Published in final edited form as:

Microsc Microanal. 2011 October ; 17(5): 817–826. doi:10.1017/S1431927611011913.

The Three-Dimensional Point Spread Function of Aberration-Corrected Scanning Transmission Electron Microscopy

A.R. Lupini¹ and N. de Jonge²

¹Oak Ridge National Laboratory, Materials Science and Technology Division, Oak Ridge, TN 37831-6064, USA

²Vanderbilt University School of Medicine, Department of Molecular Physiology and Biophysics, TN, Nashville, 37232-0615, USA

Abstract

Aberration-correction reduces the depth of field in scanning transmission electron microscopy (STEM) and thus allows three-dimensional imaging by depth-sectioning. This imaging mode offers the potential for sub-Ångstrom lateral resolution and nanometer-scale depth sensitivity. For biological samples, which may be many microns across and where high lateral resolution may not always be needed, optimizing the depth resolution even at the expense of lateral resolution may be desired, aiming to image through thick specimens. Although there has been extensive work examining and optimizing the probe formation in two-dimensions, there is less known about the probe shape along the optical axis. Here the probe shape is examined in three-dimensions in an attempt to better understand the depth-resolution in this mode. Examples are presented of how aberrations change the probe shape in three-dimensions, and it is found that off-axial aberrations may need to be considered for focal series of large areas. It is shown that oversized or annular apertures theoretically improve the vertical resolution for 3D imaging of nanoparticles. When imaging nanoparticles of several nanometer size, regular STEM can thereby be optimized such that the vertical full width at half maximum approaches that of the aberration corrected STEM with a standard aperture.

Keywords

STEM; aberration correction; 3D; nanoparticles; residual aberrations

INTRODUCTION

Traditionally, the depth of field in an electron microscope was two orders of magnitude larger than the lateral probe size, because the convergence angle of the electron beam had to be kept small to minimize the effect of the lens aberrations on the spatial resolution. Progressing towards the goal of improved two-dimensional (2D) resolution, aberration-correction (Haider, et al., 1998; Kisielowski, et al., 2008; Krivanek, et al., 1999) has brought about a significant increase in opening angles that can be used in a scanning transmission electron microscope (STEM) with a consequent reduction in the depth of field. The reduced depth of field enables the possibility of depth sectioning, i.e., recording a series of images while changing the focus to obtain “slices” through a thick, three-dimensional (3D) sample. These images can be used to reconstruct a 3D model of the sample (Borisevich, et al., 2006a; de Jonge, et al., 2010; Mobus & Nufer, 2003; van Benthem, et al., 2005b). Several papers have demonstrated that a depth sectioning technique can provide useful 3D information about single atoms, for example whether heavy atoms are definitely inside or on the surface of a lighter material (Allen, et al., 2008; van Benthem, et al., 2005a). It should be noted that the depth sectioning approach for locating single atoms is primarily limited to

amorphous samples, since channeling of electrons in aligned crystalline samples can lead to further uncertainty (Fertig and Rose, 1981, Borisevich, et al., 2006b). Another limitation of this technique is that although the lateral resolution can be sub-Ångstrom, the vertical resolution is still of the order of nanometers (Borisevich, et al., 2006a). Double-focusing or confocal techniques have been implemented to reduce this limitation (Cosgriff, et al., 2008; Einspahr & Voyles, 2006). Since nanoparticles are frequently used as specific protein labels (Xiao, et al., 2003), an application where a nanometer-scale depth resolution would be valuable would be to locate nanoparticles of heavy atoms in 3D on a relatively large biological structure that could be several microns in size (de Jonge, et al., 2010; Dukes, et al., 2011). In this application, extremely high lateral resolution may not be essential, meaning that optimizing the depth resolution, even at the expense of the lateral probe-size could be advantageous. There has been extensive work considering how aberrations affect probe formation in 2D (for example (Barth & Kruit, 1996; Mory, et al., 1985; Williams, et al., 1992)), however there has been less work examining the probe formation in 3D (Borisevich et al. 2006a, Behan et al. 2009, Xin and Muller 2009). There has also not been much work published examining the experimental factors degrading the achievable depth resolution, such as off-axis aberrations, probe semi-angle, nanoparticle size, scan distortion, image noise, radiation damage, etc. Here, a calculation method is provided for the 3D PSF including all relevant axial aberrations. The achievable vertical resolution will be calculated as a function of probe semi-angle for different nanoparticle sizes. The possible effects of scan distortion will be discussed briefly. This study will address ideal samples containing nanoparticles on rigid support structures, such that the images can be recorded with a sufficiently large signal-to-noise ratio; the effects of radiation damage will not be evaluated. Finally, ideas will be discussed to optimize the depth resolution for practical samples when sacrificing some of the lateral resolution.

CALCULATION METHODS

Analytical Model of the Point Spread Function

The basic concepts needed to understand the 3D point spread function (PSF) will first be discussed. The intensity on the z-axis for an aberration-free probe will be determined by diffraction and is given by Born and Wolf (Born & Wolf, 1959) as (see Figure 1):

$$I(z) = I_0 \left(\frac{\sin(u/4)}{u/4} \right)^2. \quad (1)$$

Where I_0 is the intensity at the geometrical focus and the dimensionless variable u defined by:

$$u = \frac{2\pi}{\lambda} \left(\frac{a}{f} \right)^2 z. \quad (2)$$

In this equation, λ is the wavelength, a is the aperture size (radius), f is the focal length and a/f is thus the aperture angle α . The depth of field due to diffraction Δz_{diff} will depend on the square of the aperture angle, typically:

$$\Delta z_{diff} = 1.77 \lambda / \alpha^2. \quad (3)$$

This equation represents the theoretical resolution of the depth sectioning technique, and the precise choice of numerical factor depends on the criterion used to define the depth of field.

