

# Validation and Generalization of a Method for Precise Size Measurements of Metal Nanoclusters on Supports

B. W. Reed,<sup>1</sup> D. G. Morgan,<sup>2,3</sup> N. L. Okamoto,<sup>2,4</sup> A. Kulkarni,<sup>2</sup> B. C. Gates,<sup>2</sup> and N. D. Browning<sup>1,2</sup>

## Affiliations

<sup>1</sup>Materials Science and Technology Division, Lawrence Livermore National Laboratory, Livermore, CA 94550

<sup>2</sup>Department of Chemical Engineering and Materials Science, University of California, Davis, CA 95616

<sup>3</sup>CryoEM Facility, Indiana University, Bloomington, IN 47405

<sup>4</sup>Department of Materials Science and Engineering, Kyoto University, Sakyo-ku, Kyoto, Japan

## Corresponding Author Contact Information:

Bryan W. Reed

Lawrence Livermore National Laboratory

7000 East Avenue, L-356

Livermore, CA 94551

Phone: (925) 423-3617

Fax: (925) 423-7040

Email: reed12@llnl.gov

## Abstract

We recently described a data analysis method for precise ( $\sim 0.1$  Å random error in the mean for a 200 kV instrument with a 3 Å FWHM probe size) size measurements of small clusters of heavy metal atoms on supports as imaged in a scanning transmission electron microscope, including an experimental demonstration using clusters that were primarily triosmium or decaosmium. The method is intended for low signal-to-noise ratio images of radiation-sensitive samples. We now present a detailed analysis, including a generalization to address issues of particle anisotropy and biased orientation distributions. In the future, this analysis should enable extraction of shape as well as size information, up to the noise-defined limit of information present in the image. We also present results from an extensive series of simulations designed to determine the method's range of applicability and expected performance in realistic situations. The simulations reproduce the experiments quite accurately, enabling a correction of systematic errors so that only the  $\sim 0.1$  Å random error remains. The results are very stable over a wide range of parameters. We introduce a variation on the method with improved precision and stability relative to the original version, while also showing how simple diagnostics can test whether the results are reliable in any particular instance.

**PACS Codes:** 36.40.-c, 61.16.Bg, 82.65.Jv

## Keywords:

Scanning Transmission Electron Microscopy (STEM); Microscopic methods, specifically for catalysts and small particles; Data processing/image processing

## I. Introduction

Metal nanoparticles are ubiquitous and essential components in many industrial catalysts [1,2]. The catalytic action of nanoparticles is usually enhanced by reducing the particle size, which increases the fraction of the metal atoms that are surface atoms and accessible to reactant molecules. Also, for some reactions the catalytic activity per surface atom is enhanced by nano-size effects [2,3]. Thus, there is a strong motivation for investigation of the smallest metal particles, stabilized on the surfaces of supports. Investigations of such supported particles have been facilitated by advances in electron microscopy, especially aberration-corrected high-angle annular dark field (HAADF) scanning transmission electron microscopy (STEM), [4,5], which is capable of atomic resolution and single-atom sensitivity, especially in the case of heavy atoms on a low-density support.

Yet the high electron probe current densities enabled by aberration correction have also highlighted the limited stability of the sample under the influence of the electron beam. Although high-resolution, high signal-to-noise ratio (SNR) images can now be rapidly acquired, the motion of small particles and of the atoms within them can be so extreme that image interpretation becomes nearly impossible. This limitation pertains especially to particles weakly bonded to a support, as in many catalysts. This limitation forces the STEM operator to reduce the probe current, the scan resolution, and/or the dwell time in order to reduce the dose to the specimen. This inevitably reduces the SNR, often to levels at which it is difficult to know whether a given feature is real or random noise. Developments in biological and cryogenic TEM, and associated data analysis techniques, have yielded enormous progress on this problem, [6-8] for example by developing ways to combine data from a large number of nominally identical particles. These techniques have only rarely been adapted to non-biological samples.

There is thus a need for straightforward image analysis techniques appropriate to low-SNR HAADF STEM images of small particles and clusters of heavy atoms. We recently [9] introduced such a technique that uses a combination of curve fitting and progressive blurring to produce a robust estimate of particle size distributions corrected for all known natural and artificial blurring effects. Our method is similar in some mathematical respects to established particle-tracking algorithms [10-12], especially in the effects of SNR on the curve fits. Yet it also differs in some essential ways (e.g. we are far more interested in size, mass, and morphology than in precise position). The method was developed on a test case involving MgO-supported  $\text{Os}_3$  and  $\text{Os}_{10}\text{C}$  clusters with associated CO ligands, measured with a 200 kV STEM with 3 Å FWHM probe size. The resulting distribution of root-mean-square radii ( $r_{\text{RMS}}$ ) for isolated single particles agreed with independent measurements to within  $\sim 0.2\text{\AA}$  or better, even with a relatively small sample size of only 30 particles total. The larger particles seemed to produce  $r_{\text{RMS}}$  values somewhat larger than expected, but this result was of relatively low statistical significance (a discrepancy of 1.8 to 2.4 standard deviations, depending on the theoretical model). The random error in the mean was of order 0.1 Å. The relevant quantitative results are summarized in the first three columns of Table 1.

In the present article we expand on this test case. First, we explicitly examine the mathematical assumptions underlying the method and see how we may generalize the formalism, for example by examining the role of anisotropy. We suggest how the method could be extended to estimate not just the scalar  $r_{\text{RMS}}$  but also the full second-rank tensor equivalent for anisotropic particles, and we also derive constraints on the degree to which an anisotropic orientation

distribution can bias the results. We also highlight what is probably the least obvious assumption in the derivation, namely, that a curve fit to the  $r_{RMS}$  estimates from progressive blurring of the same image should produce a meaningful result. The validity of this least-squares fit cannot be derived from the usual maximum-likelihood concepts, but must instead be based on empirical or semi-empirical tests.

Second, in order to supply such tests, we present results from a series of simulations—synthetic data sets—that were designed to produce images nominally equivalent to our measurements. For these synthetic images, we knew exactly what each atomic cluster truly looks like, so that we knew what answer *should* have been produced by the analysis algorithm. This let us address directly some questions left open in our previous work:

1. For the SNR and SBR (signal-to-background ratio) of our experiments, should the algorithm have performed as well as it appears to have performed?
2. Does the progressive-blurring curve fit perform its intended function, namely, to produce an optimized estimate of particle size that minimizes the effects of noise and background?
3. Is there an alternative to the linear progressive-blurring fit that makes better use of the information gathered in the set of blurred images?
4. Is there a systematic error introduced by the algorithm, and does this error vary in a predictable way with parameters such as particle size? If so, can the estimate of actual particle sizes be improved?
5. Can we make use of additional information, such as the estimated particle masses and the patterns in the progressive-blurring analysis, to understand more about the particles?

In light of the simulation results, we claim that all of these questions can be answered in the affirmative, although with some qualifications.

## II. Mathematical Methods

### A. Overview

The general idea and mathematical formalism of the convolution-blurring/curve-fitting algorithm are described in previous work [9] and summarized here. Suppose that we have a HAADF STEM image with clearly identifiable bright spots representing isolated small atomic clusters on a relatively slowly varying background, and that those clusters are drawn from a statistical distribution of perhaps a few different sizes and shapes. We are interested in a precise measurement of the size of each kind of cluster, as well as an ability to distinguish various clusters on the basis of their size in order to produce population statistics of the various species.

To do this, we curve fit each particle image with a Gaussian peak plus a polynomial background to extract an estimate of its  $r_{RMS}$  (as well as its total "mass," defined below, which can also be used to help classify particles into species). This estimate may be subject to bias arising from the background signal, from random noise peaks, from the fact that the particle itself is not a Gaussian distribution of mass density, and from various point spread functions that blur the image. In order to try to average out the effects of some of these biases, we deliberately blur the image—digitally—using a Gaussian kernel with width  $\sigma_{gb}$ . Small values of  $\sigma_{gb}$  may be more prone to random noise spikes and limited quality of the curve fit, whereas larger values may be more influenced by effects from the background and, eventually, from the fact that the measured size will be dominated by the blur. A curve fit of the extracted  $r_{RMS}^2$  versus  $\sigma_{gb}^2$  results

in an estimate of what would have been measured in a noise-free measurement with zero artificial blurring. This measurement is further corrected for systematic effects arising from the known point spread function including contributions from the probe size, vibrations, and focus errors. Focus errors are minimized by taking a through-focus series and analyzing, for each particle, its image at the focus value for which it appeared smallest.

## B. RMS size metrics

Because of its well-behaved mathematical properties, we chose the root-mean-square (RMS) radius  $r_{RMS}$ , weighted by the experimentally relevant detection cross section  $S_i$  for each atom, as an appropriate measure of particle size. The definition is

$$r_{RMS} = \left( \frac{1}{m} \sum_i S_i |\vec{r}_i|^2 \right)^{1/2}, \quad (1)$$

where  $\vec{r}_i = (x_i, y_i, z_i)$  is the position of atom  $i$  in center-of-mass coordinates and  $m$  is the "mass" (i.e., total signal per molecule), defined as

$$m = \sum_i S_i. \quad (2)$$

The  $S_i$  factor can be calculated from the atomic mass, TEM accelerating voltage, and angular range of the detector by integrating the output of freely available software packages [13].

