A phantom-based forward projection approach in support of model-based iterative reconstructions for HAADF-STEM tomography

S.V. Venkatakrishnan [a], L.F. Drummy [b], M.A. Jackson [c], C.A. Bouman [a], J.P. Simmons [b], M. De Graef [d],*

[a] Purdue University, West Lafayette, IN 47907, USA
[b] Air Force Research Laboratory, Wright Patterson AFB, OH 45433, USA
[c] BlueQuartz Software, Springboro, OH 45066, USA
[d] Carnegie Mellon University, Pittsburgh, PA 15213, USA

ARTICLE INFO

ABSTRACT

We introduce a forward model for the computation of high angle annular dark field (HAADF) images of nano-crystalline spherical particles and apply it to image simulations for assemblies of nano-spheres of Al, Cu, and Au with a range of sizes, as well as an artificial bi-sphere, consisting of solid hemispheres of Al and Cu or Al and Au. Comparison of computed intensity profiles with experimental observations on Al spheres at different microscope accelerating voltages provides confidence in the forward model. Simulated tomographic tilt series for both HAADF and bright field (BF) images are then used to illustrate that the model-based iterative reconstruction (MBIR) approach is capable of reconstructing sphere configurations of mixed atomic number, with the correct relative reconstructed intensity ratio proportional to the square of the atomic number ratio.

1. Introduction

The field of materials science has seen a dramatic increase in the use of X-ray or electron-based tomographic studies of materials. Despite the availability of advanced materials–characterization tools, rapid and sensitive detectors, and massive computational resources, there is still a dire need for accurate physical models and the associated algorithms that can assist the user in (1) predicting what the data should look like, given a model of the material system, and (2) extracting all available information from an acquired data set. For instance, high angle annular dark field (HAADF) electron tomography is used to reconstruct nanoscale objects in 3D (e.g., [1]), but to-date such reconstructions are mostly qualitative instead of quantitative. In medical X-ray tomography applications (for instance, dual energy CT-scans [2]), tomographic reconstruction results in a quantitative 3D map of the object's density distribution; one can quantitatively identify bone, tissue, empty spaces, fluids, and so on. In medium resolution HAADF tomography, on the other hand, there is no clear understanding of what the quantity is that is being reconstructed; since the HAADF signal is considered to be proportional to Z^2, with Z the atomic number, one should ask the question: *can we actually reconstruct Z^2 as a function of position in the sample?* It appears from the recent HAADF literature (e.g., [3]) that the electron tomography community has not yet answered this apparently simple question. Some progress has been made at the atomic length scale [4], where 3D reconstructions of nano-particles are now within the realm of possibilities, but at the larger length scale (tens of nanometers to microns, e.g., the relevant length scales for many modern materials applications) no reports of quantitative HAADF-based reconstructions can be found. This example illustrates that today's modern data processing algorithms in electron microscopy are not necessarily being employed to the fullest extent. Extracting all possible information from a data set requires not only algorithms for the analysis of the reconstructions but also predictive (forward) algorithms so that microstructure models can be compared to actual data sets. In 3D TEM and SEM studies, such algorithms are still rare and in this contribution we describe a forward modeling approach for HAADF-STEM tomography that may ultimately make this powerful technique more quantitative.

Model-based iterative reconstruction (MBIR) algorithms have emerged as a mathematical and algorithmic framework for integrating physical models of materials and devices with experimentally measured data to form quantitative inversions of 3D material parameter volumes [5]. The MBIR framework formulates the problem of data inversion as an estimation problem, in which

* Corresponding author. Fax: +1 412 268 7596.
E-mail addresses: svenkata@purdue.edu (S.V. Venkatakrishnan), lawrence.drummy@us.af.mil (L.F. Drummy), mike.jackson@bluequartz.net (M.A. Jackson), bouman@purdue.edu (C.A. Bouman), Jeff.Simmons@wpafb.af.mil (J.P. Simmons), degraef@cmu.edu (M. De Graef).

the unknown quantity is the image or volume to be reconstructed. The MBIR problem typically then reduces to an optimization with terms representing the match of the measured data to the theoretical prediction and the known and statistical ensemble properties of the material. MBIR is a powerful framework because it allows for incorporation of general nonlinear physics-based forward models, joint estimation of unknown physical parameters (e.g., instrument calibration parameters), and both hard and soft constraints resulting from material properties and statistical material characteristics.

While conventional image reconstruction methods (e.g., filtered back projection (FBP) or simultaneous iterative reconstruction technique (SIRT) [6,7]) depend on linearity assumptions, MBIR does not require such approximations, and can produce quantitatively accurate reconstructions in a wide range of scenarios. In many cases, the use of more complex and accurate physical models leads to nonlinear forward models, e.g., surface-connected voids modify serial sectioning BSE observations due to what is essentially an occlusion process; and the attenuation of the bright-field beam changes the HAADF scattering amplitude. Both of these phenomena represent nonlinear forward dependencies, which can be fully incorporated in MBIR methods and the corresponding cost function can be solved using a range of mathematical tools, such as multiresolution/multigrid methods [8], adjoint differentiation [9], and Fréchet differentiation [10]. A key advantage of MBIR methods is that they can accommodate limitations in experimental systems by estimating calibration (i.e., hyper-) parameters automatically as part of the reconstruction process. In addition, diverse information regarding the known physical properties of a sample, along with its ensemble statistical properties, can also be integrated into the MBIR approach. Reconstruction regularity can be imposed through Bayesian prior modeling of local and even global material statistics.

In the longer-term, MBIR offers the potential for dramatic improvements through joint optimization of sensing hardware and reconstruction algorithms. For example, the acquisition speed of limited view time-varying samples can be dramatically increased by designing the microscope to acquire along sparse time-varying viewing directions. Through the use of sparse manifold constraints, the resulting 4D reconstruction can still be well-posed, even with large reductions in the total quantity of acquired data. Similar system optimization can be used to integrate multi-modal data, e.g., bright and dark field information.

One of the essential ingredients of the MBIR approach is the ability to compute the forward projection of an object. In other words, for a given object, and a given acquisition modality, the MBIR approach relies on the availability of a physics-based algorithm that can predict the images or spectra that would be generated for this modality and this object. This is not a new concept, and the microscopy community has for several decades now computed high resolution lattice images based on a model (the crystal structure) and an imaging modality (a parallel or a converged probe and knowledge of the microscope aberrations). However, such a forward modeling approach is less common for other imaging modalities, such as HAADF-STEM tomography. In the remainder of this paper, we will describe a relatively simple physics-based forward model for HAADF-STEM signal generation and apply it to three "phantom" samples. These are digital samples with an idealized and quantified structure that can be used to verify the validity and accuracy of tomographic reconstruction algorithms, in the same way that the well known Shepp–Logan phantom, which mimics the density variations of the human head and brain, is the standard for testing X-ray computed tomography algorithms. The phantom samples are a collection of aluminum nano-spheres, mixed aluminum and copper or aluminum and gold spheres, and a single Al/Cu or Al/Au bi-sphere.

