Lawrence Berkeley National Laboratory

Recent Work

Title
Optimizing disk registration algorithms for nanobeam electron diffraction strain mapping

Permalink
https://escholarship.org/uc/item/5pc374hf

Authors
Pekin, TC
Gammer, C
Ciston, J
et al.

Publication Date
2017-05-01

DOI
10.1016/j.ultramic.2016.12.021

Peer reviewed
Optimizing Disk Registration Algorithms for Nanobeam Electron Diffraction Strain Mapping

Thomas C. Pekin\textsuperscript{a,b}, Christoph Gammer\textsuperscript{c}, Jim Ciston\textsuperscript{b}, Andrew M. Minor\textsuperscript{a,b}, Colin Ophus\textsuperscript{b,*}

\textsuperscript{a}Department of Materials Science and Engineering, University of California, Berkeley, Berkeley, USA, 94720
\textsuperscript{b}National Center for Electron Microscopy, Molecular Foundry, Laurence Berkeley National Laboratory, Berkeley, USA, 94720
\textsuperscript{c}Erich Schmid Institute of Materials Science, Jahnstrasse 12, Leoben, Austria, 8700

Abstract

Scanning nanobeam electron diffraction strain mapping is a technique by which the positions of diffracted disks sampled at the nanoscale over a crystalline sample can be used to reconstruct a strain map over a large area. However, it is important that the disk positions are measured accurately, as their positions relative to a reference are directly used to calculate strain. In this study, we compare several correlation methods using both simulated and experimental data in order to directly probe susceptibility to measurement error due to non-uniform diffracted disk illumination structure. We found that prefiltering the diffraction patterns with a Sobel filter before performing cross correlation or performing a square-root magnitude weighted phase correlation returned the best results when inner disk structure was present. We have tested these methods both on simulated datasets, and experimental data from unstrained silicon as well as a twin grain boundary in 304 stainless steel.

Keywords: nanobeam electron diffraction, strain measurement

1. Introduction

Strain and its spatial distribution is important for a greater understanding of many of the relevant engineering materials currently in use. In most modern silicon devices, strain is an important parameter used to modify the properties of the device itself [1]. Likewise in metallic specimens, understanding strain and its evolution under deformation will help further the understanding and predictive capabilities of the field. While many techniques of measuring strain exist [2, 3, 4, 5, 6, 7, 8, 9], scanning convergent nanobeam electron diffraction (NBED) is attractive for a number of reasons. First, NBED strain mapping offers the potential of very high accuracy in strain measurement. Independent diffraction patterns
are recorded with high reciprocal space resolution for each probe position, which limits the spatial resolution to the probe size. On a modern scanning transmission electron microscope (STEM), this probe size can easily be below one nanometer while still maintaining a small enough convergence angle to display full diffraction patterns. This can be compared with geometric phase analysis (GPA) strain mapping, in which real space images are acquired with very high spatial resolution, but do not directly sample reciprocal space. Additionally, GPA strain maps are necessarily limited to a small field of view (FOV) as atomic columns must be resolved to make accurate measurements. With NBED this is not a problem and FOV is usually limited by the storage space of the data acquisition system or the sample stability. With the introduction of high speed direct electron detectors, a large number of diffraction patterns can be obtained from a single sample, covering a very large field of view without concerns for sample drift or other instabilities [10, 2, 11, 12, 13].

Cooper et al., have noted that NBED strain measurements can lose accuracy due to non-uniform disk intensity [12]. This non-uniformity is due to experimental limitations such as sample bending, dynamical effects, or imperfect alignment, resulting in more complicated data sets. The effects of dynamical contrast can be dealt with both during and after the experiment. During the experiment, the electron beam can be precessed around the central axis to obtain a radially averaged diffraction pattern. This has been shown to reduce dynamical contrast and return diffraction patterns that can be analyzed as if they were kinematical [14, 15]. However, in return for more easily processed diffraction patterns, acquisition speed is slowed to the speed at which the beam can be precessed (~0.1 s per pattern versus 0.0025 s per pattern with a Gatan K2-IS camera), and the spot size increases as microscope aberrations have a larger effect on off axis beams. Alternatively, after the data has been acquired, choices in disk position measurement, lattice fit to disk positions, and image downsampling provide a range of options to optimize the data obtained from malformed disks and recover strain information. Here we present an overview of the strain mapping technique itself and experimental options, tested on both simulated data as well as relevant experimental results.

