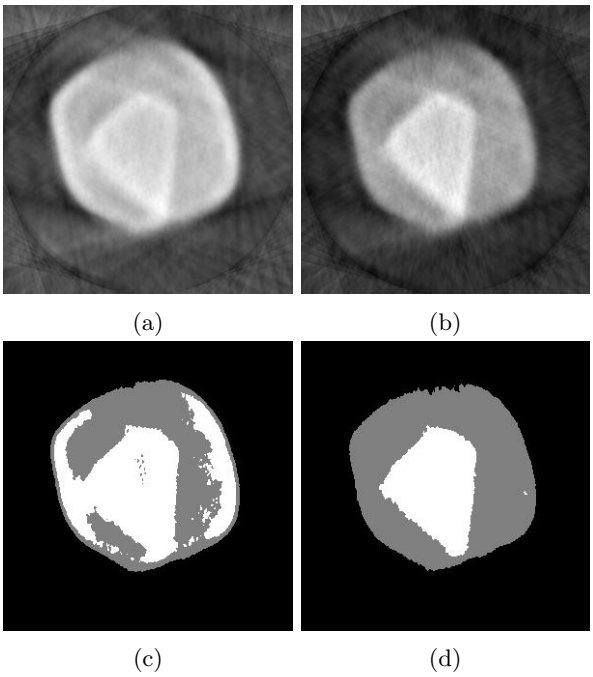
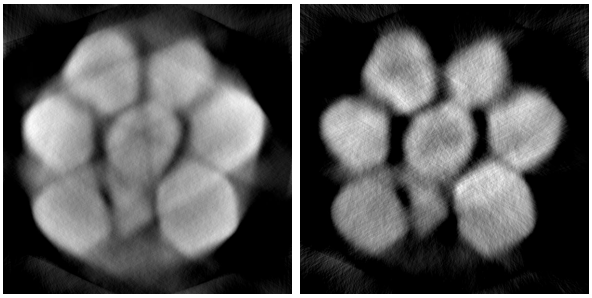
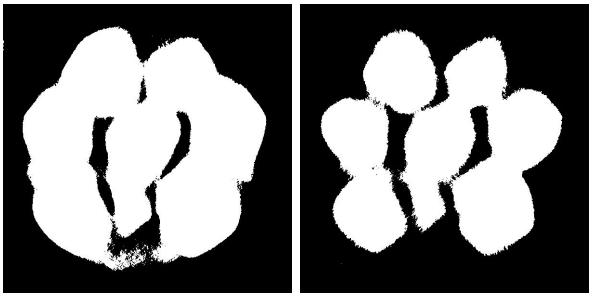


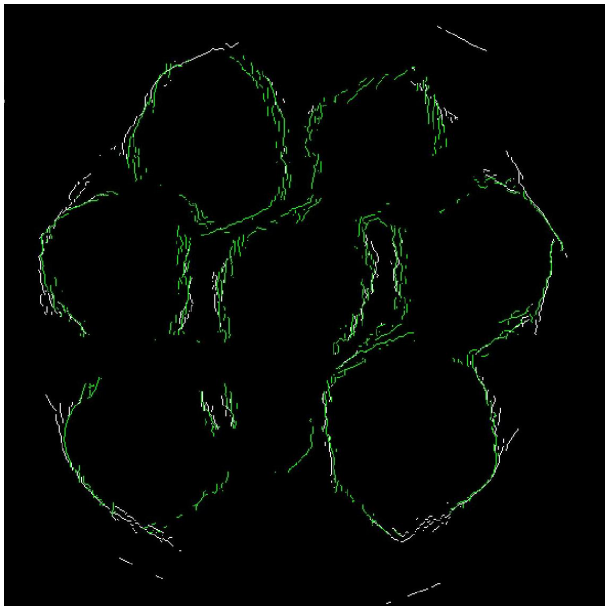
Figure 7: (a) and (b): SIRT reconstructions of the Au-Ag nanoparticle from the nonlinear projections and corrected projections. (c) and (d): Binary images segmented based on the reconstruction images (a) and (b) respectively.



(a) (b)



(c) (d)



(e)

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).

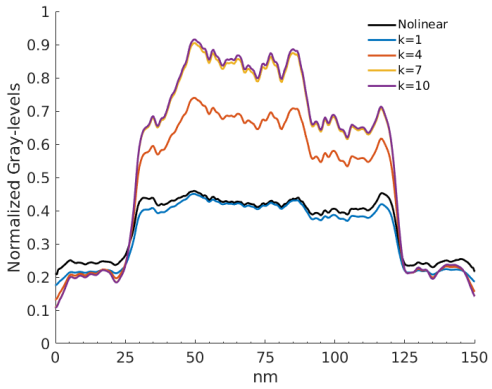


Figure 8: Cross-section line profiles of the SIRT reconstructions of the Au-Ag nanoparticle at different iterations.