A value of 2 corresponds to the distance to the first minimum (a Rayleigh-like criterion), a value of 1.77 corresponds approximately to the full width at half maximum (FWHM), and a value of 0.5 corresponds to a 20% change in intensity criterion (HW80%M). The choice of the appropriate measure might depend on whether the goal is to obtain the very sharpest image, or just to detect the presence of an out-of-focus feature.

Since the opening angles available in corrected instruments (Kisielowski, et al., 2008; Krivanek, et al., 2008; Lupini, et al., 2009) are up to a factor of four larger compared to the uncorrected case, equation (3) reveals that there exists the potential for over an order of magnitude reduction in the depth of field. Obviously this equation does not include a finite object and so it represents an optimal depth resolution for point-like objects. Similarly, it might be possible to locate the center of an extended intensity distribution more accurately than the width, meaning that the precision with which an isolated object can be located might depend on the signal to noise ratio more than this limit. Furthermore, this equation has neglected other aberrations that might affect the depth of field and so a more detailed calculation is required.

Numerical Calculation of the 3D PSF

The probe shape was calculated as follows. It was assumed that the beam-defining aperture is illuminated by an electron wave of uniform intensity with a phase that depends on the aberrations. The probe profile was calculated by propagating the electron wave from the aperture plane to the object plane by means of a Fourier transform and taking the intensity. In order to perform this calculation, a step-like function $H(\theta)$ was used to define the shape of the aperture, where θ is the angle to the optic axis. For a conventional aperture, this will take the value 1 below the maximum illumination angle and 0 above this angle. An annular aperture can be modeled in a similar way by additionally setting the aperture function to 0 below an inner angle.

A function $\chi(\Theta)$ was used to describe the aberrations, where the vector Θ was used to include the azimuthal angle φ as well as the axial angle θ , because some aberrations depend on azimuthal angle, where:

$$\Theta = (\vartheta_x, \vartheta_y) = (\theta \cos(\varphi), \theta \sin(\varphi)). \quad (4)$$

Here the function $\chi(\Theta)$ is defined as the distance from the aberrated wavefront to an ideal unaberrated wavefront, such that the phase change is given by:

$$\Phi(\Theta) = 2\pi\chi(\Theta)/\lambda. \quad (5)$$

The aberration naming convention given by Krivanek (Krivanek, et al., 1999) was used. The aberration function is given as the sum over terms allowed by symmetry in the expansion:

$$\chi(\theta, \varphi) = \sum_n \sum_{m=0}^{n+1} \left(C_{nma} \frac{\theta^{n+1}}{n+1} \cos(m\varphi) + C_{nmb} \frac{\theta^{n+1}}{n+1} \sin(m\varphi) \right). \quad (6)$$

In this expression, the order n describes how the aberration increases off-axis (as a function of θ) and the multiplicity m indicates the rotational symmetry (how many times the aberration repeats as it is rotated about the axis). The sum over m is restricted to even integers when n is odd and vice versa. The multiplicity for rotationally symmetric terms is defined as zero and can be omitted. The a or b indicates the sine or cosine term (a rotation)

and can also be omitted if not needed. Thus defocus can be C_1 or C_{10} and spherical aberration C_s could also be called C_3 or C_{30} . As an example, the phase change due to defocus is given by:

$$\Phi = \frac{2\pi}{\lambda} \frac{1}{2} C_1 \theta^2. \quad (7)$$

Note that there are several different naming schemes for aberrations, for example (Haider, et al., 2000; Kirkland, et al., 2006). In these different notations, the functions describing the aberrations are usually very similar, meaning that the particular choice of notation is not important here. There can be differences in the choice of numerical denominators, meaning some care needs to be taken when directly comparing aberration values from different schemes (Erni, 2010).

Thus, a numerical array was created containing an aperture defined by $H(\theta)$ with a phase depending on the aberrations and took the Fourier transform to obtain the probe intensity I as:

$$I(\mathbf{R}) = \left| \int H(\theta) \exp \left(\frac{2\pi i}{\lambda} (\chi(\theta) - \Theta \cdot \mathbf{R}) \right) d\Theta \right|^2 \quad (8)$$

In order to calculate the probe at different planes, the focus was changed such that the probe at a plane a distance z from the object plane is given by:

$$I(\mathbf{R}, z) = \left| \int H(\theta) \exp \left(\frac{2\pi i}{\lambda} \left(\chi(\theta, \phi) + \frac{1}{2} z \theta^2 - \Theta \cdot \mathbf{R} \right) \right) d\Theta \right|^2. \quad (9)$$

This last equation gives the probe intensity in 3D for discrete planes near to the focused plane. Figure 1 shows the analytical intensity on axis from equation (1) compared to the numerical result from equation (9) and demonstrates a good agreement. Note that this calculation includes the effect of diffraction and aberrations, but neglects some other factors that can be important experimentally, such as limited coherence, effective source size and instabilities. Since the probe was calculated in a 3D array, a chromatic focal spread could be included by convolving with a vertical blur, in just the same way that a source size is often included using a lateral convolution. For this work it was assumed that these factors could be reduced enough that neglecting them is reasonable. The focal spread from temporal incoherence (of the order of 1–10 nm) is typically smaller than the focal spread due to the lateral extent of the nanoparticles, but might sometimes increase the DOF. In this case, the contribution could potentially be reduced using chromatic aberration correction, monochromation, or ultrafast STEM.