Our ultimate goal is to create an MBIR approach for the quantitative reconstruction of HAADF-STEM tomography data sets, using BF data as the normalization, while taking into account the fact that BF data is usually complicated due to elastic scattering contributions. The first step towards solving the general problem is to take a simple model for HAADF image formation (the volume to be reconstructed is assumed to be represented by HAADF scatter coefficients at every point in space) and to combine it with a prior model (which is essentially a smoothness constraint) to formulate tomographic reconstruction as a maximum-a-posteriori probability (MAP) estimation problem. In this approach, the object to be reconstructed is described entirely by the HAADF scatter coefficients, and the BF signal is ignored completely; we have shown [11] that such a simple forward projection model already provides a significant improvement over the more traditional SIRT-based reconstruction approaches. In the next step, anticipating that a proper accounting for elastic contributions to the BF and HAADF signals would be necessary, we explored an MBIR approach to pure BF reconstructions; in our model [12] we deal with elastic (Bragg) scattering as anomalies in the data (a sudden large change in the signal restricted to certain sample tilt angles). This approach was also found to be superior to that of the more conventional reconstruction methods. In the present contribution, we present a first attempt at combining BF and HAADF signals using a more complex model for the HAADF signal, but we simplify the BF signal to one that is affected by absorption only, so that we can determine the best way to incorporate BF normalization into the MBIR reconstruction approach. As will be shown in what follows, this simplified approach already has the potential to provide quantitative reconstructions in which the reconstructed scatter coefficients have a realistic dependence on the atomic number. Finally, in ongoing research, we plan to combine all our prior approaches and quantitatively account for all scattering contributions to both BF and HAADF signals. One of the potential additional benefits could be the ability to also determine the crystallographic orientation of the sample (multiple grains or particles), since the BF signal will be properly modeled using an elastic scattering approach. It should be noted that the BF signal, which will be used to normalize the HAADF signal, is an incoherent BF signal.

The structure of this paper is then as follows: first, in Section 2, we describe the phantom structures and the basic scattering processes that will be taken into account in the forward projection model. In Section 3, we describe for each of the phantoms how the HAADF-STEM tomographic signals are computed. These simulated data sets are then used to test a recent new HAADF-STEM reconstruction algorithm. The effects of bright field normalization of the HAADF signal are explored in Section 4. Although the BF and HAADF forward models presented here do not formally consider Bragg scattering, we find that BF normalization of the HAADF signal produces quantitative tomographic reconstructions. Section 5 concludes this paper.

2. Generation of the phantom structures

2.1. The multi-sphere phantoms

Since a sphere is a relatively simple shape in terms of geometry and projections, all three phantoms will be based on spheres. In this section, we briefly describe a simple method to generate a phantom consisting of a semi-random arrangement of spheres with a range of sizes. We start from an algorithm for the random space-filling tiling of a plane with identical but differently sized shapes (a technique borrowed from statistical geometry [13]) and extend it to 3-D. Consider a 3-D box with volume V which we wish

to fill completely with spheres (i.e., a tiling with, in the limit, no empty space left over). If the first sphere that we place randomly inside the box has volume V_0, then we can fill the box with a series of spheres of ever decreasing size, to fill all the gaps. The volume of the i-th sphere is given by $V_0 g(i)$, where $g(i)$ is a suitably chosen function. In the limit of a complete tiling, we must have

$$V = V_0 \sum_{i=0}^{\infty} g(i), \tag{1}$$

so that the function $g(i)$ must sum to a finite value V/V_0. In principle there should not be any overlap between spheres, but in practice, to mimic the experimental microstructure, we will allow a very small amount of overlap.

We select the following function for $g(i)$:

$$g(i) = i^{-c}, \tag{2}$$

where $c \in \mathbb{R}$. For $c > 1$, the resulting series converges, and defines the Riemann zeta function $\zeta(c)$:

$$\zeta(c) = \sum_{i=0}^{\infty} \frac{1}{i^c}. \tag{3}$$

Optimal values for c range from 1.2 to about 1.4.

To create a realistic configuration of nano-spheres, with empty spaces in between, and aggregates of spheres of different sizes, we take a box of dimensions (b_x, b_y, b_z), with b_z significantly smaller than the other two dimensions to reflect the fact that, in a real sample, the spheres would be spread out in a relatively thin layer on a support membrane. Spheres are randomly positioned, with volumes decreasing according to the above choice of $g(i)$, but we impose a few additional constraints. First of all, to mimic the fact that the spheres lie on top of a planar support film, we do not allow spheres to approach the $z=0$ plane any closer than their radius; secondly, we allow spheres to overlap by a small number of voxels, to make sure that they are in tight contact. Thirdly, we introduce a plateau in the sphere size plot, so that there are a predefined number of spheres with identical size before we continue the decrease of the function $g(i)$. Finally, to mimic the wide open spaces typically observed in between particle clusters in experimental data, we eliminate the first N_s spheres (which are the largest ones due to our choice of $g(i)$) from the data set after it has been created. The series is then truncated after N spheres (in this paper $N=4000$). As illustrated in Fig. 1, this algorithm generates a rather realistic arrangement of spheres. The mixed aluminum–copper or aluminum–gold sphere phantoms can be generated easily by randomly assigning atomic number 13 or 29 (or 79) and the appropriate normal absorption parameter to each sphere.

2.2. The Al/Cu bi-sphere phantom

The simplest of the phantom structures used for our work is the bi-sphere, a simple sphere of radius R for which the two hemispheres have different atomic numbers and absorption parameters. At zero sample tilt, the top half of the sphere has atomic number Z_1 and normal absorption parameter μ_1, whereas the bottom half has parameters Z_2 and μ_2. If we assume that the z-axis points upwards, and the y-axis is the tilt axis, then the equation of the mid-plane for a given tilt angle τ is given by $z=\gamma x$, with $\gamma = \tan\tau$.

In the absence of Bragg scattering, the forward scattered (BF) image is relatively easy to compute analytically. The path length inside the sphere as a function of the image pixel position (x,y) for a tilted sphere is given by
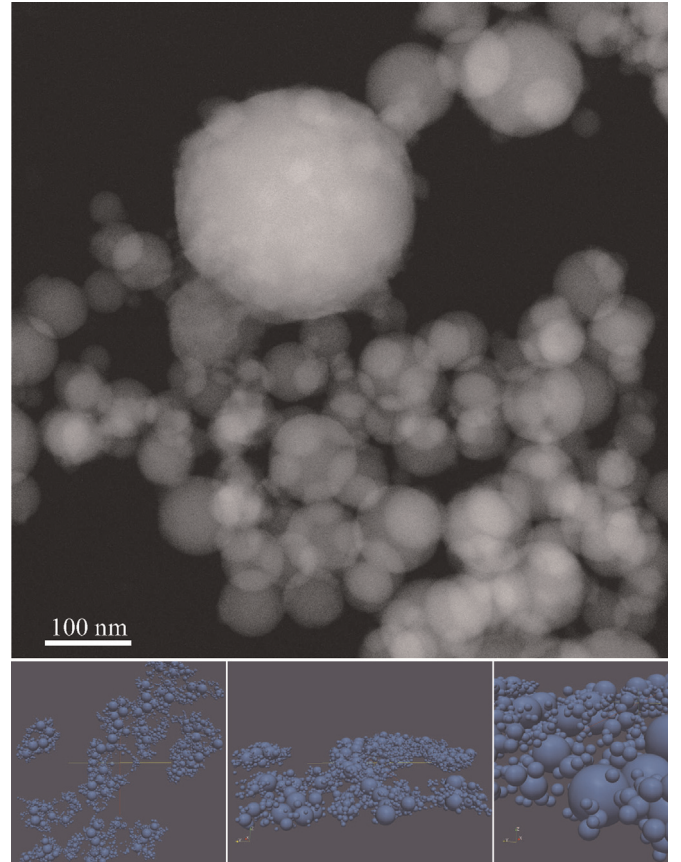

Fig. 1. (top) Experimental HAADF image of Al spheres (300 kV) and (bottom) three different views of a phantom of randomly positioned spheres.

$$\lambda(x, y) = 2\sqrt{R^2 - x^2 - y^2}\,\big[\,(\mu_1 H(\gamma) + \mu_2 H(-\gamma))H(-x) + (\mu_2 H(\gamma) + \mu_1 H(-\gamma))H(x)\,\big] \tag{4}$$

when $x^2 + y^2 \leq R^2$ and $(1 + \gamma^2)x^2 + y^2 \geq R^2$; the function $H(x)$ is the unit step function, which equals 1 when the argument is positive, and 0 when it is negative. When $(1 + \gamma^2)x^2 + y^2 \leq R^2$ then we have

$$\lambda(x, y) = -x\gamma(\mu_1 - \mu_2) + (\mu_1 + \mu_2)\sqrt{R^2 - x^2 - y^2}\,;$$

$\lambda(x, y)$ vanishes everywhere else.