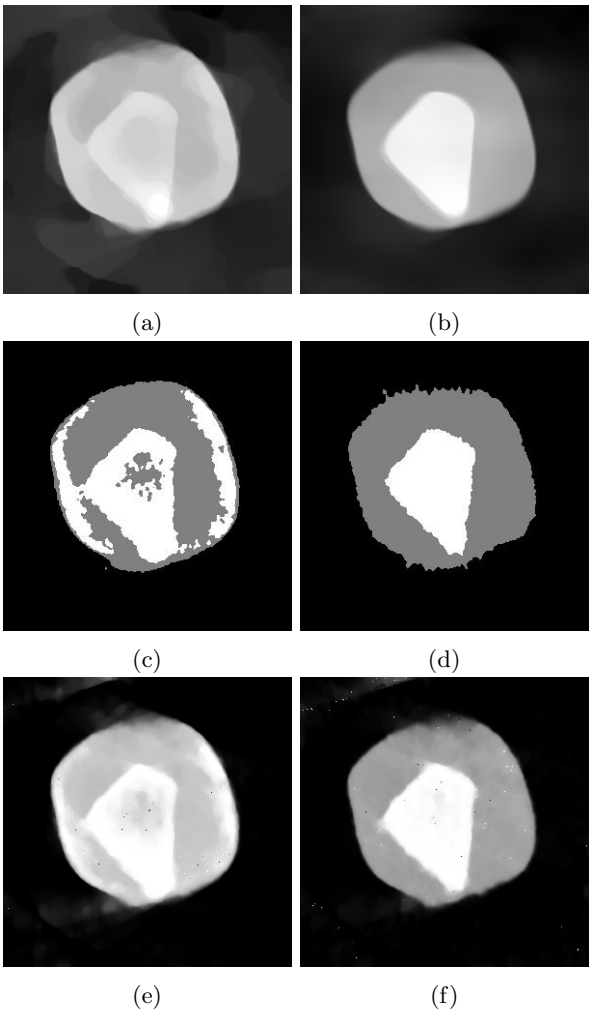


Figure 9: (a)/(b), (c)/(d) and (e)/(f) are the TV-min, DART and TVR-DART reconstructions of the Au-Ag nanoparticle from projections before/after the correction respectively.

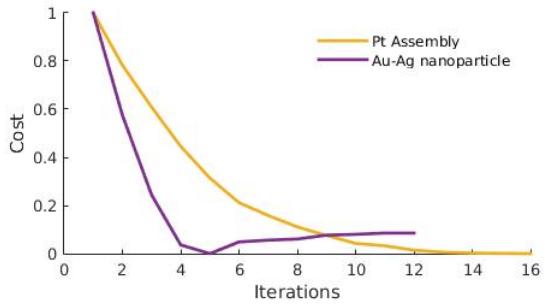


Figure 10: The residuals of cost function (Eq. 10) w.r.t. iterations for the two experimental datasets.

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:

$$err = \frac{1}{K}\sum_{e}^{K}\sum_{j}^{N} \parallel s_{ej} - g_{ej} \parallel / \sum_{j}^{N} g_{ej}, \qquad (10)$$

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.

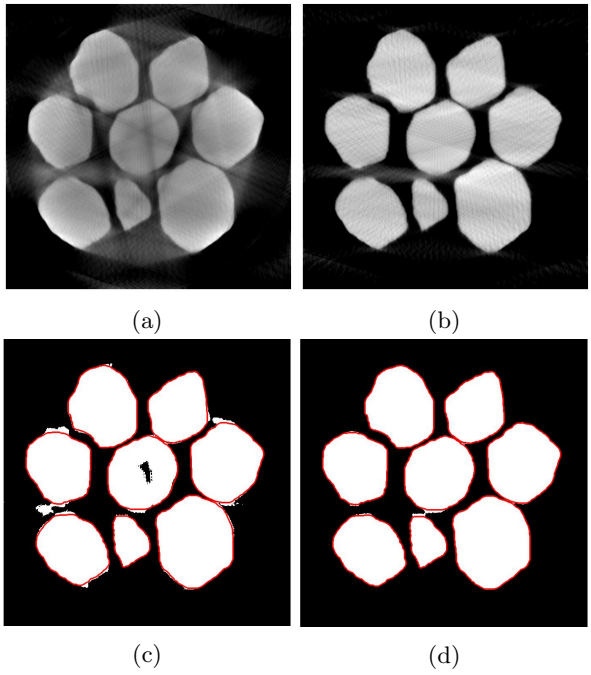


Figure 11: (a) and (b): SIRT Reconstruction images of the nanoparticle assembly phantom simulation before and after the nonlinearity correction. (c) and (d): Binary images segmented based on (a) and (b) respectively. The red contour shows the shape of the phantom.

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 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)).