Calculation Details

There are several potential pitfalls to performing this calculation numerically, which need to be considered to provide reliable results. The wavefront was propagated from the aperture plane to the probe plane by use of a Fast Fourier transform (FFT), which treats the calculation as periodic (e.g. (Press, et al., 1988)), meaning that it is equivalent to a tiled, infinite, array of apertures generating an array of probes. These probes should be far enough apart that they do not interfere with the calculation. Additionally, the sampling in real-space depends upon the reciprocal of the angular range used. Thus fine sampling in real-space requires a large angular range. The problem is that the size of the array used for the numerical calculation is limited, meaning that a larger angular range requires either more

elements or coarser sampling in reciprocal space. Clearly fine sampling in both real and reciprocal space can require very large arrays (see the related example in (Borisevich, et al., 2006b)). Finally, the phase is limited to 2π , meaning that if the phase changes by more than this value between adjacent elements numerical errors can be expected. Since the limiting effects of aberrations are considering here, these errors will be of particular interest. The following relation can be used as rule of thumb for when phase errors are expected:

$$|\chi(\Theta_1) - \chi(\Theta_2)| 2\pi / \lambda \geq 2\pi. \quad (10)$$

Where $\chi(\Theta_1)$ is the aberration function at an element Θ_1 and the adjacent element is at Θ_2 . If it is assumed that the distance between these two elements $d\Theta$ is small then:

$$\chi(\Theta_2) = \chi(\Theta_1) + d\Theta \cdot \nabla \chi \quad (11)$$

Thus for a purely radial change, the criterion becomes:

$$\frac{\partial \chi}{\partial \theta} d\theta \geq \lambda. \quad (12)$$

Some insight can be gained into this equation by considering two limiting cases. If the aberration function is dominated by defocus C_1 then numerical errors are expected approximately at some maximum angle $\theta_{\max}$ where:

$$\theta_{\max} \geq \lambda / (C_1 d\theta). \quad (13)$$

If the aberration function is dominated by spherical aberration C_3 , then:

$$\theta_{\max}^3 \geq \lambda / (C_3 d\theta). \quad (14)$$

Obviously, these equations do not include any aberration balancing (e.g. defocus is normally used to balance C_3), or non-round terms, and the criterion could arguably be changed by a numerical factor, which means that these criteria are only approximate. These limits can be investigated by examining the value of $d\theta$.

The element spacing in reciprocal space $d\theta$ is related to the number of elements used N and the maximum angle along the axis θ_{range} as:

$$d\theta = \theta_{\text{range}} / N. \quad (15)$$

Therefore, although a simple method to increase $\theta_{\max}$ is to increase the number of elements, these equations reveal that for a spherical aberration limited calculation, doubling the useable aperture requires a factor of 8 increase in the number of elements, because $\theta_{\max}$ depends on the cube root of θ_{range} . Since the numerical calculation is two dimensional for each focal step, this increase rapidly leads to inconveniently large arrays. The conclusion is that even when the numerical calculations are reliable for simple tests, it is still possible to drive them out of range by increasing the aberrations or using too large apertures.

RESULTS AND DISCUSSION

Optimal Focus

In order to determination of the probe size in 3D the optimal focus position was first calculated. Most TEM users will be familiar with using underfocus to balance spherical aberration (and other aberrations can balance in similar ways (Krivanek, et al., 2008)), meaning that the location of the optimal focus plane is not always obvious. Figure 2 shows the effect of spherical aberration on the on-axis probe intensity for a fixed aperture size. It can be seen that the spherical aberration shifts the location of the maximum vertically, as it partially balances the defocus, and that the intensity of this maximum becomes progressively reduced as the aberration is increased. It can also be seen that the depth of field increases but becomes quite difficult to define, since the intensity consists of quite a small central maximum with tails that spread over a larger focal range. To consider this problem more quantitatively a method was developed to determine the most focused plane in the presence of arbitrary aberrations.

The obvious method to determine the optimal focus is to search for the maximum probe intensity on axis. However, any numerical errors in the calculation (such as rounding errors or the sampling difficulties described above) could occasionally produce unexpected intensity spikes in a single pixel that would confound such a search. A more reliable method appeared the calculation of the full width containing 50% of the current d_{50} at each plane from the above calculation and to search for a minimum. The d_{50} was measured by taking an integral of the current contained within a particular diameter until the diameter contains 50% of the total current in that plane. In this procedure the current was integrated in a circular pattern, which is convenient but may be a poor measure for a stigmatic probe that is much smaller along one axis than another. Since the probe intensity was calculated on a grid with finite spacing, a linear interpolation (horizontally only) was performed to improve the accuracy of the diameter, which reduces the error from the real-space sampling. Figure 3 shows the width of the probe calculated in this way for the same conditions as in Figure 2.

To test this process for locating the minimum diameter, the numerically calculated probe diameter was compared with those calculated by the analytic formulae given by others (Barth & Kruit, 1996). Figure 4 illustrates this process for the selected aberration values. At small aperture angles the agreement is good. At larger angles, sampling errors occur, as discussed above. Thus it can be concluded that this method to find focus is acceptable, but can be limited by the sampling limitations that were derived. For this example, 1024 elements and a maximum angle of 50 mrad were used, meaning that equation (10) gives a limit of approximately 25 mrad for $C_3=1.3$ mm. Doubling the number of elements in each of the 2 dimensions (4 times more elements in total) in the calculation increases the range of angles that can be used up to 31 mrad. Another important consideration is that the focus range of the calculation needs to be suitably chosen to contain the focused plane and any other planes of interest. Although it should be possible to estimate and automatically adjust this range, in this work it was manually adjusted.

