A model based reconstruction technique for depth sectioning with scanning transmission electron microscopy

W. Van den Broek *, S. Van Aert, D. Van Dyck
EMAT, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerp, Belgium

ARTICLE INFO

Keywords:
Atomic resolution tomography
HAADF STEM
Depth sectioning
Algebraic reconstruction technique

ABSTRACT

Depth sectioning in high angular annular dark field scanning transmission electron microscopy is considered a candidate for three-dimensional characterization on the atomic scale. However at present the depth resolution is still far from the atomic level, due to strong limitations in the opening angle of the beam. In this paper we introduce a new, parameter based tomographic reconstruction algorithm that allows to make maximal use of the prior knowledge about the constituent atom types and the microscope settings, so as to retrieve the atomic positions and push the resolution to the atomic level in all three dimensions.

1. Introduction

With depth sectioning in high angular annular dark field scanning transmission electron microscopy (HAADF STEM) the depth of field is made smaller than the specimen thickness such that by adjusting the defocus, different heights within the specimen can be imaged. This is made possible by recent developments in aberration correction, allowing the convergence semi-angle $\alpha_0$ of the electron beam to increase. This improves the lateral resolution as well as the depth resolution of the coherent contribution to the probe intensity, since these are approximately equal to $0.61/\lambda_0$ and $2\lambda/\alpha_0^2$, respectively, with $\lambda$ the electron wavelength. The sample is optically sliced by recording a set of images, each with a focus value that has been changed by a fixed amount. Such images are known as optical sections. This focus series forms a three-dimensional (3D) data stack through which one can take horizontal or vertical slices, as is illustrated in Fig. 1.

Thus far results from depth sectioning in EM are interpreted by analyzing horizontal and vertical slices through the 3D data stack. Experimental demonstrations are restricted to imaging heavy atoms in a lighter surrounding: vertical slices of the data stack show elongated bright streaks centred on the atom positions, superposed on a much weaker background. In [1] van Benthem et al. demonstrate the detection of single Hf atoms in an environment of amorphous $\text{SiO}_2$. In [2] a Pt atom on a carbon support is detected, while in [3] the detection of Au atoms in a Si nanotube is demonstrated.

An underlying assumption of depth sectioning with HAADF STEM, often unspoken, is that the image formation is incoherent:

- The image of the sample is the sum of the images of the individual atoms.
- The image of a single atom is a projection of the convolution of the probe intensity with an object function.

As such, a heavy atom within the depth of field produces a bright and localized image while the images of lighter atoms outside the depth of field are more dim and more smeared out; a heavy atom in a surrounding of lighter atoms is thus imaged as a bright signal superposed on a non-distinct background. This is illustrated in Fig. 2.

Rather than interpreting the individual images of the focus series directly, we will use them to find estimates of the positions of all the atoms in the sample. The reconstruction algorithm as introduced in [4] for atomic resolution tomography will be used. This algorithm is based on the multiplicative simultaneous iterative reconstruction technique [5,6], and makes extensive use of prior knowledge. In this way depth sectioning becomes very similar to tomography with the images at various defoci playing the role of the tomographic projections.

The projection requirement as given in [7] states that the image intensity must be proportional to the integral of a local property of the sample along the beam direction. From Fig. 2 it can be seen that in HAADF STEM with a large convergence angle this requirement is violated since atoms of the same element produce different images depending on them being in focus or not. For reconstruction techniques on the atomic scale this requirement is needlessly strict and can best be restated: The image intensity of any two atoms must equal the sum of the image intensities of the...
individual atoms. Hartel et al. show in [8] that this is easily accomplished for atoms separated in a direction normal to the beam by choosing a detector with a high angular width. For atoms only separated in the direction parallel to the beam also a high inner detector radius is necessary if the distance is below approximately 1 nm. Also coherent dynamical effects must be excluded. This means that the sample must be tilted out of zone axis, or must be amorphous.