The quantity  $r_{RMS}$  is by definition isotropic and is thus a scalar invariant property of the cluster. When the cluster is actually viewed in a STEM, we will see a projection onto the x-y plane of the mass density for some particular orientation, convolved with various point spread functions (PSF). Assuming we can correct for the PSFs, we can deduce from the image the contributions (which we call  $\sigma_{obj-x}^2$  and  $\sigma_{obj-y}^2$ ) to the squared  $r_{RMS}$  arising from the  $x$  and  $y$  components. Since  $r_{RMS}^2 = \sigma_{obj-x}^2 + \sigma_{obj-y}^2 + \sigma_{obj-z}^2$ , we have to make some assumptions about the unmeasured  $\sigma_{obj-z}^2$  if we want to extract the orientation-independent measure of particle size  $r_{RMS}$ .

In an experiment, one would measure a large number of clusters and seek to average the results. Since the RMS measures of size which we are using add in quadrature (a fact which we use extensively throughout this work), it is most convenient to take the means in quadrature as well. We define the orientation ensemble average  $\langle f \rangle$  as the mean of some quantity  $\langle f \rangle$  over a large number of measurements of independent particles drawn from some random distribution of orientations. In reference [9], we assumed that the ensemble of particles was statistically isotropic, that is, that all orientations of the particles were equally likely. This assumption means that the various RMS widths could not depend on the choice of coordinate axes, so that we can delete the  $x$ ,  $y$ , and  $z$  subscripts for each of the ensemble averaged quantities:

$$\langle r_{RMS}^2 \rangle = \langle \sigma_{obj-x}^2 + \sigma_{obj-y}^2 + \sigma_{obj-z}^2 \rangle = \langle \sigma_{obj-x}^2 \rangle + \langle \sigma_{obj-y}^2 \rangle + \langle \sigma_{obj-z}^2 \rangle = 3 \langle \sigma_{obj}^2 \rangle \quad (3)$$

In other words, if the sampling is isotropic, then we may take the experimentally accessible quantities  $\langle \sigma_{obj-x}^2 \rangle^{1/2}$  or  $\langle \sigma_{obj-y}^2 \rangle^{1/2}$  as proxies, statistically equivalent (aside from a constant factor of  $3^{1/2}$ ) to the desired orientation-independent measure of particle size  $r_{RMS}$ .

This leaves the problem of subtracting off the various contributions to the PSF (e.g. probe size, thermal vibrations, focus errors, and deliberate gaussian blurring.), which may be approximated as simple convolutions with various point spread functions, each of which will have its own RMS width  $\sigma_Q$ . One of the convenient properties of RMS measures of size is that they always add in quadrature for any convolution effects,[14] regardless of the specific functional forms involved. We thus arrive at essentially equation (2) from [9], relating the ideal measured RMS width to the intrinsic width from the size of the object:

$$\sigma_{meas-x}^2 = \sigma_{obj-x}^2 + \sigma_{probe-x}^2 + \sigma_{vib-x}^2 + \sigma_{focus-x}^2 + \sigma_{gb-x}^2 \quad (4)$$

(and similarly for the y components).

Thus, as long as all of these blurring effects are sufficiently precisely quantified, they simply contribute a systematic error that is easily corrected. Evaluation of the contributions of uncertainties in these RMS widths to the final uncertainty in  $r_{RMS}$  is an elementary exercise in error propagation [15] and will not be considered further in this paper, except to note that the relative error in  $r_{RMS}$  is greatly magnified if any of these  $\sigma_Q$ 's is similar to it in magnitude. Thus, unsurprisingly, the algorithm will do a poor job of measuring the sizes of particles that are smaller than the experimental point spread function. We do not mean to leave the impression that these effects are unimportant; in fact, an accurate calibration of all of the contributions to the PSF is essential to obtaining good results, and precise measurements of the PSF can be challenging in practice.

### C. Isotropy and Anisotropy

In general, neither the particles nor the ensemble will be isotropic, and we now consider how to extend our previous method [9] to these cases. The RMS measures of size derive from the scaled second moment tensor, defined as

$$M_{jk} = \frac{1}{m} \sum_i S_i r_{i,j} r_{i,k} \quad (5)$$

(again, in the center of mass frame, and defining, for example,  $r_{i,x} = x_i$ ). So, for example,  $\sigma_{obj-x}^2 = M_{xx}$ . As for any symmetric, real, positive definite matrix, there will be a rotation  $(x, y, z) \rightarrow (x', y', z')$  that diagonalizes  $M_{jk}$ , yielding non-negative eigenvalues  $A$ ,  $B$ , and  $C$  that we are free to label in nondecreasing order, so that

$$M'_{jk} = \begin{bmatrix} \sigma_{obj-x'}^2 & 0 & 0 \\ 0 & \sigma_{obj-y'}^2 & 0 \\ 0 & 0 & \sigma_{obj-z'}^2 \end{bmatrix} \equiv \begin{bmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{bmatrix}, 0 \leq A \leq B \leq C, \quad (6)$$

and because the trace of a tensor is independent of orientation, it happens that

$$r_{RMS}^2 = TrM = TrM' = A + B + C. \quad (7)$$

The double-projection  $\sigma_{obj-x}^2$  of this second moment onto any x-axis must lie between  $A$  and  $C$  (inclusive).

We can now explore the assumption that the experimental ensemble average  $\langle f \rangle$  is drawn from an isotropic distribution of particle orientations. The worst-case scenario for this assumption would be when (1) one of the eigenvalues is very different from the other two, and (2) the particles are much more likely to lie in some orientations than others with respect to the electron probe.

Consider the limiting case of a flat disk-like particle with  $A = 0$  and  $B = C$ . If such particles always lie flat on the substrate, then the user will always observe  $\sigma_{obj-x}^2 = \sigma_{obj-y}^2 = C$ , to within experimental precision. If the user then assumes that the ensemble is isotropic, this would be equivalent to assuming that  $A$  is also equal to  $C$ , and the resulting  $r_{RMS}$  will be overestimated by a factor of  $(3/2)^{1/2} = 1.225$ . At the other extreme, if these disks are always viewed on edge, then the user will obtain  $\sigma_{obj-x}^2 + \sigma_{obj-y}^2 = C$ , and the assumption of isotropy will yield

$r_{RMS}^2 = 3C/2$ , when the correct answer is  $2C$ , so that  $r_{RMS}$  is now underestimated by a factor of  $(3/4)^{1/2} = 0.866$ . Similarly, in the case of an extremely rod-like particle ( $A = B = 0$ ,  $C > 0$ ), the cases are bound by the unlikely case of all particles aligned with the electron beam (so that the user measures  $r_{RMS} = 0$ ), and all particles lying flat on the support (with  $r_{RMS}$  overestimated by  $(3/2)^{1/2}$ ).

It is possible to check for these worst-case scenarios by several methods. First, the user often has some *a priori* knowledge about the particle morphologies (e.g., whether they might be extremely oblate, or extremely prolate, or roughly isotropic). The preceding paragraph gives examples of how to estimate the sampling bias, or at least an upper bound of the sampling bias, in such cases. Second, the user is not merely passively applying an algorithm but is actually looking at the images and selecting each particle for analysis. Thus, the user will notice whether the images of the particles are noticeably more or less isotropic on average than would be expected for their expected morphologies and can thereby compensate if it should be determined that the particles are (for example) usually lying flat. Such a hypothesis can easily be checked by tilting the sample and re-imaging. This procedure can be made more quantitative by use of curve fit functions that explicitly include anisotropy, such as an elliptical Gaussian characterized by three independent second moments  $\sigma_{xx}$ ,  $\sigma_{yy}$ , and  $\sigma_{xy} = \sigma_{yx}$ . This will also help in cases where the experimenter has no idea of the expected morphology. Pronounced anisotropy will show up in these second moments, such that for many particles ( $\sigma_{xx} - \sigma_{yy}$ ) and/or  $\sigma_{xy}$  will differ significantly from zero.

Third, the experimenter can tilt the sample about two orthogonal axes and repeat the experiment on the very same particles, thus providing enough data to estimate the *entire*  $M_{ij}$  tensor for each particle. This proposed tomographic technique would give the  $A$ ,  $B$ , and  $C$  for every particle along with an estimate of each particle's three-dimensional orientation. This procedure would be most effective if the substrate supporting the particles were very thin, the available tilt angles very large, and the sample extremely stable. Deriving the relevant formulae from the equations we have already given would be a straightforward application of linear algebra. Whether the experiment itself can actually be performed to the required precision, and the reconstruction performed in a sufficiently noise-tolerant way, are open questions.

