

BAYESIAN TOMOGRAPHIC RECONSTRUCTION FOR HIGH ANGLE ANNULAR DARK FIELD (HAADF) SCANNING TRANSMISSION ELECTRON MICROSCOPY (STEM)

Singanallur Venkatakrishnan¹, Lawrence Drummy², Michael Jackson³, Marc De Graef⁴, Jeff Simmons², and Charles Bouman¹

¹ Purdue University, School of Electrical and Computer Engineering, West Lafayette, IN

² Air Force Research Lab, Wright-Patterson Air Force Base, Dayton, OH

³ BlueQuartz Software, Springboro, OH

⁴ Carnegie Mellon University, Department of Material Science and Engineering, Pittsburgh, PA

ABSTRACT

HAADF-STEM data is increasingly being used in the physical sciences to study materials in 3D because it is free from the diffraction effects seen in Bright Field STEM data and satisfies the projection requirement for tomography. Typically, reconstruction is performed using Filtered Back Projection (FBP) or the SIRT algorithm. In this paper, we develop a Bayesian reconstruction algorithm for HAADF-STEM tomography which models the image formation, the noise characteristics of the measurement, and the inherent smoothness in the object. Reconstructions of polystyrene functionalized Titanium dioxide nano particle assemblies show results that are qualitatively superior to FBP and SIRT reconstructions, significantly suppressing artifacts and enhancing contrast.

Index Terms— Electron tomography, dark-field, Bayesian methods

1. INTRODUCTION

The last decade has seen a growing interest in the use of HAADF-STEM tomography to study materials (e.g. [1, 2]). A typical acquisition involves focusing an electron probe on the material and measuring the electrons scattered into an annular range of 50 – 300 mrad as show in Fig. 1(a). The electron beam is raster scanned and at each point a measurement is made, to eventually obtain a projection image of the object. The object is then tilted and the process is repeated. Thus, at the end of the acquisition a set of projection images is obtained corresponding to each tilt of the object. In most cases, due to mechanical constraints the object can only be tilted in the range of approximately $\pm 70^\circ$. In summary, HAADF-STEM tomography can be classified as a parallel beam, limited angle tomography modality. In practice, the object shifts during tilting and this is compensated by aligning the acquired images. More details of HAADF-STEM acquisition and pre-processing can be found in [3].

Most efforts reported in the literature use FBP or SIRT [4] to perform the inversion of the aligned projection images. However, due to the limited and noisy nature of the data, the resulting reconstructed volumes can show significant artifacts. Furthermore, algorithms such as SIRT require the selection of an ad-hoc stopping criterion; otherwise the reconstructed image will typically diverge or become excessively noisy.

Bayesian reconstruction algorithms have enabled significant qualitative and quantitative improvements in other tomographic modalities like X-Ray CT [5] motivating the extension to HAADF-STEM tomography. We begin by developing a forward model for the image formation in terms of the scatter coefficients, the number

of transmitted electrons and a tilt specific offset added by the microscope. While complex models for HAADF-STEM image formation [6] have been developed, the wide spread use of FBP and SIRT in producing reconstructions at medium resolutions (≈ 1 nm), indicates that a simple linear model for the image formation, implicit in those algorithms, is also useful. Next, we model the noise as a Gaussian with variance proportional to the mean value of the measurement to capture the Poisson characteristics of the signal [7]. We use a Generalized Gaussian Markov Random Field (GGMRF) [8] prior for the scattering coefficients as it has been shown to have good edge preserving characteristics and hence is suitable for materials containing strongly scattering grains embedded in a weakly scattering support material. Finally, we formulate the MAP estimation problem and develop an algorithm to minimize the corresponding cost function. Experimental results show that the Bayesian approach helps to effectively suppress artifacts and reduce noise compared to SIRT and FBP.