In Section 2 we present the reconstruction technique, which explicitly takes into account the spatial distribution of the image of a single atom, and thereby the microscope settings that cause this distribution: the high tension, the spherical aberration, the defocus, the higher order aberrations and the objective aperture. Also the atomic nature of the object is explicitly taken into account. In Section 3 the incoherent image formation process is made explicit. We explain how the coherent point source, the incoherent extended source size and the object function of the atoms combine to yield a point spread function. In Section 4 this technique is tested on a simulation of an amorphous silicon layer with various levels of noise. In Section 5 the method and the simulation results are discussed and in Section 6 the conclusions are drawn.

2. Reconstruction technique

2.1. Modelling the projection process

In the algebraic reconstruction technique (ART) [5,6,9] the object \( f \) is overlaid with a grid and within each square, \( f \) is assumed to be constant. The pixels of \( f \) are indexed with one variable \( j = 1, 2, \ldots, K \), as in Fig. 2. The values of the projections \( p \) are indexed with one variable \( i \) too. If there are \( N \) projection angles with \( n \) pixels per projection, \( i \) runs from 1 to \( M = nN \). According to [5] every projection value \( p_i \) can be written as the weighted sum of the pixels of \( f \). The projections are then a set of linear equations:

\[
\sum_{j=1}^{K} w_{ij} f_j = p_i, \quad i = 1, 2, \ldots, M.
\] (1)

There are various ways of choosing the weights, and in [4] we proposed a new definition. For every pixel \( i \) in the projections and every voxel \( j \) in the object, we define the weight \( w_{ij} \) as the value in \( i \) of the image of a single atom positioned in \( j \), normalized to a total intensity of 1, i.e. \( \sum_j w_{ij} = 1 \). This is illustrated in Fig. 2. The image of an atom is determined by its atomic number and the microscope parameters, these weights therefore bring prior knowledge in the reconstruction, i.e. make the algorithm model based. The spatial extent of an atom image is now accounted for completely by the weights, the object \( f \) therefore only contains Dirac functions at the atom positions and zeros in between. Envisage an object \( f \) like in Fig. 2. It is composed of an atom in voxel \( j \) and one in voxel \( l \). Its image intensity is a sum of the image intensities of both individual atoms. The projection value \( p_i \) is given by (1) and reduces to \( w_{ij} f_j + w_{il} f_l \) because \( f \) is composed of Dirac functions only.

Usually, reconstruction algorithms calculate 2D images out of 1D line projections, and 3D reconstructions by treating the 2D projections as a set of line projections normal to the tilt axis and the object as a stack of 2D slices normal to the tilt axis. The underlying assumption is that neighbouring slices are imaged independently. This assumption is invalid for atomic resolution due to the spatial extent of the image of a single atom in the two dimensions normal to the beam direction. Therefore an algorithm was written that does the 3D reconstruction directly out of the 2D projections.

2.2. Maximum likelihood

Eq. (1) is more compactly written as

\[
W f = p.
\] (2)

with \( f = (f_1, \ldots, f_K)^T \) the unknown object \( f \) written as a vector, \( p = (p_1, \ldots, p_M)^T \) the vector of the expectation values of the projections and \( W \) the \( M \times K \) matrix containing the weights \( w_{ij} \). All pixels in the projections are assumed to be statistically independent and to suffer from Poisson noise. Therefore the probability \( P(q|f) \) of obtaining a set of measurements \( q \) defined by the vector \( q = (q_1, \ldots, q_M)^T \) is given as

\[
P(q|f) = \prod_{i=1}^{M} \frac{(W f_i)^6}{q_i!} \exp[-(W f_i)].
\] (3)