C_3 -limited Depth of Field

One interesting observation is that in Figure 3, far from the optimal focus, the change of the probe width with focus is not affected by the spherical aberration (i.e., the lines are essentially parallel). The reason for this is that the change of the aberration function with focus is given by:

$$\frac{\partial \chi}{\partial C_1} = \frac{1}{2} \alpha^2. \quad (16)$$

The somewhat surprising implication is that although spherical aberration shifts the location of the minimum probe diameter vertically and reduces the maximum probe intensity, the effect on the vertical spread is less direct and depends heavily on the choice of aperture. However, it should be clear that the spherical aberration does affect the depth of field, because the aperture size α is usually chosen based on the geometrical aberrations. For the uncorrected case of a C_3 -limited resolution, the aperture size follows as (formula from (Krivanek, et al., 2008)):

$$\alpha_{optC3} = \sqrt{2}(\lambda/C_3)^{1/4}. \quad (17)$$

Using the depth of field from equation (3) (with a factor of 2, to eliminate numerical constants from the result) gives:

$$\Delta z_{C3} = (C_3\lambda)^{1/2}. \quad (18)$$

Thus an approximate limit for the depth of field was derived based on the spherical aberration when the aperture size is selected to be optimal for 2D imaging. As an example, the depth of field for a standard 300 kV STEM with $C_3 = 1.3$ mm from equation (14) would be about 50 nm. For comparison, the depth of field from equation (3) for an aberration-corrected machine with a 30 mrad aperture would be just less than 4 nm.

Residual Aberrations

After correcting the spherical aberration there will always be some aberrations that remain due to limitations in the optics, small drifts over time, or just to the finite accuracy of the measurement and correction process. These are known as residual aberrations and provide an important practical limit to the 2D resolution (Haider, et al., 2000). It is, therefore, also important to consider the effect other aberrations on the depth of field.

The first non-round aberration to consider is astigmatism (C_{12}). Figure 5 shows the probe intensity as a function of defocus for several different values of astigmatism. This figure reveals that as the astigmatism is increased the intensity on axis decreases and eventually becomes double peaked. It is also interesting to note that even 1–2 nm of astigmatism gives a noticeable decrease in the maximum intensity, even though it has only a small effect on the width of the intensity on-axis.

In the example shown in Figure 5, when the astigmatism is 5 nm, the intensity on-axis has peaks at ± 5 nm, and so on. It therefore seems reasonable to infer that when present, such peaks will be located at $\pm C_{12}$ (i.e. the distance between the peaks is $2C_{12}$). Figures 5 and 6 reveal that these peaks correspond to line-like crossovers in perpendicular directions. Interestingly the depth of field from equation (3) is about 4 nm, and these peaks appear in Figure 5 when C_{12a} is above 2 nm, so it appears that this behavior occurs when the astigmatism is larger than about half the depth of field.

Figure 5 also shows the difficulties in defining a depth of field simply from looking at the intensity on axis when the astigmatism is significant. Examination of Figure 6 suggests that the d_{50} diameter could provide a more reliable route to define the depth of field, at least until the probe becomes so distorted that a single clear maximum is hard to define.

This discussion of the effect of aberrations on the probe size and shape in 3D suggests that it will be useful to have a visualization of the probe shape when dominated by different aberrations to provide a qualitative method to identify their different effects. Figure 7 shows

a x-z section through the probe for selected values of the first few aberrations (spherical aberration, astigmatism, coma, three-fold astigmatism). It can be seen that spherical aberration spreads the intensity along the axis, eventually giving a distribution with a secondary maximum. This observation could be very important for a simple interpretation of an experimental focal series in order to determine if a double-peaked experimental result corresponds to two discrete objects. Astigmatism gives cross-overs in perpendicular directions, which will be familiar to most STEM users who have tried to focus and stigmatize on single atoms. If astigmatism is present, images of atoms will appear to be stretched in perpendicular directions either side of focus. Coma creates a curved, crescent-like PSF, moving the center of the probe away from the axis.

Scan Distortion

One application of a depth-sectioning technique is to image a biological sample that might span several microns (de Jonge, et al., 2010; Dukes, et al., 2011), in which certain protein species are specifically labeled with heavy nanoparticles (Xiao, et al., 2003). In this application, it is important to understand the change in the appearance of nanometer-scale heavy-metal labels with focus and to know whether the PSF varies over the scanned area. Figure 8 shows a frame from a 3D STEM dataset of a flat test sample consisting of small (~5 nm diameter) gold particles on a carbon film. The data was recorded on a JEOL 2010F running at 200 kV, equipped with a CEOS aberration corrector (convergence angle 27 mrad). The first processing step is to align the dataset to remove the effect of drift. Several methods exist to perform this step, including commercial methods that can compensate for minor distortions, but here only drift correction was used. A simple script was used to measure the cross-correlation between subsequent frames. Because the lateral shifts of sections between the images in the focal series were not constant over the whole image, a number of smaller patches were cut from the data and the average shift used to align the frames. Multiple passes of this process were used to obtain a good alignment. Figure 8 shows the shift measured in this manner and a cross-section through the data after alignment. The (lateral) drift measured in this way over the 100 frame series was ~12 nm, giving an average drift around 0.12 nm/frame, or around 0.1 Ångstrom per second.