2. MEASUREMENT MODEL

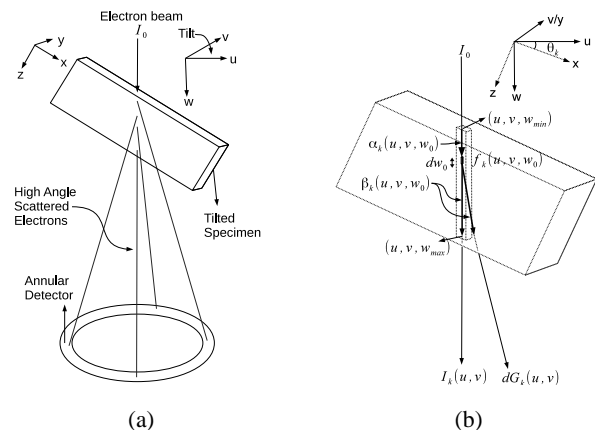


Fig. 1. (a) HAADF-STEM measurement setup (b) Equivalent model.

The goal of HAADF-STEM tomography is to reconstruct the HAADF scatter coefficients (units of nm^{-1}) denoted by $f(x, y, z)$ at every point in space. If (x, y, z) is the reference frame of the object and (u, v, w) is the reference frame for the electron source (See Fig. 1), then any function of space can be reparameterized

so that $f_k(u, v, w) = f([u, v, w]R_{\theta_k})$ where R_{θ_k} is an orthonormal rotation of the spatial coordinates by an angle θ_k . For tomographic reconstruction we require measurements of the projection integral $\int_{-\infty}^{+\infty} f_k(u, v, w)dw$ through the object for every tilt θ_k and every point (u, v) . We begin by describing how this measurement can be obtained from the HAADF-STEM signal. Let I_0 be the source electron flux in units of electrons per nm^2 . Let $\mu_k(u, v, w) = \mu([u, v, w]R_{\theta_k})$ be the electron attenuation coefficient in units of nm^{-1} at (u, v, w) . If we consider a beam which travels along a path given by (u, v, w) for $w \in [w_{min}, w_{max}]$ for the k^{th} tilt angle θ_k as shown in Fig. 1(b), then the electron flux transmitted through the object along this line can be modeled using Beer's Law and is given by

$$I_k(u, v) = I_0 \exp \left\{ - \int_{w_{min}}^{w_{max}} \mu_k(u, v, w)dw \right\}$$

where $\mu_k(u, v, w)$ accounts for all mechanisms that cause attenuation including HAADF scatter. If we consider a specific point (u, v, w_o) along the path of the transmitted electron beam, the attenuation can be broken into two components, the component before the point is reached and the component after the point is past. More specifically let us define the following three functions,

$$\begin{aligned} \alpha_k(u, v, w_o) &= \exp \left\{ - \int_{w_{min}}^{w_o} \mu_k(u, v, w)dw \right\} \\ \beta_k(u, v, w_o) &= \exp \left\{ - \int_{w_o}^{w_{max}} \mu_k(u, v, w)dw \right\} \\ \gamma_k(u, v) &= \alpha_k(u, v, w_o)\beta_k(u, v, w_o) \\ &= \exp \left\{ - \int_{w_{min}}^{w_{max}} \mu_k(u, v, w)dw \right\}. \end{aligned} \quad (1)$$

Then the transmitted flux is given by

$$I_k(u, v) = I_0 \gamma_k(u, v). \quad (3)$$

At a given point (u, v, w_o) ($w_o \in [w_{min}, w_{max}]$) along the line we assume that a fraction of the incident attenuated beam is scattered and this scattered beam then propagates to the detector. Since the electrons are only scattered by small angles in the range of 50 – 300 mrad we approximate the path of the HAADF scattered electrons to be along the line (u, v, w) for $w \in [w_o, w_{max}]$. Therefore the total HAADF scatter can be derived as

$$G_k(u, v) = \int_{w_{min}}^{w_{max}} I_0 \alpha_k(u, v, w_o) f_k(u, v, w_o) \beta_k(u, v, w_o) dw_o$$

Using (1), we get

$$G_k(u, v) = I_0 \gamma_k(u, v) \int_{w_{min}}^{w_{max}} f_k(u, v, w)dw. \quad (4)$$

Putting equation (3) and (4) together, we obtain the following relationship

$$\frac{G_k(u, v)}{I_k(u, v)} = \int_{w_{min}}^{w_{max}} f([u, v, w]R_{\theta_k})dw. \quad (5)$$

So the normalized quantity $G_k(u, v)/I_k(u, v)$ is an estimate of the tomographic projection of the HAADF scatter coefficients at angle θ_k and at position (u, v) . We note that this process of normalizing the measured signal by the attenuated beam is conceptually similar to the attenuation correction in Positron Emission Tomography.