This function is called the joint probability density function of the observations. It is a function of the observations \( q \). The maximum likelihood (ML) method for estimating the object vector \( f \) is as follows. The available observations \( q \) are
substituted in the probability density function. Since the observations are numbers, the resulting expression depends only on the elements of the object vector \( f \). The elements of \( f \), the hypothetical true values, are now considered to be variables. To express this, they are replaced by \( t = (t_1, \ldots, t_L) \). The logarithm of the resulting function, \( \ln(p(q, t)) \), is called the log-likelihood function. The ML estimate \( f_{ML} \) of the object vector \( f \) is defined as the vector that maximizes the log-likelihood as a function of the object:

\[
\hat{f}_{ML} = \arg \max_t \ln(p(q, t)),
\]

i.e. it is the object that is most likely to have produced the projections \( q \) at hand. It is proven in [6] that \( f_{ML} \) satisfies

\[
f_{ML} = f_{ML} W^T \frac{q}{W f_{ML}},
\]

where arithmetic operations between vectors are elementwise, and \( W \) is normalized such that each column sum is 1. Expectation maximization [6] is the simplest iterative method that solves (5):

\[
f^{k+1} = f^k W^T \frac{q}{W f^k}, \quad k = 0, 1, \ldots .
\]

This requires an input \( f^0 \), here it is set to a constant vector with a value equal to the average of \( f \), which is estimated as the sum over all of \( q \)'s pixels divided by the number of projection angles and the number of voxels in \( f \). Comparing this with the simultaneous iterative reconstruction technique in [5] shows that expectation maximization essentially is a multiplicative version of this method. The joint probability density function is also defined for distributions other than the Poisson distribution, thus the concept of ML is not restricted to Poisson noise.

We propose a heuristic modification that explicitly takes the atomic nature of the sample into account. One expects \( f \) to be composed of individual Dirac functions separated by the vacuum, so every voxel with a value below half of the average of the non-zero voxels is likely to belong to the vacuum. Hence we set its value to zero and keep it zero in all subsequent iterations. The loss in intensity is compensated for by multiplying the values of the other voxels with a constant. This is repeated for every iteration \( k \). It is observed that this greatly accelerates convergence.

3. Image Formation

The image formation will be assumed to be incoherent. The 3D object function \( O(x, y, z) \) and the 3D electron probe intensity distribution \( h(x, y, z) \) are convolved in the two dimensions normal to the beam, the image intensity is then proportional to a projection.

\[
\text{value to zero and keep it zero in all subsequent iterations. The loss in intensity is normalized such that each column sum is 1. Expectation maximization [6] is the simplest iterative method that solves (5):}
\]

\[
f^{k+1} = f^k W^T \frac{q}{W f^k}, \quad k = 0, 1, \ldots .
\]

This requires an input \( f^0 \), here it is set to a constant vector with a value equal to the average of \( f \), which is estimated as the sum over all of \( q \)'s pixels divided by the number of projection angles and the number of voxels in \( f \). Comparing this with the simultaneous iterative reconstruction technique in [5] shows that expectation maximization essentially is a multiplicative version of this method. The joint probability density function is also defined for distributions other than the Poisson distribution, thus the concept of ML is not restricted to Poisson noise.

We propose a heuristic modification that explicitly takes the atomic nature of the sample into account. One expects \( f \) to be composed of individual Dirac functions separated by the vacuum, so every voxel with a value below half of the average of the non-zero voxels is likely to belong to the vacuum. Hence we set its value to zero and keep it zero in all subsequent iterations. The loss in intensity is compensated for by multiplying the values of the other voxels with a constant. This is repeated for every iteration \( k \). It is observed that this greatly accelerates convergence.

The coherent point source contribution to the probe intensity is given by [10,11]

\[
h_c(r, z, \Delta f) = \frac{4\pi}{k_0} \int_{0}^{\infty} \exp(-iz(k, z, \Delta f)j_0(2\pi kr)k^2 dk .
\]

with \( r^2 = x^2 + y^2 \), \( k = \lambda/\lambda \) is the absolute value of the spatial frequency in the horizontal plane, \( k_0 \) is the maximum spatial frequency set by the objective aperture that limits the convergence angle to \( z_0 \) and \( j_0 \) is the zeroth order Bessel function of the first kind. The effect of the incoherent extended source size \( d_s \) is described by the function [12]

\[
S(r) = \frac{1}{2\pi d_s^2} \exp \left[ -\frac{r^2}{2d_s^2} \right].
\]

The functions \( h_c \) and \( S \) combine to give the probe intensity [13]:

\[
h(r, z, \Delta f) = h_c(r, z, \Delta f) \otimes_{xy} S(r).
\]

where \( \otimes_{xy} \) denotes a convolution in the (x,y)-planes.