2. Theory

2.1. Measuring Lattice Vectors from Nanobeam Electron Diffraction Patterns

Strain measurement via diffracted peak location is a well understood result of Bragg’s law, and has been successfully performed using several different experimental techniques [16, 10, 11]. For NBED, the first and most important step is proper data acquisition. While the sample does not need to be on a perfect zone axis, it must be close enough to have ideally several orders of diffraction disks illuminated. For every pixel in the reconstructed strain image, an entire diffraction pattern must be recorded, shown schematically in Figure 1a. From each of these patterns, the disk positions are extracted and stored as an \((x, y)\) location in reciprocal space, usually to subpixel precision. After all the disk positions are recorded, they are used to find the local lattice vectors at every probe position. This is done by solving the system of linear equations

\[
BL = P \tag{1}
\]
Figure 1: Strain measurements from nanobeam electron diffraction (NBED). (a) Experimental geometry showing a single NBED measurement. The correlation of the measured diffraction pattern with a center disk template produces an image with sharp peaks at each disk position. Different correlation methods are applied to synthetic disks with (b) no internal structure, (c) disks with signal on opposite edges and (d) disks with signal along a single disk edge. Each method shows an example correlation image, as well as the horizontal and vertical error (divided by disk radius) as a function of counts in the ideal disk without internal structure. Red ellipses show the best fit standard deviations for an elliptic Gaussian function on all peaks.
for $L$, where $L$ is the matrix made up of two lattice vectors defined by pixel lengths from the (000) spot, $P$ is a matrix of every disk position in pixels, and $B$ is a matrix of every disk position in normalized lattice vectors. Equivalent rows of $P$ and $B$ should correspond to the same diffraction disk for each diffraction disk registered. Often, this calculation is overdetermined, as there ideally will be many more disk positions than lattice vectors. If the solution is overdetermined, the fit accuracy can be improved by using weighted least squares, where the weights are equal to the correlation peak value. This calculation is carried out for every diffraction pattern in the dataset. In addition, a reference lattice $L_0$ is computed using the disk positions from either a single real space pixel, or the mean of the disk positions for a subset of pixels (the reference region of the dataset).

Once the lattice vectors have been calculated for every diffraction pattern, it is simple to calculate matrix strain using $L_0^T = L$, where $L_0$ is the reference lattice, $T$ is the transformation matrix, and $L$ is the current lattice for the pixel in question. If infinitesimal strain theory is assumed, the resulting strain matrix (and infinitesimal rotation) is simply

$$\begin{bmatrix}
\varepsilon_{xx} & \frac{1}{2} (\varepsilon_{xy} - \theta) \\
\frac{1}{2} (\varepsilon_{yx} + \theta) & \varepsilon_{yy}
\end{bmatrix} = T - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad (2)$$

where $\varepsilon_{xx}$ is the strain in the $x$ direction, $\varepsilon_{yy}$ is the strain in the $y$ direction, $\varepsilon_{xy}$ and $\varepsilon_{yx}$ are the shear, and $\theta$ is the lattice rotation. Finally, we note that the above expression describes lattice strain in reciprocal space. Because NBED is a reciprocal lattice measurement, the signs of $\varepsilon_{xx}$, $\varepsilon_{yy}$ and $\varepsilon_{xy}$ (but not the sign of $\theta$) must be flipped in order to produce the correct strain measurement of the real space sample deformation.