Further we note that Bright Field electron tomography is similar to transmission tomography and Dark Field tomography is similar to emission tomography. Since this normalization value $I_k(u, v)$ is not measured at present we assume that it is a constant for each tilt. That is

$$I_k(u, v) = I_k \quad (6)$$

This will occur when $\mu(x, y, z)$ is a constant and the material is of constant thickness. We also observe that the HAADF-STEM signal at each tilt is offset by a value, d_k (counts). We model the i^{th} measurement (corresponding to the electron source at (u_i, v_i)) at tilt k by a Gaussian random variable $g_{k,i}$, with mean

$$\mathbb{E}[g_{k,i}] = \iint G_k(u, v) h_i(u, v) dudv + d_k$$

where $h_i(u, v)$ is a kernel which averages the electron flux over the area of the i^{th} pixel. If f is a discretized version of $f(x, y, z)$ organized as a $M \times 1$ vector, where M is the total number of voxels and, A_k is a $P \times M$ projection matrix for tilt k , where P is the number of measurements per tilt then using (4) and (6)

$$\begin{aligned} \mathbb{E}[g_{k,i}] &= I_k \iint \left(\int_{w_{min}}^{w_{max}} f_k(u, v, w)dw \right) h_i(u, v) dudv + d_k \\ &= I_k (A_k f)_i + d_k \end{aligned}$$

where $(A_k f)_i$ is the i^{th} entry of the vector $A_k f$. The variance of each measurement is given by

$$\text{Var}[g_{k,i}] = \sigma_k^2 \mathbb{E}[g_{k,i}]$$

where σ_k^2 is a parameter used to model the noise variance at tilt k and $\mathbb{E}[g_{k,i}]$ accounts for the Poisson characteristics of the measurement. We assume that all the measurements are conditionally independent. If $\mathbb{E}[g_{k,i}] \approx g_{k,i}$, $\Lambda_k = \text{diag} \left(\frac{1}{g_{k,1}}, \dots, \frac{1}{g_{k,P}} \right)$, $g_k = [g_{k,1}, \dots, g_{k,P}]^t$, $g = [g_1^t \dots g_N^t]^t$, $I = [I_1, \dots, I_N]$, $d = [d_1, \dots, d_N]$ and $\sigma^2 = [\sigma_1^2, \dots, \sigma_N^2]$ then

$$\begin{aligned} p(g|f, I, d, \sigma^2) &= \left(\prod_{k=1}^N \frac{1}{(2\pi\sigma_k^2)^{\frac{P}{2}} |\Lambda_k|^{-\frac{1}{2}}} \right) \\ &\exp \left\{ -\frac{1}{2} \sum_{k=1}^N \frac{1}{\sigma_k^2} \|g_k - I_k A_k f - d_k \mathbb{1}_{\Lambda_k}\|^2 \right\} \end{aligned} \quad (7)$$

where N is the total number of tilts.

3. PRIOR MODEL

We use a GGMRF [8] model for the probability density of f . If χ is the set of all neighboring voxels (26 point neighborhood), w_{ij} is a weighting kernel which is inversely proportional to the distance between voxel i and voxel j , normalized so that $\sum_{j \in \partial i} w_{ij} = 1$, ∂i is the set of all neighbors of voxel i , then

$$p(f) = \frac{1}{Z} \exp \left\{ -\frac{1}{p\sigma_f^p} \sum_{\{i,j\} \in \chi} w_{ij} |f_i - f_j|^p \right\} \quad (8)$$

where Z is a normalizing constant and p and σ_f are GGMRF parameters. Typically $1 < p \leq 2$ is used to ensure convexity of the subsequent MAP cost function. When p is close to 1 it corresponds to strong edge preserving reconstructions while $p = 2$ corresponds to smooth reconstructions.

