

Published in final edited form as:

*J Micro Nanolithogr MEMS MOEMS*. 2016 ; 15(3): . doi:10.1117/1.JMM.15.3.034001.

## Advancing X-ray scattering metrology using inverse genetic algorithms

Adam F. Hannon\*, Daniel F. Sunday, Donald Windover, and R. Joseph Kline\*

National Institute of Standards and Technology, Materials Science and Engineering Division, 100 Bureau Drive, Gaithersburg, MD 20899

### Abstract

We compare the speed and effectiveness of two genetic optimization algorithms to the results of statistical sampling via a Markov chain Monte Carlo algorithm to find which is the most robust method for determining real space structure in periodic gratings measured using critical dimension small angle X-ray scattering. Both a covariance matrix adaptation evolutionary strategy and differential evolution algorithm are implemented and compared using various objective functions. The algorithms and objective functions are used to minimize differences between diffraction simulations and measured diffraction data. These simulations are parameterized with an electron density model known to roughly correspond to the real space structure of our nanogratings. The study shows that for X-ray scattering data, the covariance matrix adaptation coupled with a mean-absolute error log objective function is the most efficient combination of algorithm and goodness of fit criterion for finding structures with little foreknowledge about the underlying fine scale structure features of the nanograting.

### Keywords

X-ray scattering; nanostructure metrology; genetic algorithm; covariance matrix adaptation evolutionary strategy; differential evolution; Markov chain Monte Carlo

## 1 Introduction

The recent progression of information processing technology moving toward device components with smaller and smaller feature sizes has required the advent of novel metrology methods for characterizing the fine details of those nanostructures. Traditional photolithographic fabrication methods have started to reach fundamental limits in the size of features they can produce<sup>1-3</sup> and the next generation methods such as extreme ultraviolet lithography<sup>4,5</sup> have many lingering issues needing to be addressed to become economically viable<sup>6,7</sup>. Thus, alternative lithographic methods such as block copolymer (BCP) directed self-assembly (DSA)<sup>8-13</sup>, direct write focused ion or electron beam<sup>14,15</sup>, or other multistep patterning photolithography techniques<sup>16-18</sup> are necessary to continue the trend predicted by Gordon Moore of device density doubling every couple of years<sup>19</sup>. In particular, BCP DSA can be quite prone to defects<sup>20</sup> and poor long-range order<sup>21,22</sup> compared to traditional

\*Correspondence to adam.hannon@nist.gov and joe.kline@nist.gov.

lithography, and as a result better metrology techniques for measuring the overall order need to be implemented. Regardless of the fabrication method chosen, precisely characterizing the structures made at these sub-10 nm dimensions is critical for ensuring quality control and device performance.

Current methods for characterizing defects in thin film BCPs include a combination of optical defect detection with scanning electron microscopy (SEM)<sup>23</sup>, scatterometry or optical critical dimension (OCD) metrology<sup>24</sup>, and X-ray scattering metrology using critical dimension small angle X-ray scattering (CDSAXS)<sup>25–30</sup> with resonant soft X-rays<sup>31–35</sup>. SEM is most useful for detecting defects related to pattern registration, where deviations from the patterned template are visible at the surface. However, simulation studies have shown that for BCP thin films, deviations in the through-film morphology can be present below the top surface<sup>36</sup> and thus not be discernable by scanning electron microscopy. CDSAXS on the other hand has demonstrated the capability for detecting periodic buried structure morphologies that differ from a desired uniform grating<sup>34</sup> but requires an inverse algorithm to convert the scattering intensity data into a meaningful real space structure. If the goal is having fast turn-around non-destructive characterization of detailed internal structures (with many model parameters), then traditional optimization refinement methods are not practical due to the long run times necessary to solve and test simulated scattering data iteratively and the very real possibility of identical-cost, degenerate structural solutions (common to scattering analysis). Thus an optimization algorithm that converges to the global best fit quickly and consistently is desired. Supplemental knowledge of the underlying structure can expedite this process and has to be used to some degree in model design, but such sample knowledge is not always readily available.

Previous studies<sup>33–35</sup> have used a Markov chain Monte Carlo (MCMC) algorithm<sup>37</sup> to determine the parameter set for the given structure model that best fits the measured scattering data. This approach requires a relatively good initial guess with tight parameter search bounds for the structure parameters and requires multiple independent chain runs to ensure the algorithm has convergence. However, the potential for fabrication errors require larger parameter search bounds to ensure the solution found is the actual structure; thus such an approach may not be able to distinguish a potential faulty sample. Approaches have also been tried that use massive computing resources with parallelization and highly refined grid based models with a reverse MCMC<sup>38</sup>, but such an approach is limited by the availability of the computing resources. Genetic and evolutionary based algorithms have shown the capability to search large parameter spaces with wide bounds successfully. These algorithms mimic biological evolution using the model parameter sets as the encoding genetic information that is then processed through some kind of mixing strategy over many generations until the optimal parameter set evolves from an initial set of randomly generated parameters<sup>39</sup>. The mixing strategy can be simple such as in a differential evolution (DE)<sup>40</sup> approach or more complicated such as in the covariance matrix adaptation evolutionary strategy (CMAES)<sup>41–43</sup>. These methods are better suited to search a wider parameter space than stochastic algorithms like the MCMC. Previous work<sup>44,45</sup> used an MCMC-like approach to determine the positions of post motifs directing the self-assembly of BCPs into complex patterns and required many runs that were averaged over space to find the most optimal solutions that yielded the desired target structure morphology. Similar work was

performed using chemical spot motifs using a CMAES algorithm that showed convergence to global solutions was possible consistently within a reasonable number of algorithm generations<sup>46,47</sup>.

In this work, we compare these methods of inverse structure determination from CDSAXS measurements using MCMC, DE, and CMAES algorithms. First, the algorithms are compared by fitting the parameters of model target structures with simulated intensity profiles, starting from random initial parameters and typical parameter search ranges. Next, a set of experimental data from a silicon nanograting made using a BCP mask pattern transfer was used to compare the three algorithms, which were also compared with the results previously found using this data and only an MCMC approach. Additionally, a series of objective functions are investigated for each algorithm to see how the high dynamic range of scattering intensity measurements affects convergence of model parameters. Similar high dynamic range data is encountered in other problems where inverse solutions are needed such as X-ray or neutron reflectivity<sup>48</sup>, ellipsometry<sup>49,50</sup>, scatterometry/OCD metrology<sup>51</sup>, or seismic surveying<sup>52,53</sup>, thus the methods presented should be applicable in those fields as well. The results of these various runs are compared to infer which algorithm and objective function combination is best suited for optimized fitting of scattering data in terms of both refinement speed and convergence robustness.

## 2 Methods

### 2.1 Inverse Algorithms

The three different methods tested were the MCMC, DE, and CMAES algorithms. Here we detail general parameters and procedures that are shared between the three methods. Details of each algorithm, how the internal parameters for each algorithm were chosen and optimized, and any nuances each algorithm presents to the problem of inverse structure determination are provided in the appendices (see Appendix A).

For all the inverse algorithms considered, the goal is to find a set of model parameters that best parameterizes the real space structure of the sample investigated via CDSAXS intensity measurements. The model parameter set contains  $N_{\text{Shape}}$  parameters describing the geometric positions of the defining features in the periodic structure of the thin film and the relative magnitude of the scattering length density,  $SLD \equiv \rho(\vec{r})$ , of the material as a function of those geometric positions. Additionally, three intensity scaling parameters are included to properly scale the simulated intensities with the measured experimental intensities and account for interfacial roughness, giving the total model parameters needed for a given structure as  $N_{\text{Param}} = N_{\text{Shape}} + 3$ . Model parameter indices are denoted as  $\kappa = 1, \dots, N_{\text{Param}}$ . These model parameters are stored in a vector  $\vec{P}_{G,C}$  where the index  $G$  represents the current iteration in the case of MCMC or generation for the genetic algorithms and  $C$  is either the chain number in the MCMC algorithm or individual index of the population in the genetic algorithms.

A general outline of the methodology is shown in Fig. 1. The approach is initiated by choosing which inverse optimization algorithm to use, selecting what kind of geometric model to use (and thus how many  $N_{\text{Param}}$  are needed), initiating model parameters within the

relevant bounds, picking the kind of objective function to be used to measure goodness of fit  $\Omega$ , importing the experimental (or simulated) target intensity profile, and simulating an intensity profile for a test model parameters set. The main algorithm then commences as follows: 1) The scattering intensity profile for each model parameter set  $\vec{P}_{G,C}$  within the population (for the genetic algorithms) or for the current step of the chain set (for the MCMC) is simulated yielding  $I_{\text{Sim},G,C}(\vec{q})$ , the simulated scattering intensity as a function of the reciprocal space vector  $\vec{q} = [q_x, q_z]$  for  $N_q$   $q_x$  and  $q_z$  values. 2) The goodness of fit  $\Omega$  is calculated for  $I_{\text{Sim},G,C}(\vec{q})$  with the experimental or target intensity profile  $I_{\text{Tar}}(\vec{q})$ . 3) Depending on the exact algorithm used, a comparison is made between  $\Omega$  values of previous iterations/generations, amongst  $\vec{P}_{G,C}$  in the current population, or a best found “most fit”  $\vec{P}_{\text{MF}}$  and a decision is made on how to proceed with the algorithm. 4) For the genetic algorithms, if the population has bottlenecked into a constant  $\Omega$  solution not less than  $\Omega_{\text{Crit}}$  (based on the average change in the goodness of fit over a fixed number of generations (chosen here as 20) staying below a predefined small value (chosen here as  $10^{-2}$ )), the population is reseeded keeping the best model parameter set  $\vec{P}_{\text{MF}}$  but reinitializing all other model parameter sets. This step generally occurs after the solution parameter set is already very close to the best solution, thus the aggressiveness of the step is only for fine refinement of the solutions. 5) The algorithm being used is implemented to modify  $\vec{P}_{G,C}$  based on the decision of the  $\Omega$  comparison. These steps are repeated each time using the modified  $\vec{P}_{G,C}$  in the first step until either a convergence criterion is reached (i.e. a predetermined value for  $\Omega = \Omega_{\text{Crit}}$  that is deemed an acceptable solution;  $\Omega_{\text{Crit}}$  is usually determined by running a few preliminary algorithm runs and seeing what the  $\Omega$  values starts to converge towards such that only a few accepting events occurs within several hundred generations, otherwise  $\Omega_{\text{Crit}}$  is just set to 0) or for a preset number of generations/iterations.

## 2.2 Modeling Scattering Length Density

In order to simulate scattered intensity profiles, an appropriate model must be used to convert  $\rho(\vec{r})$  into a shape profile that can be defined in a set of parameters of size  $N_{\text{Param}}$ . In general, the simulated intensity is written as

$$I_{\text{Sim}}(\vec{q}) = I_S I_0(\vec{q}) e^{-DW^2(\vec{q} \cdot \vec{q})} + I_{\text{BK}}, \quad (1)$$

where  $I_S$  is an intensity scaling parameter,  $DW$  is the Debye-Waller factor,  $I_{\text{BK}}$  is the background intensity, and  $I_0(\vec{q})$  is the amplitude squared of the Fourier transform of  $\rho(\vec{r})$  such that

$$I_0(\vec{q}) = \left| \int \rho(\vec{r}) e^{-i\vec{q} \cdot \vec{r}} d\vec{r} \right|^2. \quad (2)$$

For this study, a model using stacks of trapezoids was used. In the model, a system of  $M$  trapezoids is used to represent  $\rho(\vec{r})$  each with their own  $SLD = \rho_m$ , bottom width  $W_{B,m}$ , top width  $W_{T,m}$ , height  $H_m$ , and position of the bottom left corner in  $x$  and  $z$  ( $x_m, z_m$ ). Here  $m$  is

an index for the  $m$ th trapezoid. In principle each trapezoid is defined by 6 parameters, but generally constraints are assumed with trapezoids in the same stack having the same parameters defined for adjoining sides and one trapezoid has a corner defined as the origin, thus the actual total number of parameters will be less than  $6M$ . Additional details of how the intensities for the trapezoid model simulated intensities are calculated are in the appendices (see Appendix B). Fig. 2 shows how example composite shape profiles are developed from the trapezoid model with both single column (SC) and double columns (DC) of trapezoidal stacks. Details of how these model parameters are related and determined are also discussed in the appendices (see Appendix B).

### 2.3 Objective Function – Goodness of Fit

In order to find an optimal set of parameters producing a simulated intensity profile which most closely matches the experimental or target structure intensity profile, an objective function that properly captures the goodness of fit needs to be selected. The symbol  $\mathcal{Q}$  is defined as the measure of the goodness of fit for a general objective function. Each specific objective function then has its own designated symbol. Depending on the data being sampled and compared, various objective functions may be appropriate. For the intensity data examined in CDSAXS experiments, the sampled data can cover orders of magnitude in values suggesting some logarithmic objective function may be appropriate. At the same time, Poisson noise is expected to dominate the measured intensity values in terms of random error, suggesting a  $\chi^2$  objective function might be appropriate. Based on these facts, we chose to use three functions in the algorithm comparison study:  $\chi^2$ , a mean-absolute error logarithmic function, and a mean-absolute error function as a control. It should be noted that a parameter set that gives a minimum using one objective function does not necessarily give a minimum for other objective functions; thus the proper selection of an objective function that precisely embodies the goodness of fit for all the data is essential to inverse structure determination.

**2.3.1  $\chi^2$  goodness of fit**—The  $\chi^2$  goodness of fit objective function is a standard objective function that compares the square difference of values in the measured data and the simulated data normalized to the expected model value (treated here as the simulated value). In our intensity simulation studies, the following form of  $\chi^2$  was used:

$$\chi_{G,C}^2 = \frac{1}{N_q} \sum_{\vec{q}} \frac{\left( I_{\text{Sim},G,C}(\vec{q}) - I_{\text{Tar}}(\vec{q}) \right)^2}{\left| I_{\text{Sim},G,C}(\vec{q}) \right|}. \quad (3)$$

$\chi^2$  works best for cases where the error in the data obeys a Poisson distribution.

**2.3.2 Mean-absolute error log goodness of fit**—The mean-absolute error log (MAE<sub>log</sub>) goodness of fit objective function is a standard objective function that compares the difference of the logarithm of the values in the measured data and the simulated data<sup>54</sup>. In our intensity simulation studies, the following form of this objective function was used:

$$\varepsilon_{G,C} = \frac{1}{N_q - 1} \sum_{\vec{q}} \left| \log_{10} I_{\text{Sim},G,C}(\vec{q}) - \log_{10} I_{\text{Tar}}(\vec{q}) \right|. \quad (4)$$

This objective function is best suited for data that has values that span many orders of magnitude which is true for scattering intensity data where the primary peak can be many orders of magnitude larger than the higher order peaks<sup>40</sup>. When compared with a mean-squared error type function, this function has been shown to be less sensitive to statistical noise<sup>40</sup>.

**2.3.3 Mean-absolute error goodness of fit**—The mean-absolute error (MAE) objective function is another common cost function which incorporates elements of both  $\chi^2$  and  $\text{MAE}_{\log}$ . The MAE objective function can be appropriate for X-ray diffraction measurements when there is important information in the lower order peaks (i.e. the largest intensities)<sup>40</sup>. Mathematically this function is written as:

$$\psi_{G,C} = \frac{1}{N_q} \sum_{\vec{q}} \left| I_{\text{Sim},G,C}(\vec{q}) - I_{\text{Tar}}(\vec{q}) \right|. \quad (5)$$

This function is essentially the sum of the absolute value of the residuals between the data and simulated fit data.