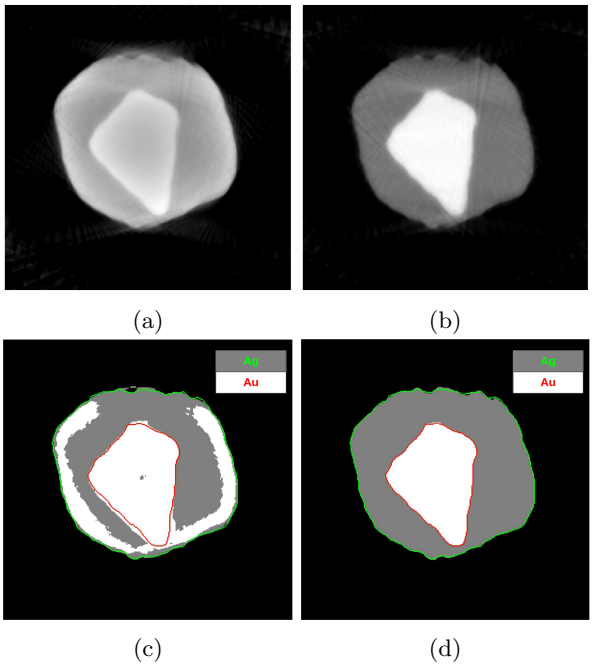


Figure 12: (a) and (b): SIRT reconstructions of the Au-Ag nanoparticle phantom simulation before and after the nonlinearity correction. (c) and (d): Binary images segmented based on (a) and (b) respectively. The red and green contours show the shape of the phantoms of Au and Ag respectively.

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 overestimated 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 by noting 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 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%

facts caused the loss of contrast between Au and Ag, even though Au and Ag have a large difference in atomic number. As a result, many pixels were misclassified in the binary images (Fig. 7(b)). The algorithm corrected the experimental data and enhanced 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 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).

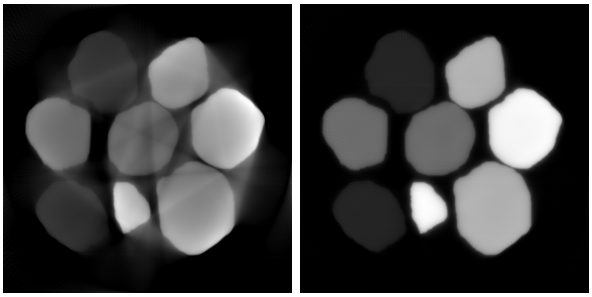
For the third simulation, we see cupping artifacts (Fig. 13(a)) with features observed in the previous two cases. First, there are dark streaks and underestimated gray levels. Second, the contrast between different chemical compositions is blurred. These artifacts were corrected after applying the correction 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 [19]. 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.

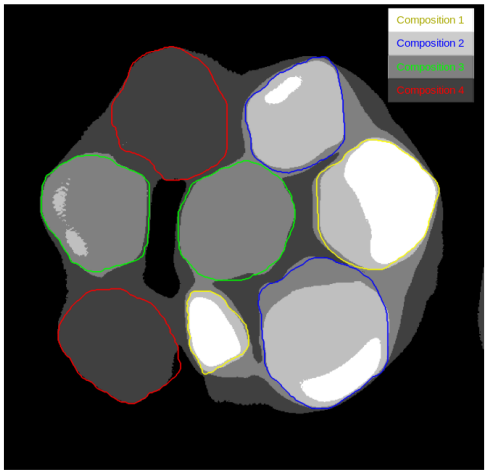
4. Conclusion

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.

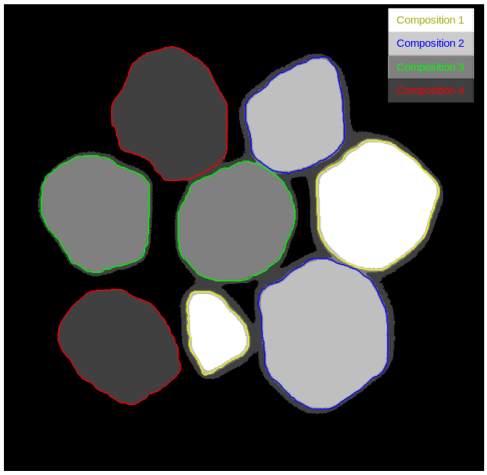
We showed that the algorithm is an effective tool in achieving better tomographic reconstructions. It successfully corrected the nonlinear damping effects and the subsequent cupping artifacts in three cases where one, two and four chemical compositions are present respectively. The correction is useful 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.



(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] 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.

[11] G. Van Gompel, K. Van Slambrouck, M. Defrise, 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] 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] E. V. de Casteele, D. V. Dyck, J. Sijbers, E. Raman, An energy-based beam hardening model in tomography, Physics in Medicine & Biology 47 (23) (2002) 4181.

[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] 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] D. B. Williams, C. B. Carter, The transmission electron microscope, Springer, 1996.