The object function of a single atom in an amorphous sample is approximated by a Gaussian function with a standard deviation \( \sigma_{ap} \) equal to that of the atomic potential given in [11]. The object function \( O_j(x, y, z) \) of an atom in position \( (x_j, y_j, z_j) \) is given as

\[
O_j(x, y, z) = \frac{1}{\sigma_{ap}^2(2\pi/3)^{1/2}} \exp \left[ -\frac{(x - x_j)^2 + (y - y_j)^2 + (z - z_j)^2}{2\sigma_{ap}^2/3} \right].
\]

This simple approximation is in the right order of magnitude and that is sufficient to demonstrate our algorithm. The object function of the entire specimen is given as

\[
O(x, y, z) = \sum_j O_j(x, y, z).
\]

The image intensity \( I(x, y, \Delta f) \) for a certain defocus can then be written as

\[
l(x, y, \Delta f) = \int_{-\infty}^{\infty} h(x, y, z, \Delta f) \otimes_{xy} \delta O_j(x, y, z) dz,
\]

\[
= \sum_j \int_{-\infty}^{\infty} h(x, y, z, \Delta f) \otimes_{xy} O_j(x, y, z) dz,
\]

(15) shows that the restated projection requirement is fulfilled: the image intensity is the sum of the intensities of the images of the individual atoms.

To calculate the weights \( w_{ji} \) in (1), let the subscript \( i \) denote the pixels in the image, and let \( j \) denote the voxels in the object, then for a certain defocus \( \Delta f \),

\[
w_{ji} = \int_{-\infty}^{\infty} h(x_j, y_j, z, \Delta f) \otimes_{xy} O_j(x, y, z) dz,
\]

with \( O_j \) the object function of an individual atom (see (12)) placed in the centre of the \( j \)th voxel. \( w_{ji} \) is the image of a single atom positioned in voxel \( j \).

The Rayleigh resolution criterion is valuable to estimate the resolution if the depth sections are being interpreted by analyzing horizontal and vertical slices through the 3D data stack. The Rayleigh resolution \( d_R \) is defined as the distance between the maximum of the point spread function an its first minimum [14]. For point spread functions that do not have a minimum, \( d_R \) can be redefined as the distance between two atoms for which the image intensity between them drops to 80% of the maximum.

The lateral resolution can be estimated as \( 0.61 \lambda / \sigma_0 \), which is the Rayleigh criterion for the diffraction limited case, for the depth resolution \( d_{Rz} \) equals \( 2\lambda / \sigma_0^2 \) [2]. For the settings used in section 4,
these values are 0.17 and 6.66 Å, respectively. The Rayleigh resolution for the complete image of one atom, i.e. for one of the terms of (15), must be numerically calculated: the lateral resolution is 0.84 Å and the depth resolution is 18.48 Å.

4. Simulation

4.1. Simulation conditions

The idea will be demonstrated with a simulation on amorphous Si. The object $f$ measures 6.0 × 6.0 Å in the horizontal directions and is 50.1 Å thick. The voxels are 0.3 Å wide, yielding a data stack of $20 \times 20 \times 167$. Within this volume 73 Si atoms are randomly distributed with an average distance of 2.70 Å. In Fig. 3 slices through this data stack are given.

The images will be generated with the model in Section 3. The microscope is assumed to be corrected for chromatic aberrations and to be limited by a fixed seventh order aberration $C_7 = 10$ cm, while the lower order aberrations are set to optimal values. The accelerating voltage $U$ is 200 kV. The optimal settings for such a microscope are given in [10] and repeated here in Table 1. The size of the extended source image in (10) is set to $d_S = 0.25$ Å. The object function of an individual Si atom is given in (12), with $\sigma_{ap} = 0.24$ Å. In Fig. 4 the weights $W$ are shown for $z = 0$ and $\Delta f$ ranging from $-50$ Å to $+50$ Å, the same table can be used for other values of $z$ as well.