A striking observation from the final panel of Figure 8 is that there is a clear distortion in the PSF over the field of view. While the PSF in the middle of the image looks symmetric, the maximum intensity can be seen to bend towards the edges at either side of the field of view. It is known that present aberration correctors all suffer from a limited field of view, and it is expected that the aberrations increase the further the probe is off-axis (for example, see (Rose, 2008) for a description of off-axial aberrations and the steps necessary to reduce their effect in corrected systems). Comparing the shape of the PSF apparent in Figure 8 to the calculated probe shapes in Figure 7 suggests that the coma increases towards the edges of the field of view. It seems plausible that off-axial coma is present, because off-axial coma can be introduced when minimizing C_5 in an aberration-corrected STEM, or from slightly misaligned scan coils. Present automatic alignment procedures do not routinely measure off-axial aberrations, although this feature could potentially be included in the future. This result is important because deconvolution algorithms are beginning to be used to reduce the spreading of the PSF in 3D datasets (e.g. de Jonge *et al.* 2010). Since Figure 8 reveals that the PSF is not a constant over the whole dataset, off-axial aberrations may need to be taken into account when attempting to deconvolve focal series spanning a large field of view.

Calculating the Depth of Field

At this point, the depth of field has to be defined for a more general probe, where the aperture is not the optimal size from equation (17) and other aberrations may be present. Figure 2 already suggests that looking purely at the intensity on axis does not lead to an easy

way to define this depth, since there is a narrow central maximum with rather more extended tails. Figure 7 showed that this difficulty could be even worse for non-round aberrations, where the intensity can shift away from the center or the value on axis can become double peaked. Since the minimum d_{50} was used to define the best focus, a reasonable definition of the depth of field will be the distance between the planes where the d_{50} is twice its minimum. Another choice could have been to use the distance between the planes where the d_{50} is a factor of $\sqrt{2}$ larger than its minimum – corresponding to a doubling of the area. The doubling of d_{50} criterion is convenient, because for zero C_3 this definition matches well to the FWHM of a purely diffraction limited probe from equation (3).

The depth resolution for a conventional STEM, without an aberration corrector, was calculated as well, with $C_3 = 1.3$ mm, $C_5 = 0.15$ m, at 300 kV. Figure 9 (triangles) shows that for low angles the numerically calculated depth of field follows the diffraction-limited FWHM of equation (3), until the aberrations become significant for larger apertures. The minimum is near an aperture size of 8–9 mrad, which is close to the optimal value for the 2D probe size. This result is not surprising since the optimal aperture size for 2D is selected to balance between diffraction and C_3 limited cases. Furthermore, since the depth of field was calculated based on when the probe diameter is twice its minimum, it seems reasonable that an increase in this minimum diameter is likely to increase the depth of field.

The axial resolution for large nanoparticles

It was noted above that the optimal axial resolution can only be achieved for point-like objects and Figure 8 clearly shows that the nanoparticles are visible over a larger range than would be expected from equation (3) or the calculated depth of field. For extended nanoparticles with diameter d , where d is larger than the minimum 2D probe size or PSF, a geometrical argument suggests that the observed intensity will be extended over:

$$\Delta z_d \approx \sqrt{\frac{3}{2}} \frac{d}{\alpha}, \quad (19)$$

which has been experimentally verified (de Jonge, et al., 2010) and given with a prefactor of $\sqrt{3/2}$ (Xin and Muller 2009). It can thus be expected that the axial profile of the 5 nm diameter gold nanoparticles should extend out to 222 nm at 27 mrad, which corresponds reasonably well to Figure 8. The extra intensity from the nanoparticles spreads out, since the probe is convergent. This means that the out-of-focus nanoparticles will still be detectable over a longer focus range, and that they significantly increase the background, which in turn will affect detection of other labels. Notably, the numerical calculations take no account of signal to noise, which could be a limiting factor in practice. A similar increase in background that affected the detectability of dopant atoms was reported previously (van Benthem *et al.* 2005).

Equation (19) shows that the axial resolution can be improved by increasing the aperture angle. Current aberration-correction technology allows for imaging at angles up to about $\alpha = 41$ mrad (de Jonge, et al., 2010), under which conditions a 5 nm-diameter nanoparticle would then extend out to about 150 nm. Two nanoparticles of 5 nm diameter placed above each other at exactly the same lateral (x, y) position cannot then be distinguished unless they are axially (z) spaced by more than ~150 nm at this aperture angle. Fortunately, a sample containing nanoparticles, such as labeled proteins in cells (Dukes, et al., 2011), rarely contains nanoparticles at exactly the same lateral position, and most will be laterally separated. Isolated nanoparticles are therefore essentially a special case of the reconstruction problem. Once two nanoparticles differ in lateral position by only a few pixels, their 3D location can be determined with high precision using combination of noise filtering,

deconvolution, and position determination algorithms (as commonly used in wide field microscopy (Pawley, 1995)). It has been demonstrated that the 3D positions of nanoparticles in a whole cell samples could be determined with a precision of 3 nm (de Jonge, et al., 2010).

Since equation (19) does not depend upon the 2D probe size directly, it appears that one area where an uncorrected microscope might have the potential to provide useful depth-sectioning resolution would be where the labels are themselves significantly larger than the optimal 2D resolution. In this case, by using an oversized aperture, it should be possible for the increase in convergence angle to outweigh the increased probe size due to the aberrations. For example, in the proposed imaging mode an aperture could be selected such that the minimum 2D probe size is just smaller than the labels. In order to better consider this tempting idea, the case of 1 nm particles used as labels was numerically examined.