## 2.4 Algorithm Comparison Methodology

To test the three algorithms and their ability to converge to acceptable parameter solutions for target scattered intensities, two approaches were devised. In the first approach, various target shape profiles built from trapezoids were constructed using defined parameter sets and used to create a simulated  $I_{\text{Tar}}(\vec{q})$  for testing the algorithms. Poisson noise typical to that of measured data was added to the simulated structure intensities to make the data more realistic and avoid having sharp features in the solution space (see Appendix C for how this noise was added to the simulated data). In the second approach, experimental data previously studied using an MCMC algorithm was fit with trapezoid models using all three algorithms. The success of the different algorithm and objective function combinations were measured based on two primary criteria: the magnitude of the best goodness of fit converged by that algorithms  $\Omega_{\text{Best}}$  and the normalized time to converge  $t_{\text{Conv}}/\tau_{\text{Gen}}$  to within a given target goodness of fit value  $\Omega_{\text{Tar}}$ .  $t_{\text{Conv}}$ , the absolute time for an algorithm to converge within  $\Omega_{\text{Tar}}$ , is the number of generations/iterations to converge  $G_{\text{Conv}}$  multiplied by  $\tau_{\alpha}$ , the average CPU time per generation/iteration for a given algorithm  $\alpha$ , such that  $t_{\text{Conv}} = G_{\text{Conv}} \tau_{\alpha}$ . This value is normalized by the average time of a CMAES generation  $\tau_{\text{Gen}}$  (CMAES generation time was chosen arbitrarily for the normalization as any algorithm could have been used) such that  $\tau_{\text{Gen}} = \tau_{\text{CMAES}} \cong 0.95 \tau_{\text{DE}} \cong 1.17 \tau_{\text{MCMC}}$ . In other words, a single step or generation in all three algorithms takes nearly an identical amount of time.  $\Omega_{\text{Tar}}$  is defined for the simulated structures as 1.1 times the theoretical goodness of fit  $\Omega_{\text{Conv}}$  for an exact match (found by removing the simulated Poisson noise from the simulated intensities) and 1.1 times  $\Omega_{\text{Best}}$  for the experimental data. This corresponds to a 10 % range of error in the

converged goodness of fit values which will be shown to be reasonable for the parameters examined in the results and discussion section. The algorithms were run using MATLAB<sup>®</sup> (R2014b) code on a system with dual 8 core processors (Intel<sup>®</sup> Xeon<sup>®</sup> E5-2687W v2) with 3.4 GHz clock speed<sup>†</sup>.

### 3 Results and Discussion

Here we present the results of using all three algorithms and objective functions. We then discuss which algorithm and objective functions worked best for the different data. For the genetic algorithms, the population size was always fixed to  $p_{\text{size}}=96$  for consistent comparison. 96 parallel chains were used in the MCMC runs to be comparable with the genetic algorithms.

#### 3.1 Simulated Structure Results

For the target simulated data case, an initial test was done using a single trapezoid with  $N_{\text{param}} = 6$  and refinements were performed varying these  $N_{\text{param}}$ . Fig. 3 shows the parameters for this structural model and the intensity data produced for fitting by the algorithms. Bounds for fitting of the model parameters were chosen for each parameter and are detailed in the appendices (see Appendix C). To correspond to previous studies of such systems<sup>33–35</sup>, only a selection of data was used to fit the parameters, with five constant  $q_x$  values being used to extract the intensity as a function of  $q_z$  for those five constant values (see Fig. 3(d)). Each slice each contained 1001 data points for each  $I_{\text{tar}}(q_z)$  for a total  $N_q = 5005$  among the 5  $q_z$  slices.  $q_z$  was varied from  $\approx -1.0451 \text{ nm}^{-1}$  to  $1.0451 \text{ nm}^{-1}$  in steps of  $0.00209 \text{ nm}^{-1}$  which is a similar resolution and range of values to those measured experimentally. Experimentally this  $q_z$  range is a function of the  $q_x$  resolution and angle step size as the experimental data is reconstructed from  $q_{xz}$  slices at different X-ray incidence angles. The  $q_x$  values for the slices were chosen to have a good range in intensity order of magnitude to reflect how such differences are encountered in experiments. Background noise was simulated as a constant plus Poisson distributed random noise, emulating detector read out noise behavior. Once the target data was simulated, our inverse methodology was implemented with parameters initiated randomly for the given population in the case of CMAES or DE or chain set for MCMC and the algorithm cycled until a solution close to the actual parameters was found with  $\Omega$  saturating within the  $\Omega_{\text{tar}}$  value. Fig. 4 shows a representative set of fits using the CMAES algorithm with  $\mathcal{E}$  objective function for a DC structure with  $M=2$  ( $N_{\text{param}} = 13$ ). The test fits, the target simulated data, and the residuals between the fit data and target simulated data are all plotted.

For the case of a single trapezoid with dimensions defined in Fig. 3,  $N_{\text{Runs}} = 10$  runs were performed using all three algorithms each with the three objective functions for a maximum of 10000 generations each run. To quantify the performance of the different algorithm and objective function combinations,  $t_{\text{Conv}}/\tau_{\text{Gen}}$  and  $\Omega_{\text{Best}}$  were compared. These values are

<sup>†</sup>Disclaimer

Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

shown for the single trapezoid structure in Table 1. The  $t_{\text{Conv}}/\tau_{\text{Gen}}$  values here are averaged over the 10 runs with the  $\Omega_{\text{Best}}$  values being the minimum observed for all 10 runs. The values for  $\Omega_{\text{Tar}}$  here were chosen such that the goodness of fit had converged within 10 % of the theoretical  $\Omega_{\text{Conv}}$  value (i.e.  $\Omega_{\text{Tar}}=1.1\Omega_{\text{Conv}}$ ) found using the intensities with noise as  $I_{\text{Tar}}$  and without noise as  $I_{\text{Sim}}$ . This convention was used for all three objective functions, so since each objective function scales goodness of fit differently, strict comparisons should only be made within the same objective function basis when considering convergence rates. The CMAES clearly outperformed in  $t_{\text{Conv}}/\tau_{\text{Gen}}$ , followed by the DE and then MCMC. In terms of  $\Omega_{\text{Best}}$ , both genetic algorithms reached similar best values while the MCMC tended to not refine exactly to the same value. However, the values converged towards were below or close enough to  $\Omega_{\text{Conv}}$  to be deemed an acceptable solution. For the  $\chi^2$  and  $\mathcal{E}$  objective functions, the actual  $\Omega_{\text{Best}}$  values found were even better than  $\Omega_{\text{Conv}}$ . This just means the parameters converged toward are slightly different than the actual parameter set without noise but are still acceptable solutions.

To examine how the algorithms perform with increasing  $N_{\text{Param}}$ , additional simulations were performed with SC trapezoid stacks of size  $M=2$  to 4 ( $N_{\text{Param}}=8,10,$  and 12) as well DC stacks of one and two trapezoids with fixed column widths ( $N_{\text{Param}}=9$  and 13) using the  $\mathcal{E}$  objective function. 10 runs were performed for each case for a maximum of 200000 generations. Fig. 5 shows plots of  $\Omega_{\text{Best}}$  versus  $t/\tau_{\text{Gen}}$  for the different numbers of parameters with the  $\Omega_{\text{Tar}}$  lines added to see where the target goodness of fit values are reached for each algorithm along with the corresponding shape profile used to simulate the intensity data. The CMAES algorithm clearly outperforms the DE and MCMC by  $\approx 1$  to 2 orders of magnitude in speed for all  $N_{\text{Param}}$ . The DE algorithm generally descends faster to start relative to the CMAES, but the CMAES always at some point descends in goodness of fit rapidly overtaking the DE and MCMC. This observation is due to the way the CMAES converges once the covariance matrix is well developed after several generations<sup>43</sup>. As  $N_{\text{Param}}$  increases, the CMAES and DE continues to converge consistently toward the lowest  $\Omega$  values, while the MCMC starts to take exponentially longer to get to those lowest  $\Omega$  values after converging below  $\Omega_{\text{Tar}}$ . Examining the target shape profiles (dotted black lines) with the solution shape profiles (colored lines), the solution shape profiles found match exactly for all targets except for the SC  $M=4$  with the MCMC where there is a slight mismatch in the placement of the bottom of the middle trapezoid (however, the overall shape matches well) and for the DC  $M=2$  with the MCMC where there are a couple of slight shifts in the trapezoid dimensions. The MCMC would likely eventually converge better for these cases with many more generations, but seeing each generation takes around 1 s to 2 s to run and 200000 generations or  $\approx 4.6$  d of computation time were already performed compared to the CMAES converging within an hour, continuing to run the algorithm for such further refinement was deemed unnecessary. Also, the DE takes longer to finally converge for the DC  $M=1$  case likely due to the search space for this structure having a local solution minimum in which the algorithm is staying. Additional discussion of the performance of the different algorithms with  $N_{\text{Param}}$  and number of columns plus all goodness of fit trajectory plots are given in the appendices (see Appendix D).



### 3.2 Experimental Structure Results

For experimental comparison, a data set with intensity  $I_{\text{Tar}}(q_z)$  for four  $q_x$  values that was used previously for determining a structure with multiple chain runs of an MCMC algorithm<sup>35</sup> was chosen to compare the efficiency and effectiveness of the different algorithms. The previous study found a stack of  $M=3$  trapezoids to yield an appropriate fit for the silicon lines formed from the pattern transfer of a block copolymer template, but for comparison with that study stacks of size  $M=1$  to 5 were additionally tested. The parameters that yielded the lowest goodness of fit solutions found in this study using the CMAES algorithm with  $\mathcal{E}$  objective function are shown and compared in Table 2. These results show the CMAES yielded a similar solution but with a lower  $\mathcal{E}$  value for  $M=3$  (normalized here by  $N_q - 1 = 2792$ ), slightly better fit with  $M=4$  trapezoids, and slightly worse fit with  $M=5$ . Additional details of these comparisons are in the appendices (see Appendix D).

To examine the speed of the algorithms in the experimental case, plots of the goodness of fit  $\Omega$  values as a function of  $t/\tau_{\text{Gen}}$  for the runs that yielded  $\Omega_{\text{Best}}$  upon completion are shown in Fig. 6. From the plots, the CMAES converged to a solution  $\approx 1$  to 2 orders of magnitude in time faster than the MCMC and less than an order of magnitude faster than the DE for the experimental structure for all objective functions. Fig. 6 also shows plots of the best shape profile fits (colored lines) for different combinations of algorithm and objective function used compared with the previously found solution (dashed black lines). The CMAES fits matched well with the previous study for all objective functions, the DE relatively well, and the MCMC only well for the  $\mathcal{E}$  objective function. These discrepancies are likely due to a combination of the  $\chi^2$  and  $\psi$  weighing the first order slice too much compared to the higher ordered slices while the  $\mathcal{E}$  objective function weighs them more evenly, as shown in Table 3, and the MCMC not completely converging. Since the higher order peaks contain the information on the finer details of the periodic nanostructures (i.e.  $d_{\text{feature}} \propto q^{-1}$ ), knowing how the objective function used weighs the calculated goodness of fit is important in deciding if a parameter solution set is an appropriate fit to the measured data. Thus, the fact that the  $\mathcal{E}$  function weighs the higher order peaks more appropriately in addition to giving a similar shape profile to the previously reported solution for all the algorithms gives credence to it being the best objective function for analyzing X-ray scattering data.

When comparing between algorithms for these simulations, there does not appear to be any correlation in a particular peak slice having a better fit with a particular algorithm; only when changing the objective function used do such discrepancies arise. Based on the fact the  $\mathcal{E}$  objective function gave the best match to the previous work, that objective function is expected to fit this kind of scattered intensity data most accurately in terms of weighing residual contributions from different peaks with large order of magnitude differences. For both the simulated structure and experimental data algorithm tests, the CMAES was determined to be the best algorithm to use for fast convergence to an accurate parameter set solution with basically no need to have an accurate starting solution population set.

## 4 Conclusion

The results of both the experimental and simulated structure studies demonstrate the CMAES algorithm converged most rapidly and reliably to an acceptable solution, the DE did so reliably as well, but at a slower rate, and the MCMC mostly converged to an acceptable solution but at the slowest rate. Higher structure model parameters resulted in longer computation time for all algorithms and slightly higher  $\Omega_{\text{Best}}$  values for the DE and MCMC compared to the CMAES but still gave acceptable solutions for the highest parameter numbers tested (i.e.  $\Omega_{\text{Best}} < \Omega_{\text{Tar}}$ ). For the experimental data tested, the CMAES with  $\mathcal{E}$  objective function gave the best match with the previously determined model parameters from an MCMC study<sup>35</sup>.

Seeing that the CMAES algorithm generally had a faster absolute time to converge to solutions and always had the lowest observed  $\Omega_{\text{Best}}$  values for the various objective functions, the algorithm shows great promise in advancing X-ray scattering metrology. With the insight gained from this study of genetic algorithms in analyzing CDSAXS data, future studies will be able to utilize these methods to analyze data quickly and efficiently with little foreknowledge of the internal structure. This current study only examined parameterized structures with up to 13 independent parameters, so future studies should examine the efficiency of the algorithms at shape models with  $N_{\text{Param}}$  in the range of 20 to 100. There may be further algorithm enhancements than those posited here such as hybrid algorithms (e.g. using the CMAES to find the relative location of the parameter solution and honing in on better solutions with a local MCMC algorithm with tighter parameter bounds or Levenberg-Marquardt algorithm<sup>55</sup>), making structure reconstruction more tractable and deterministic. Overall, the results of this study show the CMAES algorithm will enhance X-ray scattering in turn-around time for inverse complex structure determination.

## Acknowledgments

This research was performed while the author held a National Research Council Research Associateship Program award at the National Institute of Standards and Technology. The experimental data used in this work was collected at the DuPont-Northwestern-Dow Collaborative Access Team (DND-CAT) beamline which is located at the Advanced Photon Source (APS) Sector 5. E.I. DuPont de Nemours & Co., Northwestern University, and the Dow Chemical Company all supported DND-CAT. The Department of Energy (DOE) Contract No. DE-AC02-06CH11357 supported the use of the APS which is an Office of Science User Facility for the DOE Office of Science at Argonne National Laboratory. Steven Weigand and Denis Keane assisted with the data collection at sector 5-ID-D and their aid is greatly appreciated.