Now let us consider the best-case scenario, which comes up more often than might be expected: The case in which, so far as the second moments are concerned, the particles are isotropic ( $A = B = C$ ). Obviously this will apply for spherical particles, but it also applies to

particles with cubic, tetrahedral, and icosahedral symmetry, which also includes octahedra, cuboctahedra, dodecahedra, and symmetrically faceted versions of any of these. Small atomic clusters with such symmetry happen to be very common; one of our experimental test cases [9], for example, was primarily a tetrahedral cluster of 10 Os atoms. In these situations, it does not matter if the experimental ensemble is isotropic, because the isotropy lies in the particles themselves. In other words, if  $M_{ij}$  is an isotropic tensor, then

$$\langle \sigma_{obj-x}^2 \rangle = \langle \sigma_{obj-y}^2 \rangle = \langle \sigma_{obj-z}^2 \rangle = \sigma_{obj-x}^2 = \sigma_{obj-y}^2 = \sigma_{obj-z}^2 = A = B = C \quad (8)$$

*even if all of the particles have the same orientation.*

In this rather common best-case scenario, the isotropic-ensemble assumption plays no role in the analysis. The fact that every measurement from every single particle should (within error) produce exactly the same value of  $\sigma_{obj-x}^2$  also reduces the number of measurements required to establish a precise estimate of  $r_{RMS}$ , because the fractional error in the estimated  $\langle r_{RMS} \rangle$  will be roughly equal to the fractional RMS variation of the measured  $\sigma_{obj-x}$  values divided by the square root of the number of particles included in the average (we are using the concept of standard deviation of the mean [15], to which we return below).

#### D. Estimating $\sigma_{obj}$ from two-dimensional fits

We next examine the problem of extracting a best-guess  $\sigma_{obj}$  from the series of two-dimensional fits performed with different values of the artificial Gaussian blurring  $\sigma_{gb}$ . These fits result in a series of estimates of the RMS widths in the x and y directions  $\sigma_{est}(\sigma_{gb})$ . When plotted on quadratic axes, this function appears linear with a slope close to 1. We will refer to such a plot as a blur-variance plot. We then correct this estimate for all known systematic effects, following Equation (7) and taking  $\sigma_{est}$  to approximate the ideal measured value  $\sigma_{meas}$ :

$$\frac{r_{RMS}^2}{3} \approx \sigma_{obj}^2 \approx \sigma_{corr}^2 \equiv \sigma_{est}^2 - \sigma_{probe}^2 - \sigma_{vib}^2 - \sigma_{focus}^2 - \sigma_{gb}^2, \quad (9)$$

and similarly for the y direction, and where we have introduced  $\sigma_{corr}$ . In this equation, we have taken the particles and/or the ensemble to be isotropic for the sake of simplicity, but the equations can be extended to the anisotropic case using the methods of the preceding section.

In practice,  $\sigma_{corr}$  will not exactly equal  $\sigma_{obj}$  and will vary with  $\sigma_{gb}$  for each particle. These variations will come from all of the nonidealities in the curve fit, including random noise, inadequate background subtraction, and the fact that the particles themselves are not shaped like the idealized fit functions. Previously, we supposed that these errors would vary with  $\sigma_{gb}$ , and that a curve fit would determine an optimized estimate of what  $\sigma_{obj}$  would have been in the absence of these nonidealities. After trying various forms, we found that a linear fit of  $\sigma_{est}^2$  against  $\sigma_{gb}^2$  produced satisfactory results in our experimental test case, reproducing the expected particle sizes to an accuracy of order 0.1-0.2 Å. In the present work, we revisit this procedure in order to understand both its efficacy and its limitations. First we will subtract off all known blurring effects and also apply the factor of 3 before doing the fit (i.e., we will perform a fit of

the estimated  $r_{RMS}^2 = 3\sigma_{corr}^2$  against  $\sigma_{gb}^2$ ). This procedure produces a mathematically equivalent result while revealing the  $\sigma_{gb}$  dependence of the systematic errors.

This progressive-blurring curve fit is difficult to justify in purely theoretical terms. The variations with  $\sigma_{gb}$  *cannot* be treated as independent random noise, because they all derive from reanalyses of a single data set. Therefore the maximum-likelihood principle that underlies most curve fits does not directly apply. For data sets with extremely high signal-to-noise ratios (SNR) and signal-to-background ratios (SBR), it hardly matters, and the entire Gaussian blurring procedure is an unnecessary complication, as the particle sizes could just be read directly off the images. But what of the far more interesting case in which the SNR and/or SBR is small? Does the curve fit procedure produce an improved estimate of the particle size relative to what would be obtained with a simpler procedure? This question can be approached with a combination of simulation and experiment. We have already presented an experimental test case showing the apparent effectiveness of the method [9]. In the remainder of this paper, we present results from synthesized data, nominally equivalent to those obtained in the experiments, in order to understand the artifacts that may arise from the analysis algorithm itself.

The results suggest that the apparent success of the experimental study was not merely fortuitous. Simulations of the same particles, with the same probe sizes, scan resolutions and approximate SNR and SBR, show statistical outcomes remarkably similar to those of the experiment. The largest deviations between the experimental and the expected values probably arose from consistent, quantifiable systematic errors, most importantly from the deviation of the particle shapes from the idealized fit functions but also from imperfect background subtraction. This result suggests that future applications of the method could correct for such errors, producing estimates of mean particle sizes accurate to better than 0.1 Å. It also shows that the coupling of the “blind” curve-fitting technique with simulation and with other knowledge about the sample is significantly better than the fitting technique alone. Note that the experiments were carried out with samples synthesized specifically to incorporate predominantly clusters of selected sizes, and the typical catalyst sample is not nearly as uniform as our samples. The fact that we expected our samples to consist of just two predominant particle morphologies, and that the larger of these was close to isotropic, facilitated interpretation and reduced the number of measurements required for a confident interpretation.

### III. Synthetic Data Sets

#### A. Methods

Figure 1 illustrates the procedure used for the synthetic data set analysis. Figure 1(a) shows an atomic model of an Os<sub>10</sub>C tetrahedral cluster (the models also included CO ligands, but these are omitted for clarity). To simulate an image, we first choose a random orientation for the cluster (molecule) (Figure 1(b)). Rotations were chosen from a distribution that is uniform over the space of all three-dimensional rotations. This procedure can be carried out, for example, by choosing the four elements of a rotation quaternion [16] from Gaussian distributions, then normalizing the quaternion to unit magnitude.

The algorithm for simulating the STEM image is a simple projection of the atomic signal strength  $S_i$ , convolved with the instrumental point spread function (Figure 1(c)). The  $S_i$  were obtained from a standard database [13], with the differential cross section integrated over the 35 to 90 milliradian collection annulus used in the experiments. The point spread function was



taken to be a Gaussian with RMS radius 1.27 Å, which was approximately the probe size used in the experiments. This simple projection method is based on the assumption that each atom scatters independently; cross terms from coherent scattering between pairs of atoms are neglected.

We tested this assumption with simulations performed using Kirkland's STEM simulation code [17]. STEM images were simulated using our instrumental parameters, first with a single Os atom and then with a pair of Os atoms separated by 2.85 Å. After scaling to the same peak intensity, the 2-atom image differed from the sum of two single-atom images by only a very small amount (~0.3% or less in the vicinity of the atoms). This difference is negligible in comparison with the measurement noise, which was more than 10% of the peak height in the relatively low-dose conditions used in the experiment. As more atoms are involved in the scattering, the two-atom cross terms become far more numerous than the single-atom terms, but the effect should still be negligible when most of the signal is coming from only 10 atoms. The simple-projection approximation may have to be revisited if the same method is used on much larger, crystalline clusters.

A real STEM image of a particle on a support layer will include noise from various sources, as well as background signal from the support. The support layer is simulated as a polynomial in  $x$  and  $y$  designed to produce a magnitude and spatial scale comparable to the actual background variation in the experiments (Figure 1(d)). This model was appropriate for the thin MgO substrates used in our experiments, as imaged in a relatively low-resolution instrument; the substrate in our images looks quite smooth on the scale of a few Å. The particle-size measurement algorithm is expected to work only if the particles of interest are significantly more localized and peaked than are the variations in the background signal, or for particles at the edge of the substrate. More sophisticated background subtraction methods (e.g. carefully targeted Fourier filtering) will likely have to be used in cases where the image shows atomic-resolution information coming from the substrate, and especially in the case of crystalline substrates imaged in an aberration-corrected STEM. Exploring this issue would be a significant effort in itself and is beyond the scope of the present report.

Varying the degree and magnitude of the polynomial background terms allows the user to test the limitations of the method. We simulated two kinds of noise: Shot noise (also called counting statistics or Poisson noise) and Gaussian white noise. For the shot noise, the user can specify the image intensity  $I_0$  equal to the signal from one detected electron. The software then applies a Poisson distribution to determine the number of detected electrons in each pixel, scaled appropriately via  $I_0$ . This procedure results in an RMS noise per pixel of  $(I(x,y)I_0)^{1/2}$ , where  $I(x,y)$  is the noise-free image intensity. The Gaussian white noise of RMS magnitude  $I_G$  is added last. We almost always set  $I_G > I_0$  to minimize the effect of artificial quantization from the Poisson algorithm.

The above steps (Figures 1(b-d)) are repeated 50 times to produce a population of images, each with its own random cluster orientation and random noise. Each image is then analyzed using progressive Gaussian blurring (Figure 1(e)) and two-dimensional curve fitting (Figure 1(f)), with seven different values of  $\sigma_{gb}$  typically being used. Thus, a full synthetic data set includes 350 two-dimensional fits. For each fit,  $\chi^2$  is minimized over the full set of parameters, using an uncertainty of  $(I_{fit}(x,y)I_0 + I_G^2)^{1/2}$  to scale the residuals [9], thus including the random contributions from both the shot noise and the Gaussian white noise. The optimization uses the Nelder-Mead simplex direct search [18] as implemented in the MATLAB software environment. We found that the fit was quite robust except in some cases in which the blur was large enough

that the algorithm could mistake background intensity as part of the cluster. This produces characteristic large jumps in the corrected blur-variance curves, as we will show.