The BF image intensity according to Beer's law is then given by

$$I_{BF}(x, y) = I_0 e^{-\lambda(x,y)}.$$

Fig. 2 shows a few frames of a simulated tilt series based on the equation above for a bi-sphere of radius 60 nm, and absorption parameters $\mu_1 = 2\pi/2000\ \text{nm}^{-1}$ and $\mu_2 = 2\pi/800\ \text{nm}^{-1}$. The tilt angles are (from left to right) 85°, 70°, 50°, 20°, and 0°.

3. HAADF forward projection model

The HAADF signal stems primarily from Rutherford scattering of the incident electrons. The differential scattering cross section for scattering of individual electrons by atoms with atomic number Z is given by

Fig. 2. Bright field tilt series for a bi-sphere with absorption parameters specified in the text.

$$\frac{d\sigma}{d\Omega} = \left(\frac{Ze^2}{16\pi\epsilon_0 E_{kin}}\right)^2 \frac{1}{\sin^4\frac{\beta}{2}},$$

where $\epsilon_0 = 8.854 \times 10^{-12}$ F/m is the permittivity of vacuum, Ω the solid angle, E_{kin} is the (relativistic) kinetic energy of the electron, and β the scattering angle measured with respect to the forward direction. For 200 kV electrons (the relativistically corrected voltage is 239,139 V) and Al atoms, the pre-factor is given by 3.83×10^{-28} m², or 3.83 barns; for 300 kV electrons (388,062 V relativistically), the differential cross section pre-factor equals 1.92 barns.

For a given Rutherford scattering event on a single atom with a scattering angle in the range $[\beta_1, \beta_2]$, the total scattering cross section is then given by the integration (denoting the pre-factor by b and working in spherical coordinates):

$$\sigma = b\int_0^{2\pi} d\varphi \int_{\theta_1}^{\theta_2} d\theta \frac{\sin\theta}{\sin^4\frac{(\pi-\theta)}{2}} = 4\pi b\left(\frac{1}{\sin^2\frac{\beta_i}{2}} - \frac{1}{\sin^2\frac{\beta_o}{2}}\right).$$

The annular detector inner and outer angles β_i and β_o are expressed as $\theta_1 = \pi - \beta_o$ and $\theta_2 = \pi - \beta_i$. For the annular detector used in typical experiments, the relevant angles are $\beta_i \approx 60$ mrad and $\beta_o \approx 5\beta_i = 300$ mrad, leading to a scattering cross section in this angular range of around $13,400b$ for a single atom. For a given detector geometry, this quantity must be multiplied by the incident beam current density (about 10^6 A/cm²) and the pixel dwell time (10 μs–1 ms) to obtain the total signal generated at a scattering point inside a conical shell tied to the annular detector.

It should be noted that this angular dependence results in a very uneven distribution of electrons across the detector. Fig. 3 (a) shows the integrand (logarithmic scale) of the cross section integral above as a function of the angular range (expressed in terms of θ), for $\beta_i = 3.5°$ and $\beta_o = 17.5°$. This steep angular variation means that the numerical integration over θ will need to be performed using smaller integration steps near the inner edge of the detector and larger steps near the outer edge. We use the following harmonic series to determine the step sizes $\Delta\beta_j$ for this integration: $\Delta\beta_j = \alpha h^{N-j-1}$, where N is the number of integration intervals, h a constant in the range [0, 1], and

$$\alpha = (\beta_o - \beta_i)\frac{h-1}{h(h^N-1)}. \tag{5}$$

It is easily verified that this expression spans the entire angular range:

$$\sum_{j=1}^N \Delta\beta_j = \beta_o - \beta_i.$$

The constant h has been set to the value 0.9 for all simulations reported in this paper.

The total signal acquired by the annular detector consists of electrons that were Rutherford scattered at any point inside the Al spheres. The Rutherford signal contributing to a single "detector pixel" must be attenuated to account for the path length traversed by the scattered electron inside the original sphere plus all other spheres that this electron traversed. In this model, we ignore multiple Rutherford scattering events.

Consider the case illustrated in Fig. 3(b): a Rutherford scattering event occurs in the top sphere, and the event location is taken as the origin. The electrons that travel inside the gray cone in sphere 1 can reach the annular detector at the bottom of the drawing. The number of such electrons is determined by (1) the depth inside sphere 1 at which the scattering event occurs; (2) the distance that the scattered electron travels inside sphere 1, and
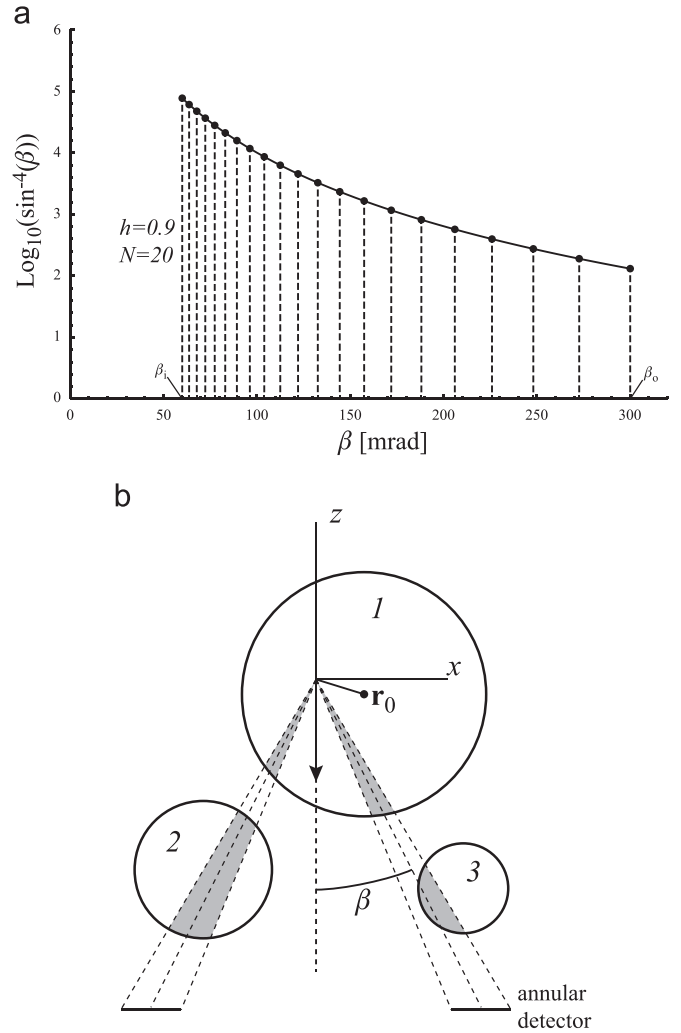
Fig. 3. (a) Integrand of the Rutherford scattering cross section for a HAADF detector angular range of 60–300 mrad; the vertical dashed lines show the integration subdivisions for the harmonic series discussed in the text. (b) shows how Rutherford electrons scattered in one sphere may intersect one or more additional spheres on their trajectory to the HAADF defector.

(3) the distance(s) that the electron travels inside other spheres.