4. MAP ESTIMATION

The posterior probability of the parameters (f, I, d, σ^2) given the data g is

$$p(f, I, d, \sigma^2 | g) = \frac{p(g|f, I, d, \sigma^2)p(f)p(I, d, \sigma^2)}{p(g)} \quad (9)$$

where the parameter f is assumed to be independent of (I, d, σ^2) . Further we constrain $f \geq 0$ and the arithmetic mean of I to be equal to \bar{I} . The constraint on I is present to prevent the subsequent optimization algorithm from diverging to unreasonable values of the scatter coefficients. \bar{I} is arbitrary but affects the scaling of f and therefore the choice of σ_f . The MAP estimate is given by

$$(\hat{f}, \hat{I}, \hat{d}, \hat{\sigma}^2) = \underset{f \geq 0, I \in \Omega, d, \sigma^2}{\operatorname{argmin}} \left\{ -\log p(g|f, I, d, \sigma^2) - \log p(f) \right\}$$

where $\Omega = \left\{ I \in \mathbb{R}^N : \frac{1}{N} \sum_{k=1}^N I_k = \bar{I} \right\}$ and we have assumed uniform priors for the $p(I, d, \sigma^2)$. Using (7), and (8) we obtain

$$(\hat{f}, \hat{I}, \hat{d}, \hat{\sigma}^2) = \underset{f \geq 0, I \in \Omega, d, \sigma^2}{\operatorname{argmin}} c(f, I, d, \sigma^2)$$

where

$$c(f, I, d, \sigma^2) = \frac{1}{2} \sum_{k=1}^N \frac{1}{\sigma_k^2} \|g_k - I_k A_k f - d_k \mathbb{1}\|_{\Lambda_k}^2 \quad (10)$$

$$+ \frac{1}{2} \sum_{k=1}^N \log \left((2\pi\sigma_k^2)^P |\Lambda_k|^{-1} \right) + \frac{1}{p\sigma_f^p} \sum_{\{i,j\} \in \mathcal{X}} w_{ij} |f_i - f_j|^p$$

5. OPTIMIZATION ALGORITHM

While the cost function $c(f, I, d, \sigma^2)$ is convex in f (when $1 < p \leq 2$) it is not jointly convex in (f, I, d, σ^2) . We adapt the ICD algorithm [9] to minimize the cost function. The basic structure of the optimization routine is to repeatedly perform the following 3 steps until convergence is achieved.

1. $\hat{f} \leftarrow \underset{f_1 \geq 0, \dots, f_M \geq 0}{\operatorname{argmin}} c(f, I, d, \sigma^2)$
2. $(\hat{I}, \hat{d}) \leftarrow \underset{I \in \Omega, d}{\operatorname{argmin}} c(\hat{f}, I, d, \sigma^2)$
3. $\hat{\sigma}^2 \leftarrow \underset{\sigma^2}{\operatorname{argmin}} c(\hat{f}, \hat{I}, \hat{d}, \sigma^2)$

The algorithm is terminated if the relative change in the magnitude of the reconstruction is less than a preset threshold. While each of the above steps reduces the cost function there is no guarantee it will converge to a global minimum. Therefore, it is important to have reasonable initial estimates for the parameters. For Step 1 we use the update equations as in [5] to find the minimum of the cost function with respect to each voxel. For Step 2 we turn the constrained optimization problem into an unconstrained one by using Lagrange multiplier λ . Rewriting the first summation of the cost function in (10) as a sum of quadratics in $[I_k, d_k]$ and dropping terms which do not involve I_k, d_k gives us a new cost function corresponding to the unconstrained optimization problem,

$$\begin{aligned} \tilde{c}(I, d, \lambda) = & \frac{1}{2} \sum_{k=1}^N \left([I_k \ d_k] Q_k \begin{bmatrix} I_k \\ d_k \end{bmatrix} - 2 [I_k \ d_k] b_k + c_k \right) \\ & + \lambda \left(\frac{1}{N} \sum_{k=1}^N I_k - \bar{I} \right) \end{aligned}$$