To calculate the axial resolution for the imaging of nanoparticles the PSF was mathematically convolved with the object and the FWHM of the obtained axial intensity profile was determined. The FWHM calculated this way represents an estimate of the axial resolution obtained on a practical sample including nanoparticles, while the DOF represents the case of a sample consisting of point objects. A spherical nanoparticle was approximated as a 2D disc with intensity proportional to the thickness of a sphere of the same diameter. This approximation is assumed to be valid where the spread from the shadowing effect is significantly larger than the object thickness. The numerically calculated convolved FWHM values are shown in Figure 9 (crosses) for a nanoparticle diameter of 1 nm. The calculation followed eq. (19) up to a semi-angle of 10 mrad (crosses in Figure 9), suggesting that the used calculation method is correct. At larger angles the FWHM increases due to the effects of aberrations. For angles larger than 16 mrad the vertical profile has a narrow maximum with an extended tail, making the FWHM rather inaccurate, so those points were excluded.

Figure 9 shows that the minimum FWHM on axis (140 nm) for a 1 nm label occurs at an aperture size of about 10–11 mrad, which is slightly larger than the optimal size for 2D imaging (a 9 mrad aperture gives 150 nm vertical FWHM). Further calculations suggest that this trend of favoring slightly oversized apertures continues for larger particles. For example, by the same measure, an aperture of about 13 mrad is optimal for 2 nm particles (220 nm vs 280 nm FWHM), and about 17 mrad for 5 nm particles (405 nm vs 665 nm FWHM). Thus it appears to be possible to get an improvement in the depth resolution by using an oversize aperture when the goal is to locate finite size nanoparticles. The improvement in vertical FWHM is relatively small when compared to the potential order of magnitude improvement in the depth of field offered by an aberration-corrector.

Another interesting imaging mode to consider involves using an annular aperture. In this case, because the low-angle beams are excluded, it is possible to use more defocus to balance the spherical aberration (or higher-order aberrations). Thus the position of the optimal focus will be shifted and the aperture outer angle can be increased. Figure 9 also includes a calculation of an annular aperture with a size from $0.8a$ to a . The depth of field (DOF) for the case of the imaging of point objects (dots) is larger (worse axial resolution) than using the regular aperture for small angles, but at angles above 10 mrad, the situation is reversed. A relatively small benefit can be obtained by using the annular aperture at 15 mrad.

For the annular aperture the vertical FWHM of the probe after convolution with the 1 nm-diameter nanoparticles (plusses in Figure 9) provides an interesting result. At large angles, the vertical FWHM is improved by almost a factor of two (60 nm at $\alpha = 13$ mrad) relative to the regular aperture of the same size. Thus although the extra aberration balancing allowed

by using an annular aperture did not significantly improve the depth of field, it did improve the vertical resolution for the imaging of a 1 nm-diameter nanoparticle (at the expense of 2D probe size). A similar result was found for 5 nm particles, with an annular aperture allowing a 240 nm FWHM at $\alpha = 21$ mrad (which is not much worse than our corrected STEM example in Figure 8). Hence, although the depth of field is degraded for point-like objects, using a large annular aperture theoretically offers benefits for depth-sectioning with labels that are larger than the optimal 2D probe size. Note that the collection angle of the ADF detector may also have to be adjusted in an experiment with larger probe angles to maintain the high-angle condition needed for Z-contrast imaging.

A further interesting question is then to consider using an oversized aperture for extended labels in a corrected STEM. In that case, the limit to the aperture size will be set by higher-order aberrations, such as C_5 instead of C_3 . Figure 10 shows the depth of field calculated for an aberration-corrected microscope with a residual C_5 of 10 nm, all other aberrations zero, and contains the results for both regular and annular apertures (0.8α to α).

Although the minimum depth of field of 4 nm occurs close to the optimal 2D aperture (27.5 mrad), the minimum vertical FWHM for 1 nm objects is 35 nm and occurs at a larger aperture size (37.5 mrad). This trend continues for larger particles (124 nm at $\alpha = 50$ for 5 nm particles). As for the uncorrected microscope, the best depth of field that can be achieved with the annular aperture is worse than with a regular aperture. For extended objects, the annular aperture again appears to offer a reduction in the vertical FWHM of a 1 nm object, with a minimum of 23 nm (at $\alpha = 42.5$ mrad). A useful analogy is to 2D Z-contrast imaging, where using an oversized aperture and a defocused probe can allow increased resolution at the expense of contrast (Nellist and Pennycook 1998). Oversized or annular apertures might be considered for use when the labels being imaged are larger than the optimal 2D probe size.

Conclusions

A method was developed for calculating the STEM probe shape in 3D, and this method was used to numerically investigate the effect of aberrations on the depth of field in a STEM. It was found that the on-axis intensity was not a reliable parameter to measure the optimal focus plane, whereas numerically searching for the minimum probe diameter in 3D gave a good match to the analytical probe size in test cases. The reliability of the numerical results was investigated and it was found that the phase change between calculation elements needs to be considered, which placed limits upon the aperture sizes that could be calculated in the presence of aberrations. A formula was derived to estimate the depth of field in the presence of spherical aberration and a method was given to calculate the depth of field for cases where an analytic estimate is not available. It was found that other residual aberrations, such as astigmatism, can affect the depth of field and examples were presented to aid in the visual identification of these effects. Images from a focal series of gold nanoparticles showed significant distortions off-axis that resemble off-axial coma. This suggests that off-axial aberrations will need to be considered for experimental focal series of large areas, which may be of considerable importance for the deconvolution of such data series.

For extended objects, where the depth resolution is likely to be dominated by the object size and the convergence angle, it was necessary to include the object into the resolution calculation. In this case, it was found that using a slightly oversized or an annular aperture can offer a reduction in the vertical FWHM. Through the use of such apertures, calculations suggest that it is possible to optimize the vertical FWHM of nanoparticles in an uncorrected STEM to values approaching that of a corrected STEM with the standard aperture, even though the DOF for point-objects is much worse. In future, the use of different ratio annular

apertures could be explored and it would be interesting to confirm the reduced vertical FWHM experimentally.