To obtain the path length we need a simple expression for the length of a line intersecting a sphere. If we assume an origin outside a sphere with radius R and located at $\mathbf{c}$, then it is easy to show that the intersecting points along a line with unit direction vector $\hat{\ell}$ are given by $\lambda_{\pm}\hat{\ell}$, where

$$\lambda_{\pm} = \hat{\ell}\cdot\mathbf{c} \pm \left[(\hat{\ell}\cdot\mathbf{c})^2 + R^2 - c^2\right]^{1/2},$$

and $c^2 = |\mathbf{c}|^2$. Intersections will only occur if the expression under the square root is positive; if it is zero, then the line is tangent to the sphere and the intersection length vanishes. The intersection length $\Delta\lambda$ is thus given by

$$\Delta\lambda = 2\left[(\hat{\ell}\cdot\mathbf{c})^2 + R^2 - c^2\right]^{1/2}.$$

If the origin lies inside the sphere (as is required for the computation of Rutherford scattering events), then the portion of the line length inside the sphere starting at the scattering event is given by

$$\Delta\lambda = \lambda_+ = \hat{\ell}\cdot\mathbf{c} + \left[(\hat{\ell}\cdot\mathbf{c})^2 + R^2 - c^2\right]^{1/2}.$$

For numerical convenience, we express the unit vector $\hat{\ell}$ in

spherical coordinates as $\hat{\ell} = (\cos\varphi\sin\theta, \sin\varphi\sin\theta, \cos\theta)$. We discretize the annular detector both in the radial and azimuthal directions, with step sizes $d\theta$ and $d\varphi$. This leads to a total of

$$N_d = \left\lfloor \frac{2\pi}{d\varphi} \right\rfloor \times \left\lfloor \frac{(\theta_{\max} - \theta_{\min})}{d\theta} \right\rfloor,$$

where $\lfloor \cdots \rfloor$ indicates the nearest smaller integer.

The final expression for the HAADF signal strength (i.e., number of electrons counted) at a given beam position (x,y), with an incident beam current density j_s and exposure time Δt is given in integral form by

$$
\begin{aligned}
& I(x, y) \\
&= bj_s\Delta t \sum_{i \in \mathcal{N}(x,y)} \int_{z_{i,\text{in}}}^{z_{o,\text{out}}} dz_i \int_0^{2\pi} d\varphi \int_{\pi-\beta_0}^{\pi-\beta_i} d\theta \frac{\sin\theta}{\sin^4\left(\frac{\pi-\theta}{2}\right)} \times \\
& \exp\left[-\mu\left(z_i + \sum_{j=1}^{i-1}(z_{j,\text{out}} - z_{j,\text{in}}) + \lambda_i(\varphi,\theta) + \sum_k \Delta\lambda_k(\varphi,\theta)\right)\right].
\end{aligned}
$$

In this expression, $\mathcal{N}(x, y)$ is the set of spheres whose center, when projected along the z-axis, lies less than a distance $R_{\max}$ from the image pixel (x,y), and z_i is the distance along the z-axis measured from the entrance point $z_{i,\text{in}}$ of the beam in sphere i. The first sum in the exponential covers the attenuation of the incident beam traveling through spheres that precede sphere i. The factor $\lambda_i(\varphi, \theta)$ is the distance traveled through sphere i in the direction φ, θ after the Rutherford scattering event at depth z_i; the final sum covers the attenuation of the scattered electrons when they travel through other spheres that happen to overlap the Rutherford "conical volume" described by $(\varphi, [\beta_i, \beta_0])$.

Since we will employ a discrete description of the detector, the integrals must be replaced by summations. The step size along the beam direction is represented by Δz, the azimuthal step size by $\Delta\varphi$, and the step size along θ is $\Delta\theta$. The summations then become

$$
\begin{aligned}
& I(x, y) \\
&= bj_s\Delta t \Delta z \Delta\varphi \sum_{i \in \mathcal{N}(x,y)} \sum_{m=1}^{m_i} \exp \\
& \left[-\mu\left(m\Delta z + \sum_{j=1}^{i-1}(z_{j,\text{out}} - z_{j,\text{in}})\right)\right] \times \sum_{p=0}^{p_\theta} \frac{\Delta\theta_p \sin(\theta_p)}{\sin^4\left(\frac{\pi-\theta_p}{2}\right)} \\
& \sum_{q=1}^{N_\varphi} \exp\left[-\mu\left(\lambda_i(q\Delta\varphi, \theta_p) + \sum_k \Delta\lambda_k(q\Delta\varphi, \theta_p)\right)\right],
\end{aligned}
\tag{6}
$$

where m_i is the number of steps Δz needed to traverse sphere i at location (x,y); p_θ is the number of subdivisions of the θ angular range of the detector (see Eq. (5); N_φ is the number of azimuthal subdivisions, and θ_p is short hand notation for $\theta_p = \theta_1 + p\Delta\theta$. For the mixed sphere case, the parameters b and μ become dependent on the atomic number Z, and b must be placed inside the first summation.

Fig. 4(a) and (b) shows HAADF images of a 510 nm diameter spherical Al particle at 300 kV (a) and 80 kV (b) microscope accelerating voltages. Images were acquired on an FEI Titan microscope operating at 80 kV and at 300 kV at spot size 7, STEM mode, camera length 100 mm, with a Fischione HAADF STEM detector. In Fig. 4(c), simulated radial intensity profiles are shown for several values of the absorption length ξ in $\mu = 2\pi/\xi$, computed using Eq. (6). Azimuthally averaged intensity profiles for the Al sphere are superimposed as a dashed white wedge on the figure; note that near the center of the sphere, the profiles are noisy due to the small number of data points in the average. All profile intensities are normalized with respect to the intensity at the center of the sphere, and the radial distance is normalized by the sphere radius. The 300 kV experimental profile coincides nearly perfectly with the simulated curve labeled $\xi=500$, while the 80 kV result lies between the curves labeled 325 and 350. The agreement between the normalized simulated and experimental intensity profiles indicates that Eq. (6) provides a reasonable forward model for the HAADF signal from spherical particles.
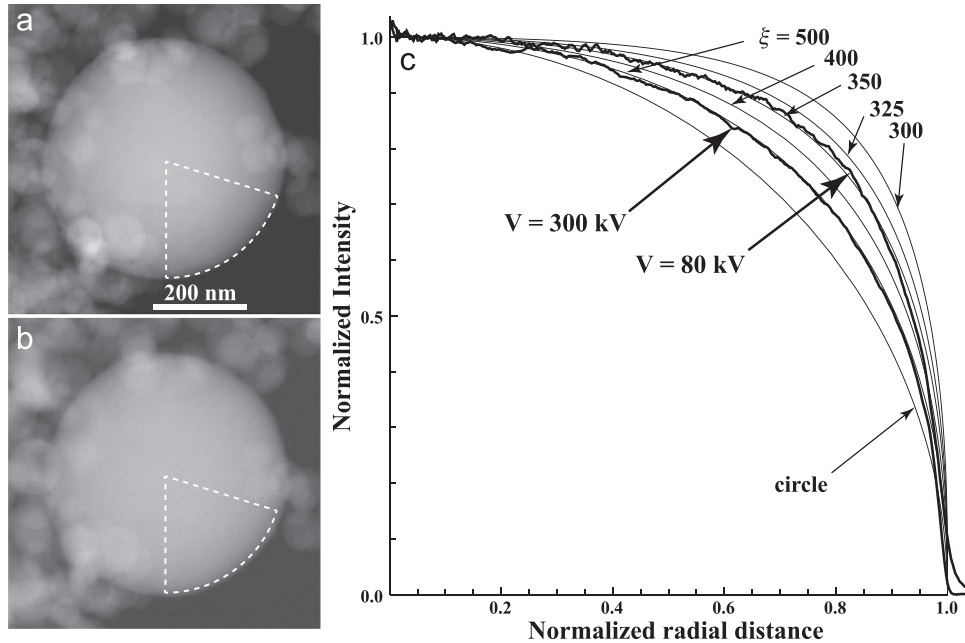

Fig. 4. (a) 300 kV and (b) 80 kV experimental HAADF images from a 510 nm diameter Al spherical particle. The thin continuous lines in (c) are normalized radial intensity profiles for the indicated values of ξ in the absorption parameter $\mu=2\pi/\xi$; the normalized experimental profiles, azimuthally averaged over the indicated wedge (dashed white lines in (a) and (b)), for the two accelerating voltages are superimposed as thick lines in (c).