Although our uncertainty scaling function does overestimate the noise of the blurred images, we have found that this does not greatly affect the final results (the effect of using different noise models is shown in a later section). As usual for this kind of curve fitting, the shot noise can produce a bias in the fit parameters (Chapter 12 of Bevington [15]). The background function used in the fit need not be of the same polynomial order as the background function used in the image simulation, and we varied these parameters to gauge their effects on the final result. The results are usually robust as long as the background model is of high enough order to reduce the background residuals essentially to random noise but not so high-order as to subtract out the particles along with the background.

## B. Results: Baseline simulations

We start with a baseline simulation, intended to produce results as similar as possible to our experiments [9]. We used models of  $\text{Os}_3(\text{CO})_{12}$  and  $[\text{Os}_{10}\text{C}(\text{CO})_{24}]^{2-}$  as determined from EXAFS spectroscopy and X-ray diffraction crystallography and an understanding of the chemistry of the cluster synthesis. For brevity, we refer to these structures as  $\text{Os}_3$  and  $\text{Os}_{10}\text{C}$ , respectively, with the presence of CO ligands being understood. Ideal  $r_{\text{RMS}}$  values were calculated from these models using equations (1-3). Very slightly revised (because of gradual evolution of the set of best locally available data) sensitivities  $S_i$  and atomic coordinate models were used in comparison with our previous publication, resulting in shifts of the ideal values by a few hundredths of an angstrom (which reflects the precision of the theoretical ideal  $r_{\text{RMS}}$  in the experiment).

In the experiments we were trying to mimic, typically the pixel-to-pixel noise was on the order of 10–30% of the peak heights, and the background varied by ~30–60% of the peak heights over a characteristic distance scale on the order of 10–20 Å (see Figure 3 of [9] for an example). The scan resolution ranged from 0.42 to 0.63 Å per pixel. Our baseline test case (Figure 1(d) is an example) was chosen to lie in the middle of these ranges, with  $I_G$  set equal to ~20% of the maximum image intensity, background varying by ~40–50% of the peak height over a lateral distance of 10 Å, and a scan resolution of 0.5 Å. These values are rough, but fortunately (as shown below), the final results are not particularly sensitive to the exact SNR and SBR in the parameter range we were most interested in.

In terms of relevant experimental parameters, these noise and background levels correspond to shot noise from detecting ~100 electrons per pixel (which can be estimated from the probe current, dwell time, atomic density, scattering cross sections, and annular collection angle range) and an effective mass-thickness (i.e. scattering density  $S_i$  per unit area) for the substrate on the same order as that for the particle. In our case, the  $\text{Os}_{10}\text{C}$  clusters had 10 Os atoms in a cross-sectional area of ~15 Å<sup>2</sup> lying on an MgO substrate with 0.054 MgO units per cubic angstrom. If we assume a  $Z^2$  scattering dependence, we find that the SBR is ~1 for a substrate thickness of 340 Å. This kind of very rough estimate will help to determine the experimental conditions needed to apply our methods. In practice, shot noise is a lower-bound noise estimate and the actual number of detected electrons may need to be much larger. It is the variation in background intensity that is most important, as opposed to its absolute value.

The results of the baseline simulations are shown in Table 1 (including data derived from our original publication [9]). Throughout this work, we report particle sizes in the format  $r_{\text{RMS}} \pm$

$\sigma_{dist}$  ( $\sigma_{mean}$ ), where  $r_{RMS}$  is the root-mean-square result from the entire ensemble,  $\sigma_{dist}$  is the sample standard deviation of the distribution of measured  $r_{RMS}$  values for different particles, and  $\sigma_{mean}$  is the estimated standard error of the mean (i.e., the estimated precision to which  $r_{RMS}$  has been determined [15], calculated as  $\sigma_{dist}/N^{1/2}$ ).

The results are quite encouraging. The  $r_{RMS}$  values obtained from the experiments and the simulations are completely indistinguishable, the deviations being a fraction of the standard error for both cases. The close match happened to be better than chance for these two test runs; typical precisions on the order of a few hundredths of an angstrom are more the norm, consistent with the calculated  $\sigma_{mean}$  values. The distribution widths  $\sigma_{dist}$  are somewhat different between the experiment and the simulation, and the difference is significant even considering the small size of the set of experimental results (only 20 clusters total for the two groups).

We speculate on possible causes for this difference. For the (anisotropic)  $Os_3$  clusters, the cluster orientations in the experiment may have been biased, which could explain the larger variance in the simulation for this case. Also, the SNR and SBR values for the simulation, although close to those in the typical experiment, were not exactly the same in detail. Finally, all of the simulations in each set of 50 images were performed with the same background. Variations in the background are likely to cause variations in systematic error arising from incomplete background subtraction, which would increase the variance in the calculated radii. It may be that the (more isotropic)  $Os_{10}C$  clusters revealed this effect.

Note the difference in systematic errors for the two cluster sizes used here. The smaller cluster yields the more accurate result, with the experimental, simulated, and ideal values indistinguishable within the error bars. The larger cluster ( $\sim 2.3$  times the probe size) is consistently found by the algorithm to be  $\sim 6\text{--}7\%$  larger than it really is, for both the experiments and the simulations. Analysis of the normalized residuals (presented below) suggests a likely reason: The Gaussian fit function does a better job of matching the target function when the object is not much larger than the instrumental point spread function. Thus, the RMS size of the measured intensity distribution should be a closer match to the RMS size of the Gaussian fit function as the object gets smaller. For larger objects, there is significant intensity variation in the image that is not captured by the Gaussian fit function, and the two RMS sizes tend to differ by a fairly consistent amount that depends on the size and shape of the particle.

There are two ways of correcting for this limitation. One is to calibrate the systematic error by using synthetic data sets, as we have done here, so that we may correct the  $r_{RMS}$  values yielded by the curve fits. If, in our original work [9], we had performed the synthetic data simulations presented here, we would have expected the  $Os_{10}C$  clusters to yield a size measurement that was too large by  $\sim 6\%$ . Thus we would have corrected the measured value of  $3.11 \text{ \AA}$  to a value of  $2.92 \text{ \AA}$ , which is well within uncertainty of the "ideal" value of  $2.909 \text{ \AA}$ . The other method would be to use a more complicated fit function to capture more of the variation in the measured image than is captured by either the simple or the elliptical Gaussian. We briefly attempted this approach and found that the particle size measurement results were not dramatically improved, whereas the computation time increased significantly.

This approach may yet be useful in the "blind" case in which the likely atomic structure of the clusters is unknown, so that the calibration method is more difficult to perform. It may also be useful for extracting information about the shape and orientation of each particle in addition to its isotropic  $r_{RMS}$  size, for cases in which the SNR and SBR are high enough that the images do contain such information. We introduce examples below that will help to identify when such information may be available.

### C. Improvements to the Algorithm

A crucial step in the algorithm is the extraction of an  $r_{RMS}$  estimate for each particle from its blur-variance curve (Figure 2). In our first efforts, we performed the simplest procedure, namely, linear extrapolation of the sampled blur-variance curve to zero blur (Figure 2(a)). This procedure entails a linear curve fit to a set of data points that are not statistically independent. The result depends on the arbitrary choice of  $\sigma_{gb}$  values, and low-blur data points were accorded more weight than they perhaps should have been accorded. Also, our original method did not make use of most of the information in the blur-variance curves. Although this first version of the method produced good results, we were motivated to overcome these shortcomings and produce a more mathematically well founded, robust, and precise version of the method. We have done this, and the new technique is illustrated in Figure 2(b-d).

Now, instead of the uncorrected blur-variance plot in Figure 2(a), we introduce a fully corrected blur-variance plot in Figure 2(b), in which all known convolution effects have been subtracted out. As a result, the mean slope of the corrected blur-variance curves is very nearly zero, and there is no reason to suppose that the extrapolation to zero blur is any better an estimate of  $r_{RMS}$  than is, say, a value chosen from the middle of the curve. We have used cubic splines to interpolate the measured points and yield smooth curves. Rarely (as is visible on the right-hand side of Figure 2(b)) at high blurs the curve jumps discontinuously, which means that the nonlinear two-dimensional curve fit has probably jumped from one local minimum of  $\chi^2$  to another. But for the most part the curves seem fairly well behaved.

Next consider Figure 2(c), showing curves from a selected group of 10 particles along with a dotted horizontal line showing the ideal result (which we can calculate exactly, since these are simulations). The curves separate roughly into three bands (as is more obvious when all 50 curves are shown in Figure 2(b)). The middle band, which includes the majority of the particles, consists of mostly horizontal curves that meander through a relatively small range of values (roughly 8–11 Å<sup>2</sup> in this example). For most of these curves, it matters little which point we choose as the optimized  $r_{RMS}$  estimate for that particle, as there is little consistent bias.

On the other hand, most of the curves in the upper band—i.e. those curves that give anomalously high  $r_{RMS}$  values over most of the domain—have large negative initial slopes and positive curvatures, while the opposite is true for many of the curves in the lower band. In both cases, the best estimate of  $r_{RMS}$  is obtained for some positive  $\sigma_{gb}$ , the zero-blur extrapolated value being quite far from the ideal value.