[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] 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] 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 $Z^{\sim 2}$, 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,

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

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

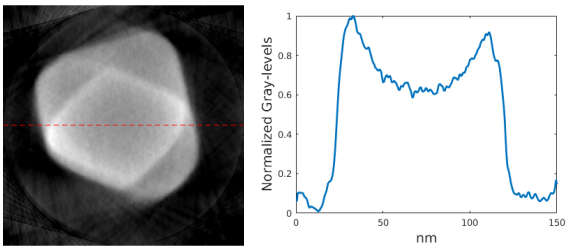
It is important to correct the nonlinear effects and the subsequent cupping artifacts for three reasons. First of all, compositional analysis based on gray levels becomes difficult when the cupping artifacts occur, as gray levels are not proportional anymore to density and atomic numbers. Second, morphological analysis based on segmentation of reconstruction images is hindered by the cupping artifacts. Some straightforward segmentation methods, e.g. Otsu's method [9], require that for each chemical composition there should be one constant gray level. Third, the nonlinear effects limit applying advanced reconstruction algorithms to address a critical issue of ET – the missing wedge artifacts caused by the limited tilt range of the sample. Algorithms such as total variation minimization [10] reduce the missing wedge artifacts by incorporating prior knowledge i.e. sparsity of the unknown sample. Nevertheless, these algorithms have an even stronger requirement for the linear forward model which is inaccurate due to the nonlinear effects.

Despite these shortcomings of using uncorrected data, there are few publications addressing the nonlinearity issue in ET [5, 6]. Nonlinear effects are usually ignored or mitigated during image acquisition by increasing the inner angle of the HAADF detector but at the cost of losing signal strength [5]. An alternative to adjusting the acquisition parameters is to correct the measured data in a post-processing step by linearizing the projection data, provided that the incident beam intensity is known [6]. The method described here requires only the HAADF signal, consequently, it can be applied to correct cupping artifacts in many existing datasets acquired in a conventional manner. The mathematical model of nonlinearity and the concept of linearization in [6] are also used in this paper, which will be explained in Section 2.1.

Here, we propose an iterative algorithm to automatically correct the nonlinear effects and the cupping artifacts. It does not require the extra measurement of the incident beam intensity as in [6]. Instead, it automatically models the nonlinear effects given the projection data. The algorithm iteratively searches for the minimal distance between the acquired projections and the nonlinear re-projections of chemical compositions by varying the nonlinear model and the reconstruction image, so as to estimate a nonlinear relationship between the measured HAADF-STEM intensities and sample thickness for all chemical compositions. The algorithm contains the following steps in every iteration: first a reconstruction image with continuous gray levels is made; then the image is segmented into several binary images, each of which corresponds to a chemical composition; after that, the nonlinear effects are modeled by minimizing the projection distance; based on the model, the projection data is linearized at last. A similar concept of iterative correction has been introduced to correct beam hardening artifacts for X-ray computed tomography [11].

Our approach is only applicable to samples consisting of several chemical compositions with uniform densities, such as homogeneous or core-shell particles. It is assumed that for these samples the volumetric distributions of the compositions can be approximated well by segmenting the reconstructed image based on gray levels and that this segmentation improves as the correction model applied to the measured data becomes more accurate. In fact, this kind of samples are commonly studied in materials science. For example, the samples typically studied in the context of discrete tomography [12, 13] match the requirements.

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



(a) (b)

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.

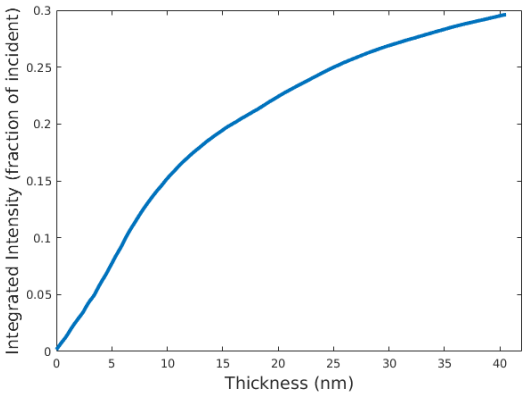


Figure 2: Normalized HAADF signal intensity w.r.t the thickness of Au slabs mistilted 10 degrees from the [100] zone axis about the <100> axis, simulated using the multislice method [5]. The accelerating voltage is 200 kV, the convergence angle is 10 mrad and the detector angular range is 50 - 250 mrad.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

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, 14], it is illustrated that the HAADF detector collects electrons complementary to the electrons scattered to angles smaller than its inner detector angle. 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

$$p_t = I_0 \exp(-\sum_e^K \mu_e t), \tag{1}$$

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:

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

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

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

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

We also assume the chemical compositions are not mixed and voxels are small enough to resolve every chemical composition, which means that in each voxel only one element is present. As stated in Introduction, this algorithm is applied to samples with uniform density. Thus we assume that each chemical composition is either present (1) or absent (0) in each voxel. The distribution of chemical composition e is described by binary variables s_{ej}, where $j = 1, \ldots, N$ is the index of voxel.

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

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

where $i = 1, \ldots, M$.

2.2. The Correction Algorithm

The basis of the correction algorithm is to estimate the nonlinear relationship of Eq. 3 based on the reconstructed distributions of chemical compositions. The procedures of the automatic correction algorithm are given in the flowchart (Fig. 3). The correction is realized iteratively through the following steps: (1) make a reconstruction image based on the linear model from the projections; (2) segment the reconstruction into a series of binary images, one for each chemical composition; (3) estimate the parameters of the nonlinear model in Eq. 3 given the projections and the binary images; (4) reduce the nonlinearities in the projections through a rescaling of the intensities based on the nonlinear model.