2.2. Methods for Determining Diffraction Disk Positions

It is clear from the above discussion that the resolution of any resulting strain measurement depends on the accuracy and precision of the diffracted disk position measurement. If only whole pixel disk shifts were to be used, an artificial lower bound would be placed on the accuracy of the local strain measurements, effectively binning the measured strain values from a continuous distribution to a discrete one. Therefore, in order to obtain the most accurate strain measurements, subpixel shifts in disk position must be measured. While many subpixel registration algorithms exist [17, 18, 19, 20], several considerations must be taken into account when choosing which to use. Due to the incredibly large size of the datasets obtained, both in file size and number of disk positions to be measured, only the fastest algorithms are considered. For example, while other authors have used edge detection via Canny or Prewitt filters followed by circle fitting to find the center of the disk [8], that approach is very computationally intensive, requiring several iterative steps per disk in order to resolve a measurement with subpixel accuracy. Additionally, that method relies on accurate filter thresholds which can vary with the type of diffracted disk structure observed, decreasing general robustness and ease of use.

Radial gradient maximization is another method used [21, 8, 22], in which concentric circles or ellipses are placed around an estimated disk center and their rotational averages are calculated. Since most disks have sharp edges, when the difference between the rotational
average of concentric shapes are a maximum, the concentric shapes are properly placed
around the center of the disk. Through an iterative process, the estimated center is shifted
until this maximum is found. However, this method performs best when there are whole
disks without missing portions or strong contrast changes along the disks’ edges, which is
not always experimentally possible. Finally, this is an iterative method, and while it is faster
than edge detection [22], it still requires several computational steps per disk.

Cross correlation with respect to image registration is a fast technique in which a template
and base image are aligned via a convolution-like operation. While intuitively this might
seem computationally intensive, the Cross Correlation Theorem reduces the problem to a
matrix element multiplication the size of the image, two discrete Fourier Transforms (DFTs)
and a single inverse (fast) Fourier Transform (IFFT). For images \( g_1 \) and \( g_2 \) with DFTs of
\( G_1 \) and \( G_2 \),
\[
\mathcal{F}\{g_1 \otimes g_2\} = \frac{G_1 \circ G_2^*}{|G_1 \circ G_2^*|^n}
\]  
where \( \otimes \) represents the cross correlation operation, \( \mathcal{F} \) denotes the discrete Fourier transform,
\( ^* \) indicates the complex conjugate, \( \circ \) is the element-wise product of the two matrices, and
\( 1 - n \) is the power to which the Fourier coefficients are weighted.

Standard cross correlation is performed when \( n \) is set equal to zero. In this case, the
intensity distribution in the real space image is fully taken into account. In the Fourier
transforms, this manifests itself as a large weight put into low spatial frequency signals,
resulting in a center-of-mass like registration between the image and the template.

Phase correlation is the natural extension of Eq. 3, where \( G_1 \circ G_2^* \) is normalized element-
wise to unit magnitude, thereby only retaining the phase information. This occurs when \( n \)
is set equal to 1. Phase correlation results are insensitive to image intensity, but also more
sensitive to noise, sharp edges, and other high frequency components that previously had
low image intensity. Finally, both cross and phase correlation can be seen as the limits of
a more general correlation, in which image intensity is weighted by some arbitrary fraction
\( n \) from 0 to 1. In this study, we have tested correlation measurements using \( n = 0.5 \), which
we hereafter refer to as the "hybrid" correlation method.

After taking the IFFT of \( \mathcal{F}\{g_1 \otimes g_2\} \), the resulting matrix is the computed correlation,
where the pixel of highest intensity corresponds to the shift required to align one image to
the other. Two primary methods exist to increase the accuracy of the alignment to subpixel
values. The first is curve-fitting the correlation surface to find the function’s maximum
[23, 24]. This method can require iteratively curve-fitting a function (usually parabolic or
Gaussian) to the correlation in the spatial domain and can be made arbitrarily accurate,
but is computationally intensive.

The second, faster, algorithm involves Fourier upsampling in the frequency domain and
is well described in [18, 25]. Using the original correlation peak as its center, the method
directly computes subpixel IFFT values a fixed range around the brightest pixel with an
arbitrary upsampling factor \( \kappa \). This computation is performed as the matrix multiplication
of three matrices and is very computationally efficient. The drawback of this method when
compared to curve fitting is the lack of peak shape information and granularity of peak
location measurement dependent on \( \kappa \). However, to improve the final peak location, a last
step is performed in which a simple parabola is fit to the nearest neighbors of the upsampled peak location, giving the measurement a final slight improvement. Due to the lack of a two step iterative curve-fitting step, we observed Fourier upsampling to be approximately three times faster when compared to fitting a Gaussian curve to the peak location for subpixel registration.