## References

1. The International Technology Roadmap for Semiconductors. 2013. <http://www.itrs2.net/itrs-reports.html>
2. Sanders DP. Advances in patterning materials for 193 nm immersion lithography. Chem Rev. 2010; 110:321–360. DOI: 10.1021/cr900244n [PubMed: 20070116]
3. Moors R, et al. Extreme-Ultraviolet Specifications: Tradeoffs and Requirements. J Micro/Nanolithogr, MEMS, MOEMS. 2012; 11:021102.
4. Stulen RH, Sweeney DW. Extreme ultraviolet lithography. IEEE J Quantum Electron. 1999; 35(5): 694–699. DOI: 10.1109/3.760315
5. Gwyn CW, et al. Extreme ultraviolet lithography. J Vac Sci Technol B. 1998; 16:3142–3149. DOI: 10.1116/1.590453

6. Benschop J, et al. Extreme ultraviolet lithography: Status and prospects. *J Vac Sci Technol B Microelectron Nanom Struct.* 2008; 26(6):2204–2207. DOI: 10.1116/1.3010737
7. Tomie T. Tin laser-produced plasma as the light source for extreme ultraviolet lithography high-volume manufacturing: history, ideal plasma, present status, and prospects. *J Micro/Nanolithography, MEMS MOEMS.* 2012; 11(2):021109.doi: 10.1117/1.JMM.11.2.021109
8. Herr DJC. Directed block copolymer self-assembly for nanoelectronics fabrication. *J Mater Res.* 2011; 26:122–139. DOI: 10.1557/jmr.2010.74
9. Kim SO, et al. Epitaxial self-assembly of block copolymers on lithographically defined nanopatterned substrates. *Nature.* 2003; 424:411–414. DOI: 10.2494/photopolymer.20.511 [PubMed: 12879065]
10. Stoykovich MP, et al. Directed assembly of block copolymer blends into nonregular device-oriented structures. *Science.* 2005; 308:1442–1446. DOI: 10.1126/science.1111041 [PubMed: 15933196]
11. Ruiz R, et al. Density multiplication and improved lithography by directed block copolymer assembly. *Science.* 2008; 321:936–939. DOI: 10.1126/science.1157626 [PubMed: 18703735]
12. Bitá I, et al. Graphoepitaxy of self-assembled block copolymers on two-dimensional periodic patterned templates. *Science.* 2008; 321:939–943. DOI: 10.1126/science.1159352 [PubMed: 18703736]
13. Yang JKW, et al. Complex self-assembled patterns using sparse commensurate templates with locally varying motifs. *Nat Nanotechnol.* 2010; 5:256–260. DOI: 10.1038/nnano.2010.30 [PubMed: 20228786]
14. Gierak, J. *Lithography.* John Wiley & Sons, Inc; Hoboken, New Jersey: 2013. Focused Ion Beam Direct-Writing; p. 183-232.
15. Winston D, et al. Scanning-helium-ion-beam lithography with hydrogen silsesquioxane resist. *J Vac Sci Technol B Microelectron Nanom Struct.* 2009; 27(6):2702–2706. DOI: 10.1116/1.3250204
16. Chawla JS, et al. Patterning challenges in the fabrication of 12 nm half-pitch dual damascene copper ultra low-k interconnects. *Proc SPIE.* 2014; :9054 905404.doi: 10.1117/12.2048599
17. Van Veenhuizen, M., et al. Demonstration of an electrically functional 34 nm metal pitch interconnect in ultralow-k ILD using spacer-based pitch quartering. *2012 IEEE Int. Interconnect Technol. Conf. IITC;* 2012; 2012. p. 1-3.
18. Villarrubia JS, et al. Scanning electron microscope measurement of width and shape of 10nm patterned lines using a JMONSEL-modeled library. *Ultramicroscopy.* 2015; 154:15–28. DOI: 10.1016/j.ultramicro.2015.01.004 [PubMed: 25747180]
19. Moore GE. Cramming More Components Onto Integrated Circuits. *Proc IEEE.* 1998; 86:82–85. DOI: 10.1109/JPROC.1998.658762
20. Bencher C, et al. Directed self-assembly defectivity assessment. Part II. *Proc SPIE.* 2012:8323, 83230N.
21. Cheng JY, et al. Fabrication of nanostructures with long-range order using block copolymer lithography. *Appl Phys Lett.* 2002; 81(19):3657–3659.
22. Cheng JY, et al. Templated self-assembly of block copolymers: Top-down helps bottom-up. *Adv Mater.* 2006; 18:2505–2521. DOI: 10.1002/adma.200502651
23. Delgadillo PR, et al. Defect source analysis of directed self-assembly process. *J Micro/Nanolithogr, MEMS, MOEMS.* 2013; 12(3):031112.
24. Dixit DJ, et al. Metrology for block copolymer directed self-assembly structures using Mueller matrix-based scatterometry. *J Micro/Nanolithogr, MEMS, MOEMS.* 2015; 14(2):021102.
25. Jones RL, et al. Small angle X-ray scattering for sub-100 nm pattern characterization. *Appl Phys Lett.* 2003; 83(19):4059–4061. DOI: 10.1063/1.1622793
26. Hu T, et al. Small angle X-ray scattering metrology for sidewall angle and cross section of nanometer scale line gratings. *J Appl Phys.* 2004; 96(4):1983–1987. DOI: 10.1063/1.1773376
27. Wang C, et al. Small angle X-ray scattering measurements of lithographic patterns with sidewall roughness from vertical standing waves. *Appl Phys Lett.* 2007; 90(19):1–4. DOI: 10.1063/1.2737399

28. Jones RL, et al. Pattern fidelity in nanoimprinted films using critical dimension small angle X-ray scattering. *J Microlithogr Microfabr Microsystems*. 2006; 5(1):013001.doi: 10.1117/1.2170550
29. Wang C, et al. Characterization of correlated line edge roughness of nanoscale line gratings using small angle X-ray scattering. *J Appl Phys*. 2007; 102(2):024901.doi: 10.1063/1.2753588
30. Wang, C., et al. *Thin Solid Films*. Vol. 517. Elsevier B.V; 2009. Small angle X-ray scattering measurements of spatial dependent linewidth in dense nanoline gratings; p. 5844-5847.
31. Virgili JM, et al. Analysis of order formation in block copolymer thin films using resonant soft X-ray scattering. *Macromolecules*. 2007; 40(6):2092–2099.
32. Wang C, et al. Defining the nanostructured morphology of triblock copolymers using resonant soft X-ray scattering. *Nano Lett*. 2011; 11(9):3906–3911. DOI: 10.1021/nl2020526 [PubMed: 21805981]
33. Sunday DF, et al. Three-dimensional X-ray metrology for block copolymer lithography line-space patterns. *J Micro/Nanolithography, MEMS, MOEMS*. 2013; 12(3):031103.doi: 10.1117/1.JMM.12.3.031103
34. Sunday DF, et al. Determination of the Internal Morphology of Nanostructures Patterned by Directed Self Assembly. *ACS Nano*. 2014; 8(8):8426–8437. DOI: 10.1021/nn5029289 [PubMed: 25075449]
35. Sunday DF, et al. Template-polymer commensurability and directed self-assembly block copolymer lithography. *J Polym Sci Part B Polym Phys*. 2015; 53(8):595–603. DOI: 10.1002/polb.23675
36. Liu CC, et al. Chemical Patterns for Directed Self-Assembly of Lamellae-Forming Block Copolymers with Density Multiplication of Features. *Macromolecules*. 2013; 46(4):1415–1424.
37. Mosegaard K, Sambridge M. Monte Carlo analysis of inverse problems. *Inverse Probl*. 2002; 18(3):R29–R54.
38. Sarje, A.; Li, X.S.; Hexemer, A. High-Performance Inverse Modeling with Reverse Monte Carlo Simulations. 2014 43rd International Conference on Parallel Processing; 2014. p. 201-210.
39. Mitchell, M. *Computers & Mathematics with Applications*. Vol. 32. MIT Press; Cambridge, MA: 1996. An introduction to genetic algorithms.
40. Wormington M, et al. Characterization of structures from X-ray scattering data using genetic algorithms. *Phil Trans R Soc Lond A*. 1999; 357(1761):2827–2848. DOI: 10.1098/rsta.1999.0469
41. Hansen N, Ostermeie a. Completely derandomized self-adaptation in evolution strategies. *Evol Comput*. 2001; 9(2):159–195. DOI: 10.1162/106365601750190398 [PubMed: 11382355]
42. Hansen N, Müller S, Koumoutsakos P. Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES). *Evol Comput*. 2003; 11(1):1–18. DOI: 10.1162/106365603321828970 [PubMed: 12804094]
43. Hansen, N.; Kern, S. Evaluating the CMA Evolution Strategy on Multimodal Test Functions. *Proc. 8th Int. Conf. Parallel Probl. Solving from Nat. - PPSN VIII*; 2004. p. 282-291.
44. Hannon AF, et al. Inverse design of topographical templates for directed self-assembly of block copolymers. *ACS Macro Lett*. 2013; 2(3):251–255. DOI: 10.1021/mz400038b
45. Hannon AF, et al. Optimizing topographical templates for directed self-assembly of block copolymers via inverse design simulations. *Nano Lett*. 2014; 14:318–325. DOI: 10.1021/nl404067s [PubMed: 24328687]
46. Qin J, et al. Evolutionary pattern design for copolymer directed self-assembly. *Soft Matter*. 2013; 9:11467–11472. DOI: 10.1039/c3sm51971f
47. Khaira GS, et al. Evolutionary Optimization of Directed Self-Assembly of Triblock Copolymers on Chemically Patterned Substrates. *ACS Macro Lett*. 2014; 3:747–752. DOI: 10.1021/mz5002349
48. Ulyanekov A, Omote K, Harada J. Genetic algorithm: Refinement of X-ray reflectivity data from multilayers and thin films. *Phys B Condens Matter*. 2000; 283(1–3):237–241. DOI: 10.1016/S0921-4526(99)01972-9
49. Fernandes VR, et al. Multi-objective genetic algorithm applied to spectroscopic ellipsometry of organic-inorganic hybrid planar waveguides. *Opt Express*. 2010; 18(16):16580–16586. DOI: 10.1364/OE.18.016580 [PubMed: 20721048]

50. Kudla A. Application of the genetic algorithms in spectroscopic ellipsometry. *Thin Solid Films*. 2004; 455:804–808.
51. Raymond C. Overview of Scatterometry Applications in High Volume Silicon Manufacturing. *AIP Conf Proc*. 2005; 788:394–402.
52. Boschetti F, Dentith MC, List RD. Inversion of seismic refraction data using genetic algorithms. *Geophysics*. 1996; 61(6):1715–1727. DOI: 10.1190/1.1444089
53. Aleardi M. Seismic velocity estimation from well log data with genetic algorithms in comparison to neural networks and multilinear approaches. *J Appl Geophys*. 2015; 117:13–22.
54. Sivia, DS.; Skilling, J. *Data Analysis: A Bayesian Tutorial*. 2. Oxford University Press; Oxford, UK: 2006.
55. Marquardt DW. An algorithm for least-squares estimation of nonlinear parameters. *J Soc Ind Appl Math*. 1963; 11(2):431–441.
56. Bernstein H, Bernstein C. Evolutionary origin of recombination during meiosis. *Bioscience*. 2010; 60(7):498–505.

## Appendix A: Inverse Algorithm Details

### A.1 MCMC Algorithm Description and Optimization

In the MCMC algorithm<sup>37</sup>, a set of chains are started from randomly seeded parameters  $\vec{P}_{G,C}$  where the indices  $G$  and  $C$  represent the iteration number and chain number, respectively. It should be noted the MCMC algorithm is inherently not a parameter search algorithm but a statistical distribution sampling method, thus the use and modification of the algorithm here as a search method instead should be emphasized.  $N_{\text{chains}}$  defines the total number of Markov chains and was fixed at a value of 96 for this study. Each parameter set chain can be initiated from a uniform random number distribution with each parameter being bound by an upper and lower bound based on the physical range of the model parameter or using *a priori* knowledge on the approximate value of the parameters. Alternatively, the values for  $\vec{P}_{G,C}$  can be initiated using a previously found  $\vec{P}_{G,C}$  that is known to be close to the solution for faster convergence (this approach was previously done in prior CDSAXS studies). For the simulations performed, only randomly initiated parameters were used for comparison with the genetic algorithms where the parameters were also initiated randomly. During the algorithm, when  $\Omega$  is calculated, a decision on how to proceed is made using a Metropolis-Hastings criterion<sup>37</sup>. Whenever  $\Omega_{G,C} < \Omega_{G-1,C}$ , the current  $\vec{P}_{G,C}$  is accepted as the working model parameter set for that chain. Also, if the criterion

$$\text{rand}([0, 1]) < e^{-\frac{(\Omega_{G,C} - \Omega_{\text{MF}})^2}{2}} \quad (6)$$

is satisfied, that  $\vec{P}_{G,C}$  is also accepted as the working model parameter set for that chain. Here  $\Omega_{\text{MF}}$  is the “most fit” goodness of fit value found thus far amongst all  $N_{\text{chains}}$  chains. If neither of these criteria are satisfied (i.e.,  $\Omega$  did not decrease locally or the weighted probability to accept a higher  $\Omega$   $\vec{P}_{G,C}$  did not fall in the bounds of a uniform random distribution number selected), then the previous  $\vec{P}_{G-1,C}$  is set for the working model parameter set for that chain.

After determining whether to use  $\vec{P}_{G,C}$  or  $\vec{P}_{G-1,C}$  as the working model parameter set for each chain, random walk moves are made based on the following considerations. A set of random move vectors  $\vec{V}_{\text{move},C}$  are added to  $\vec{P}_{G,C}$  such that

$$\vec{P}_{G+1,C} = \vec{P}_{G,C} + \vec{V}_{\text{move},C}, \quad (7)$$

where

$$\vec{V}_{\text{move},C} = (\vec{B}_{\text{Upper}} - \vec{B}_{\text{Lower}}) \cdot \vec{R}_C \quad (8)$$

with

$$\vec{R}_C(\kappa) = \frac{\text{rand}([0.5, 0.5])}{(\sigma_{\text{Step}}(1 + \text{rand}([0, 1])))}. \quad (9)$$

Here  $\vec{B}_{\text{Upper}}$  and  $\vec{B}_{\text{Lower}}$  are vectors of size  $N_{\text{Param}}$  containing the upper and lower bounds of each corresponding parameter in  $\vec{P}_{G,C}$ , respectively, and  $\sigma_{\text{Step}}$  is a step size parameter that is either set to a constant value or can be constrained to change as the algorithm converges (for simplicity a constant value was used in this study; optimization of the step size is discussed in the next subsection). If the addition of this move vector to an individual parameter  $\kappa$  would result in the value of that parameter being outside the defined bounds, the plus sign in the equation is changed to a minus sign to keep the parameter within the bounds. If changing the sign still results in the parameter moving outside of the bounds (i.e. the move size is bigger than the parameter bound range, usually only encountered for small  $\sigma_{\text{Step}}$ ), the parameter is set equal to the closest bound. We chose a value  $\sigma_{\text{Step}} \cong 32$  based on optimization tests described below, though this can be further optimized depending on  $N_{\text{Param}}$  and  $\Omega_{G,C} - \Omega_{\text{MF}}$ .

To ensure the MCMC algorithm was optimized, algorithm runs were performed at different values of  $\sigma_{\text{Step}}$  over a range of orders of magnitude for a given simulated structure for 1000 iterations. To decide the optimal  $\sigma_{\text{Step}}$ , both the average minimum goodness of fit values converged towards and the rate of acceptance between steps  $f_{\text{Acc}}$  (for both lowering from the previous step and the Metropolis-Hastings criterion for accepting a parameter set with a higher goodness of fit value). An example plot of these various values versus  $\sigma_{\text{Step}}$  is shown in Fig. 7 for the single trapezoid target structure. Of note, the average goodness of fit value of solutions found was minimized at  $\sigma_{\text{Step}} \cong 32$  while the acceptance rate value started to saturate around  $\sigma_{\text{Step}} \cong 3.2 \times 10^4$ . Since the study is more concerned with obtaining lower global fidelities (and not maximizing the amount of acceptance events) a step size of  $\sigma_{\text{Step}} = 32$  was chosen for all runs performed. The target structure used here was a single trapezoid

structure with  $W_{\text{Bot}} = 30$  nm,  $W_{\text{Top}} = 10$  nm,  $H = 15$  nm,  $I_{\text{Exp}} = 2$ ,  $I_{\text{Bk}} = 0.5$ , and  $DW = 0.5$  nm.