Before we explain the steps explicitly, we establish 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

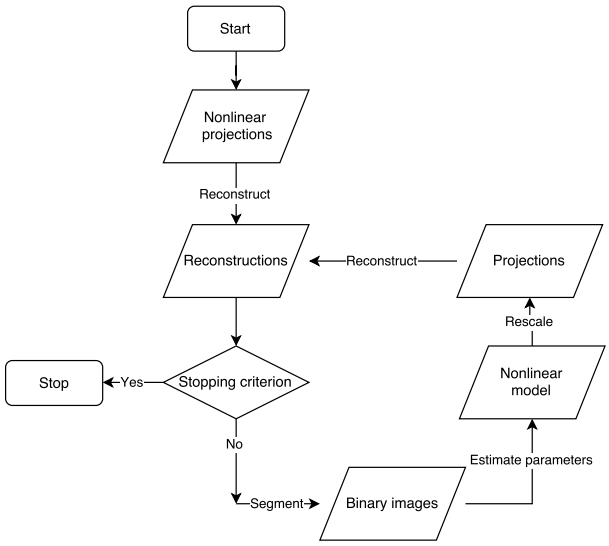


Figure 3: Flowchart of the correction algorithm

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 $\boldsymbol{W} = \{w_{ij}\}$, $\boldsymbol{\mu} = \{\mu_e\}$ 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

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

The reconstruction is computed using the simultaneous iterative reconstruction technique (SIRT) [8] which solves the following least-squares problem:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \parallel \mathbf{p}_{lin} - \mathbf{W}\mathbf{x} \parallel_2^2 . \tag{6}$$

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

The input for this step is a set of "linearized" projections p_{lin}. For the first iteration, they are just the acquired projections. For the other iterations, they are adopted as the projections that have been rescaled in the previous iteration, which will be explained in Step 4.

Step 2: Segmentation

The binary images are then determined by segmenting the reconstruction image $\mathbf{x}$. As gray levels are related to atomic numbers, we segment the SIRT reconstruction by global thresholding. The thresholds for the segmentation are determined by solving the following optimization problem:

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

The solution of this problem is found by straightforward (brute-force) sampling of the space of thresholds, each time evaluating the cost function. In practice, the thresholds are sampled from the minimum to the maximum of gray levels of the SIRT reconstruction in Step 1.

The first iteration is again an exception since parameters have not yet been estimated and the objective function cannot be computed. Thus, the above segmentation method is not applicable. Instead, the thresholds are determined using Otsu's method which finds optimal thresholds based on the gray level histograms [9].

Step 3: Nonlinear parameters estimation

Given the binary images, we can update the free parameters of the nonlinear model $I_0, p_b, \boldsymbol{\mu}$ by minimizing the objective function, which is a nonlinear regression problem. This nonlinear regression problem is solved using the Nelder–Mead method [15]. To improve the stability of the regression, the three parameters are estimated separately and iteratively in an inner loop:

For $l = 1 : L$

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

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

Step 4: Projection intensities rescaling

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

$$\mathbf{p}_{lin}^{'} = \log \frac{I_0 + p_b - \mathbf{p}}{I_0}, \tag{9}$$

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

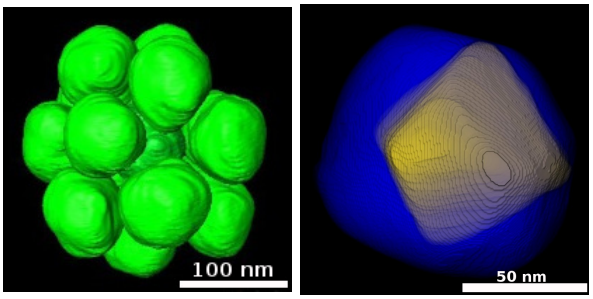
3. Experiments and Simulations

We report the correction of cupping artifacts for two sets of experimental data and three phantom simulations. The experimental data show strong nonlinear effects because the samples consist of thick metallic materials. Two phantom simulations resembling the experimental data were performed, as ground-truth is missing for quality assessment of the reconstruction image due to the lack of other measurement methods. In addition, a phantom of four chemical compositions was simulated to investigate the robustness of the algorithm when more chemical compositions are present, as the experimental samples consist of only one or two chemical compositions.