where $Q_k = \begin{bmatrix} (A_k \hat{f})^t \tilde{\Lambda}_k (A_k \hat{f}) & (A_k \hat{f})^t \tilde{\Lambda}_k \mathbb{1} \\ (A_k \hat{f})^t \tilde{\Lambda}_k \mathbb{1} & \mathbb{1}^t \tilde{\Lambda}_k \mathbb{1} \end{bmatrix}$, $b_k = \begin{bmatrix} b_{k,1} \\ b_{k,2} \end{bmatrix} = \begin{bmatrix} g_k^t \tilde{\Lambda}_k A_k \hat{f} \\ g_k^t \tilde{\Lambda}_k \mathbb{1} \end{bmatrix}$, $c_k = g_k^t \tilde{\Lambda}_k g_k$ and $\tilde{\Lambda}_k = \frac{1}{\sigma_k^2} \Lambda_k$. $\tilde{c}(I, d, \lambda)$ can be minimized with respect to (I, d, λ) by taking the gradient and setting it to zero. If $Q_k^{-1} = \tilde{Q}_k = \begin{bmatrix} \tilde{q}_{k,11} & \tilde{q}_{k,12} \\ \tilde{q}_{k,21} & \tilde{q}_{k,22} \end{bmatrix}$, we obtain

$$\hat{\lambda} = \frac{\sum_{k=1}^N (\tilde{q}_{k,11} b_{k,1} + \tilde{q}_{k,12} b_{k,2}) - N \bar{I}}{\frac{1}{N} \sum_{k=1}^N \tilde{q}_{k,11}}$$

$$\begin{bmatrix} \hat{I}_k \\ \hat{d}_k \end{bmatrix} = Q_k^{-1} \left(b_k - \frac{1}{N} \begin{bmatrix} \hat{\lambda} \\ 0 \end{bmatrix} \right). \quad (11)$$

For Step 3 we can take the gradient of (10) with respect to σ^2 and set it to zero. This gives us the optimal update for each σ_k^2 as

$$\hat{\sigma}_k^2 = \frac{\|e_k\|_{\Lambda_k}^2}{P} \quad (12)$$

where $e_k = g_k - \hat{I}_k A_k \hat{f} - \hat{d}_k \mathbb{1}$ and P is the number of measurements for each tilt.

6. EXPERIMENTAL RESULTS

In order to evaluate our approach, we compare our algorithm with FBP and SIRT from a popular electron microscopy software package, IMOD [10]. The data acquired is of a ≈ 150 nm thick sample of polystyrene functionalized Titanium dioxide nano particle assembly [11]. The sample is tilted in steps of 1° from $\pm 50^\circ$ to $\pm 70^\circ$ and in steps of 2° from -50° to $+50^\circ$ to acquire 87 images with pixels of size 0.343 nm \times 0.343 nm. The data is aligned post acquisition using a cross correlation method. We use a ≈ 350 nm \times 350 nm section of the projection images for reconstruction.

The FBP and SIRT reconstructions are performed with voxels of size 0.343 nm \times 0.343 nm \times 0.343 nm. The filter parameters for FBP are chosen to produce the most visually appealing results. The particles of interest in this data set are approximately cylindrical with diameter 18 nm and height 40 nm [11]. Thus in order to reduce computation, Bayesian reconstruction is performed with voxels of size (3×0.343) nm \times (3×0.343) nm \times (3×0.343) nm. The parameter σ_f in the Bayesian method is chosen for the best visual quality of reconstruction with the GGMRf shape parameter $p = 1.2$. The algorithm is initialized using a 4 stage multi-resolution approach. At the coarsest resolution the offset parameter vector d is initialized by a least squares fit of the average count in each view k to $\frac{1}{\cos(\theta_k)}$. We choose $\frac{1}{\cos(\theta_k)}$ because it is proportional to the measured counts of a homogeneous material of constant thickness at tilt k . More specifically, if G is a $N \times 1$ vector containing the average counts of the data at each tilt, and $D = \begin{bmatrix} \frac{1}{\cos(\theta_1)} & \dots & \frac{1}{\cos(\theta_N)} \\ 1 & \dots & 1 \end{bmatrix}^t$ then the least squares estimate is given by $[\phi_1 \ \phi_2]^t = (D^t D)^{-1} (D^t G)$ and the initial value of d is set to $\phi_2 \mathbb{1}$. The initial value of I is set to $\bar{I} \mathbb{1}$, $\sigma^2 = \mathbb{1}$ and $f = 0$ at the coarsest resolution. At the very first iteration at the coarsest scale we perform a fixed number (10) of iterations over f to prevent the algorithm from converging to an unreasonable result. In this experiment we set $\bar{I} = 20000$ counts. The value of σ_f