## A.2 DE Algorithm Description and Optimization

For the DE algorithm<sup>40</sup>, a population set of size  $p_{\text{size}}$  of  $\vec{P}_{G,C}$  is initiated and the inverse algorithm progresses based on mutation and crossover of these individual parameter sets.  $p_{\text{size}}$  was set to 96 in this study. Here the indices  $G$  and  $C$  represent the current generation of the population and individual member index label, respectively. In the algorithm, after each  $\vec{P}_{G,C}$  has a corresponding  $\Omega_{G,C}$  calculated, all members in the current generation population have their  $\Omega_{G,C}$  compared against each other and the best value of  $\Omega$  found thus far. If any individual  $\vec{P}_{G,C}$  yields an  $\Omega_{G,C}$  value better than the previous best, that  $\vec{P}_{G,C}$  is set as the new best  $\vec{P}_{\text{MF}}$  (here MF stands for “most fit”).

Regardless if a new best  $\vec{P}_{\text{MF}}$  is found or not, modification of the population for the next generation is accomplished as follows.  $\vec{P}_{\text{MF}}$  is kept separate during the mutation and crossover to ensure the best solution is always in the population pool (known as the “Genghis Khan” approach). A mutation step occurs by taking  $\vec{P}_{G,C}$  with the lowest  $\Omega_{G,C}$ , designated  $\vec{P}_{G,C_{\text{MF}}}$  in the current population and differential mixing of two other population members chosen at random are added to the parameter vector modulated by a mutation constant  $K_{\text{mut}}$ . In vector form this is simply

$$\vec{P}_{G+1,C} = \vec{P}_{G,C_{\text{MF}}} + K_{\text{mut}} \left( \vec{P}_{G,A} - \vec{P}_{G,B} \right) \quad (10)$$

where  $A$  and  $B$  are two random integers in the range from 1 to  $p_{\text{size}}$  with conditions  $A \neq B$  and  $A \neq C_{\text{MF}}$ . Following the mutation step, a crossover step is implemented going through all  $N_{\text{Param}}$  elements of each  $\vec{P}_{G,C}$  vector in the population with  $C = C_{\text{MF}}$  (i.e.  $C = [1, p_{\text{size}}]$ ) and the criterion  $\text{rand}([0,1]) < K_{\text{cro}}$  is used to determine if the  $j$ th parameter of  $\vec{P}_{G,C}$  should be swapped with the  $j$ th parameter of  $\vec{P}_{G,C_{\text{MF}}}$ . This is analogous to biological meiosis<sup>56</sup>. In the studies presented  $K_{\text{mut}} = 0.1$  and  $K_{\text{cro}} = 0.4$  based on optimization runs described below. Following the mutation and crossover steps, the new population  $\vec{P}_{G+1,C}$  is generated and the algorithm continues using this new population.

To optimize the DE crossover and mutation parameters, a scan of runs using  $K_{\text{mut}} \in [0.1, 2.0]$  and  $K_{\text{cro}} \in [0.1, 1.0]$  both in steps of 0.1 were performed for a given structure. The goodness of fit  $\mathcal{E}$  and acceptance fraction  $f_{\text{Acc}}$  were both calculated for all combinations of  $K_{\text{mut}}$  and  $K_{\text{cro}}$  examined after  $N_{\text{Gen}} = 1000$  generations. Plots of  $\mathcal{E}$  and  $f_{\text{Acc}}$  against the varied parameters are shown in Fig. 8. As seen in the figure,  $\mathcal{E}$  was minimized around  $K_{\text{mut}} \cong 0.1$  and  $K_{\text{cro}} \cong 0.4$  and  $f_{\text{Acc}}$  was maximal around  $K_{\text{mut}} \cong 0.1$  and  $K_{\text{cro}} \cong 0.5$ . Unlike the MCMC,  $f_{\text{Acc}}$  represents only situations where the global goodness of fit decreased since the DE is a true search algorithm and not a stochastic sampler. Thus, the parameters that yield the minimum  $\mathcal{E}$  values should be expected to correlate well with the values that yield the maximal  $f_{\text{Acc}}$  which they appear to do. Since the lowest  $\mathcal{E}$  are desired in the algorithm runs, the combination of parameters  $K_{\text{mut}} \cong 0.1$  and  $K_{\text{cro}} \cong 0.4$  were chosen for further runs.

### A.3 CMAES Algorithm Description

The CMAES algorithm<sup>41-43</sup> starts analogously to the DE algorithm with a population set of size  $p_{\text{size}}$  of  $\vec{P}_{G,C}$  being initiated. The algorithm progresses by calculating a mean weighted parameter vector  $\vec{P}_{G,\text{Mean}}$  for each generation that is weighted from the top  $\lambda_{\text{mix}}$  best individuals in the population with respect to their  $\Omega$  values. As in the DE algorithm, the indices  $G$  and  $C$  here again represent the current generation of the population and individual member index label, respectively. Again and analogous to the DE algorithm, each  $\vec{P}_{G,C}$  has a corresponding  $\Omega_{G,C}$  calculated after intensity simulations are performed with all members in the current generation population having their  $\Omega_{G,C}$  compared against each other and the best value of  $\Omega$  found thus far. If any individual  $\vec{P}_{G,C}$  yields an  $\Omega_{G,C}$  value better than the previous best, that  $\vec{P}_{G,C}$  is set as the new best  $\vec{P}_{\text{MF}}$ . Detailed below are all the equations that go into determining the covariance matrix  $\mathbf{V}$  and mixing equations for the CMAES.

$N_{\text{Param}}$  is defined as the number of parameters used in the model that are to be optimized to best fit the data.  $p_{\text{size}}$  is the population size of parameter sets.  $p_{\text{Opt}}$  is the lower bound optimal  $p_{\text{size}}$  needed to get convergence of the CMAES algorithm within  $G_{\text{Opt}} = 1000N_{\text{Param}}^2$  generations such that

$$p_{\text{Opt}} = 4 + \lfloor 3 \ln(N_{\text{Param}}) \rfloor. \quad (11)$$

Using large  $p_{\text{size}}$  can enhance convergence for a given algorithm run (i.e. lower  $G_{\text{Opt}}$ ) but at the cost of computational time. For consistency between the DE and CMAES algorithms, in this study  $p_{\text{size}}$  was fixed to 96 for both algorithms which is well above  $p_{\text{Opt}}$  for any parameter set considered (the largest  $N_{\text{Param}}$  explored was  $N_{\text{Param}} = 13 \rightarrow p_{\text{Opt}} = 11$ , much smaller than 96).

$\lambda_{\text{mix}}$  is the number of individuals mixed during the CMAES crossover/mutation mixing step.  $\lambda_{\text{mix}}$  is set equal to  $\lambda_{\text{Opt}}$  such that

$$\lambda_{\text{Opt}} = \lfloor p_{\text{size}}/2 \rfloor, \quad (12)$$

so in this study  $\lambda_{\text{mix}} = 48$ .  $\sigma$  is the parameter coordinate wise standard deviation step size used to scale the mixing of parameter sets for the next generation.  $\sigma$  is set to an initial value of  $\sigma_1 = 5$  and is updated according to the scheme detailed below.  $\sigma$  is reset to  $\sigma_1$  if it surpasses a value  $\sigma_{\text{Max}}$  which for this study was set as  $10^{20}$  and the corresponding evolutionary parameters reset to avoid numerical overflow.

$$\phi_C = \frac{\ln\left(\frac{p_{\text{size}}+1}{2C}\right)}{\sum_{C=1}^{\lambda_{\text{mix}}} \ln\left(\frac{p_{\text{size}}+1}{2C}\right)}, \quad (13)$$



$\phi_C$  is the crossover probability weight for an individual  $C$  with the individuals in the population up to  $\lambda_{\text{mix}}$ . This is calculated once at the beginning of the algorithm and then remains the same throughout when used in calculating  $P_{G,\text{Mean}}$ .

$$\lambda_{\text{Eff}} = \frac{\left(\sum_{C=1}^{\lambda_{\text{mix}}} \phi_C\right)^2}{\sum_{C=1}^{\lambda_{\text{mix}}} \phi_C^2}, \quad (14)$$

$\lambda_{\text{Eff}}$  is the effective number of individuals mixed in each generation used in calculating several of the evolutionary strategy updating parameters.

$G$  is the current generation number which is increased each generation until reaching a preset value  $G_{\text{Max}}$ . When  $p_{\text{size}} = p_{\text{Opt}}$  this maximum value should not be larger than

$$G_{\text{Opt}} = 1000 N_{\text{Param}}^2 \quad (15)$$

assuming the problem is bound correctly. In practice  $G_{\text{Max}}$  was set to values  $\approx 40000$  for consistency between runs, but because an additional criterion checking if  $\Omega < \Omega_{\text{crit}}$  is used to end the algorithm run meaning this  $G_{\text{Max}}$  is not always reached. During the algorithm, when objective function goodness of fit values are compared, the most fit individual  $G_{\text{MF}}$  is kept for reference along with the generation that individual was found  $G_{\text{MF}}$  and the corresponding parameter set  $P_{\text{MF}}$ .

The following parameters are used in updating the covariance matrix  $V$  and step size  $\sigma$ . They are  $t_V$ , the time constant for the accumulation of  $V$ ,  $t_\sigma$ , the time constant for the accumulation of  $\sigma$ ,  $R_1$ , the rate for the update of the 1<sup>st</sup> rank update of  $V$ ,  $R_{\lambda_{\text{mix}}}$ , the rate for the update of the  $\lambda_{\text{mix}}^{\text{st}}$  rank update of  $V$ , and  $\delta_\sigma$ , the damping coefficient for the generational update of  $\sigma$  which is  $\approx 1$ . These parameters remain the same throughout an entire CMAES run.

$$t_V = \frac{4 + \frac{\lambda_{\text{Eff}}}{N_{\text{Param}}}}{N_{\text{Param}} + 4 + \frac{2\lambda_{\text{Eff}}}{N_{\text{Param}}}}, \quad (16)$$

$$t_\sigma = \frac{\lambda_{\text{Eff}} + 2}{N_{\text{Param}} + \lambda_{\text{Eff}} + 5}, \quad (17)$$

$$R_1 = \frac{2}{(N_{\text{Param}} + 1.3)^2 + \lambda_{\text{Eff}}}, \quad (18)$$

$$R_{\lambda_{\text{mix}}} = \min \left( \left[ \begin{array}{c} 1 - R_1, \\ \frac{2 \left( \lambda_{\text{Eff}} - 2 + \frac{1}{\lambda_{\text{Eff}}} \right)}{(N_{\text{Param}} + 2)^2 + \lambda_{\text{Eff}}} \end{array} \right] \right), \quad (19)$$

$$\delta_{\sigma} = 1 + 2 \max \left( \left[ \begin{array}{c} 0, \\ \sqrt{\frac{\lambda_{\text{Eff}} - 1}{N_{\text{Param}} + 1}} - 1 \end{array} \right] \right) + t_{\sigma}. \quad (20)$$

The following parameters are internal algorithm updating parameters that are updating during every generation. They include  $\vec{P}_{G,V}$ , the parameter evolution path for  $V$ ,  $\vec{P}_{G,\sigma}$  the parameter evolution path for  $\sigma$ ,  $B$ , the eigenvector matrix of the covariance matrix  $V$  initialized as an identity matrix of size  $N_{\text{Param}}$  by  $N_{\text{Param}}$ ,  $D$ , the diagonal eigenvalue matrix of the covariance matrix  $V$  again initialized as an identity matrix, and  $X_N$ , the expectation value constant from of normally distributed random values of length  $N_{\text{Param}}$  used in the generational update of  $\sigma$ . Thus the covariance matrix is given as

$$V \equiv B D^2 B^T, \quad (21)$$

where  $T$  is the matrix transpose operator. In updating the parameter evolution paths, the inverse square root of the covariance matrix  $V^{-1/2}$  is a useful quantity to calculate during each step given as

$$V^{-1/2} \equiv B D^{-1} B^T. \quad (22)$$

$X_N$  is given as

$$X_N = \sqrt{N_{\text{Param}}} \left( 1 - \frac{1}{4N_{\text{Param}}} + \frac{1}{21N_{\text{Param}}^2} \right). \quad (23)$$

$\vec{P}_{G,\text{Mean}}$  is calculated from  $\phi_C$  and  $\vec{P}_{G,C}$  where  $\vec{P}_{G,C}$  is sorted with parameter sets in increasing goodness of fit value.

$$\vec{P}_{G,\text{Mean}} = \sum_{C=1}^{\lambda_{\text{mix}}} \phi_C \vec{P}_{G,C}^{\text{Sorted}}. \quad (24)$$

The parameter evolution path vectors during each generation are updated as follows:

$$\vec{P}_{G+1,\sigma} = (1-t_\sigma) \vec{P}_{G,\sigma} + \sqrt{t_\sigma(2-t_\sigma)\lambda_{\text{Eff}}} V^{-\frac{1}{2}} \frac{(\vec{P}_{G,\text{Mean}} - \vec{P}_{G-1,\text{Mean}})}{\sigma}, \quad (25)$$

$$\vec{P}_{G+1,\nu} = (1-t_\nu) \vec{P}_{G,\nu} + h_\sigma \sqrt{t_\nu(2-t_\nu)\lambda_{\text{Eff}}} \frac{(\vec{P}_{G,\text{Mean}} - \vec{P}_{G-1,\text{Mean}})}{\sigma}, \quad (26)$$

where

$$h_\sigma = \begin{cases} 1 & \text{if } \frac{(\vec{P}_{G+1,\sigma} \cdot \vec{P}_{G+1,\sigma})}{N_{\text{Param}} \left( \frac{2G}{1-(1-t_\sigma)\lambda_{\text{mix}}} \right)} < 2 + \frac{4}{N_{\text{Param}} + 1} \\ 0 & \text{otherwise} \end{cases}. \quad (27)$$

The covariance matrix for the next generation  $\mathbf{V}_{G+1}$  is then given as

$$\mathbf{V}_{G+1} = (1-R_1 - R_{\lambda_{\text{mix}}}) \mathbf{V}_G + R_1 \left( \vec{P}_{G+1,\nu} \times \vec{P}_{G+1,\nu}^T + (1-h_\sigma)t_\nu(2-t_\nu)\mathbf{V}_G \right) + R_{\lambda_{\text{mix}}} \mathbf{W}(\phi_C \mathbf{I}) \mathbf{W}^T, \quad (28)$$

where

$$\mathbf{W} = \frac{1}{\sigma} [\vec{P}_{G,1} - \vec{P}_{G-1,\text{Mean}}; \vec{P}_{G,\lambda_{\text{mix}}} - \vec{P}_{G-1,\text{Mean}}] \quad (29)$$

and  $\sigma_{G+1}$  for the next generation is given as

$$\sigma_{G+1} = \sigma_G e^{\frac{t_\sigma}{\delta_\sigma}} \left( \frac{\vec{P}_{G+1,\sigma}}{|\vec{P}_{G+1,\sigma}|_{X_N}} - 1 \right). \quad (30)$$

The population of parameters for the next generation  $\vec{P}_{G+1,C}$  is then given as

$$\vec{P}_{G+1,C} = \vec{P}_{G,\text{Mean}} + \sigma_{G+1} \mathbf{B} \text{Undiag}(\mathbf{D} \cdot \eta \mathbf{I}), \quad (31)$$

where  $\eta$  is a vector of size  $N_{\text{Param}}$  with random numbers drawn from a normal distribution with standard deviation 1 and the Undiag(\*) function outputs the diagonal values of a square matrix as a vector. The values used for  $\mathbf{B}$  and  $\mathbf{D}$  in updating  $\vec{P}_{G+1,C}$  are only updated when the following condition is satisfied:

$$\text{If } G - G_E > \frac{P_{\text{size}}}{10 N_{\text{Param}} (R_1 + R_{\lambda_{\text{mix}}})}, \quad (32)$$

where  $G_E = 0$  to start and  $G_E$  is set to the current  $G$  each time  $\mathbf{B}$  and  $\mathbf{D}$  are updated. When such an update occurs, the covariance matrix  $\mathbf{V}$  is updated by the following scheme to ensure the matrix is symmetric.