4. Simulated BF+HAADF-STEM data sets and reconstructions

4.1. Image simulations

For each of the phantom structures introduced in Section 2, HAADF-STEM as well as BF tilt series were computed using the image models described in the previous section. A microscope accelerating voltage of 300 kV was used for all simulations. Table 1 shows a number of parameters for each tilt series: number of image pixels ($n \times n$), pixel size (Δx), tilt range ($[-\theta_{max}, \theta_{max}]$ in 1° increments), and the volume and type of reconstruction that was carried out (either based on the HAADF simulations only, or using both BF and HAADF signals). All tomographic reconstructions were carried out using the OpenMBIR public domain package [14]. For all image simulations the HAADF detector had an acceptance range of [3.5°, 17.5°]. The HAADF detector was segmented in 20 radial and 90 azimuthal sectors, with $h=0.9$ in Eq. (5). An incident beam current of 10^6 C/cm²/s was used for all simulations, with a pixel dwell time of 1 ms. Normal absorption lengths were estimated using the Weickenmeier–Kohl absorptive form factors [15] and are 843 nm for Al, 1028 nm for Si, 420 nm for Cu, and 236 nm for Au (at 300 kV). The integration step size along the beam direction was taken to be 0.5 nm.

Fig. 5 shows selected results of the tilt series simulations; all phantom data sets are made available via the OpenMBIR web site [14]. Fig. 5(a) and (b) are for the bisphere phantom, Al–Cu on the left and Al–Au on the right. The label on the top right of each image indicates the detector intensity range for BF and HAADF images. The particular tilt angle for the displayed images is shown in the lower right corner of the BF images. The increase in intensity on one side of the sphere is due to the fact that for a tilted bisphere, two regions with different atomic number and different absorption lengths contribute to the HAADF signal. Fig. 5(c) and (d) are BF, HAADF pairs for a pure Al sphere phantom, with a maximum sphere size of 50 nm, and two different tilt angles of 0° and 60°. The bottom row (Fig. 5(e) and (f)) shows identical simulation results with half of the Al spheres randomly replaced by Cu spheres (e) or Au spheres (f). Note that the maximum HAADF intensity increases with increased atomic number, as expected. Note also that the HAADF signal is about four orders of magnitude smaller than the incident intensity of the BF signal, in agreement with the Rutherford scattering probability, indicating that HAADF and BF signals are properly scaled with respect to each other. This is important, since the BF signal will be used as a normalization signal for the HAADF tomographic reconstruction.

4.2. Tomographic reconstructions

4.2.1. Reconstruction algorithm summary

The HAADF-STEM tomographic reconstructions are carried out using the MBIR approach [11]. The optimal reconstructions are obtained as

$$\hat{f} \leftarrow \arg\min_{f \geq 0} \frac{1}{2} \| g - Af \|_\Lambda^2 + s(f) \tag{7}$$

where g is the acquired HAADF data (with or without BF normalization) organized as an $M \times 1$ vector, A is an $M \times N$ forward projection matrix, f represents all the unknown voxels organized as an $N \times 1$ vector and Λ is an $M \times M$ diagonal matrix with each entry inversely proportional to the noise variance of the HAADF measurement; $s(f)$ corresponds to a Markov random field prior model that penalizes differences between adjacent voxels. In particular we use a q-GGMRF model for $s(f)$ which has the form

$$s(f) = \sum_{\{j,k\} \in \mathcal{N}} w_{jk} \rho(f_j - f_k)$$

$$\rho(f_j - f_k) = \frac{\left| \frac{f_j - f_k}{\sigma_f} \right|^2}{c + \left| \frac{f_j - f_k}{\sigma_f} \right|^{2-p}}$$

where $\mathcal{N}$ is the set of pairs of neighboring voxels (e.g., a 26 point neighborhood), and p, c and σ_f are q-GGMRF parameters. The weights w_{jk} are inversely proportional to the distance between voxels j and k, normalized to 1. The parameter p controls the smoothness of the edges in the reconstruction while σ_f controls the tradeoff between resolution and noise. For the results in this paper, we set $p=1.2$ and adjust σ_f for the best visual quality in the reconstructions. The value of c is set to 0.001, which makes the function $\rho(\cdot)$ have bounded second derivatives, thereby simplifying the overall optimization.

Minimizing the above cost function is a computationally challenging problem. We use the iterative coordinate descent (ICD) algorithm [16] which lowers the cost function with respect to each pixel one at a time. In order to reduce the computational complexity of the 1-D optimizations we construct a substitute function [17] to the original function and minimize this new function. The substitute function is constructed so that it is computationally inexpensive to minimize and minimizing the substitute lowers the original cost. The algorithm is further sped up by using a multi-resolution initial condition. This method solves the optimization problem at a coarser scale (larger voxels) and uses the low resolution reconstruction to initialize the subsequent finer resolution reconstruction [18]. Since we start with a better initial condition, the convergence of the algorithm to the minimum is rapid at the computationally expensive finer scales.

4.2.2. Bisphere phantom reconstruction results

Fig. 6 shows a summary of the Al–Cu bisphere reconstruction results. The top row shows reconstructed intensity histograms (on a logarithmic scale) for the partial tilt series (a) and the full tilt series (b). The solid line represents the normalized BF+HAADF reconstruction, whereas the dashed line uses only the HAADF data. The reconstructed intensities were scaled to the interval [0,1] before the histogram calculation. Note that both BF+HAADF curves show two peaks, presumably corresponding to signals from the Al

Table 1
Phantoms, simulation parameters and reconstruction modes for all the tilt series generated using the image models introduced in Section 3.

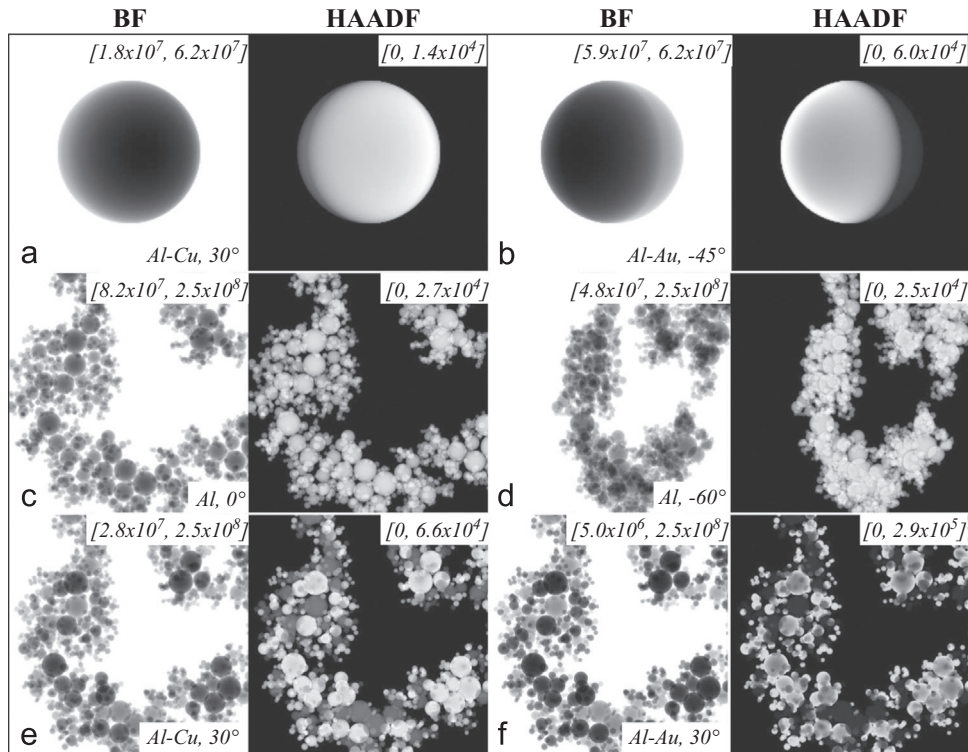| Phantom	n	Δx (nm)	θ_{max}	Reconst. volume	HAADF	BF+HAADF
Al–Cu bisphere	200	1	75°, 90°	$200 \times 132 \times 200$	✓	✓
Al–Au bisphere	200	1	75°	$200 \times 132 \times 200$	✓	✓
Al spheres	511	2	75°	$510 \times 511 \times 200$		✓
Al, Si spheres	511	2	90°	$510 \times 511 \times 160$		✓
Al, Cu spheres	511	2	75°, 90°	$510 \times 511 \times 160$		✓
Al, Au spheres	511	2	75°, 90°	$510 \times 511 \times 160$		✓
Cu, Au spheres	511	2	75°, 90°	$510 \times 511 \times 160$		✓