set for the finest resolution is adapted for coarse resolutions as given in [12]. The stopping threshold for the algorithm is set to 0.9%. The dimensions of the reconstructed volume are set so as to account for all the voxels contributing to the projection data. In presenting the results we only show voxels that can be reliably reconstructed from the projection data i.e. at every tilt there is a projection measurement corresponding to those voxels.

Fig. 2 shows a single $x - z$ and $x - y$ slice from FBP, SIRT and the Bayesian method. Fig. 3 shows the initial values and final estimates of the parameters. We observe that in SIRT and FBP there are streaking artifacts in the $x - z$ plane of reconstruction. The Bayesian approach significantly suppresses these artifacts. The presence of material towards the top and bottom of the slice indicates model mismatch, which we believe may be possible to eliminate if we incorporate the transmission measurement. In the $x - y$ plane the effects of noise are effectively suppressed in the Bayesian technique clearly showing the Titanium dioxide nano particles against the background support material. This demonstrates the effectiveness of the method even for this particularly noisy data set. Finally we note that further work needs to be done to compare the reconstructions with ground truth data, and to test the algorithm using phantoms.

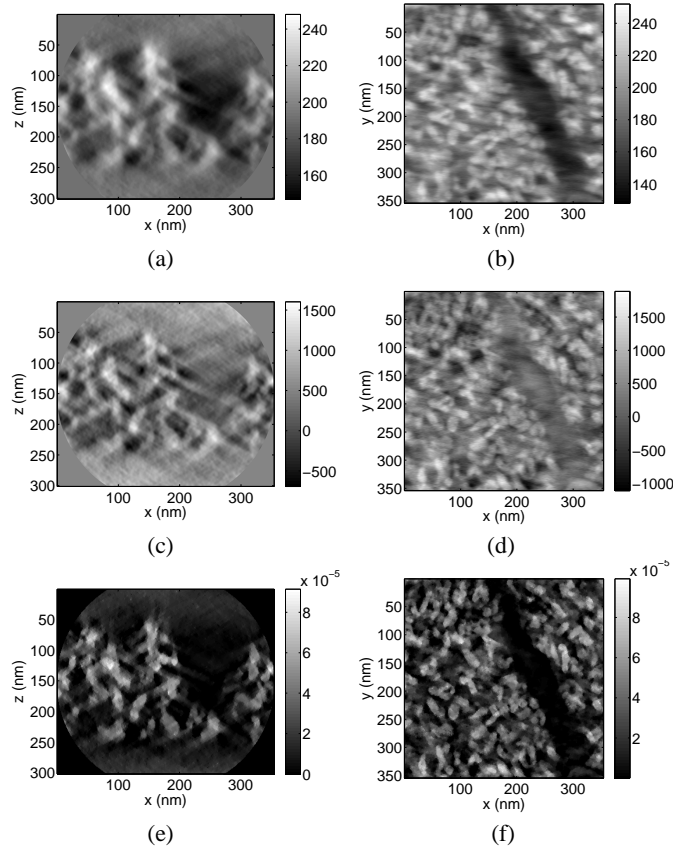


Fig. 2. Left panel shows a $x - z$ reconstructed slice and the right panel shows a $x - y$ reconstructed slice. (a) and (b) FBP reconstruction, (c) and (d) SIRT with 20 iterations, (e) and (f) Bayesian reconstruction with $p = 1.2$.