$$\mathbf{V} = \begin{bmatrix} 1 & 1 & & 1 & 1 \\ & 0 & 1 & & 1 \\ & \vdots & \ddots & & \vdots \\ 0 & 0 & & 1 & 1 \\ 0 & 0 & & & 1 \end{bmatrix} \cdot \mathbf{V}_{G+1} + \left( \begin{bmatrix} 0 & 1 & & 1 & 1 \\ & 0 & \dots & 1 & 1 \\ & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & & 0 & 1 \\ & & & \dots & \\ 0 & 0 & & 0 & 0 \end{bmatrix} \cdot \mathbf{V}_{G+1} \right)^T. \quad (33)$$

$\mathbf{B}$  and  $\mathbf{D}$  are then found by solving the eigenvalue problem

$$\mathbf{V}\mathbf{B} = \mathbf{D}^2\mathbf{B}. \quad (34)$$

## Appendix B: Details of Calculating Intensities from the Trapezoid Model

In the trapezoid model system the base intensity (square of the form factor) is given as

$$I_0(q_x, q_z) = \left( \sum_{m=1}^M \rho_m \left( \frac{B_{L,m}}{q_x} \frac{C_{L,m}}{\left(q_z + \frac{q_x}{S_{L,m}}\right)} - \frac{B_{R,m}}{q_x} \frac{C_{R,m}}{\left(q_z + \frac{q_x}{S_{R,m}}\right)} \right) \right)^2, \quad (35)$$

where

$$C_{\Theta,m} = \left( e^{-i\left(q_z + \frac{qx}{S_{\Theta,m}}\right)(z_m + H_m)} \quad -e^{i\left(q_z + \frac{qx}{S_{\Theta,m}}\right)z_m} \right) \text{ and } B_{\Theta,m} = e^{\frac{iq_x b_{\Theta,m}}{S_{\Theta,m}}}, \quad (36)$$

with  $\Theta=L$  or  $R$  for left and right, respectively, such that the  $m$ th trapezoid is defined by the intersection of the four lines

$$z = S_{L,m}x + b_{L,m}, \quad z = S_{R,m}x + b_{R,m}, \quad z = z_m, \quad \text{and} \quad z = z_m + H_m. \quad (37-40)$$

In terms of the model parameters, these slopes and intercept parameters are defined such that

$$S_{L,m} = \frac{2H_m}{(W_{B,m} - W_{T,m})} = -S_{R,m}, \quad (41)$$

$$b_{L,m} = z_m - S_{L,m}x_m, \quad (42)$$

and

$$b_{R,m} = z_m - S_{R,m}(x_m + W_{B,m}). \quad (43)$$

A general parameter vector for this model will have the following form:

$$\vec{P}_{G,C} = [W_{B,1}, W_{T,1}, H_1, x_1, z_1, \rho_1, \dots, W_{B,M}, W_{T,M}, H_M, x_M, z_M, \rho_M, I_S, DW, I_{Bk}]. \quad (44)$$

This form would yield  $6M+3$  independent parameters, but usually this number is reduced by additional constraints placed on a given model for a system. For a single trapezoid, the corner position is generally assumed to be that of the origin  $(x_1, z_1)=(0,0)$  and the *SLD* usually is assumed to have a fixed value making a single trapezoid model only have 6 independent parameters (3 geometric parameters with 3 intensity scaling parameters). For a column of  $M$  trapezoids stacked on top of each other, each trapezoid in the middle of the stack shares widths with the trapezoids above and below them with their positions being fixed by those adjoining trapezoid widths, making the total number of parameters for such a system assuming the *SLD* to be the same for each trapezoid be  $N_{\text{Param}} = 2M + 4$  as there are  $M+1$  independent widths,  $M$  independent heights, and 3 intensity scaling parameters. For a periodic array of  $c$  columns of  $M$  trapezoids with each column independent of the adjoining column and fixed column widths, the only additional parameter needed to specify each column beyond the first is the position of the lower left corner of the column, giving

$$N_{Param} = c(2M+1) + (c-1) + 3 = 2(c(M+1)+1). \quad (45)$$

More complicated models with various constraints can be made based on these ideas, but for the current manuscript only the kinds mentioned here are investigated.

## Appendix C: Parameters and Bounds for the Different Structures

Table 4 through Table 9 list all the simulated structure parameters and the corresponding upper and lower bounds imposed on them during the inverse algorithm runs. Additionally, the expected parameters from the previous experimental study with the bounds used for those runs are listed in Table 10. For Table 10, the lower bounds for the heights were set slightly above 0 to disallow  $M = 2$  or 1 trapezoid solutions (i.e. where a given  $H$  would go to 0) but also allow approach of such solutions to be possible in case such solutions were along path trajectories towards the best solutions. The parameter  $I_{Exp}$  and  $I_{Bk}$  were calculated based as effective values as the previous study used an additional set of data with  $q_z \cong 0 \text{ nm}^{-1}$  for a range of  $q_x$  from near  $0 \text{ nm}^{-1}$  to just over  $1 \text{ nm}^{-1}$  that was used to rescale the peak intensities during fitting (thus the input values in that study were different, so the values presented here are after rescaling). As a reminder from the main text, SC stands for single column and DC for double column. The width and height parameters generally had a large bound range to ensure decent parameter space coverage with a tighter bound range being imposed on  $DW$  and the intensity scaling parameters  $I_{Exp}$  ( $I_S = 10^{I_{Exp}}$ ) and  $I_{Bk}$ .  $I_{Bk}$  always had a minimum bound of 0.01 to ensure that the intensities were always greater than 0 so that  $\mathcal{E}$  did not have singularities (i.e.  $\mathcal{E} \rightarrow \infty$  as  $I_{Sim} \rightarrow 0$ ).

In addition to the tables showing the parameters and bounds for the different target structure sets, Fig. 9 through Fig. 13 show schematic diagrams of the target structure shape profiles, the calculated target intensity profile  $I_{Tar}(\vec{q})$  as a function of  $\vec{q} = [q_x, q_z]$ , and the slices from the profiles with Poisson noise added with constant  $q_x$  values used for parameter optimization by the different algorithms. Poisson noise is a random value proportional to the square root of the number of observed counts, thus the noise added was developed as

$$\xi(\vec{q}) = \sqrt{\frac{I_{Tar}(\vec{q})}{I_{Count}}} \text{rand}([-1, 1]), \quad (46)$$

where  $I_{Count}$  is the intensity contribution from a single photon scattering event. Since the simulated data is produced at a continuum approximation for the scattering with an analytical Fourier Transform,  $I_{Count}$  must be selected independent of the intensity calculation.  $I_{Count}$  should be larger than  $I_{Bk}$  assuming a decent signal-to-noise. Based on the observed experimental values of noise,  $I_{Count}$  for the simulated data was chosen to be  $I_{Count} = 6I_{Bk}$  to get similar noise levels with those observed in experiment. For each simulated intensity, a fixed noise profile was produced for that intensity and used in all runs for that target structure. These figures only show the simulated target structures. For the

experimental data, the constant value  $q_x$  slices were already extracted from the integrated intensity data and their plots can be seen in Fig. 14 where the black curves are the experimental data. An example fit (green) using the CMAES algorithm with the corresponding residuals using  $M = 3$  trapezoids is shown in Fig. 14 as well.

## Appendix D: Compiled Results of Different Target Structure Studies

### D.1 Simulated Target Structures

Fig. 15 shows plots for all  $N_{\text{Run}}$  runs of the SC  $M = 1$  trapezoid target structure goodness of fit values versus normalized time for all three objective functions and three algorithms. Fig. 16 shows plots for all  $N_{\text{Run}}$  runs of the SC  $M = 2, 3,$  and  $4$  structures as well as the DC  $M = 1$  and  $2$  structures for the  $\mathcal{E}$  objective function for all three algorithms. In both figures, there are inset images of the target structure. In general, green curves are used for the CMAES, blue curves for DE, and red curves for MCMC. The general trend seen in all cases is that CMAES takes on average about an order of magnitude less time than the DE or MCMC to start rapidly converging towards a minimum goodness of fit value. In terms of actual convergence percentage rate, the CMAES almost always converges within the number of generations tested while DE converges most of the time and MCMC converges more poorly with increasing  $N_{\text{param}}$ . Normalized times were chosen for comparison rather than absolute times because different CPU loading conditions affected the absolute times. Thus the normalized times can be compared whatever the speed of the internal CPU clock.

### D.2 Experimental Target Structure

Table 11 shows a comparison between the model parameter set solution found in a previous seeded MCMC study with simulations varying the number of trapezoids in the stack model where the CMAES algorithm was used with the  $\mathcal{E}$  objective function. This is an extended version of Table 2 from the main text. Where no parameter is comparable (because of the number of trapezoids being different) NC is placed in the table for not comparable. In the previous study, one nuance difference to account for when comparing the parameters found in that study is that a fit to a  $q_z \cong 0$  slice of data as a function  $q_x$  was also used in conjunction with the constant  $q_x$  peak slices to fit the data requiring the intensities to be rescaled to the maximum peak intensities found. This simply means that the background intensity parameter  $I_{Bk}$  and scaling exponent  $I_{Exp}$  cannot be directly compared between the studies, but all geometric shape parameters and the Debye-Waller factor are still directly comparable. The values for the previous study in the table are instead effective parameters found by taking the fit data from the previous study and rescaling the data with the same number of data points.

To explore how well the different algorithm and objective function combinations converged to appropriate parameter set solutions for the experimental data set, plots of the goodness of fit values versus  $t/\tau_{Gen}$  were made. These plots are shown in Fig. 17 with green curves for the CMAES, blue curves for DE, and red curves for MCMC. These results are from  $N_{\text{Run}} = 11$  for the CMAES and DE algorithms and  $N_{\text{Run}} = 5$  for the MCMC. Similar trends are seen

here that were seen for the simulated target structure data for the trends of when the algorithms start to converge with the CMAES being the best followed by DE and MCMC.

## Appendix E: List of Abbreviations and Symbols

<b>CDSAXS</b>	Critical dimension small angle X-ray scattering
<b>OCD</b>	Optical critical dimension
<b>MCMC</b>	Markov chain Monte Carlo
<b>DE</b>	Differential evolution
<b>CMAES</b>	Covariance matrix adaptation evolutionary strategy
<b>SC</b>	Single column
<b>DC</b>	Double column
<b><i>SLD</i> or <math>\rho</math></b>	Scattering length density
<b><math>\vec{r}</math></b>	Position vector in real space with coordinates $x$ and $z$
<b><math>N_{\text{Shape}}</math></b>	Number of independent parameters describing the shape profile in the models
<b><math>N_{\text{Param}}</math></b>	Number of independent parameters (includes $N_{\text{Shape}}$ and 3 intensity scaling parameters)
<b><math>G</math></b>	Iteration (MCMC) or generation index (DE/CMAES)
<b><math>C</math></b>	Chain number (MCMC) or population individual index (DE/CMAES)
<b><math>\vec{P}</math></b>	Parameter vector of size $N_{\text{Param}}$ for the set of parameters
<b><math>\kappa</math></b>	Parameter index
<b><math>\vec{q}</math></b>	Reciprocal space vector with coordinates $q_x$ and $q_z$
<b><math>I</math></b>	General intensity variable (subscript distinguishes type)
<b><math>I_{\text{Sim}}</math></b>	Simulated scattered intensity corresponding to a model parameter set
<b><math>I_{\text{Tar}}</math></b>	Experimental scattered intensity or target simulated scattered intensity
<b><math>\Omega</math></b>	General goodness of fit value
<b><math>\Omega_{\text{Crit}}</math></b>	Value used as one deciding factor in exiting an inverse algorithm



$\vec{P}_{MF}$	“Most fit” parameter set (i.e. model parameter set that gives the lowest $\mathcal{Q}$ found)
$I_0$	Absolute square of the form factor (unscaled simulated intensity)
$I_S$	Multiplicative scaling intensity parameter
$I_{Exp}$	Logarithm of $I_S$ in base 10
$DW$	Debye-Waller factor ( $\sim$ interfacial thickness between materials in shape model)
$I_{Bk}$	Background intensity shifting parameter
$F$	General function
$M$	Total number of trapezoids used in shape profile model
$m$	Index label for a given trapezoid in shape profile model
$W_B$	Bottom width of a given trapezoid
$W_T$	Top width of a given trapezoid
$H$	Height of a given trapezoid
$N_q$	Number of data points used in calculating goodness of fit values (i.e. number of $q_x$ and $q_z$ coordinates used)
$\chi^2$	Chi squared objective function
$\mathcal{E}$	Mean absolute error log objective function
$\psi$	Mean absolute error objective function
$\alpha$	Index for algorithm type (MCMC, DE, or CMAES)
$t_{Conv}$	Time needed for a given algorithm to converge to within a given goodness of fit value
$t$	General time variable
$\tau_{Gen}$	Time needed for a single generation or iteration to finish
$G_{Conv}$	Number of generations needed for a given algorithm to converge within a given goodness of fit value
$\tau_\alpha$	Time constant for the average time for a single generation step for a given algorithm $\alpha$
$\mathcal{Q}_{Tar}$	Goodness of fit value that needs to be reached to have a solution within desired accuracy

$\Omega_{\text{Conv}}$	Theoretical best $\Omega$ value possible if exact parameter solution is found
$P_{\text{size}}$	Number of individual parameter sets used in a given genetic algorithm run (population size in CMAES/DE or number of chains in MCMC)
$N_{\text{Runs}}$	Number of times an algorithm was used with different initial conditions for the same input data (simulated or experiment)
$W_{\text{FWHM}}$	Full width at half maximum height for a column of trapezoids. Used in comparing the widths of trapezoid stacks of varying number modeling the same data set.
	Best Subscript notation indicating a solution has the lowest found goodness of fit
$R_I$	Residual of the target and simulated intensities
$N_{\text{Chains}}$	Number of chains in MCMC
$\text{rand}([n_1, n_2])$	Function whose output is a random number drawn from a uniform distribution with minimum $n_1$ and maximum $n_2$
$\Omega_{\text{MF}}$	Lowest global goodness of fit value found during an inverse algorithm run
$\vec{V}_{\text{move}}$	Parameter move vector in MCMC
$\vec{B}_{\text{Upper}}$	Upper bound parameter vector
$\vec{B}_{\text{Lower}}$	Lower bound parameter vector
$\vec{R}$	Random contribution to $\vec{V}_{\text{move}}$
$\sigma_{\text{Step}}$	Step size in MCMC, the parameter move size is inversely proportional to this factor
$f_{\text{Acc}}$	Fraction of moves that are accepted during an algorithm run under a given criterion
$\Omega_{\text{Norm}}$ or $E_{\text{Norm}}$	Fidelity value used to normalize goodness of fit value such that it scales with $f_{\text{Acc}}$
$K_{\text{mut}}$	Mutation constant in DE
$A$ and $B$	Random population index integers in DE mixing scheme
$K_{\text{cro}}$	Crossover constant in DE
$\lambda_{\text{mix}}$	Number of individuals used in mixing in CMAES for making mean parameter set