Fig. 5. Example images from the simulated tilt series for the cases listed in Table 1. In each image pair, the labels at the top right indicate the intensity range on the BF or HAADF detector; the label at the bottom right of the BF image indicates the type of spheres used as well as the tilt angle for this particular image. The field of view is 200 × 200 nm for the bisphere images in (a) and (b) and 1022 × 1022 nm for (c)–(f).

and Cu portions of the bisphere. When the most intense peak position is determined (vertical dashed lines labeled 1 in (a) and (b)), and this position is scaled by a factor $(Z_{Al}/Z_{Cu})^2 = 0.2$, then we obtain the vertical line labeled 2. Note that this line very nearly coincides with the location of the second peak in each of the histograms. The ratio of the HAADF and incoherent BF signals overcomes the (non-linear) attenuation of the HAADF signal; both signals contain a Beer's law style attenuation, so that the normalized HAADF signal becomes nearly linear in the sample thickness. Since the HAADF signal consists of both a Z^2 contribution and a Beer's law attenuation, it should be no surprise that the normalized reconstruction properly scales the reconstructed intensities to be proportional to Z^2. For the HAADF-only reconstructions, however, no clear peaks are observed for the tilt series with the missing wedge (a); the full tilt series (b) has a small peak near normalized intensity level 0.65 and a broad peak near the corresponding level of 0.13, but neither of these peaks can be used easily for segmentation of the reconstructed data set. The BF+HAADF reconstruction, even in the case of the missing wedge, lends itself more readily to a threshold-based segmentation approach.

Fig. 6 (c) shows vertical sections (normal to the tilt axis) through the bisphere, displayed as gray-scale images with a common intensity scale inside each set of four; in all cases, Cu is at the bottom. The labels above each image refer to the slice location, in nm from the sphere center. In the full-tilt BF+HAADF case, the reconstructed interface between Al and Cu is perfectly sharp; in all other cases, varying amounts of diffuseness can be observed near the interface, in particular in the HAADF-only reconstruction with a missing wedge. In addition, this reconstruction shows significant non-uniformity in the Cu region, particularly near the sphere edge. Such non-uniformity is absent in the BF+HAADF reconstruction for the 75° tilt range, although some voxels at the surface of the hemisphere show an increased intensity relative to the internal

voxels. This non-uniformity is due to a model mismatch between the physically accurate forward model, which includes both HAADF scatter and attenuation, and the forward model contained in the MBIR reconstruction algorithm, which is a linear HAADF scatter model without attenuation. While it is conceivable to implement the more physically realistic forward model described in this paper directly into the MBIR iterative reconstruction process, here we investigate the results when the HAADF projection images are normalized by the BF image, and then reconstructed using MBIR.

Fig. 6 (d) shows line profiles normal to the interface plane through the center of the bisphere (dashed line in the inset in (d)). Dashed lines correspond to the 75° tilt series, solid lines to the full tilt series; gray lines represent the HAADF-only reconstruction, and solid lines the normalized BF+HAADF reconstruction. All intensities were scaled to the interval [0,1]. Note that the BF+HAADF reconstructions show nearly horizontal plateaus in both the Cu region (left half) and the Al region (right half). The HAADF-only reconstruction shows a significant variation of the intensity across the entire range. Profiles parallel to the interface plane, through both the Cu and Al regions, are shown in Fig. 6(e); the gray curves correspond to the Al region, black curves to Cu. Once again, the BF+HAADF reconstruction is nearly perfectly level across each region. Note that the curves on the left represent the full tilt series on the right the tilt series with a missing wedge. The results of these simulations suggest that the BF+HAADF reconstructions provide a quantitative aspect that the HAADF-only approach cannot match. In the following section therefore, we will focus only on reconstructions that employ both HAADF and BF data as input. Note also that an attempt to use a standard SIRT reconstruction algorithm was not successful with the present data sets presumably due to the rather large contrast between the two hemispheres.
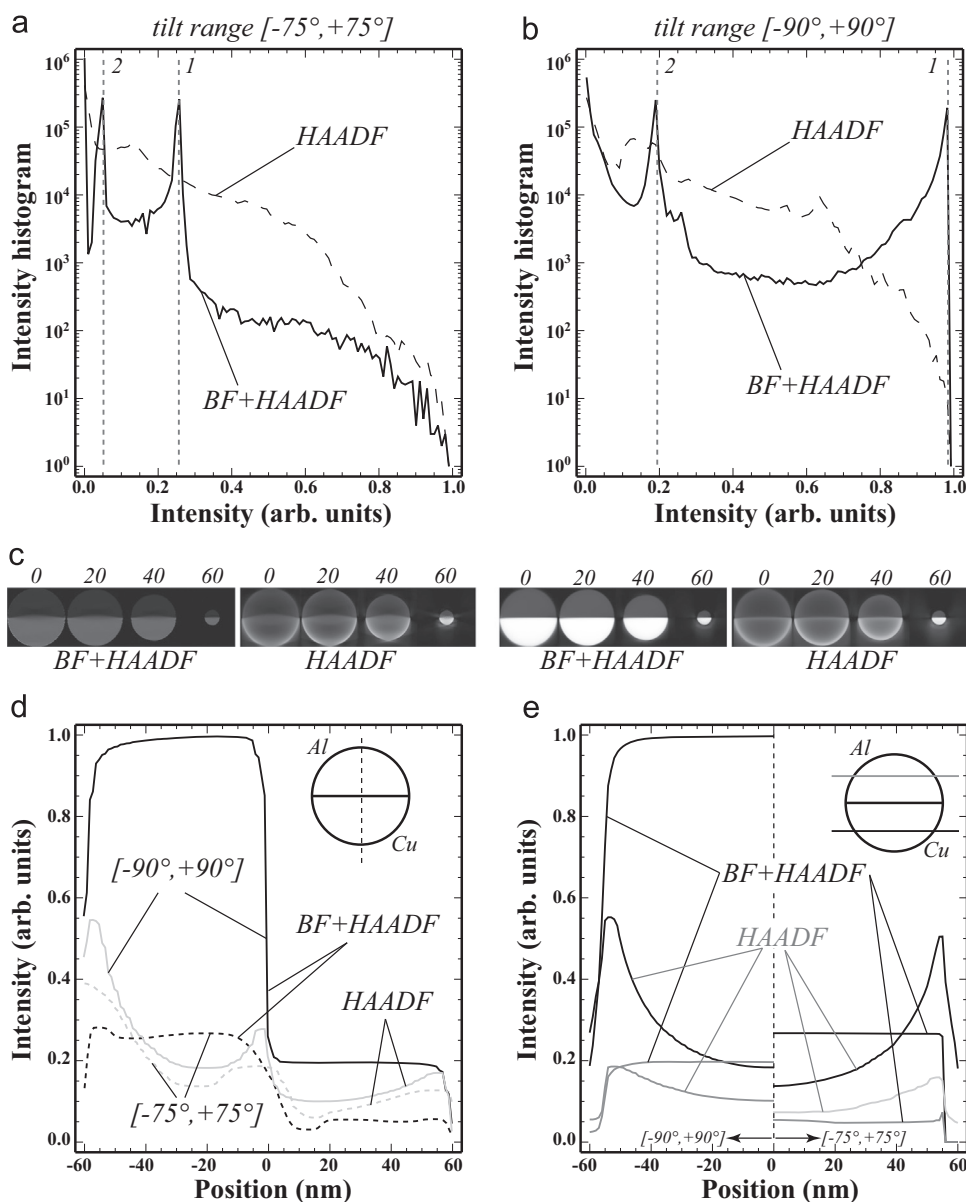
Fig. 6. Reconstructed intensity histograms (logarithmic scale) for a cubic box surrounding the Al–Cu bisphere, using HAADF-only and BF+HAADF reconstructions; (a) uses a partial tilt range of [−75°,+75°], whereas (b) shows the results for the full tilt series [−90°,+90°]. (c) shows cross sections normal to the tilt axis at the indicated distances from the center, for HAADF-only and BF+HAADF reconstructions. (d) and (e) show profiles across the bisphere, perpendicular to the interface plane (d) and parallel to the interface plane. Both partial (dashed) and full (solid) tilt series results are shown for HAADF-only and BF+HAADF reconstructions (d); in (e), gray lines correspond to profiles through the Al hemisphere, whereas black lines correspond to the Cu hemisphere.