Acknowledgments

We thank S.J. Pennycook and R. Ramachandra for discussions. Electron microscopy conducted at the Shared Research Equipment user facility at Oak Ridge National Laboratory sponsored by the Division of Scientific User Facilities, U.S. Department of Energy. This research was supported by NIH grant R01GM081801.

References

- Allen JE, Hemesath ER, Perea DE, Lensch-Flak JL, Li ZY, Yin F, Gass MH, Wang P, Bleloch AL, Palmer RE, Lauhon LJ. High-resolution detection of Au catalyst atoms in Si nanowires. *Nat Nanotechnol.* 2008; 3:168–173. [PubMed: 18654490]
- Barth JE, Kruit P. Addition of different contributions to the charged particle probe size. *Optik.* 1996; 101:101–109.
- Borisevich AY, Lupini AR, Pennycook SJ. Depth sectioning with the aberration-corrected scanning transmission electron microscope. *Proc Natl Acad Sci.* 2006a; 103:3044–3048. [PubMed: 16492746]
- Borisevich AY, Lupini AR, Travaglini S, Pennycook SJ. Depth sectioning of aligned crystals with the aberration-corrected scanning transmission electron microscope. *J Electron Microsc.* 2006b; 55:7–12.
- Born, M.; Wolf, E. *Principles of Optics.* London: Pergamon Press Ltd; 1959.
- Cosgriff EC, D'Alfonso AJ, Allen LJ, Findlay SD, Kirkland AI, Nellist PD. Three-dimensional imaging in double aberration-corrected scanning confocal electron microscopy, Part I: Elastic scattering. *Ultramicroscopy.* 2008; 108:1558–1566. [PubMed: 18639381]
- de Jonge N, Sougrat R, Northan BM, Pennycook SJ. Three-dimensional scanning transmission electron microscopy of biological specimens. *Microsc Microanal.* 2010; 16:54–63. [PubMed: 20082729]
- Dukes MJ, Ramachandra R, Baudoin JP, Jerome WG, de Jonge N. Three-dimensional locations of gold-labeled proteins in a whole mount eukaryotic cell obtained with 3 nm precision using aberration-corrected scanning transmission electron microscopy. *J Struct Biol.* 2011; 174:552–562. [PubMed: 21440635]
- Einspahr JJ, Voyles PM. Prospects for 3D, nanometer-resolution imaging by confocal STEM. *Ultramicroscopy.* 2006; 106:1041–1052. [PubMed: 16916585]
- Erni, R. *Aberration-corrected imaging in transmission electron microscopy: An Introduction.* Imperial College Press; 2010.
- Fertig J, Rose H. Resolution and contrast of crystalline objects in high-resolution scanning transmission electron microscopy. *Optik.* 1981; 59:407–429.
- Haider M, Rose H, Uhlemann S, Kabius B, Urban K. Towards 0.1 nm resolution with the first spherically corrected transmission electron microscope. *J Electron Microsc.* 1998; 47:395–405.
- Haider M, Uhlemann S, Zach J. Upper limits for the residual aberrations of a high-resolution aberration-corrected STEM. *Ultramicroscopy.* 2000; 81:163–175. [PubMed: 10782641]
- Kirkland AI, Meyer RR, Chang L-YS. Local Measurement and Computational Refinement of Aberrations for HRTEM. *Microsc Microanal.* 2006; 12:461–468. [PubMed: 19830937]
- Kisielowski C, Freitag B, Bischoff M, van Lin H, Lazar S, Knippels G, Tiemeijer P, van der Stam M, von Harrach S, Stekelenburg M, Haider M, Uhlemann S, Mueller H, Hartel P, Kabius B, Miller D, Petrov I, Olson EA, Donchev T, Kenik EA, Lupini AR, Bentley J, Pennycook SJ, Anderson IM, Minor AM, Schmid AK, Duden T, Radmilovic V, Ramasse QM, Watanabe M, Erni R, Stach EA, Denes P, Dahmen U. Detection of Single Atoms and Buried Defects in Three Dimensions by Aberration-Corrected Electron Microscope with 0.5-Å Information Limit. *Microsc Microanal.* 2008; 14:469–477. [PubMed: 18793491]
- Krivanek, OL.; Dellby, N.; Keyse, RJ.; Murfitt, MF.; Own, CS.; Szilagy, ZS. *Advances in Aberration-Corrected Scanning Transmission Electron Microscopy and Electron Energy-Loss Spectroscopy.*