$\vec{P}_{G,Mean}$	Average parameter vector in CMAES from $\lambda_{mix}$ best parameter sets
$V$	Covariance matrix used in CMAES
$p_{Opt}$	Lower bound optimal population size necessary for CMAES to converge well
$G_{Opt}$	Optimal number of generations needed for CMAES to converge to an acceptable solution when $p_{Opt}$ is used
$\lambda_{Opt}$	Optimal mixing size for CMAES algorithm
$\sigma$	Coordinate wise standard deviation step size used in CMAES mixing scheme; Also the shape parameter for the log-normal distribution
$\sigma_1$	Initial value of $\sigma$ in CMAES algorithm
$\sigma_{Max}$	Maximum value of $\sigma$ allowed in CMAES algorithm to avoid numerical overflow
$\phi$	Crossover probability weight (i.e. how much weight a parameter set gives to the mean parameter set) in the CMAES algorithm
$\lambda_{Eff}$	The effective number of individual parameter sets used in the CMAES mixing scheme based on the values of $\phi$
$G_{Max}$	Maximum number of generations allowed before algorithm terminates
$MF$	Subscript for “most fit” parameter set found within a given inverse algorithm run
$t_V$	the time constant for the accumulation of $V$
$t_\sigma$	the time constant for the accumulation of $\sigma$
$R_1$	the rate for the update of the 1 <sup>st</sup> rank update of $V$
$R_{\lambda_{mix}}$	the rate for the update of the $\lambda_{mix}^{st}$ rank update of $V$
$\delta_\sigma$	the damping coefficient for the generational update of $\sigma$
$\vec{P}_{G,V}$	the parameter evolution path for $V$
$\vec{P}_{G,\sigma}$	the parameter evolution path for $\sigma$
$B$	the eigenvector matrix of the covariance matrix $V$
$D$	the diagonal eigenvalue matrix of the covariance matrix $V$
$X_N$	the expectation value constant

$T$	is the matrix transpose operator
$h_{\sigma}$	Constant used in calculating the parameter evolution path vectors
$W$	Matrix used in calculating the covariance matrix
$\eta$	Random vector of normally distributed numbers
<b>Undiag(*)</b>	Function that takes in a square matrix and outputs the diagonal values
$G_E$	Generation index used in determining if an update to the covariance matrix should be performed
$\theta$	Index used for distinguishing the right R and left L sides of a given trapezoid
$B_{\theta,m}$	Exponential phase factor used in calculating the scattering of a given trapezoid
$C_{\theta,m}$	Exponential difference factor used in calculating the scattering of a given trapezoid
$S_{\theta,m}$	Slope of a given trapezoid side
$b_{\theta,m}$	Intercept of a given trapezoid side
$c$	Number of columns used in trapezoid stack model
$\mu$	Mean for the log-normal distribution
$PDF_{\mathcal{E}}$	Log-normal probability distribution function
$S_F$	Scaling factor for scaling log-normal probability distribution function
$PDF_{\mathcal{E}S_F}$	Rescaled log-normal probability distribution function
$\mu_I$	Arithmetic mean of the simulated intensity
$\sigma_I$	Arithmetic standard deviation of the simulated intensity
$\xi$	Poisson distributed background noise added to simulated data
$I_{\text{Count}}$	Intensity amount contributed by a single photon count in the simulated data

## Biographies

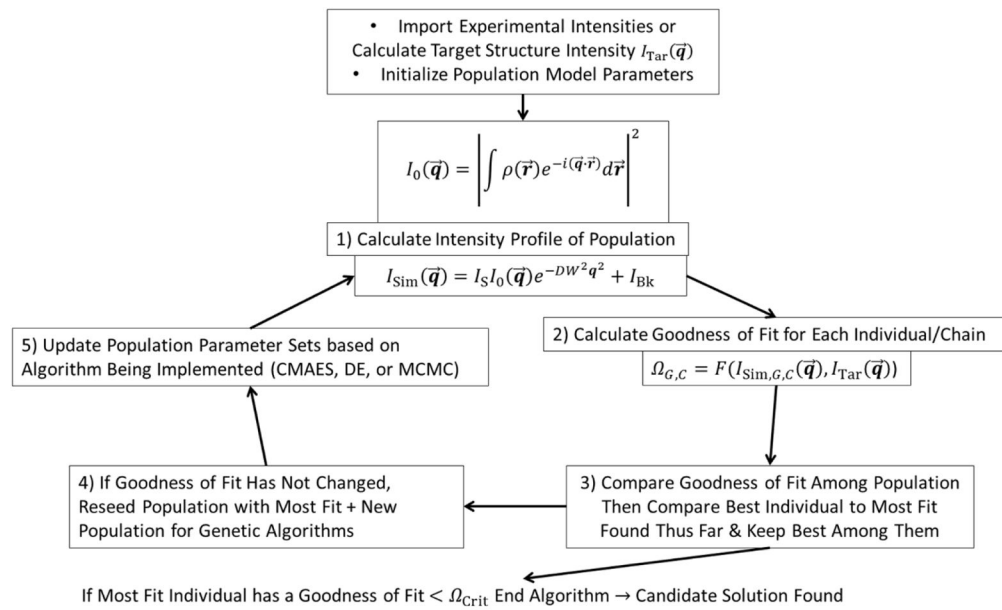
**Adam F. Hannon** is a postdoctoral research associate at NIST, where he conducts research in simulating the self-assembly of complex polymer systems (particularly block copolymer

blends) and incorporating advanced inverse algorithms and physics based models into X-ray characterization techniques. He obtained his doctor of science (ScD) degree in materials science and engineering from the Massachusetts Institute of Technology and BS degrees in both physics and polymer & fiber engineering from the Georgia Institute of Technology.

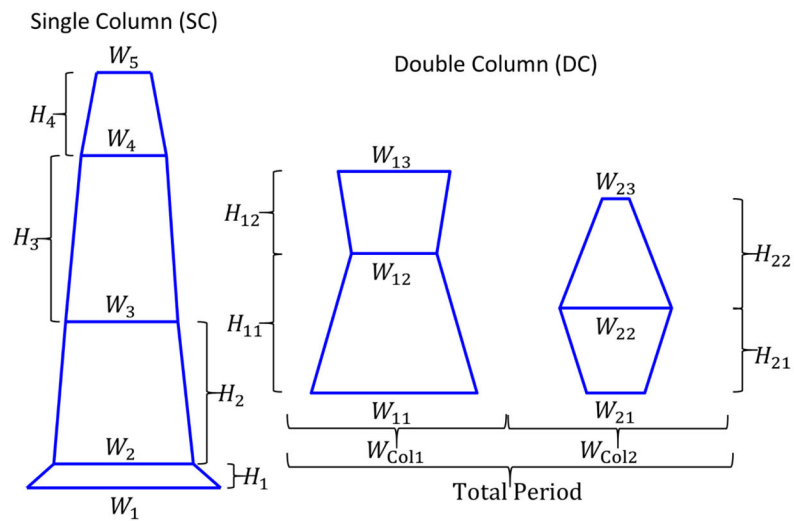
**Daniel F. Sunday** is a research scientist at NIST, where he researches X-ray characterization methods of nanostructures and thin films as well as the self-assembly of block copolymers. He obtained his PhD in chemical engineering from the University of Virginia and BS degree in chemical engineering from Carnegie Mellon.

**Donald Windover** is a physicist at NIST, where he researches the fundamental metrology limits of X-ray characterization methods when applied to complex systems. He has worked in X-ray standards development, instrumentation design, X-ray reflectometry, and wavelength and angle metrology. He obtained his PhD in Physics from Rensselaer Polytechnic Institute and his BS in Physics from the University of Maine.

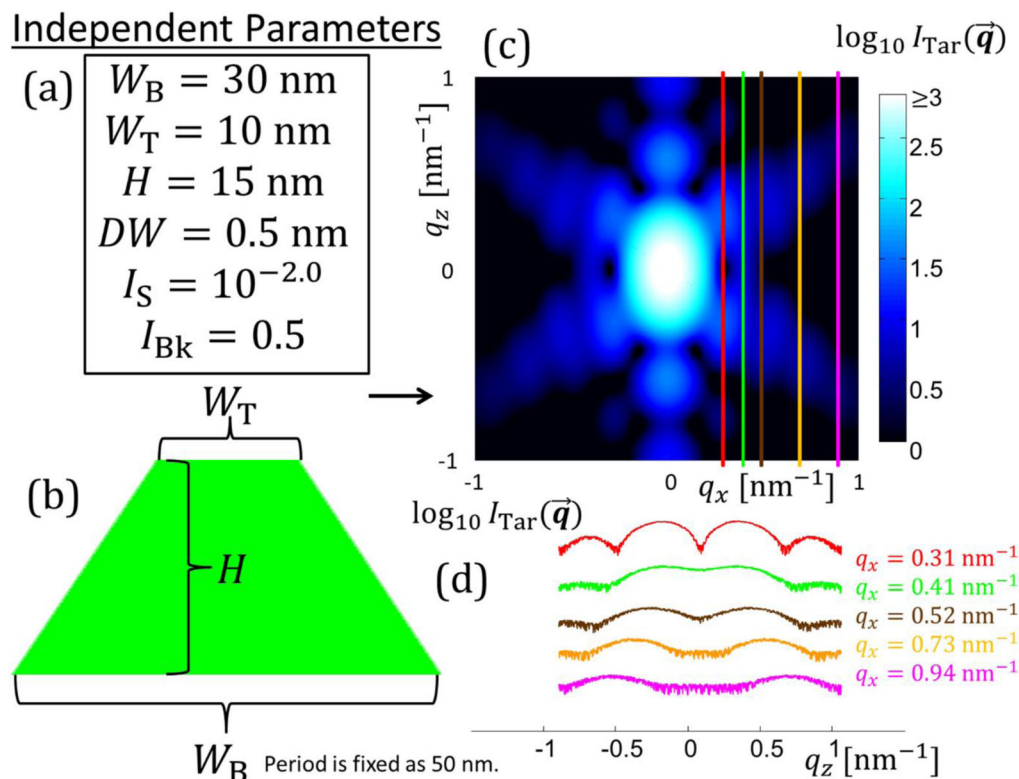
**R. Joseph Kline** is the leader of the dimensional metrology for nanomanufacturing project at NIST. He researches X-ray based dimensional metrology of nanostructures as well as X-ray structure measurements of soft matter systems. He obtained his PhD in materials science from Stanford University and BS and MS degrees in material science from North Carolina State. His publications number more than 70 articles plus four book chapters, and he has given more than 35 invited presentations. In 2012 he received the Presidential Early Career Award for Science and Engineering.



**Fig. 1.** Diagram outlining the major steps in our approach. This format is general for different objective functions and different parameter set updating algorithms (CMAES, DE, or MCMC).

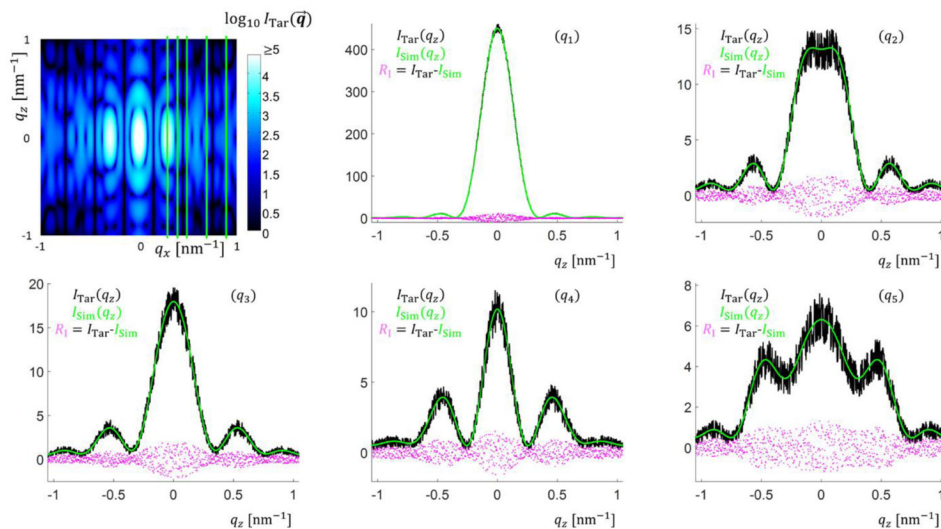


**Fig. 2.** Example trapezoid model based shape profiles used in calculating the *SLD* for the target intensities. The two main types of model parameter sets explored are those of a single column of periodic trapezoid features shown on the left (SC) and those of double columns of periodic features shown on the right (DC). Geometric model parameters are denoted in the figure.

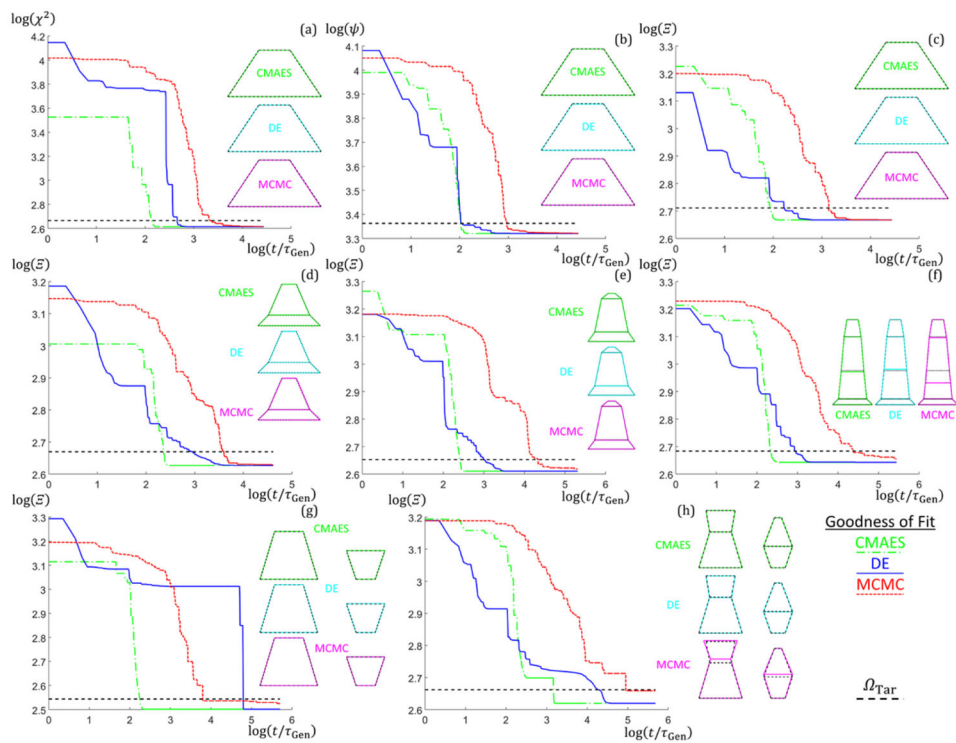
**Fig. 3.**

Schematic diagram of how known solution intensities were produced. (a) Set of independent parameters is defined for the real space structure with various geometric parameters and intensity scaling parameters. In this case three geometric parameters are required to define the bottom width  $W_B$ , top width  $W_T$ , and height  $H$  of a single symmetric trapezoid plus the three intensity scaling parameters  $DW$ ,  $I_S$ , and  $I_{Bk}$ . (b) Schematic diagram of the shape profile of the single trapezoid target structure. (c) Simulated intensity over the range of  $q_x$  and  $q_z$  values varying from  $\approx -1 \text{ nm}^{-1}$  to  $1 \text{ nm}^{-1}$ . Intensity is plotted on a log scale for better contrast. Lighter blue regions are higher intensity and darker blue to black regions are low intensities. The five colored lines are the five intensity  $q_z$  slices chosen for the inverse algorithm to use to solve for the structure. (d) Plots of the five intensity slices used in the inverse algorithm with the same coloring as from (c) with the respective  $q_x$  values colored appropriately. Scaling of the intensity values is arbitrary to get all five curves onto one plot without overlapping.