A variety of image filters can also be applied prior to correlation in order to emphasize features of interest that will aid in accurate registration. Often, this is the disk edge. A number of iterative edge detection filters exist, including the previously mentioned Prewitt and Canny filters, but they require user defined thresholds and several attempts to correctly find an edge. A simpler approach is the Sobel filter, which is a simple $3 \times 3$ kernel convolved with the image which approximates the gradient of intensity in a certain direction. Since the Sobel filter is a simple matrix, it can be applied across datasets without concern for effects caused by changing user defined thresholds, and is much faster and impartial than an iterative method. Sobel filtering has been shown to improve cross correlation accuracy of the forward scattered disk, useful in differential phase contrast imaging, but has not been robustly tested with diffracted disks containing dynamical contrast [26].

Simulated examples of correlation measurements using a circular template with perfect noisy circular disk signals (with and without regions of zero intensity) are given in Fig. 1b-d. In all cases, as the image intensity weighting factor $n$ in Eq. 3 is increased from 0 to 1, the correlation peak sharpens. In a noisy measurement, this represents a trade-off between measurement precision and robustness of the disk position measurement. Additionally, we have simulated the relative disk position error (measured displacement from true position divided by disk radius $R$) as a function of number of counts (electrons) calculated using Poisson statistics in the total circular disk area. For perfect disks, or disks with intensity modulations that have inversion symmetry, cross correlation performs the best, and approaches the the ideal measurement root-mean-square (RMS) error along one direction $\sqrt{\langle \Delta x^2 \rangle}$ of

$$\frac{\sqrt{\langle \Delta x^2 \rangle}}{R} \approx \frac{1.2}{\sqrt{\lambda R}} \quad (4)$$

where $\lambda$ is the total number of counts. For the case in Fig. 1d without inversion symmetry, cross correlation fails badly. In Figs. 1c and d, the horizontal error is approximately double the vertical error for most cases. This is because the disk edges are well-defined at the top or top and bottom of the circular boundary. Small horizontal shifts in the position measurements are not well-constrained due to the lack of horizontal edges.

Another trend apparent in Figs. 1b-d is that all of the correlation methods follow a power law that scales with the usual value of $1/\sqrt{\lambda}$, provided the number of counts are high enough relative to the disk size. Below this threshold, large position errors are always present. To prevent these errors, the diffraction space measurements should be sufficiently binned to smaller disk radii, as shown in Fig. 2.

Figs. 2a-d show that for low signal-to-noise ratios, the power law error dependence can be recovered in call cases by down-sampling the measurements, i.e. binning the resulting images. However binning also increases measurement error, and so one must be careful to
Figure 2: Effect of binning on disk position measurement error, using (a) phase correlation, (b) hybrid correlation, (c) cross correlation and (d) Sobel-filtered cross correlation, for different numbers of counts. (e) One example simulation with Sobel filtering for various numbers of counts.
limit the amount of down-sampling. Phase correlation and the Sobel-filtered cross correlation measurements are particularly sensitive to over-sampling-induced position errors. The reason why is demonstrated in Fig. 2e for the Sobel case. When the number of counts is low enough to produce separated point measurements, the edge filtering no longer produces an accurate measurement. A similar effect happens in phase correlation where erroneous high spatial frequencies increase the measurement error when separated points are apparent. We therefore recommend that all measurements are down-sampled until all disks resemble the center column of Fig. 2e or better.