4.2.3. Multisphere phantom reconstruction results

Fig. 7 shows reconstructed intensity histograms (logarithmic scale) for the partial (left column) and full (right column) tilt series. As before, intensities were normalized to the interval [0,1] and binned with a step size of 0.01. Each row corresponds to a different combination of Z_1 and Z_2: Al–Cu (top row), Al–Au (center row) and Cu–Au (bottom row). Each histogram shows two gray curves, corresponding to the reconstructed intensities inside a 3D mask for each of the sub-sets of spheres (475 spheres of type 1, 452 spheres of type 2). The solid black line is the sum of the two individual histograms. The vertical dashed line labeled "1" in each plot is located at the center of mass of the intensities for the higher atomic number spheres. The other line, labeled "2", is located at a position that is scaled by the factor $(Z_1/Z_2)^2$ from line "1".

For the Al–Cu phantom, two peaks are observed in Fig. 7(a) and (b); the position of the second vertical line, with a scaling factor of $(13/29)^2 = 0.2$, is slightly towards the right of the corresponding

peak, similar to the results for the bisphere. For the Al–Au tilt series, the reconstructed histograms, shown in Fig. 7(c) and (d), only reveal one single peak for the Au spheres. The scaling factor for the second peak is equal to $(13/79)^2 = 0.027$, which indicates that the reconstructed intensity associated with the Al spheres will occur at the very left edge of the histogram, just barely above the lowest overall intensity. As a result, the Al spheres are not well defined in the reconstruction, likely due to the high Z^2 contrast between the two materials. Image acquisition with a larger dynamic range throughout the entire imaging series would be necessary for resolving the Z-contrast between Al and Au.

When the Al spheres are replaced by Cu, the scaling factor becomes $(29/79)^2 = 0.134$, which is still fairly small. The reconstructed histograms in Fig. 7(e) and (f) do show a peak at the lower intensities, to the left of line "2"; replotting the histogram with a smaller binsize (not shown) does result in the intensity levels dropping off towards the left edge of the histogram, and the
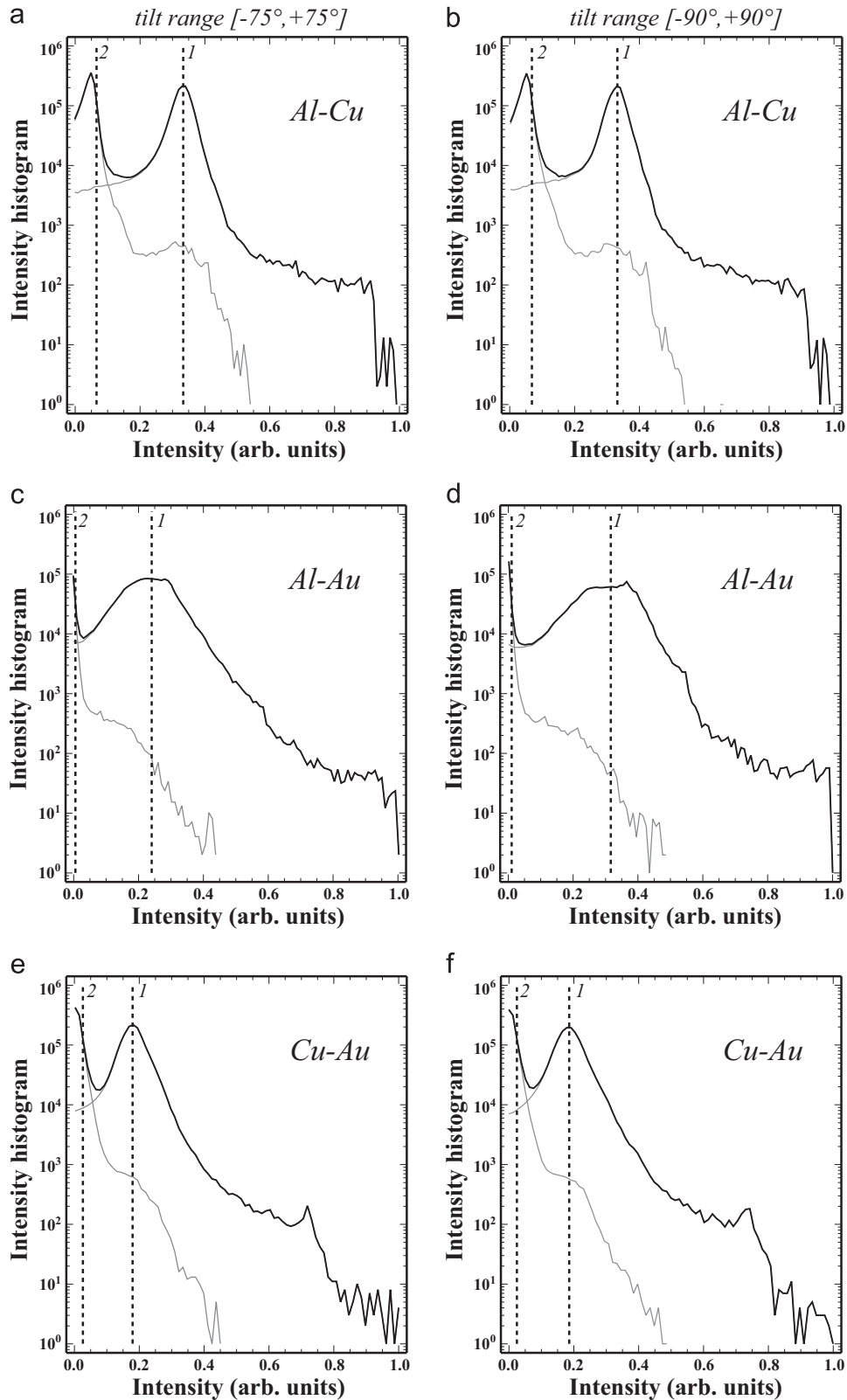
Fig. 7. Reconstructed intensity histograms (logarithmic scale) for (a) and (b) Al–Cu, (c) and (d) Al–Au, and (e) and (f) Cu–Au multi-sphere phantoms for tilt ranges of $[-75°,+75°]$ (left column) and $[-90°,+90°]$ (right column). The vertical dashed line labeled "1" corresponds to the center-of-mass of the higher atomic number peak; the position of the line labeled "2" is computed from the position of the first line by a scaling factor of $(Z_1/Z_2)_2$. Gray lines in each plot correspond to histograms obtained by masking out everything but the sphere of type Z_1 or Z_2.
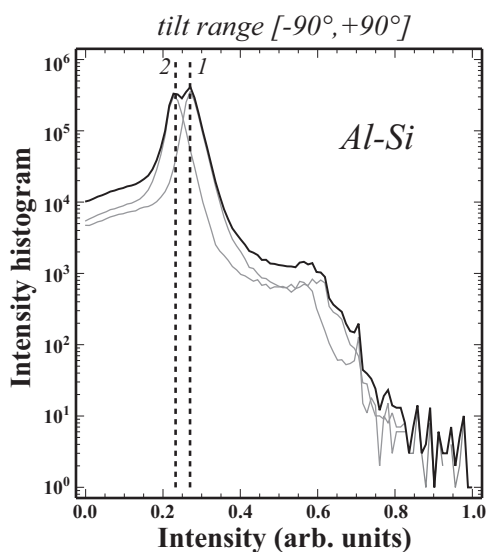
Fig. 8. Reconstructed intensity histograms (logarithmic scale) for an Al–Si multi-sphere phantom (full tilt series). The peaks corresponding to the two materials overlap substantially, but are located at the correct relative position.
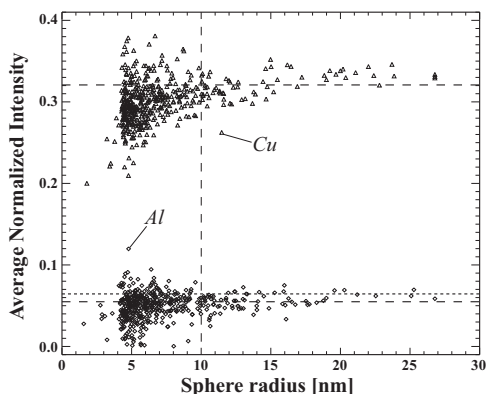