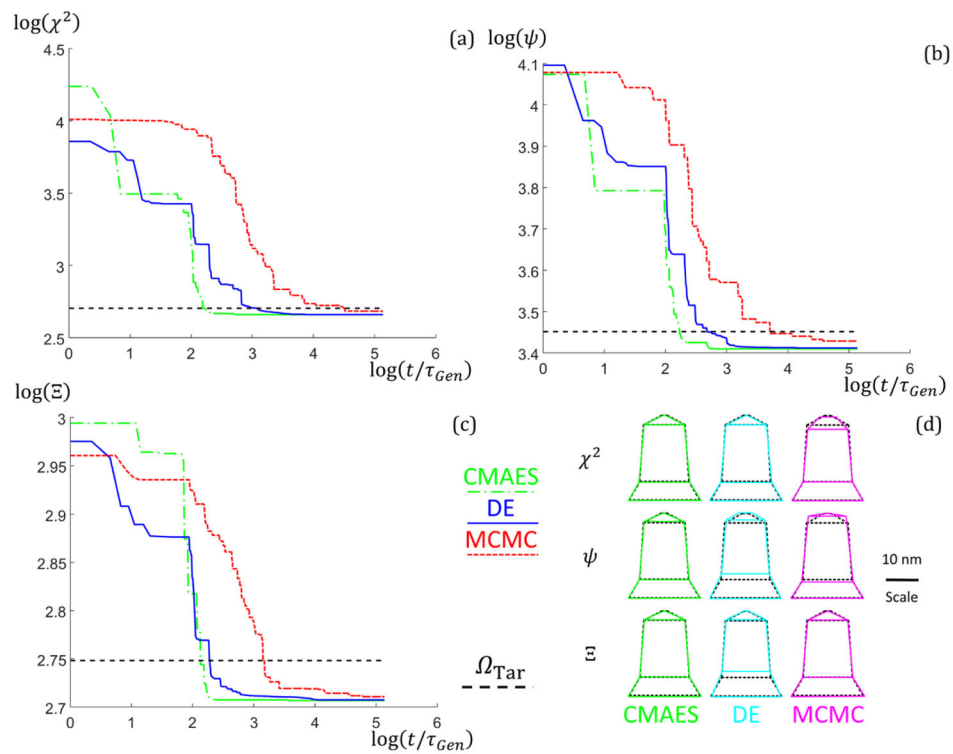




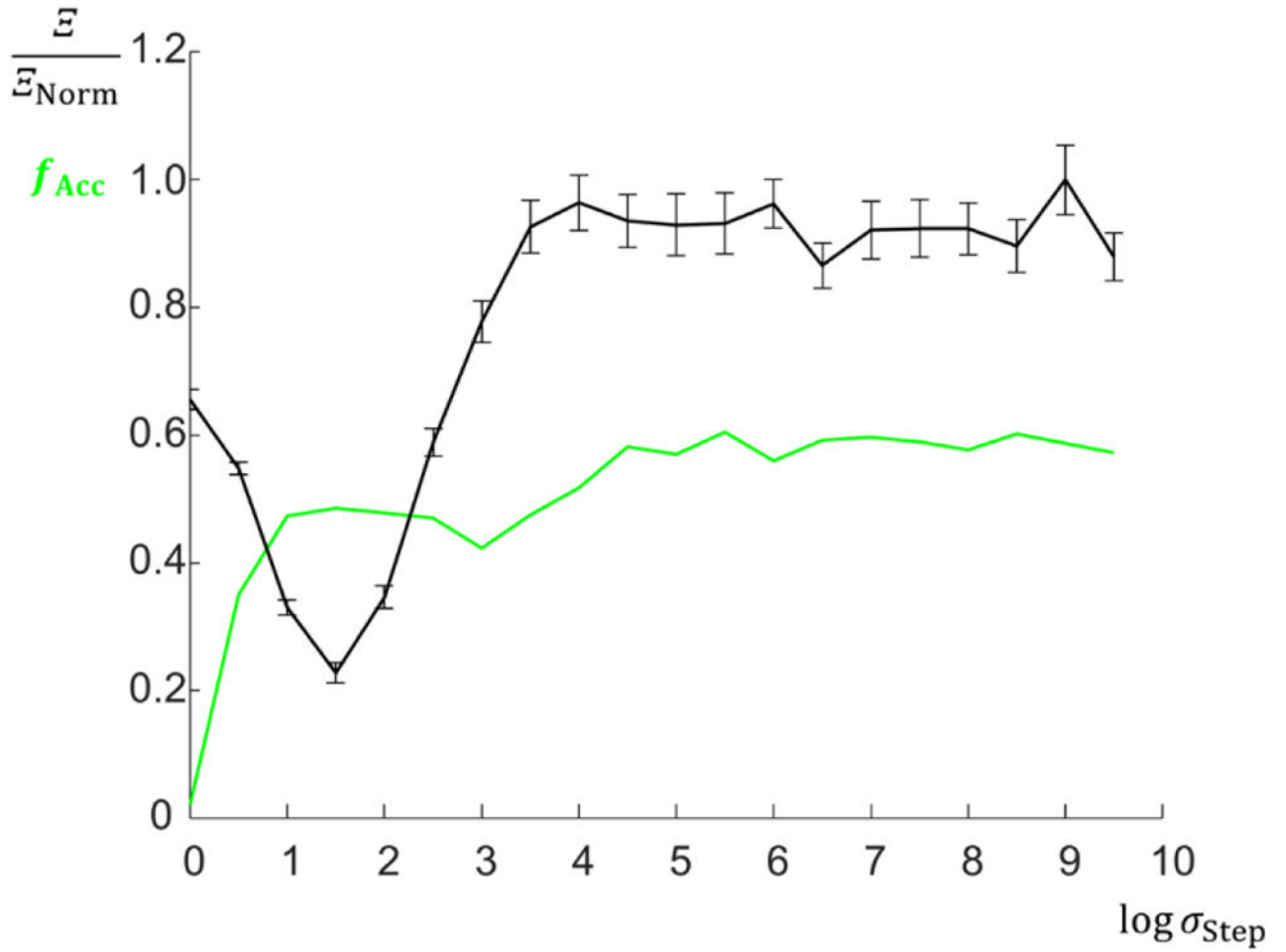
**Fig. 4.** Comparison of plots of the intensity fits found by the CMAES algorithm with  $\mathcal{E}$  function for the DC  $M=2$  target structure. Top left shows the entire simulated intensity profile with the green lines being where the  $q_z$  slices were taken. The rest of the figure shows the simulated data  $I_{\text{Tar}}$  with Poisson noise plotted in black, a fit curve  $I_{\text{Sim}}$  plotted in green, and the residuals  $R_I = I_{\text{Tar}} - I_{\text{Sim}}$  between the fit and simulated data with noise plotted in magenta ( $q_1$ ) Slice 1  $q_x = 0.31 \text{ nm}^{-1}$ . ( $q_2$ ) Slice 2  $q_x = 0.41 \text{ nm}^{-1}$ . ( $q_3$ ) Slice 3  $q_x = 0.52 \text{ nm}^{-1}$ . ( $q_4$ ) Slice 4  $q_x = 0.73 \text{ nm}^{-1}$ . ( $q_5$ ) Slice 5  $q_x = 0.94 \text{ nm}^{-1}$ .



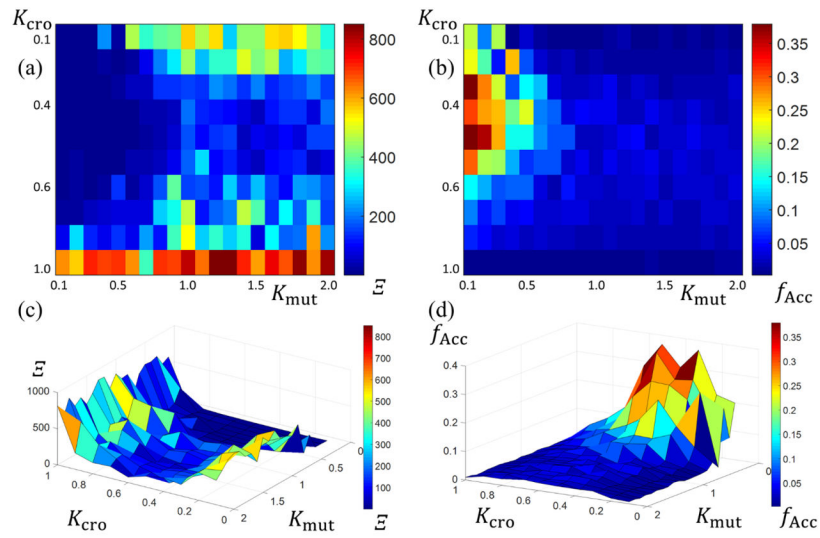
**Fig. 5.** Plots of the best (lowest goodness of fit found from up to 10 runs) run goodness of fit values versus normalized time for different number of parameter structures using the three different algorithms (CMAES in green dash dotted lines, DE in blue solid lines, and MCMC in red dotted lines). Plots are log scale.  $\Omega_{Tar}$  thresholds are shown as a black dashed line. Target shape profiles (black dashed lines) and solution shape profiles (CMAES in green, DE in cyan, and MCMC in magenta) are shown to the right of each corresponding  $\Omega$  versus  $t/\tau_{Gen}$  plot. (a–c)  $N_{Param} = 6$  SC  $M = 1$ . (a)  $\Omega \rightarrow \chi^2$  (b)  $\Omega \rightarrow \psi$  (c)  $\Omega \rightarrow \mathcal{E}$  (d–h)  $\Omega \rightarrow \mathcal{E}$  (d)  $N_{Param} = 8$  SC  $M = 2$ . (e)  $N_{Param} = 10$  SC  $M = 3$ . (f)  $N_{Param} = 12$  SC  $M = 4$ . (g)  $N_{Param} = 9$  DC  $M = 1$ . (h)  $N_{Param} = 13$  DC  $M = 2$ .

**Fig. 6.**

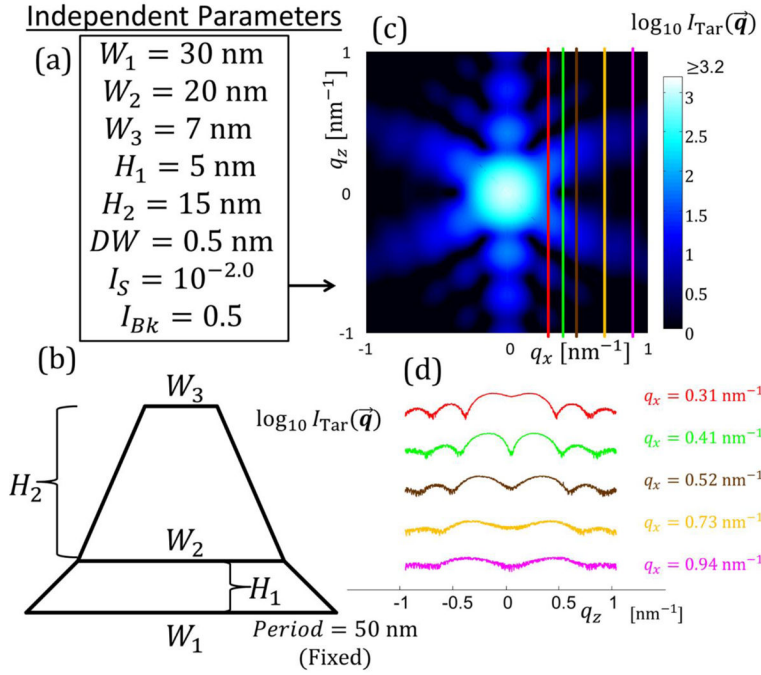
(a–c) Goodness of fit versus  $t/\tau_{Gen}$  results with  $M=3$  for the experimental data structure between the three algorithms (CMAES in green dash dotted lines, DE in blue solid lines, and MCMC in red dotted lines). Data is plotted on a log base 10 scale. Plots are best (lowest goodness of fit) data sets from 10 runs. (a)  $\chi^2$  objective function was used. (b)  $\psi$  objective function was used. (c)  $\mathcal{E}$  objective function was used. (d) Geometric structure shape profiles for the best parameter set solutions found using the different objective functions (rows top to bottom are  $\chi^2$ ,  $\psi$ , and  $\mathcal{E}$ ) and algorithms (columns left to right are green for the CMAES, cyan for DE, and magenta for MCMC) for the  $M=3$  experimental data structure. Scale is inset on the right of the figure. The overlaid dotted black line structure is the solution found from the previous MCMC study for comparison.

**Fig. 7.**

Plot of the acceptance fraction  $f_{\text{Acc}}$  (green) and average minimum goodness of fit value converged towards with standard uncertainty (black) for several MCMC algorithm runs at different  $\sigma_{\text{Step}}$  sizes. Uncertainty bars are standard uncertainty for 96 chains. The minimum goodness of fit value converged towards occurred around  $\sigma_{\text{Step}} \cong 32$ . Here the objective function  $\mathcal{E}$  was used and normalized by the number  $\mathcal{E}_{\text{Norm}}$  to get the data comparable with the acceptance fractions (i.e.  $\mathcal{E}_{\text{Norm}}$  converts the goodness of fit value to arbitrary scaled units; a value of  $\mathcal{E}_{\text{Norm}} = 3578$  was used). At large  $\sigma_{\text{Step}}$ , the algorithm accepts many parameter set moves and thus works poorly. Conversely, at small  $\sigma_{\text{Step}}$ , the parameter set moves result in almost no acceptance events and the algorithm also works poorly. Thus the medium value of  $\sigma_{\text{Step}} = 32$  where  $\mathcal{E}$  was minimized was used in further MCMC runs.



**Fig. 8.** Plots of the minimum goodness of fit values  $\mathcal{E}$  converged towards and the acceptance fraction  $f_{Acc}$  for several DE algorithm runs at different  $K_{mut}$  and  $K_{cro}$  values. (a) 2D plot of  $\mathcal{E}$  versus  $K_{mut}$  and  $K_{cro}$  with bluer colors being lower and redder colors being higher. (b) 2D plot of  $f_{Acc}$  versus  $K_{mut}$  and  $K_{cro}$  with bluer colors being lower and redder colors being higher. (c) 3D surface plot of data shown in (a). (d) 3D surface plot of data shown in (b).



**Fig. 9.** Figure analogous to Figure 3 in the main text for target structure with SC  $M=2$ . Period is fixed as 50 nm. (a) Independent parameters defined for the SC  $M=2$  structure. (b) Schematic diagram of the shape profile for the SC  $M=2$  structure. (c) The simulated target structure intensity profile in  $q_z$  versus  $q_x$  for a range of  $[-1:1] \text{ nm}^{-1}$ . Intensity is plotted on a log scale with scale shown. Colored lines are constant  $q_x$  cuts used for parameter optimization with the algorithms corresponding to the intensities shown in (d). (d) Log of target structure intensity slices used for fitting. Scale is arbitrary so the curves all fit on the same plot.