3.1. Experiments

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

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



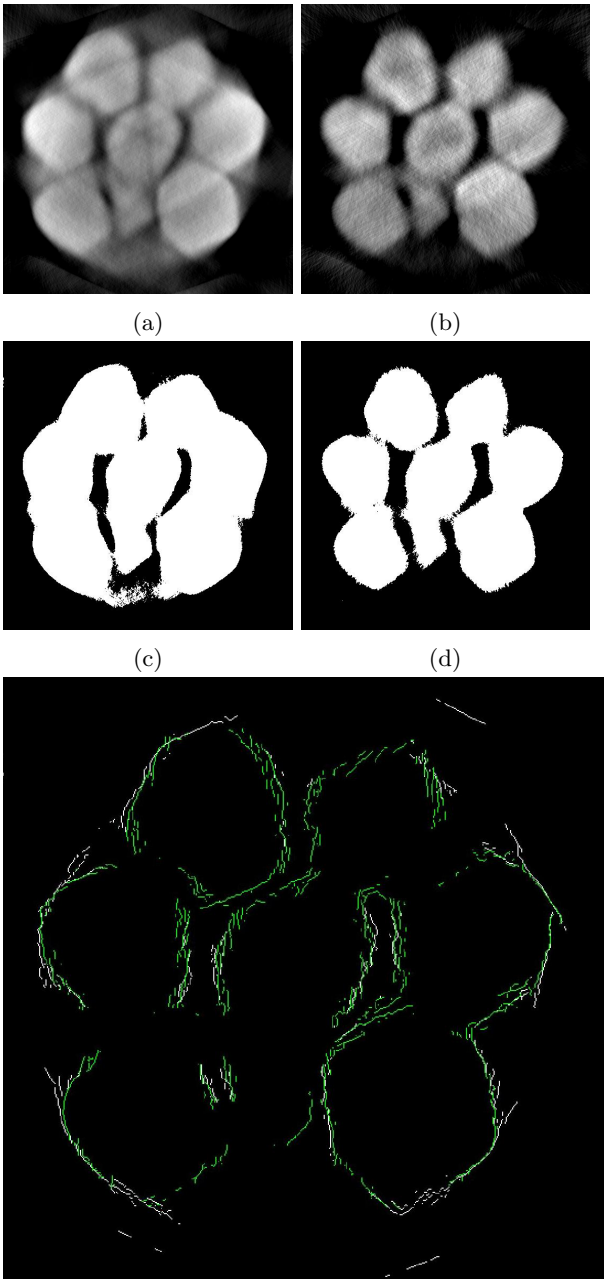
(a) (b)

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

3.1.1. Results: Nanoparticle Assembly

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

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.



(a) (b)

(c) (d)

(e)

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).

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^o to 74^o	-75^o to 75^o
projection angle increment	2^o	5^o

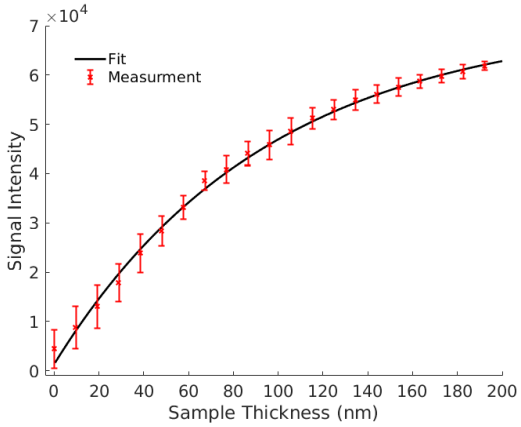


Figure 6: The nonlinear damping model fitted for projection signal intensity w.r.t. sample thickness of the nanoparticle assembly. The error bars indicate mean intensities and the standard deviations of the projection data.

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^o to 75^o. 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 incorporate 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.

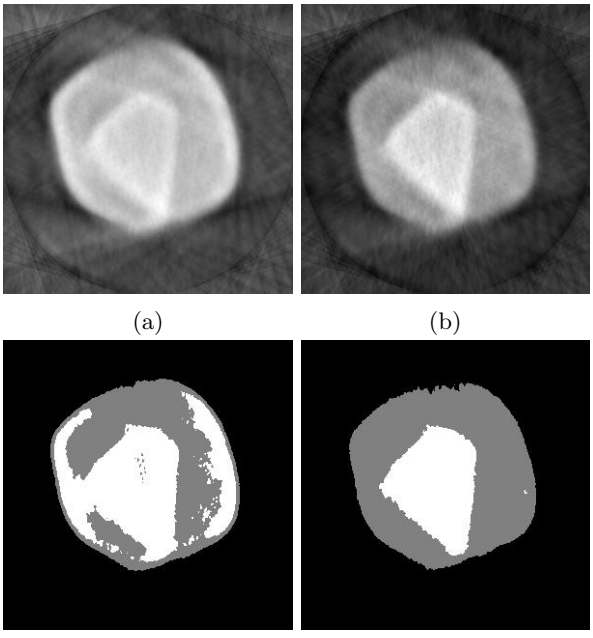


Figure 7: (a) and (b): SIRT reconstructions of the Au-Ag nanoparticle from the nonlinear projections and corrected projections. (c) and (d): Binary images segmented based on the reconstruction images (a) and (b) respectively.