The focus step is set to $-5$ Å, in agreement with the smallest experimentally achieved step we found reported in the literature [11]. This means we measured at 11 different defoci $\Delta f$, running from 0 to $-50$ Å. Various slices through the data stack are shown in Fig. 3, it is clear that directly interpreting them is far from trivial.

The through-focus series were rescaled to have mean values of 225, 900 and 3600, respectively, so that after subjecting them to Poisson noise the signal-to-noise ratios (SNRs) were 15, 30 and 60, respectively.

4.2. Results

Only the results of iterations of the iterative reconstruction algorithm with numbers equal to round $\lfloor 2^m \rfloor$ for $m = 0, 0.5, 1.5, \ldots$ were stored. In Fig. 5 the number of non-zero voxels in the reconstructions for all three SNRs is plotted against the iteration number, the relevant scale indeed is logarithmic.

Table 1

<table>
<thead>
<tr>
<th>$U$</th>
<th>$C_7$</th>
<th>$C_5$</th>
<th>$C_3$</th>
<th>$\Delta f$</th>
<th>$z_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>200 kV</td>
<td>10 cm</td>
<td>1.13 mm</td>
<td>3.54 μm</td>
<td>26.7 Å</td>
<td>86.8 mrad</td>
</tr>
</tbody>
</table>

Fig. 3. The simulated amorphous Si sample. The vertical direction is along the optical axis. (a), (b), (c) and (d) show slices 1, 7, 14 and 20, respectively, of the object $f$. The white voxels contain a Si atom. (e), (f), (g) and (h) are the corresponding slices through the data stack, i.e. through $Wf$. 
The number of atoms in the sample, i.e. the number of non-zero voxels in \( f \), is 73. For an SNR of 15 the number of non-zero voxels stabilizes at 76 from iteration 2\textsuperscript{19.5} on. We observe three atoms that are artificially split by the reconstruction, i.e. for these sites the algorithm finds two vertically separated non-zero voxels each with approximately half the intensity of a normal non-zero voxel, an example is given in Fig. 6. For an SNR of 30 the problem is less, 74 non-zero voxels are found so only one atom site is split. For an SNR of 60 no splitting is observed anymore.

Now we will look only at the atoms that were faithfully reconstructed, i.e. without splitting, and measure the distance \( d \) from their true positions. The root mean squared error (RMSE) is used as a statistical resolution criterion, which takes both

![Figure 4](image1.png)  ![Figure 5](image2.png)  ![Figure 6](image3.png)

**Fig. 4.** (a) Weights for \( z = 0 \) and \( \Delta f \) between \(-50\) and \(+50\) Å. The grey values are on a log-scale; (b) profiles of the weights in (a) on a linear scale.

**Fig. 5.** The number of non-zero voxels in the reconstruction as a function of iteration number. For a SNR of 15 it converges to 76, for a SNR of 30 it converges to 74 and for a SNR of 60 it converges to the true number of non-zero voxels, namely 73.

**Fig. 6.** (a) Slice 5 of the true object \( f \); (b) Slice 5 of the reconstruction out of measurements with a SNR of 15. The ellipse indicates the split atom. Note also the variance in the vertical positions of the other atoms; and (c) Slice 5 of the data stack \( Wf \) with Poisson noise and a SNR of 15.
accuracy and precision into account [15],

$$\text{RMSE}^2 = \frac{1}{N} \sum_{i} d_i^2,$$

(17)

where \(i\) runs over the \(N\) non-split atoms. The first observation is that the distance in the horizontal directions is always zero, there is only an uncertainty in the vertical positions. For an SNR of 15 the RMSE is 0.88 Å, for an SNR of 30 the RMSE is 0.73 Å and for an SNR of 60 the RMSE is 0.50 Å. Histograms of \(d\) are given in Fig. 7.