7. CONCLUSIONS

We developed a model for the image formation process in HAADF-STEM tomography and pointed out the need to normalize the mea-

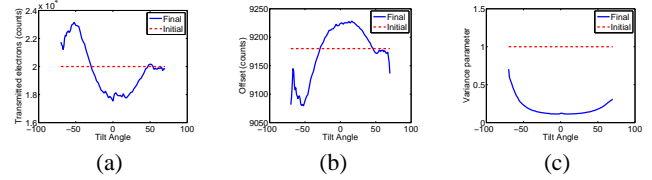


Fig. 3. Initial value and final estimates for (a) Transmitted electrons, I (b) Offsets, d and (c) Variance parameter, σ^2

sured signal by the transmitted beam. We combine our forward model with a prior model of the 3D volume and formulate an algorithm for quantitative Bayesian reconstruction of the local HAADF scatter rate per unit distance. The method also accounts for some unknown calibration parameters such as number of transmitted electrons, offset, and noise variance. Reconstructions on real data show significant qualitative improvement over FBP and SIRT from IMOD [10].

8. REFERENCES

- [1] Hanying Li, Huolin L. Xin, David A. Muller, and Lara A. Estroff, “Visualizing the 3D internal structure of calcite single crystals grown in agarose hydrogels,” *Science*, vol. 326, no. 5957, pp. 1244–1247, 2009.
- [2] Peter R. Buseck, Rafal E. Dunin-Borkowski, Bertrand Devouard, Richard B. Frankel, Martha R. McCartney, Paul A. Midgley, Mihly Psfai, and Matthew Weyland, “Magnetite morphology and life on mars,” *Proceedings of the National Academy of Sciences*, vol. 98, no. 24, pp. 13490–13495, 2001.
- [3] P.A. Midgley and M. Weyland, “3D electron microscopy in the physical sciences: the development of Z-contrast and EFTEM tomography,” *Ultramicroscopy*, vol. 96, no. 34, pp. 413 – 431, 2003.
- [4] Avinash C. Kak and Malcolm Slaney, *Principles of Computerized Tomographic Imaging*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2001.
- [5] Zhou Yu, J. Thibault, C.A. Bouman, K.D. Sauer, and J. Hsieh, “Fast model-based X-ray CT reconstruction using spatially nonhomogeneous ICD optimization,” *Image Processing, IEEE Transactions on*, vol. 20, no. 1, pp. 161 –175, January 2011.
- [6] Stephen J. Pennycook and Peter D Nellist, *Scanning Transmission Electron Microscopy*, Springer, 223 Spring Street, New York, NY 10013, USA, 2011.
- [7] M. Wernick and J. Aarsvold, *Emission Tomography*, Elsevier Academic Press, 525 B Street, Suite 1900, San Diego, California 92101-4495, USA, 2004.
- [8] C. Bouman and K. Sauer, “A generalized gaussian image model for edge-preserving map estimation,” *Image Processing, IEEE Transactions on*, vol. 2, no. 3, pp. 296 –310, July 1993.
- [9] K. Sauer and C. A. Bouman, “A local update strategy for iterative reconstruction from projections,” *IEEE Trans. on Signal Processing*, vol. 41, no. 2, pp. 534–548, February 1993.
- [10] Kremer J.R. and D.N. Mastrorarde and J.R. McIntosh, “Computer visualization of three-dimensional image data using IMOD,” *Journal Of Structural Biology*, vol. 116, pp. 71 –76, aug 1996.
- [11] Maxim N. Tchoul, Scott P. Fillery, Hilmar Koerner, Lawrence F. Drummy, Folusho T. Oyerokun, Peter A. Mirau, Michael F. Durrstock, and Richard A. Vaia, “Assemblies of Titanium Dioxide-Polystyrene Hybrid Nanoparticles for Dielectric Applications,” *Chemistry of Materials*, vol. 22, no. 5, pp. 1749–1759, 2010.
- [12] Seungseok Oh, A.B. Milstein, C.A. Bouman, and K.J. Webb, “A general framework for nonlinear multigrid inversion,” *Image Processing, IEEE Transactions on*, vol. 14, no. 1, pp. 125 –140, jan. 2005.