3. Results and Discussion

3.1. Position Measurements of Simulated Disks With Fine Structure

![Figure 3: (a) Accuracy of each correlation method as a function of electron counts per disk. The inset disk is an inverted image of the simulated diffraction disk at infinite signal to noise ratio. Accuracy was measured as the root mean square (RMS) error between the known shift and the measured shift in disk position, normalized by the disk radius.](image)

Often, the diffraction patterns recorded do not exhibit perfect kinematical scattering from an ideal sample, which would result in round, evenly illuminated diffraction disks. This is due to sample thickness, bending and other kinematical and dynamical diffraction effects, all of which increase the inner disk structure. This poses a challenge for the correlation methods previously discussed due to their reliance on specific similar features between the template and target disk. To test the effects of fine structure in the disks, one ideal disk and four imperfect disks were simulated. These are shown inset in Figure 3. For each measurement, the disk was shifted a random amount ($\Delta x, \Delta y$) and the effect of number of electron counts per disk being tested was applied using Poisson statistics. This simulated diffraction disk was then correlated with an ideal disk using one of the four methods of correlation discussed here, and difference between actual and measured shift was recorded. This process was simulated 1000 times for each signal to noise ratio level.

The results of these simulations are shown in Figure 3. For a perfect disk, it is clear that cross correlation is the best and most accurate method across all SNR levels. However,
when the disks are unevenly illuminated, cross correlation has a strong systematic error in position due to its inherent center-of-mass weighting. While cross correlation gets worse when inner structure is present, the other methods do retain their accuracy for disks with a high number of counts. This is a result of the correlation methods themselves and the type of image feature each method emphasizes. Since cross correlation does not normalize the Fourier transform, it is overly reliant on intensity variations in the image, leading to the systematic errors seen. The other methods rely much more on high frequency components of the image, namely edges, and the sharp edges present around the outer edge of the disk provide an accurate feature for registration. The results of this simulation make it clear that when inner structure appears in the diffracted disks, hybrid or Sobel filtered cross correlation should be used over pure phase or cross correlation.

### 3.2. Experimental Strain Measurements on an Unstrained Silicon Lattice

![Figure 4](image)

Figure 4: A comparison between disk registration methods on unstrained Si. a) A virtual dark field image showing thickness or bending contrast, but otherwise contrast-free. Streaks are indicative of poor camera readout during the scanning process. b) The mean diffraction pattern showing the x and y lattice vectors chosen. c-f) Line profiles of the regions marked with black rectangles. g-j) Strain maps of the area in a), highlighting differences in measured strain as a result of the correlation method used. The strain scale is from -0.6 to 0.6% strain. Inset is the root mean square value of the frame from zero strain.
Pure Si TEM support grids were purchased with 35 nm thin windows. A 50 by 50 pixel scanning nanodiffraction dataset was obtained using a convergence angle of 1.01 milliradians with a diffraction pattern image size of 512 by 512 pixels. The images were acquired with a binning of 4, and not further binned. We expect this sample to have a uniform zero strain across the field of view, which spans 250 × 250 nm. In this dataset, the central beam was obscured by a physical beam stop, preventing its use as a template. This often occurs in overly thick samples as the beam current must be raised to a level beyond the camera’s damage threshold to fully illuminate the diffracted disks. To form a template for the various correlation methods, a best-fit ellipse to the disks was used (defined as the disk with the highest normalized correlation peak). Once the peak template was created, we computed strain maps using each of the four correlation methods, compiled in Figure 4, using the mean diffraction pattern as the reference.

The results confirm the prediction from the simulations. For the standard cross correlation, large changes in strain state exist in the maps from left to right and top to bottom in the reconstructed images, and the maps are fairly free of noise. The line profile in Fig. 4c shows both low noise and an unrealistic strain gradient. Cross correlation therefore has a high precision (small changes in adjacent diffraction patterns), but has large systematic errors in accuracy due to uneven illumination in the diffracted disks. This signal could be due to changes in sample thickness, small local sample tilts due to bending, or beam tilt.

In comparison, the phase correlation strain maps are very noisy, with large discontinuous jumps in strain state from pixel to pixel. The inset numbers in Fig. 4j correspond to the root-mean-square of the image strain, and a comparison between the other methods show that phase correlation is significantly noisier. This corresponds to the relative low accuracy of phase correlation and its susceptibility to background noise, which agrees with the simulations.