These observations motivated the following modification: Instead of taking  $r_{RMS}$  for each particle as the zero-blur extrapolation of its blur-variance curve, we identify a "maximally stationary point" for each curve, defined as the smallest blur for which the absolute value of the slope reaches its minimum over the sampled domain. Ideally, this is a true stationary point—a point where the slope is zero, and a finite range of blurs will yield a nearly constant  $r_{RMS}$ . Such points do not always appear over the sampled domain, so when no such point exists, we take either an inflection point or, if none exists, the end point with the smallest absolute slope. This step is done separately for the spline-interpolated corrected blur-variance curve of each particle, yielding both an estimate of the optimal blur  $\sigma_{gb}^*$  and the best-guess radius  $r_{RMS}^*$ , which are indicated with X's in Figure 2(c). When this procedure is carried out for all 50 simulated particles (Figure 2(d)), the spread of vertical positions of the X's is, in virtually all tested cases, either less than or not significantly different from the spread of zero-blur intercepts. In other

words, the precision of the new method is as least as good as that of the old method and often significantly better.

The new maximal-stationary-point method is intended to make a best-guess optimization of the systematic and random errors when the scaling of such errors with blur is complicated and not understood in detail. If there were no such errors, then every curve in Figure 2(c) would be a straight line lying exactly on top of the ideal line. In a real measurement, the various effects of noise, background, and failure of the particle to match the idealized fit function will produce an estimated particle size somewhat different from the real size. Each of these effects will scale differently with blur. As the blur increases, the particle will look more and more like an ideal Gaussian, and noise effects will be smoothed out, but it will become harder to separate the particle from the background. So one would expect that there would be an intermediate value of blur, probably on the order of the particle size (remembering the factor of  $3^{1/2}$  difference between  $\sigma$  and  $r_{RMS}$ ), at which the combined effects of all of these errors are minimized. We hypothesize that the optimal blur usually occurs close to a crossover between one regime and another, and that such a crossover is likely to be marked by a feature such as a local minimum or inflection point. Justifying this from first principles would be extremely difficult, so instead we tested this hypothesis with Monte Carlo simulation. The results show that the new method is almost always either better than or not significantly different from the old method in terms of precision.

The new method is also more mathematically robust than the old, and it makes better use of all available data. So long as the range and density of  $\sigma_{gb}$  values is high enough to define the blur-variance curve, the result does not depend strongly on the exact set of blur values that happened to be chosen. This is not the case for the linear-fit-intercept method, for which the low-blur points can dominate the result. The shapes of the corrected blur-variance curves also carry additional information: They serve as diagnostics for identifying and mitigating the systematic errors in any given data set and can also identify extreme cases for which the results cannot be trusted. This point is explored more in the next section.

Finally, we discuss the thick dashed curve in Figure 2(d), which is the mean of all of the single-particle curves. This curve evaluates to  $(3.102 \text{ \AA})^2$  at  $\sigma_{gb} = 0$ , which (as expected) corresponds almost exactly to the value reported in the "Simulated" column of Table 1. This curve exhibits negligible random variation and even its systematic error varies very little as a function of blur (i.e., the curve is almost a horizontal line). Such a population mean square curve can be very useful in cases in which every particle can be positively identified. Unfortunately, outliers can sometimes produce  $r_{RMS}$  values as much as 20% away from the ideal, and it is quite possible to mistake one particle size for another even in nearly ideal cases.

Thus, it is better to rely on methods that do not assume that any given particle can be positively identified, as such methods will be applicable to a broader range of samples. The new maximal-stationary-point algorithm is such a method, because it produces an optimized estimate of  $r_{RMS}$  independently for every single particle. The kernel density estimator (KDE) and maximum-likelihood (ML) analysis in our original publication [9] (which we used for estimating the statistics of partially overlapping sub-populations) also adhered to this advice. These calculations were performed in such a way that it was never necessary to assume that any given particle belonged to Group I, II, or III as previously defined.

#### D. Tests of the algorithm under various conditions

To have confidence in such a data analysis method, one must know that its output is stable with respect to reasonable parameter variations. It is also important to develop diagnostics that can help to determine whether the results are trustworthy, and if not, what the likely causes are, while also being able to determine whether there might be more information in the images beyond the "mass" (i.e., total STEM signal) and RMS size of each particle.

To explore these issues, we ran a series of scenarios with various parameters for both the image simulation and the data analysis and applied these methods to both the Os<sub>3</sub> and the Os<sub>10</sub>C clusters, recording the distributions of both  $r_{RMS}$  and  $m$  for every case, using both the linear-extrapolation and the maximally stationary method (Tables 2 and 3). Each scenario represents a change of only one or two parameters at a time relative to the "baseline" scenario (scenario 0 in the tables) that we have already described. We generated the blur-variance plots and also examined normalized residual images (i.e., the difference between the generated image and the Gaussian-plus-polynomial fit function, divided by the RMS noise in each data point) for some of the two-dimensional curve fits performed with zero artificial blurring.

Some examples of these diagnostic plots are shown in Figures 3 and 4. We observed the following patterns:

1. The  $m$  and  $r_{RMS}$  values were quite stable as the parameters were varied, with some exceptions that leave clear signatures in the blur-variance and normalized residual plots.
2. The smaller Os<sub>3</sub> clusters suffered very little bias. Most of the values for both  $m$  and  $r_{RMS}$  were close to the ideal value or slightly above it, rarely deviating by more than  $2\sigma_{mean}$ .
3. The unblurred normalized residual images representing most of the Os<sub>3</sub> simulations were hardly distinguishable from random noise except in cases for which the SNR and/or SBR were unrealistically high (scenarios 10 and 11) or the instrumental resolution was much smaller than the particle (scenario 12). This result suggests that, for our experimental parameters, there is very little information present in the image beyond that already extracted from our fit parameters ( $m$  and  $r_{RMS}$ ), for the small molecules.
4. The corrected blur-variance plots for the cases of low SNR and low SBR showed only very weak patterns, with mostly horizontal curves meandering through a fairly well defined band of values. As a result, the linear-extrapolation and maximally stationary methods produced almost the same results.
5. Notable exceptions to these patterns arose in cases of incomplete background subtraction (scenario 2) and Poisson noise (scenarios 8 and 9). The blur-variance plots showed dramatic anomalies in both cases, with numerous outlying points that distorted the spline interpolations as well as consistently negative (scenario 2) or positive (scenarios 8 and 9) slopes.
6. The Os<sub>10</sub>C clusters, being larger and more massive than the others, had higher signal levels and more information on length scales above the resolution limit. Thus the Gaussian fits failed to capture all of the intensity variations, leaving a typically 3-fold-symmetric pattern in the normalized residuals. This same fact manifests in the blur-variance plots, such that at low blurs each particle was more idiosyncratic, and the curves do not maximally converge until very nearly the maximum  $\sigma_{gb}$  values have been reached. In other words, each image seems to include some information about the detailed shape and orientation of each particle, and this information could probably be partially extracted with a higher-order fit function. The elliptical Gaussian (scenario 1) produced similar results to the round Gaussian (scenario 0), since the ellipse could not capture the symmetry of the particle image.
7. In part because of the larger variation in blur-variance curves at smaller  $\sigma_{gb}$ , the maximally stationary method outperformed the linear-intercept method in almost all cases for the

Os<sub>10</sub>C clusters. Both the bias (the difference between  $r_{RMS}$  or  $m$  and its ideal value) and the uncertainty were usually smaller for the maximally stationary method. Interestingly, Poisson noise (scenarios 8 and 9) yielded an exception to this rule, as did the case of incomplete background subtraction (scenario 2). These are the very same cases that produced clear anomalies in the corrected blur-variance plots.

8. A comparison of scenarios 0 through 6, and also separately comparing scenarios 10 and 11, shows that the background has remarkably little effect on the final result, *provided that the background model is of high enough order to remove it cleanly*. If the background is not cleanly removed, the results end up turning out very badly (scenario 2). These cases can be detected by looking for patterns in the normalized residual images. Also, a persistent negative slope in the corrected blur-variance curves seems to be indicative of this problem (though this may or may not hold in general).

9. A comparison of scenarios 0, 7, and 10 shows that Gaussian white noise has little effect on the results except to increase the uncertainties when it is raised above a certain level.

10. Poisson noise, on the other hand, has a surprisingly large effect on the bias and the repeatability of the entire method. It would appear (from a comparison of scenarios 8 and 9) that it is actually slightly counterproductive to try to model the noise precisely in the  $\chi^2$  minimization. Instead, we recommend using a Gaussian white noise model for all of the two-dimensional fits. Excessive Poisson noise seems to leave a signature in the corrected blur-variance plots, namely a consistent positive slope for almost all of the particles.

#### IV. Conclusions and Future Directions

We could of course multiply examples and observations such as those in the preceding section virtually without limit. However, for present purposes we believe that we have both validated and improved upon the method reported in our previous work [9]. Let us review the questions raised at the start of this presentation and offer some answers:

1. For the SNR and SBR of our experiments, should the algorithm have performed as well as it appears to have?

Answer: Yes. Judging from extensive Monte Carlo simulations designed to produce simulated STEM images statistically equivalent to the experimental images, the precision and the quantitative degree of bias seem to be very much what we should expect.

2. Does the progressive-blurring curve fit perform its intended function, namely, to produce an optimized estimate of particle size that minimizes the effects of noise and background?