5. Discussion

In [10] it is pointed out that depth sectioning with a convergence semi-angle of \(\alpha_0\) is equivalent to tomography with a tilt series between \(-\alpha_0\) and \(+\alpha_0\). For the example in this article this means a missing wedge of \(\alpha_0 < 170^\circ\). The individual images in the focus series play the role of the projections, and their number is limited because the precision of the defocus value is limited to 5 Å. If the reconstruction process is seen as an inverse problem, these traits make it ill-posed in the sense that in general it will not have a unique solution. Nevertheless we manage to find good solutions because of the use of prior-knowledge: we assume we know the image of a single atom, and we use the fact that in our representation the object has a non-zero voxel on the atom site with only zeros in between, furthermore the noise distribution is taken to be Poissonian. These are all realistic assumptions, since in electron microscopy the beam is well characterized and simulations have been proven to describe the experiments well.

The reconstruction is not perfect however. For lower SNRs the algorithm reconstructs some atoms as pairs that are separated in the vertical direction. A possible explanation would be that the expectation maximization algorithm ended up in a local maximum. But, if we replace such a pair by a voxel with the same total intensity and place it in the middle, the log-likelihood decreases. This suggests that the expectation maximization algorithm does find the global maximum of the log-likelihood, and that the splitting is a property of that global maximum.

The number of iterations is very high as compared to the 20–40 iterations that are common practice in iterative reconstruction techniques in tomography. Conventional techniques need to estimate the values of all voxels in the reconstruction (70,400 in this case), while this technique only needs to estimate the values of the 73 non-zero voxels out of the same amount of data. Therefore the many iterations do not lead to an unacceptable noise level in the reconstruction. Take note that we do not use the number of non-zero voxels or their positions as prior knowledge, in the initial guess all voxels are assigned the same value. It may seem curious that we need more iterations if the recordings have a higher SNR. This is due to the heuristic step in the reconstruction that puts to zero every value below half the mean value of the non-zero voxels. A lower SNR in the recordings causes more fluctuations in the reconstruction, so that more voxels will have values below the threshold.

The image formation model used in this paper assumes complete incoherence:

- The image of the sample is the sum of the images of the individual atoms.
- The image of a single atom is a projection of the convolution of the probe intensity with an object function.

Only the first property is necessary to fulfill the projection requirement as restated in this article. Hartel et al. have shown in [8] that is accomplished by choosing a detector with a high angular width and a high inner detector radius if dynamical effects are excluded. The second property was assumed only to reduce the calculation time for the weights \(W\) of this method. One could also find the weights by calculating the image of a single atom for various defoci with a simulation packet that uses multi-slice or Bloch wave calculations.
6. Conclusion

In this paper we presented a method that retrieves the atom positions from a focal HAADF STEM series. This work brings depth sectioning within the field of tomography. The usual direct interpretation of the images restricts depth sectioning to finding heavy atoms in lighter surroundings. The reconstruction technique used here allows one to retrieve the positions of all atoms in the sample. With low SNR, a small percentage of the atoms gets split; for higher SNR no such behaviour was observed. The reliability of the estimates of the atom positions was quantified by their root mean squared error, in the investigated cases this was 0.88, 0.73 and 0.50 Å for SNRs of 15, 30 and 60, respectively. For the positions in the horizontal directions no deviations from the true positions were found, indicating that the reliability is better than the voxel size of 0.3 Å.

Acknowledgements

This work is carried out as part of the Condor project, a project under the supervision of the Embedded Systems Institute (ESI) and with FEI company as the industrial partner. This project is partially supported by the Dutch Ministry of Economic Affairs under the BSNI Programme.

S. Van Aert and D. Van Dyck (Project nos. G.0220.05 and G.0188.08) gratefully acknowledge financial support from the Fund for Scientific Research—Flanders (FWO).

References