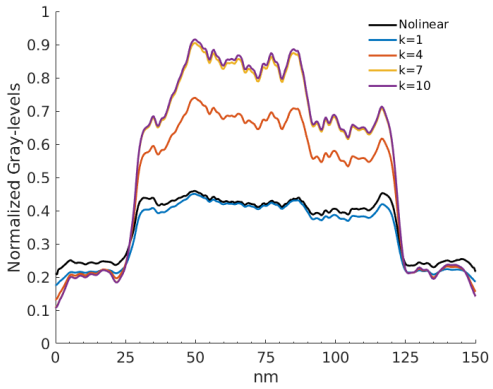


Figure 8: Cross-section line profiles of the SIRT reconstructions of the Au-Ag nanoparticle at different iterations.

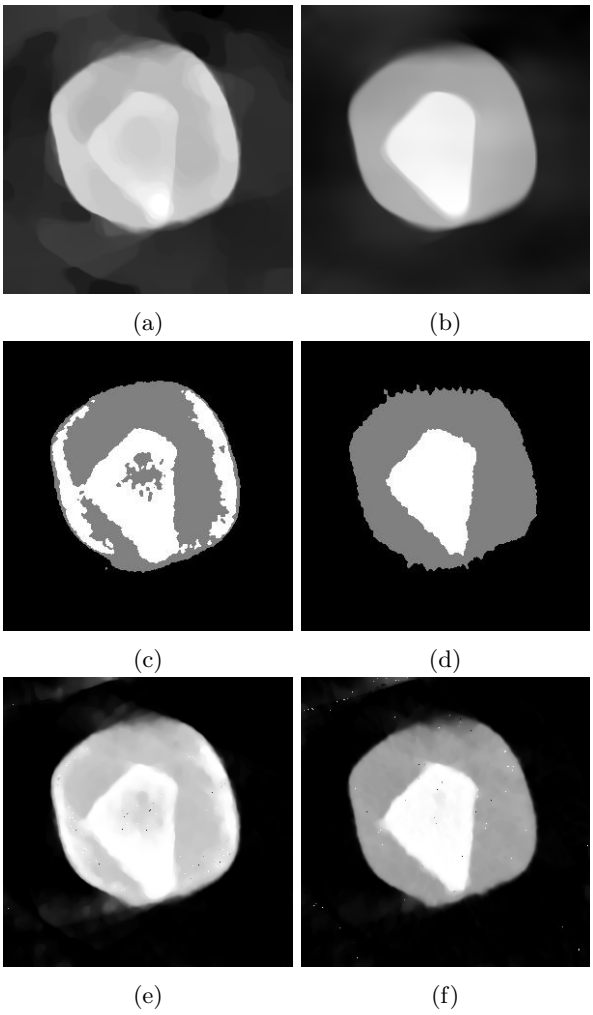


Figure 9: (a)/(b), (c)/(d) and (e)/(f) are the TV-min, DART and TVR-DART reconstructions of the Au-Ag nanoparticle from projections before/after the correction respectively.

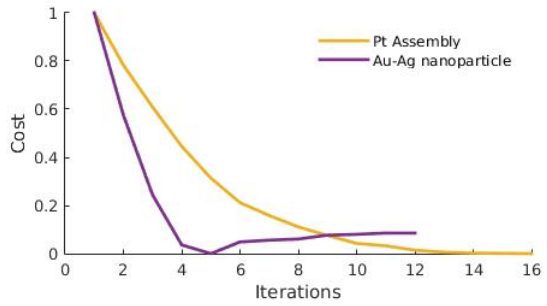


Figure 10: The residuals of cost function (Eq. 10) w.r.t. iterations for the two experimental datasets.

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:

$$err = \frac{1}{K}\sum_{e}^{K}\sum_{j}^{N} \parallel s_{ej} - g_{ej} \parallel / \sum_{j}^{N} g_{ej}, \qquad (10)$$

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.

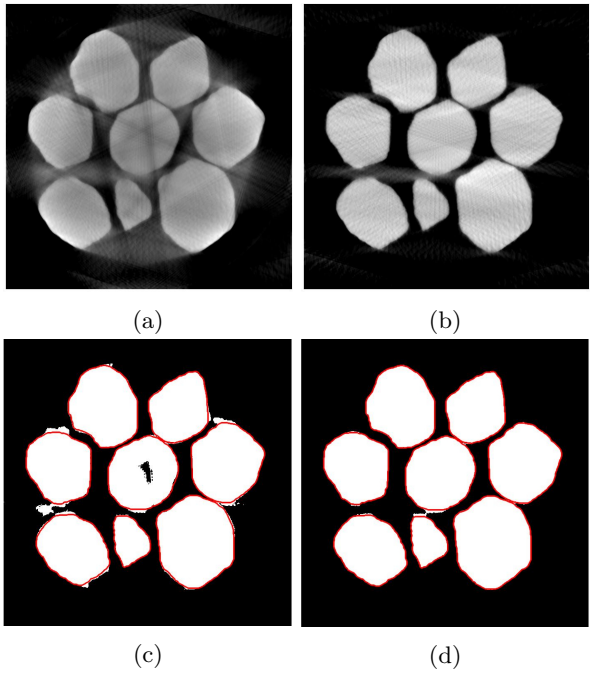


Figure 11: (a) and (b): SIRT Reconstruction images of the nanoparticle assembly phantom simulation before and after the nonlinearity correction. (c) and (d): Binary images segmented based on (a) and (b) respectively. The red contour shows the shape of the phantom.