Answer: Yes, with qualifications. For many realistic simulations, the blur-variance curves show that greater precision can indeed be obtained by including values from nonzero values of blur. However, our previous method may not have made optimal use of this information.

3. Is there an alternative to the linear progressive-blurring fit that makes better use of the information gathered in the set of blurred images?

Answer: Yes. We have introduced a maximal-stationary-point variation on our previous method that gives results either superior to or practically indistinguishable from the results of the

previous method in almost all realistic cases. This method is also more robust and more firmly grounded mathematically than is the previous method. Moreover, we have found that many of the normalized-residual and blur-variance plots show that there is information in some of the images beyond the mass and the RMS radius. We have offered some brief suggestions for identifying such cases and for extracting additional information using more complicated curve-fit functions and tilt-series acquisitions.

4. Is there a systematic error introduced by the algorithm, and does this error vary in a predictable way with parameters such as particle size? If so, can the estimate of actual particle sizes be improved?

Answer: Yes, and by correcting our previous measurements by using our current estimates of the bias we find that the match between theory and experiment is well within the bounds of random error. "Blind" application of this calibration to cases in which the particle morphologies are completely unknown is yet to be investigated.

5. Can we make use of additional information, such as the estimated particle masses and the patterns in the progressive-blurring analysis, to understand more about the particles?

Answer: Yes. We have found that the "mass" (actually total signal strength) measurements can be measured with high precision along with the RMS radii. With a sufficiently large and homogeneous sample and some correction for the bias using synthetic data sets, it should be possible to determine the number of heavy atoms per average cluster to a precision of less than one atom. However, the precision is not high enough to allow us to confidently distinguish individual clusters that differ by only one heavy atom, except perhaps in the case of very small clusters with fewer than  $\sim 5$  heavy atoms. We have also found that certain patterns in the analysis (revealed in corrected blur-variance plots) can reveal whether Poisson noise or incomplete background subtraction may be poisoning the results.

The choice of data analysis methods should always follow the maxim to make things as simple as possible, but no simpler. For data sets with large SNR and SBR, particle sizes can be easily measured directly from the images (or from simple curve fits) and corrected for instrumental point spread functions; complicated data analysis procedures are unnecessary in such cases. For more challenging data sets, the "blind" procedure with the new maximal-stationary-point method can help in both identifying and quantifying sources of error in the size determination, frequently producing errors of less than  $0.2 \text{ \AA}$  (and sometimes much better) with high confidence. In cases for which the data sets are challenging and very high precision is required, calibration of the systematic errors should be performed, for example through the use of synthetic data sets as we have used in the present work. We have shown that this procedure can yield a precision of  $\sim 0.1 \text{ \AA}$  in the mean value of  $r_{RMS}$ , even in the presence of significant noise and background, provided a reasonable number (at least  $\sim 10$ , preferably more) of nominally identical particles can be measured. Most important, this measurement can be performed at levels of electron dose that will provide little or no modification of the sample during acquisition.

## Acknowledgments



This work was supported in part by the National Science Foundation under GOALI Grant Number CTS-0500511, performed in collaboration with ExxonMobil, and by the Japan Society for the Promotion of Science for Young Scientists. This work was performed in part under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory supported by Office of Science, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering of the U.S. Department of Energy under Contract DE-AC52-07NA27344.

## References

- [1] B. C. Gates, *Chem Rev* 95 (1995) 511.
- [2] J. Guzman and B. C. Gates, *J Chem Soc, Dalton Trans* 17 (2003) 3303.
- [3] M. Arenz, S. Gilb, and U. Heiz, in *Atomic Clusters: From Gas Phase to Deposited*, ed. D. P. Woodruff, pp. 1-52, Elsevier (2007).
- [4] A. A. Herzing, C. J. Kiely, A. F. Carley, P. Landon and G. J. Hutchings, *Science* 321 (2008) 1331.
- [5] P. D. Nellist and S. J. Pennycook, *Ultramicroscopy* 78 (1999) 111.
- [6] J. Frank, *Three-dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*, 2nd ed., New York: Oxford University Press USA, 2006.
- [7] R. Henderson and P. N. T. Unwin, *Nature* 257 (1975) 28
- [8] Y. Fujiyoshi, T. Mizusaki, K. Morikawa, H. Yamagishi, Y. Aoki, H. Kihara, and Y. Harada, *Ultramicroscopy* 38 (1991) 241
- [9] N. L. Okamoto, B. W. Reed, S. Mehraeen, A. Kulkarni, D. G. Morgan, B. C. Gates, and N. D. Browning, *J Phys Chem C* 112 (2008) 1759.
- [10] R. E. Thompson, D. R. Larson, and W. W. Webb, *Biophysical Journal* 82 (2002) 2775
- [11] W. E. Moerner, *Proc. Nat. Acad. Sci.* 104 (2007) 12596
- [12] M. C. Jenkins and S. U. Egelhaaf, *Adv. Colloid Interface Sci.* 136 (2008) 65
- [13] A. Jablonski, F. Salvat, and C. J. Powell, *NIST Electron Elastic-Scattering Cross-Section Database – Version 3.0*, National Institute of Standards and Technology, Gaithersburg, MD, 2002.
- [14] C. A. Laury-Micoulaut, *Astron Astrophys* 51 (1976) 343.
- [15] P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences*, New York: McGraw-Hill, 1969.
- [16] J. H. Conway and D. Smith, *On Quaternions and Octonions*, Natick, MA: A. K. Peters, 2003.
- [17] E. J. Kirkland, *Advanced Computing in Electron Microscopy*, New York: Plenum, 1998.
- [18] Lagarias, J.C., J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions," *SIAM Journal of Optimization* 9 (1998) 112

## Figure Captions

Figure 1. Illustrating the synthetic data generation and analysis procedure. (a) Atomic model of the cluster of interest. (b)-(f) Generating and analyzing an ensemble of 50 simulated images. (b) Unbiased random rotation. (c) 2D projection and application of point spread function. (d) Addition of background and noise consistent with those of actual STEM images. (e) Progressive (7 steps for each of the 50 simulated images) artificial Gaussian blurring followed by (f) 2D curve fitting to Gaussian + polynomial background for each of the 350 resulting images. Resulting data are analyzed as in Figure 2.

Figure 2. Example blur-variance plots (from the  $\text{Os}_{10}\text{C}$ -based baseline run) illustrating the old and new algorithms. (a) Plotting the squared width of the fit peak against the square of the blur, with a linear fit to each particle, as in our previous publication [9]. The slopes are all close to 1, and the intercepts are used in calculation of the mean particle size. (b) After correcting for all known blurring effects and applying the factor of 3 to obtain the mean square radius. Curvature is now apparent, and the points for each particle are now interpolated with a cubic spline. (c) Example curves from 10 of the 50 particles, marking the maximally stationary point on each curve with an X. The theoretical ideal answer is shown as a horizontal dotted line. (d) The same, applied to all 50 particles in the simulation, with an added thick dashed curve which is the average for all of the particles.

Figure 3. Example raw (left entry in each pair) and normalized residual (right entry in each pair) images for various scenarios, providing examples of the points made in the text. See Tables 2 and 3 for quantitative results. (a) Scenario 0,  $\text{Os}_3$ . The Gaussian fit has reduced the residuals practically to random noise. (b) Scenario 0,  $\text{Os}_{10}\text{C}$ . Note the weakly apparent (but statistically significant) 3-fold pattern in the residuals corresponding to the corners of the particle in the raw image. (c) Scenario 2,  $\text{Os}_{10}\text{C}$ . Note the quadrupole moment in the background region of the residual plot: The upper left and lower right corners are significantly greater than zero, while the other two corners are significantly less. This amount of residual background is enough to skew the results. (d) Scenario 10,  $\text{Os}_3$ . The round Gaussian fails to capture much information about the particle apart from its rough size, leaving very strong residual patterns.

Figure 4. Example corrected blur-variance plots for various scenarios, illustrating some of the points made in the text. Left column is from  $\text{Os}_3$ , right from  $\text{Os}_{10}\text{C}$ . (a) Scenario 0. Scenarios 1, 3, 4, 5, 6, 7, and 12 looked qualitatively similar, but with varying degrees of systematic and random contributions to the patterns. (b) Scenario 2. The incomplete background subtraction produced enormous anomalies at higher blurs. (c) Scenario 8. Poisson noise has a surprisingly strong detrimental effect on the analysis of the smaller cluster, but this seems to leave a characteristic signature in the blur-variance graphs (a persistent positive slope) that can be used as a diagnostic.

## Tables

Table 1. Results from the baseline simulation, designed to be as similar as possible to the experiments. The fit functions were round Gaussians with quadratic backgrounds. Ensemble distributions are characterized by entries of the format  $r_{RMS} \pm \sigma_{dist} (\sigma_{mean})$ , as discussed in the text. Groups I and II from the experimental study [9] (final results reproduced here, with recalculated ideal values) are presumed to represent the Os<sub>3</sub>- and Os<sub>10</sub>C-based clusters. These groups contained 6 and 14 particles, respectively. All quantities are in angstroms. Ideal values are derived from independent EXAFS measurements of the atomic structures, with appropriate weightings  $S_i$  for the various atoms. The final column compares the means of the distributions, with uncertainty set by error propagation from  $\sigma_{mean}$ .