- In: Hawkes, PW., editor. Aberration-corrected Electron Microscopy. Academic Press; 2008. p. 121-160.
- Krivanek OL, Dellby N, Lupini AR. Towards sub-angstrom electron beams. *Ultramicroscopy*. 1999; 78:1-11.
- Lupini AR, Borisevich A, Idrobo JC, Christen HM, Biegalski M, Pennycook SJ. Characterizing the Two- and Three-Dimensional Resolution of an Improved Aberration-Corrected STEM. *Microsc Microanal*. 2009; 15:441-453. [PubMed: 19754980]
- Mobus G, Nufer S. Nanobeam propagation and imaging in a FEGTEM/STEM. *Ultramicroscopy*. 2003; 96:285-298. [PubMed: 12871795]
- Mory C, Tence M, Colliex C. Theoretical study of the characteristics of the probe for a STEM with a field emission gun. *J Microsc Spectrosc Electron*. 1985; 10:381-387.
- Nellist PD, Pennycook SJ. Subangstrom Resolution by Underfocused Incoherent Transmission Electron Microscopy. *Phys Rev Lett*. 1998; 81(19):4156-4159.
- Pawley, JB. Handbook of biological confocal microscopy. New York: Springer; 1995.
- Press, WH.; Teukolsky, SA.; Vetterling, WT.; Flannery, BP. Numerical Recipes in C The Art of Scientific Computing. Cambridge University Press; 1988.
- Rose H. Optics of high-performance electron microscopes. *Sci Technol Adv Mater*. 2008; 9:1-30.
- van Benthem K, Lupini AR, Kim M, Baik HS, Doh SJ, J-HL, Oxley MP, Findlay SD, Allen LJ, Pennycook SJ. Three-Dimensional Imaging of Individual Hafnium Atoms Inside a Semiconductor Device. *Appl Phys Lett*. 2005a; 87:034104.
- van Benthem K, Lupini AR, Kim M, Baik HS, Doh SJ, Lee JH, Oxley MP, Findlay SD, Allen LJ, Pennycook SJ. Three-Dimensional Imaging of Individual Hafnium Atoms Inside a Semiconductor Device. *Appl Phys Lett*. 2005b; 87:034104-034101-034103.
- Williams DB, Michael JR, Goldstein JI, Romig AD Jr. Definition of the spatial resolution of X-ray microanalysis in thin foils. *Ultramicroscopy*. 1992; 47:121-132.
- Xiao Y, Patolsky F, Katz E, Hainfeld JF, Willner I. "Plugging into Enzymes": Nanowiring of Redox Enzymes by a Gold Nanoparticle. *Science*. 2003; 299:1877-1881. [PubMed: 12649477]

Figure 1.

The on axis probe intensity (I) of a scanning transmission electron microscope (STEM) for a purely diffraction limited case in the vertical (z) direction from the analytical formula and calculated numerically. The intensity is normalized to the maximum. Aperture 30 mrad, 300 kV, wavelength 1.97 pm.

Figure 2.

The probe intensity from numerical calculations of the effect of spherical aberration only for increasing values of spherical aberration at a constant aperture size. Aperture 30 mrad, 300 kV, wavelength 1.97 pm. Plots normalized to same intensity.

Figure 3. The probe full width containing 50% of the current (d_{50}) versus z from numerical calculations for increasing values of spherical aberration at a constant aperture size. Aperture 30 mrad, 300 kV, wavelength 1.97 pm.

Figure 4.

The probe d_{50} as a function of the electron probe semi angle (α) calculated from analytic formulae and from numerically searching for the minimum diameter using arrays with 1024×1024 and 2048×2048 elements. $C_3 = 1.3$ mm, $C_5 = 0.15$ m, 300kV. At large angles the numerical values diverge from the analytic result.

Figure 5.

The probe intensity on axis from numerical calculations for increasing values of astigmatism. Aperture 30 mrad, 300 kV, wavelength 1.97 pm. Other aberrations zero. Plots normalized to the total intensity for zero astigmatism (not shown).

Figure 6. The probe d_{50} from numerical calculations for increasing values of astigmatism. Aperture 30 mrad, 300 kV, wavelength 1.97 pm.

Figure 7.

Examples of the effect of various aberrations on the probe shape in 3D. Shown are 2D projections in xz of the PSF. Calculated with an aperture angle of 30 mrad at 300 kV, with other aberrations zero. Zero aberrations, $C_3 = 10, 20, 30 \mu\text{m}$. $C_5 = 10, 20, 30, 40 \text{ nm}$. $C_{12a} = 2, 5 \text{ nm}$, $C_{12b} = 2, 5 \text{ nm}$, $C_{21a} = 200, 500 \text{ nm}$, $C_{21b} = 200, 500 \text{ nm}$, $C_{23a} = 200, 500 \text{ nm}$, $C_{23b} = 200, 500 \text{ nm}$.

Figure 8.

3D imaging of gold nanoparticles on a carbon support film, using a 200 kV aberration corrected STEM with $\alpha = 27$ mrad. (a) Frame showing ~ 5 nm-diameter gold particles. The frame was extracted from a 100 frame focal series. (b) The drift measured over the focal series relative to the first frame. (c) A side view of the central area of the focal series after drift-correction. Note that the vertical scale is compressed relative to the horizontal.

Figure 9.

Axial resolution D_z of standard STEM with $C_3 = 1.3$ mm, $C_5 = 0.15$ m, 300 kV as a function of angle α . For the imaging of point objects the numerically calculated depth of field (DOF) is shown, which follows the diffraction-limited full width at half maximum (FWHM) up to 10 mrad. The DOF was defined from the planes where d_{50} in lateral direction doubles. For the imaging of nanoparticles with a diameter d , the shadowing effect with $1.22d/\alpha$ needs to be taken into account. Here, the PSF was convolved with a nanoparticle with $d = 1$ nm and the FWHM is plotted. Also shown are the DOF and the convolution FWHM for an annular aperture ($0.8\alpha - \alpha$).

Figure 10.

Axial resolution of aberration corrected standard STEM with $C_5 = 10$ mm, and at 300kV as a function of angle α . The numerically calculated depth of field (DOF) is shown. For the imaging of nanoparticles with a diameter d , the shadowing effect with $1.22d/\alpha$ needs to be taken into account. Here, the PSF was convolved with a nanoparticle with $d = 1$ nm and the FWHM is plotted. Also shown are the DOF and the convolution FWHM for an annular aperture ($0.8\alpha - \alpha$).