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%

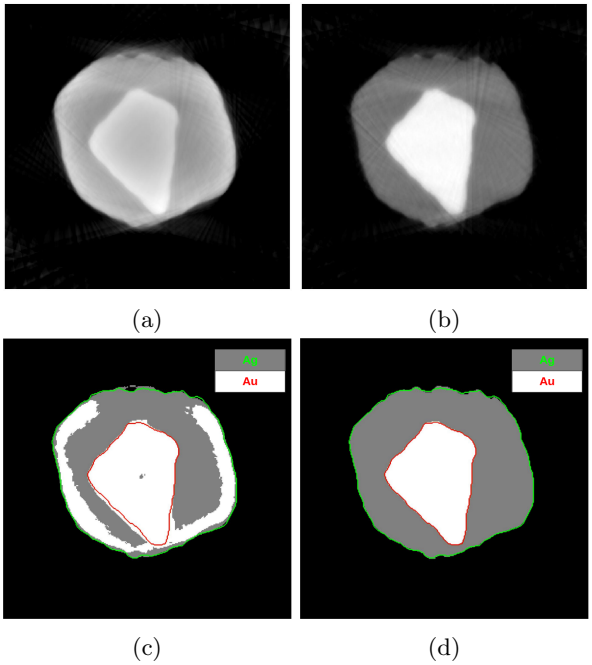


Figure 12: (a) and (b): SIRT reconstructions of the Au-Ag nanoparticle phantom simulation before and after the nonlinearity correction. (c) and (d): Binary images segmented based on (a) and (b) respectively. The red and green contours show the shape of the phantoms of Au and Ag respectively.

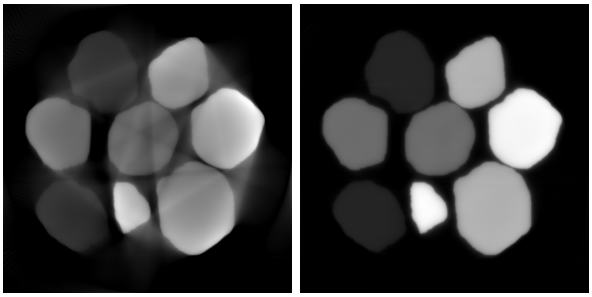
hand, as dark streaks elongated from the gaps between particles. On the other, they appear as underestimated gray levels in the interior, for which we see missing pixels in the binary image (Fig. 5(c)).

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 overestimated 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.

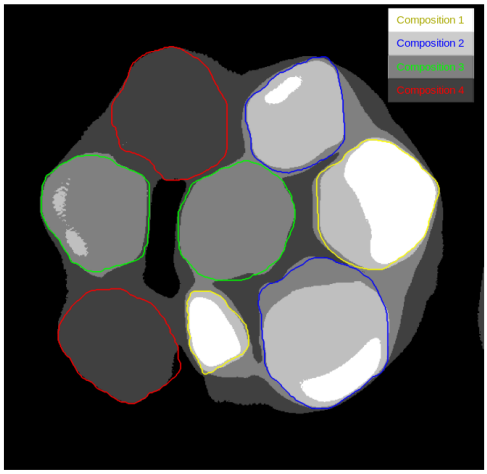
In the initial SIRT reconstruction image of the Au-Ag nanoparticle (Fig. 7(a)), the cupping artifacts caused the loss of contrast between Au and Ag, even though Au and Ag have a large difference in atomic numbers. As a result, many pixels were misclassified in the binary images (Fig. 7(b)). The algorithm 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.

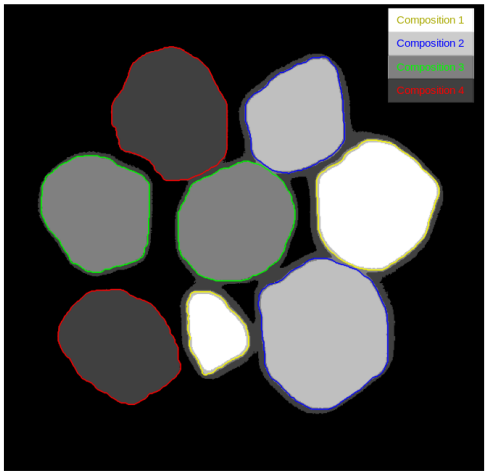
The first two phantom simulations show artifacts (in Fig. 11(a) and Fig. 12(a)) very similar to those from the experimental data, which indicates that the modeling of nonlinear effects is accurate. Both



(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.

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.

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] 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.

[11] 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.

[12] 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.

[13] 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.

[14] D. B. Williams, C. B. Carter, The transmission electron microscope, Springer, 1996.

[15] 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.

[16] 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.

[17] K. Batenburg, J. Sijbers, Adaptive thresholding of tomograms by projection distance minimization, Pattern Recognition 42 (10) (2009) 2297 – 2305.