Cluster	Ideal $r_{RMS}$	Experimental	Simulated	$r_{RMS}^{Expt} - r_{RMS}^{Sim}$
Os <sub>3</sub> -based	2.015	$2.01 \pm 0.21$ (0.09)	$2.005 \pm 0.41$ (0.058)	$0.005 \pm 0.11$
Os <sub>10</sub> C-based	2.909	$3.11 \pm 0.31$ (0.085)	$3.097 \pm 0.19$ (0.027)	$0.013 \pm 0.09$

Table 2. Results from variations on the baseline simulations on the Os<sub>3</sub>-based clusters. Entries are of the format  $mean \pm \sigma_{dist} (\sigma_{mean})$ , as discussed in the text, with all  $r_{RMS}$  values in angstroms. "Mass" values (actually total signal strength) are in arbitrary units, with Os set to 10, C to 0.15, and O to 0.26 (proportional to the  $S_i$  values). The ideal values are 2.015 angstroms and 34.02 mass units. "Linear" indicates the result obtained with the old algorithm (linear extrapolation to zero blur), while "Stationary" indicates the result obtained by evaluating each particle's spline-interpolated curve at the maximally-stationary point for  $r_{RMS}^2(\sigma_{gb}^2)$ .

Scenario	Linear $r_{RMS}$	Stationary $r_{RMS}$	Linear mass	Stationary mass
0. Baseline	$2.005 \pm 0.41$ (0.058)	$2.008 \pm 0.40$ (0.056)	$33.9 \pm 3.3$ (0.5)	$33.9 \pm 2.9$ (0.4)
1. Elliptical Gaussian	$2.113 \pm 0.35$ (0.050)	$2.103 \pm 0.36$ (0.050)	$35.2 \pm 3.1$ (0.4)	$34.9 \pm 3.1$ (0.4)
2. Linear fit to quadratic background	Failed ( $r_{RMS}^2 < 0$ )	$1.693 \pm 0.33$ (0.047)	$31.1 \pm 1.7$ (0.2)	$32 \pm 26$ (4)
3. Cubic fit to quadratic background	$2.084 \pm 0.43$ (0.061)	$2.095 \pm 0.38$ (0.053)	$34.5 \pm 3.4$ (0.5)	$34.4 \pm 2.8$ (0.4)
4. Different background shape (similar magnitude)	$2.132 \pm 0.44$ (0.063)	$2.099 \pm 0.41$ (0.059)	$34.8 \pm 3.0$ (0.4)	$34.2 \pm 2.7$ (0.4)
5. No background	$2.112 \pm 0.43$ (0.061)	$2.089 \pm 0.40$ (0.056)	$34.9 \pm 2.8$ (0.4)	$34.6 \pm 2.6$ (0.4)
6. 2X background	$2.076 \pm 0.41$ (0.058)	$2.056 \pm 0.42$ (0.059)	$34.4 \pm 3.1$ (0.4)	$34.1 \pm 2.8$ (0.4)
7. 2X Gaussian noise	$2.097 \pm 0.44$ (0.062)	$2.114 \pm 0.48$ (0.068)	$34.3 \pm 4.5$ (0.6)	$35.1 \pm 5.4$ (0.8)
8. Poisson noise added to Gaussian of similar magnitude ("mixed noise")	$1.829 \pm 0.71$ (0.101)	$2.097 \pm 0.50$ (0.071)	$35.4 \pm 3.5$ (0.5)	$36.0 \pm 4.1$ (0.6)
9. Simulate with mixed noise, fit with Gaussian noise model	$2.058 \pm 0.64$ (0.090)	$2.229 \pm 0.40$ (0.057)	$36.4 \pm 3.3$ (0.5)	$36.7 \pm 3.2$ (0.5)
10. No noise	$2.009 \pm 0.32$ (0.045)	$1.977 \pm 0.25$ (0.035)	$34.3 \pm 1.6$ (0.2)	$33.9 \pm 0.5$ (0.1)
11. No noise and no background	$2.133 \pm 0.36$ (0.051)	$2.080 \pm 0.27$ (0.038)	$34.8 \pm 1.9$ (0.3)	$34.1 \pm 0.5$ (0.1)
12. 0.5X probe diameter	$2.035 \pm 0.52$ (0.074)	$2.039 \pm 0.32$ (0.045)	$32.9 \pm 7.2$ (1.0)	$34.1 \pm 2.0$ (0.3)

Table 3. Exactly as table 2, but for the Os<sub>10</sub>C-based clusters. The ideal  $r_{RMS}$  and  $m$  values are 2.909 angstroms and 110.0 units, respectively.

Scenario	Linear $r_{RMS}$	Stationary $r_{RMS}$	Linear mass	Stationary mass
0. Baseline	$3.097 \pm 0.19$ (0.027)	$3.057 \pm 0.13$ (0.018)	$116.6 \pm 8.5$ (1.2)	$114.2 \pm 4.5$ (0.6)
1. Elliptical Gaussian	$3.190 \pm 0.19$ (0.028)	$3.154 \pm 0.19$ (0.027)	$119.3 \pm 7.8$ (1.1)	$116.9 \pm 6.6$ (0.9)
2. Linear fit to quadratic background	$2.994 \pm 0.14$ (0.020)	$2.916 \pm 0.16$ (0.022)	$108.3 \pm 2.8$ (0.4)	$106.1 \pm 4.0$ (0.6)
3. Cubic fit to quadratic background	$3.165 \pm 0.25$ (0.035)	$3.114 \pm 0.22$ (0.032)	$119.4 \pm 10.7$ (1.5)	$116.0 \pm 8.5$ (1.2)
4. Different background shape (similar magnitude)	$3.125 \pm 0.25$ (0.035)	$3.077 \pm 0.20$ (0.028)	$116.7 \pm 9.8$ (1.4)	$114.0 \pm 6.5$ (0.9)
5. No background	$3.135 \pm 0.21$ (0.030)	$3.084 \pm 0.19$ (0.026)	$119.1 \pm 8.0$ (1.1)	$115.8 \pm 6.3$ (0.9)
6. 2X background	$3.133 \pm 0.23$ (0.033)	$3.066 \pm 0.19$ (0.026)	$117.6 \pm 9.0$ (1.3)	$113.8 \pm 5.7$ (0.8)
7. 2X Gaussian noise	$3.104 \pm 0.41$ (0.058)	$3.091 \pm 0.22$ (0.032)	$119.6 \pm 13.9$ (2.0)	$115.7 \pm 10$ (1.4)
8. Poisson noise added to Gaussian	$3.173 \pm 0.24$ (0.034)	$3.213 \pm 0.20$ (0.028)	$125.1 \pm 8.6$ (1.2)	$123.4 \pm 6.9$ (1.0)
9. Simulate with mixed noise, fit with Gaussian noise model	$3.112 \pm 0.40$ (0.056)	$3.175 \pm 0.18$ (0.025)	$122.6 \pm 8.6$ (1.2)	$121.8 \pm 7.0$ (1.0)
10. No noise	$3.141 \pm 0.19$ (0.027)	$3.054 \pm 0.12$ (0.017)	$118.0 \pm 6.8$ (1.0)	$113.1 \pm 2.7$ (0.4)
11. No noise and no background	$3.167 \pm 0.18$ (0.025)	$3.072 \pm 0.11$ (0.016)	$118.5 \pm 5.9$ (0.8)	$113.3 \pm 2.3$ (0.3)
12. 0.5X probe diameter	$3.197 \pm 0.26$ (0.037)	$3.102 \pm 0.22$ (0.031)	$121.7 \pm 12$ (1.7)	$115.7 \pm 8.0$ (1.1)