Fig. 9. Average intensity versus sphere radius (in nm) for the Al–Cu multisphere phantom reconstruction (full tilt series). The horizontal long-dashed lines indicate the average intensity for the subset of spheres with radius larger than 10 nm (vertical dashed line). The short-dashed line is computed from the Cu line by scaling by a factor of $(13/29)^2$.

Cu spheres are reasonably well reconstructed.

When the atomic numbers are more similar, the reconstruction leads to overlapping intensity peaks in the histogram, as shown in Fig. 8 for a mix of Al and Si spheres; the scaling factor is now equal to $(13/14)^2 = 0.862$, and the peaks occur at the correct relative position. The gray curves indicate significant overlap between the reconstructed intensities for the Al and Si spheres; since the particle shapes are simply spherical, it should still be possible to segment the reconstruction into spheres of types Z_1 and Z_2, but, in general, more advanced segmentation approaches, such as the EM/MPM algorithm [19], may be required to handle the peak overlap.

Finally, Fig. 9 shows, for the Al–Cu multisphere phantom, the average intensity for each sphere in the data set as a function of the radius of the sphere; note that there is significantly more scatter at smaller sphere radii. Assuming that the reconstructed intensity will be more realistic for the larger spheres, the dashed horizontal lines correspond to the average normalized intensity for the Al and Cu spheres, for the subset of spheres with radius larger than 10 nm (i.e., to the right of the vertical dashed line). The short-dashed line corresponds to the location computed from the Cu line by scaling by a factor of $(13/29)^2 = 0.2$. Note that the smaller

spheres produce a lower average intensity, both for Al and Cu. For the larger Al spheres, the scaled average intensity is quite close to the expected intensity.

5. Conclusions

The results in the previous section show clearly that the model-based iterative reconstruction procedure described in [11] is capable of producing reconstructions that are nearly quantitative, due to the fact that the bright field images are used to normalize the HAADF data. Without such normalization, the reconstructed intensities do not have a straightforward interpretation, and extraction of quantitative data from a HAADF-only reconstruction appears to be problematic. The approach presented in this contribution illustrates that the combination of an accurate forward projection model for the formation of HAADF and BF images of nano-crystalline spheres combined with a model-based iterative reconstruction technique allows for the quantitative extraction of the ratio of the atomic numbers of two elemental solids, and thus provides a pathway to quantitative HAADF tomographic reconstructions. A direct reconstruction of the atomic number distribution will likely require a more accurate determination of incident beam currents and measured HAADF and BF intensities.

References

[1] I. Arslan, J.R. Tong, P.A. Midgley, Reducing the missing wedge: high-resolution dual axis tomography of inorganic materials, Ultramicroscopy 106 (2006) 994–1000, http://dx.doi.org/10.1016/j.ultramic.2006.05.010.

[2] R. Alvarez, A. Macovski, Energy-selective reconstructions in X-ray computerized tomography, Med. Phys. 21 (5) (1976) 733–744.

[3] P. Midgley, M. Weyland, Fundamentals of electron tomography, in: J. Banhart (Ed.), Advanced Tomographic Methods in Materials Research and Engineering, Monographs on the Physics and Chemistry of Materials, Oxford University Press, Oxford, UK, 2008, pp. 305–334 (Chapter 11).

[4] S. Van Aert, J. Verbeeck, R. Erni, S. Bals, M. Luysberg, D. Van Dyck, G. Van Tendeloo, Quantitative atomic resolution mapping using high-angle annular dark field scanning transmission electron microscopy, Ultramicroscopy 109 (2009) 1236–1244.

[5] J. Thibault, K. Sauer, J. Hsieh, C. Bouman, A three-dimensional statistical approach to improve image quality for multislice helical CT, Med. Phys. 34 (11) (2007) 4526–4544.

[6] M. Weyland, Electron tomography of catalysts, Top. Catal. 21 (2002) 175–183.

[7] H. Li, H. Xin, D. Muller, L. Estroff, Visualizing the 3D internal structure of calcite single crystals grown in agarose hydrogels, Science 326 (2009) 1244–1247.

[8] S. Oh, A. Milstein, C. Bouman, K. Webb, A general framework for nonlinear multigrid inversion, IEEE Trans. Image Process. 14 (2005) 125–140.

[9] A. Hielscher, A. Klose, K.M. Hanson, Gradient-based iterative image reconstruction scheme for time-resolved optical tomography, IEEE Trans. Med. Imaging 18 (1999) 262–271.

[10] J. Ye, K. Webb, C. Bouman, R. Millane, Optical diffusion tomography using iterative coordinate descent optimization in a Bayesian framework, J. Opt. Soc. Am. A. 16 (1999) 2400–2412.

[11] S. Venkatakrishnan, L. Drummy, M. Jackson, M. De Graef, J. Simmons, C. Bouman, A model based iterative reconstruction algorithm for high angle

annular dark field-scanning transmission electron microscope (HAADF-STEM) tomography, IEEE Trans. Image Process. 22 (2013) 4532–4544.

[12] Model Based Iterative Reconstruction for Bright Field Electron Tomography, vol. 8657. URL: http://dx.doi.org/10.1117/12.2013228.

[13] J. Shier, Statistical Geometry ⟨http://john-art.com/stat_geom.html⟩, 2013.

[14] M. Jackson, Open MBIR BF + HAADF-STEM Tomographic Reconstruction Software ⟨http://bluequartz.net/projects/EIM_Tomography/⟩, 2013.

[15] A. Weickenmeier, H. Kohl, Computation of absorptive form factors for high-energy electron diffraction, Acta Crystallogr. A 47 (1991) 590–597.

[16] K. Sauer, C.A. Bouman, A local update strategy for iterative reconstruction from projections, IEEE Trans. Signal Process. 41 (2) (1993) 534–548.

[17] Z. Yu, J. Thibault, C. Bouman, K. Sauer, J. Hsieh, Fast model-based X-ray CT reconstruction using spatially nonhomogeneous ICD optimization, IEEE Trans. Image Process. 20 (1) (2011) 161–175, http://dx.doi.org/10.1109/TIP.2010.2058811.

[18] M. Kamasak, C. Bouman, E. Morris, K. Sauer, Direct reconstruction of kinetic parameter images from dynamic PET data, IEEE Trans. Med. Imaging 24 (2005) 636–650.

[19] J. Simmons, B. Bartha, M. De Graef, M. Comer, Development of Bayesian segmentation techniques for automated segmentation of titanium alloy images, Microsc. Microanal. 14 (S2) (2008) 602–603.