Finally, both hybrid and Sobel filtered cross correlation offer a decent compromise between the two extremes, with the Sobel filtered results having a minimum of noise and a very uniform measure of the strain in all directions. This mirrors the simulated results, in which these were the two best methods when the diffracted disks began to show signs of inner structure.

**3.3. Experimental Measurements of an Austenitic Stainless Steel Sample**

An austenitic 304 stainless steel sample was prepared via jet electropolishing using a perchloric acid solution. A 250 by 250 pixel scanning nanodiffraction dataset was obtained using a convergence angle of 2 milliradians, with a diffraction pattern image size of 1920 by 1792 pixels and a pixel step size of 2.5 nm. The final dataset size was approximately 450 gigabytes. The data was binned by a factor of 8 to reduce dataset size to a more manageable 7 GB. The four disk registration methods were then applied to the dataset and the resulting strain maps computed via the process described previously. The results are shown in Fig. 5 with a scale showing strain from -2 to 2%. A low angle annular dark field (LAADF) image was taken at 720 mm with a collection semiangle of 4.5-22 mrad. This emphasized the presence of a number of isolated dislocations and a (111) twin boundary, which was also seen in the diffraction pattern.
Figure 5: A comparison of the four different disk registration methods and the resulting strain measurements around a twin boundary in an austenitic stainless steel. a) A HAADF image emphasizing dislocation contrast, taken prior to the NBED scan. b) A map of the summed correlation values for all diffracted peaks. The green box is the chosen zero strain reference region used in computing the strain maps. c-d) Average diffraction patterns for the right and left side of the twin boundary, showing the symmetry of the system. Additionally, x and y directions are shown in red and blue, correctly oriented with the strain maps. e-h) Strain maps computed using different correlation methods of the region of interest.
Again, there are a number of features that experimentally match the simulated data. First, in the pure phase correlation, the effects of noise and other outliers in the diffraction pattern images cause significant noise in the disk registration. This suggests that phase correlation lacks the robustness needed to ignore subtle variations in background noise.

It is harder to recognize the failure of pure cross correlation at a glance, but upon closer examination, large changes in strain are measured across flat regions of the sample (where the LAADF and ADF images have little to no change in intensity). The LAADF image does not support the dislocation density to realistically achieve a 4-6% change in strain across the FOV. Instead, this erroneous measurement can be attributed to the effect of changing inner disk illumination conditions as shown in the simulated data.

Finally, the two intermediate disk registration methods both fared much better on this dataset. While the hybrid method was noisier than the Sobel-filtered cross correlation image, the strain measurements across the image lack the large changes in strain typical to cross correlation, and measurement noise is minimized. This is apparent from the relatively smooth changes in strain and realistic values seen in their strain maps. For both the hybrid and Sobel-filtered maps, the strain fields around the individual dislocations can clearly be seen, without abnormal variations in uniformly-illuminated regions. Therefore, we found both the hybrid and Sobel-filtered cross correlation methods to provide the best strain field reconstruction of the nanobeam diffraction dataset shown here.

4. Conclusion

Four separate correlation methods were tested against simulated disks, an unstrained Si sample, and an engineering alloy. We show that when the diffracted disks have inner structure, methods such as pure phase or cross correlation often include artifacts that can be eliminated by Sobel filtered cross correlation or hybrid correlation disk registration methods. Additionally, we show that for low signal to noise disks, binning can help the accuracy of the disk position measurement when using Sobel filtered, phase or hybrid correlation. This is increasingly relevant due to advancements in detector acquisition speed, which can lower the dose per diffraction pattern for a given total beam current. Finally, we compared the four selected methods on two different experimental datasets, with similar results. Both cross and phase correlation had irrecoverable flaws (large systematic errors and a high degree of noise respectively), while hybrid and Sobel filtered cross correlation yielded optimal disk registration results, and therefore had the most accurate strain measurements.

Acknowledgements

The authors acknowledge support by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 under the Mechanical Behavior of Materials program. CG acknowledges support by the Austrian Science Fund (FWF):[J3397]. Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.
We would like to thank both Mary Scott and Julian E. C. Sabisch for assistance with sample acquisition and preparation.

References