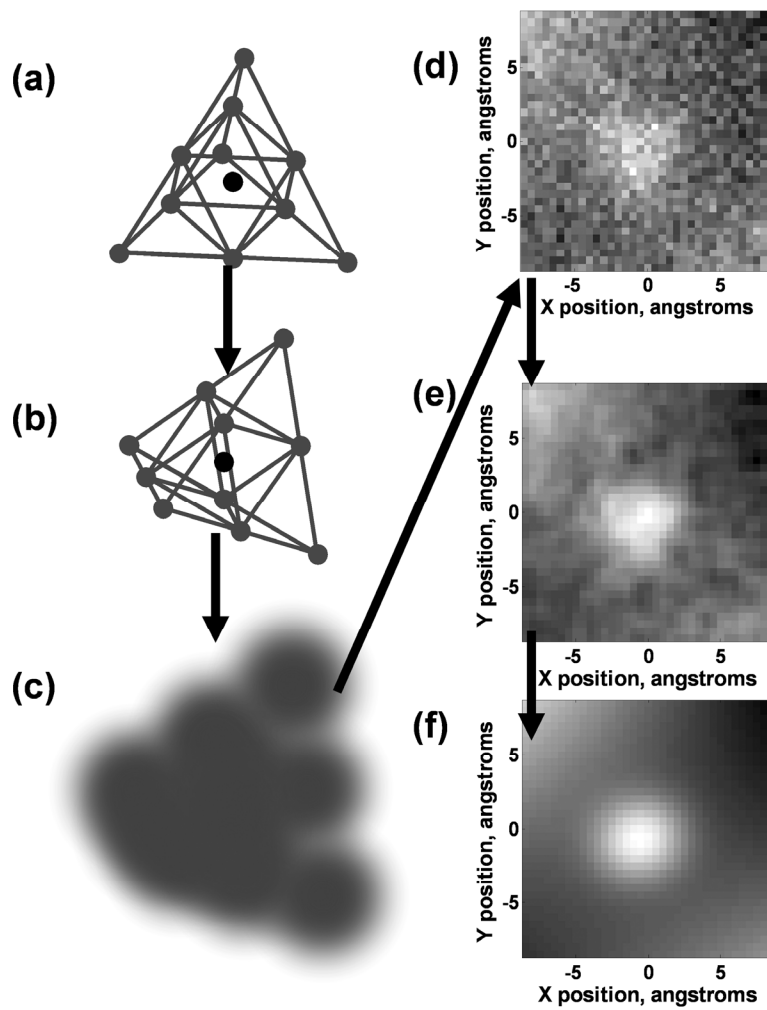


Figure 1

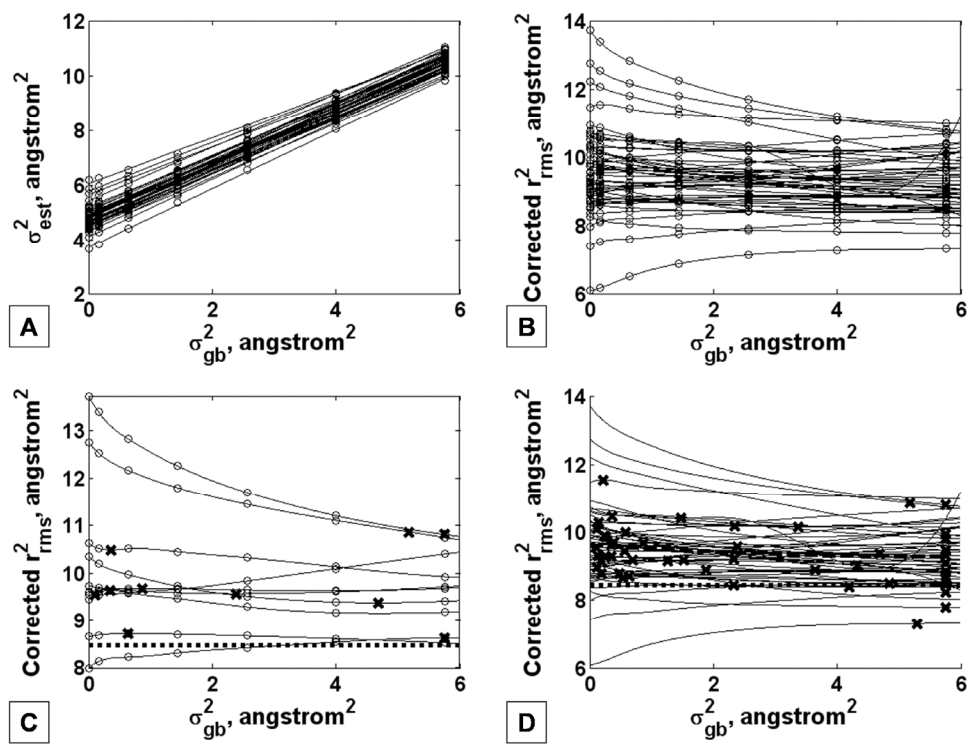


Figure 2

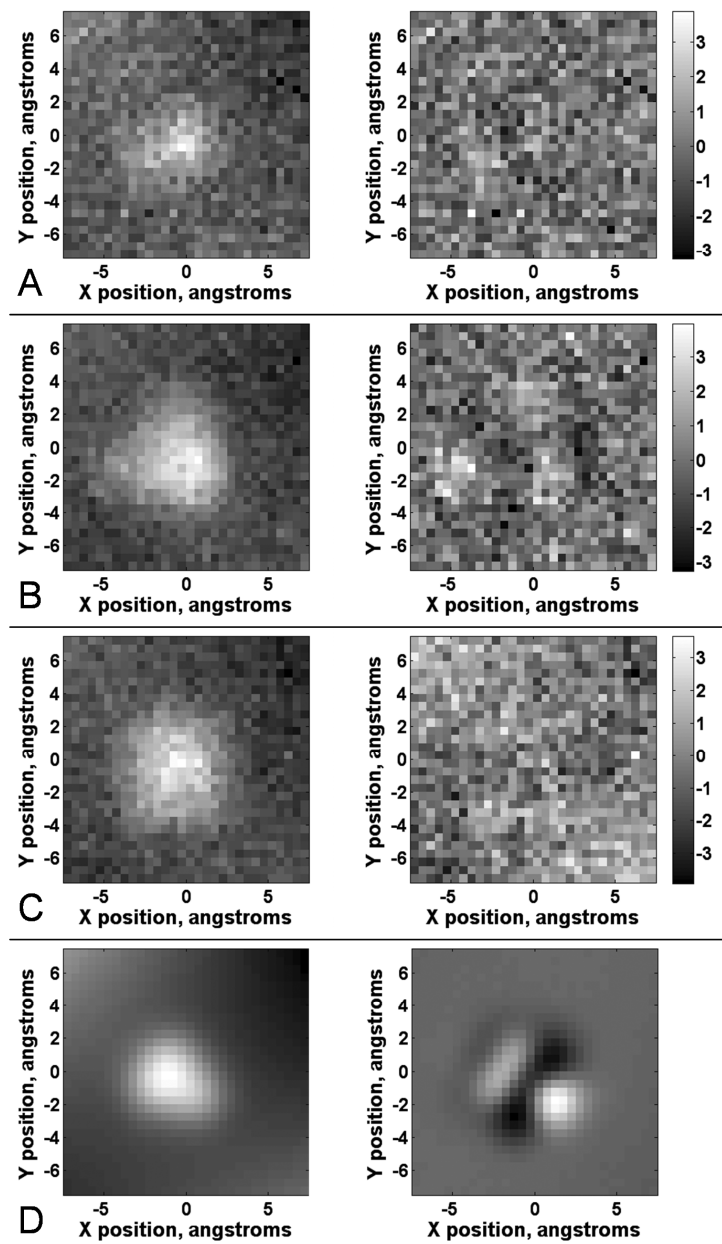


Figure 3

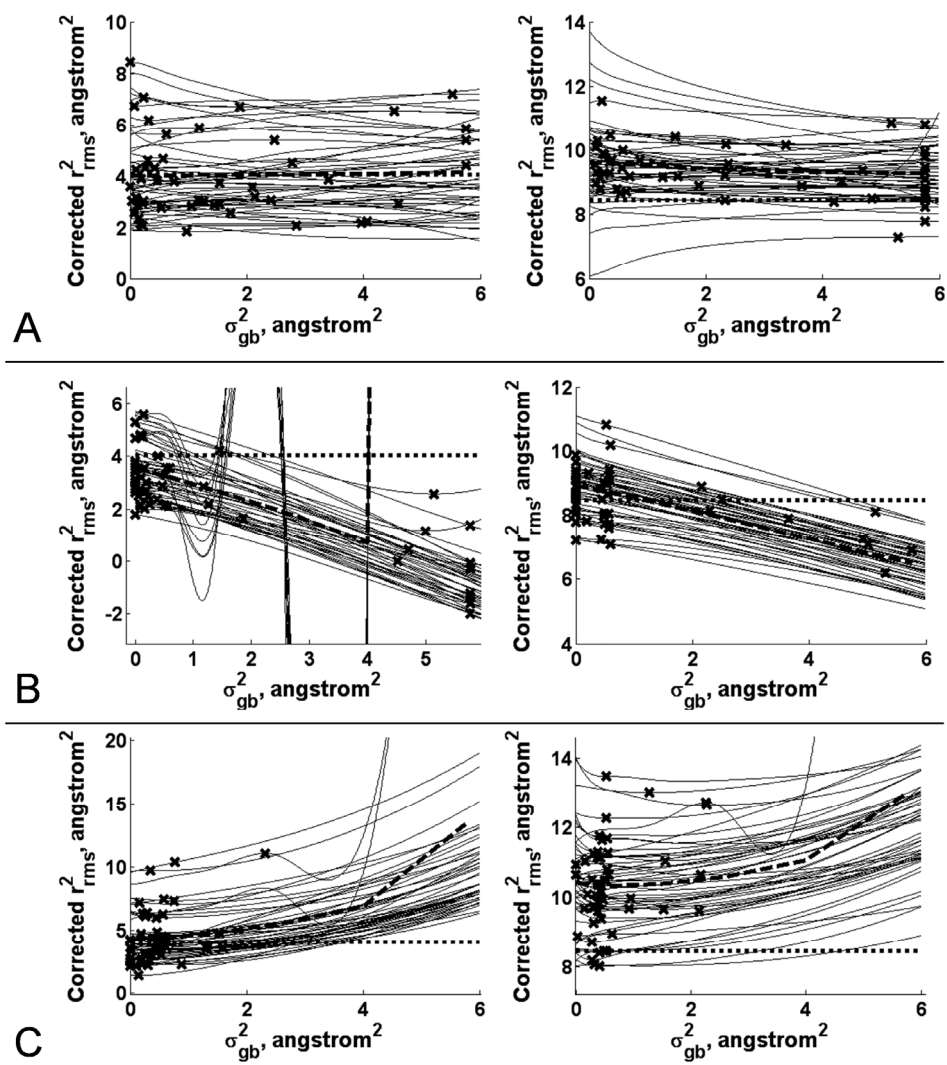


Figure 4