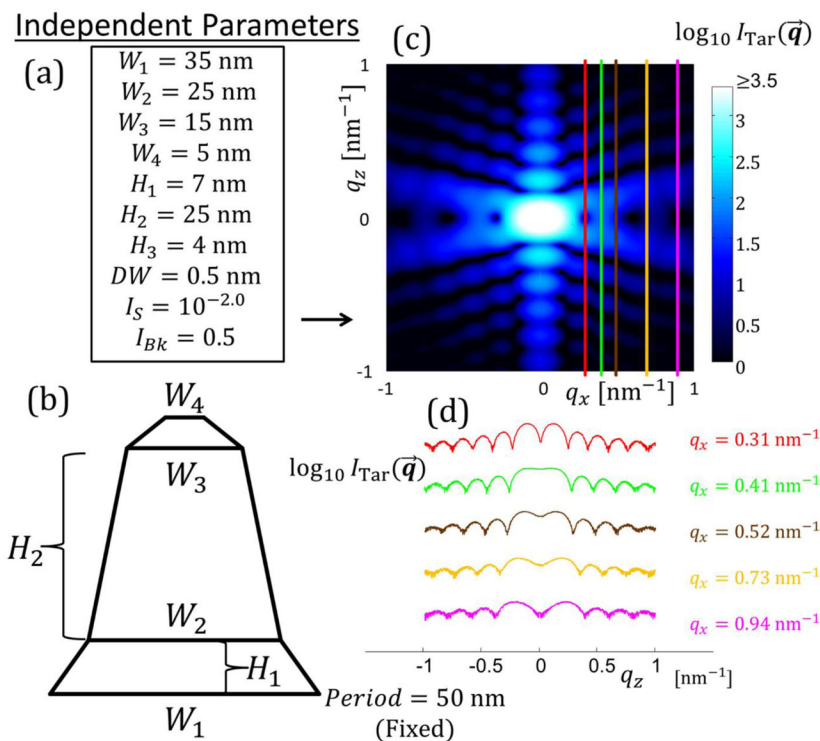
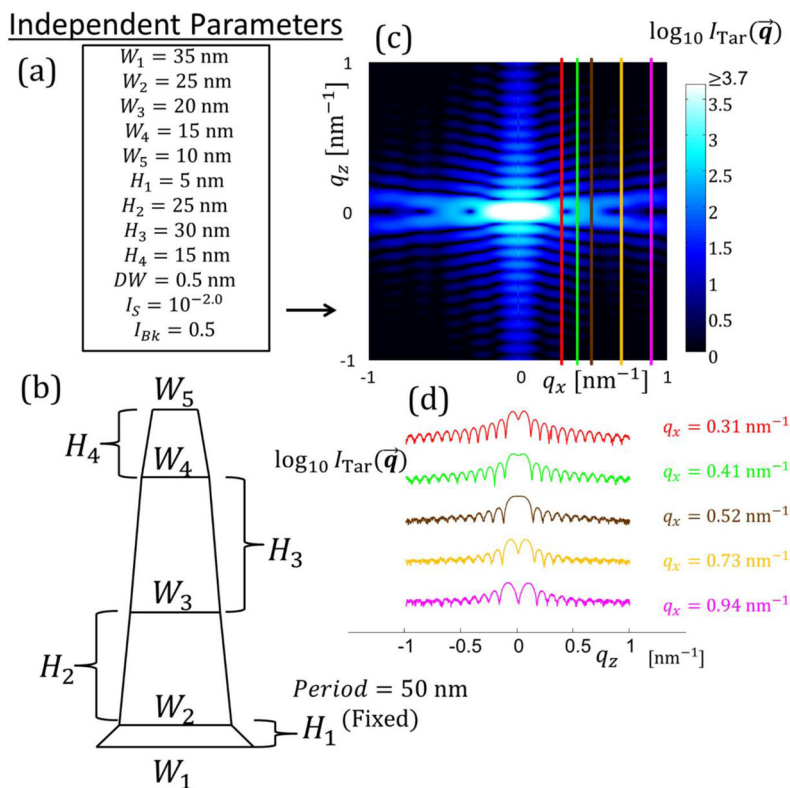
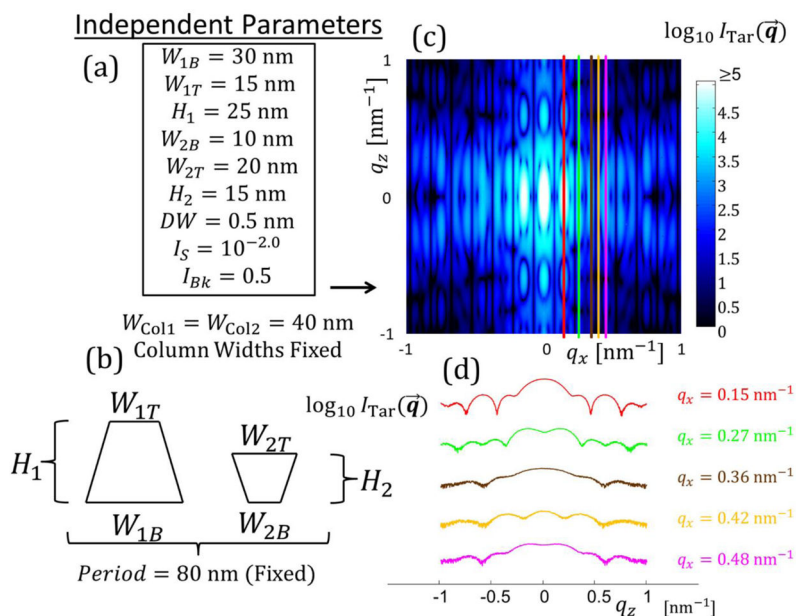
**Fig. 10.**

Figure analogous to Figure 3 in the main text for target structure with SC  $M=3$ . Period is fixed as 50 nm. (a) Independent parameters defined for the SC  $M=3$  structure. (b) Schematic diagram of the shape profile for the SC  $M=3$  structure. (c) The simulated target structure intensity profile in  $q_z$  versus  $q_x$  for a range of  $[-1:1] \text{ nm}^{-1}$ . Intensity is plotted on a log scale with scale shown. Colored lines are constant  $q_x$  cuts used for parameter optimization with the algorithms corresponding to the intensities shown in (d). (d) Log of target structure intensity slices used for fitting. Scale is arbitrary so the curves all fit on the same plot.



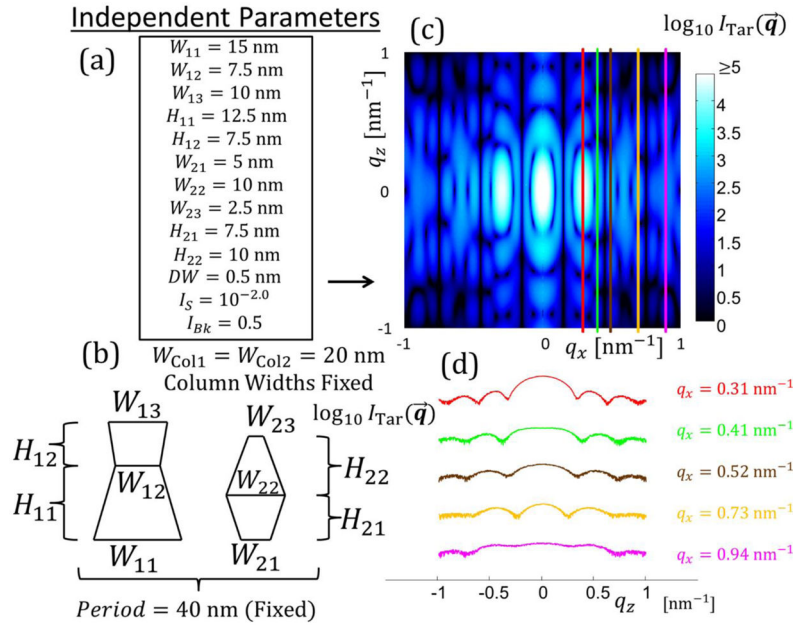
**Fig. 11.** Figure analogous to Figure 3 in the main text for target structure with SC  $M=4$ . Period is fixed as 50 nm. (a) Independent parameters defined for the SC  $M=4$  structure. (b) Schematic diagram of the shape profile for the SC  $M=4$  structure. (c) The simulated target structure intensity profile in  $q_z$  versus  $q_x$  for a range of  $[-1:1] \text{ nm}^{-1}$ . Intensity is plotted on a log scale with scale shown. Colored lines are constant  $q_x$  cuts used for parameter optimization with the algorithms corresponding to the intensities shown in (d). (d) Log of target structure intensity slices used for fitting. Scale is arbitrary so the curves all fit on the same plot.





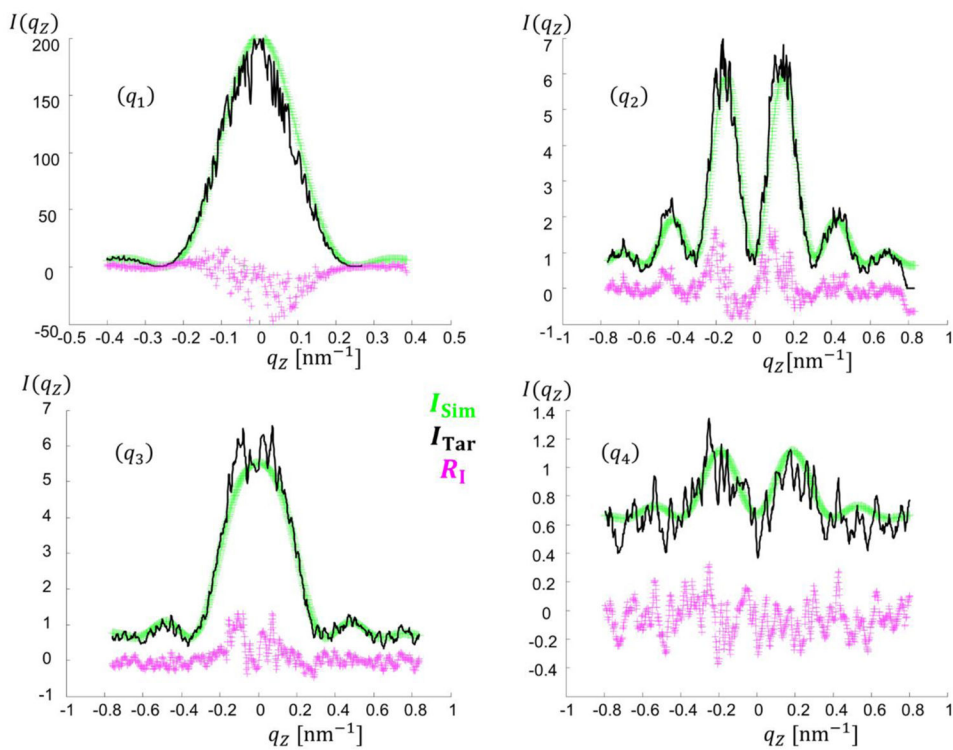
**Fig. 12.**

Figure analogous to Figure 3 in the main text for target structure with DC  $M=1$ . Period is fixed as 80 nm. (a) Independent parameters defined for the DC  $M=1$  structure. (b) Schematic diagram of the shape profile for the DC  $M=1$  structure. (c) The simulated target structure intensity profile in  $q_z$  versus  $q_x$  for a range of  $[-1:1] \text{ nm}^{-1}$ . Intensity is plotted on a log scale with scale shown. Colored lines are constant  $q_x$  cuts used for parameter optimization with the algorithms corresponding to the intensities shown in (d). (d) Log of target structure intensity slices used for fitting. Scale is arbitrary so the curves all fit on the same plot.

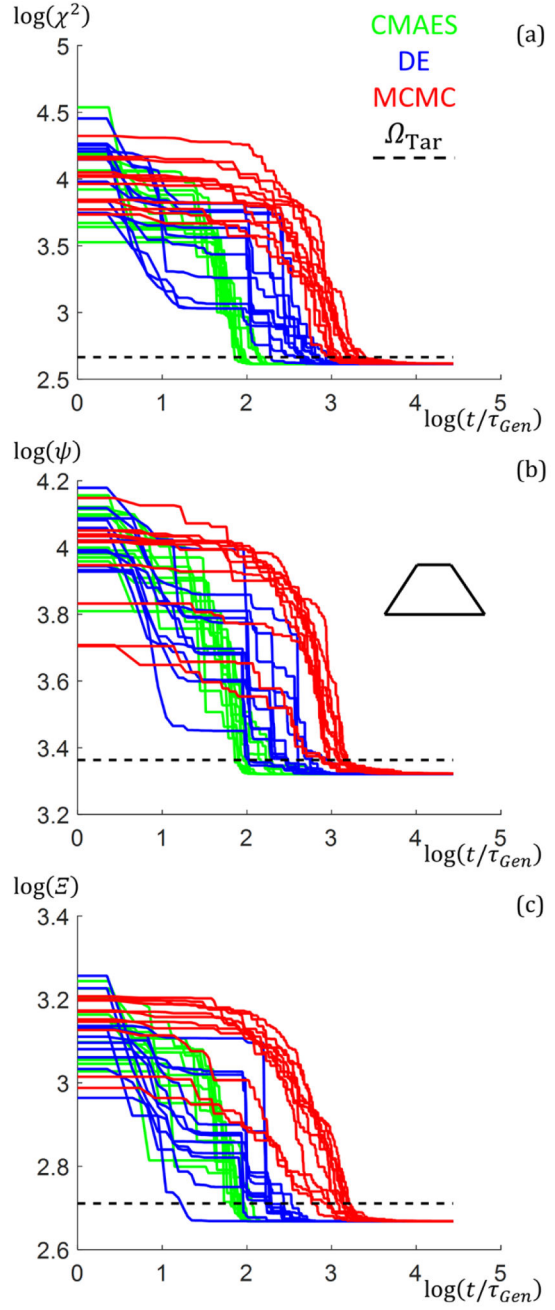


**Fig. 13.**

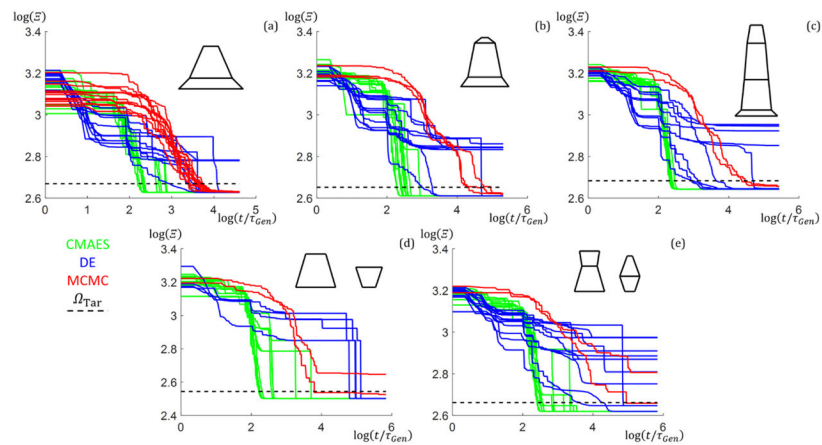
Figure analogous to Figure 3 in the main text for target structure with DC  $M=2$ . Period is fixed as 40 nm. (a) Independent parameters defined for the DC  $M=2$  structure. (b) Schematic diagram of the shape profile for the DC  $M=2$  structure. (c) The simulated target structure intensity profile in  $q_z$  versus  $q_x$  for a range of  $[-1:1] \text{ nm}^{-1}$ . Intensity is plotted on a log scale with scale shown. Colored lines are constant  $q_x$  cuts used for parameter optimization with the algorithms corresponding to the intensities shown in (d). (d) Log of target structure intensity slices used for fitting. Scale is arbitrary so the curves all fit on the same plot.



**Fig. 14.** Comparison of plots of the intensity fits found by the CMAES algorithm with  $\mathcal{E}$  function for the experimental target structure data. The figure shows the experimental data  $I_{\text{Tar}}$  plotted in black, a fit curve  $I_{\text{Sim}}$  plotted in green, and the residuals  $R_1 = I_{\text{Tar}} - I_{\text{Sim}}$  between the fit and experimental data plotted in magenta. ( $q_1$ ) Slice 1  $q_x = 0.223 \text{ nm}^{-1}$ . ( $q_2$ ) Slice 2  $q_x = 0.451 \text{ nm}^{-1}$ . ( $q_3$ ) Slice 3  $q_x = 0.677 \text{ nm}^{-1}$ . ( $q_4$ ) Slice 4  $q_x = 0.905 \text{ nm}^{-1}$ .

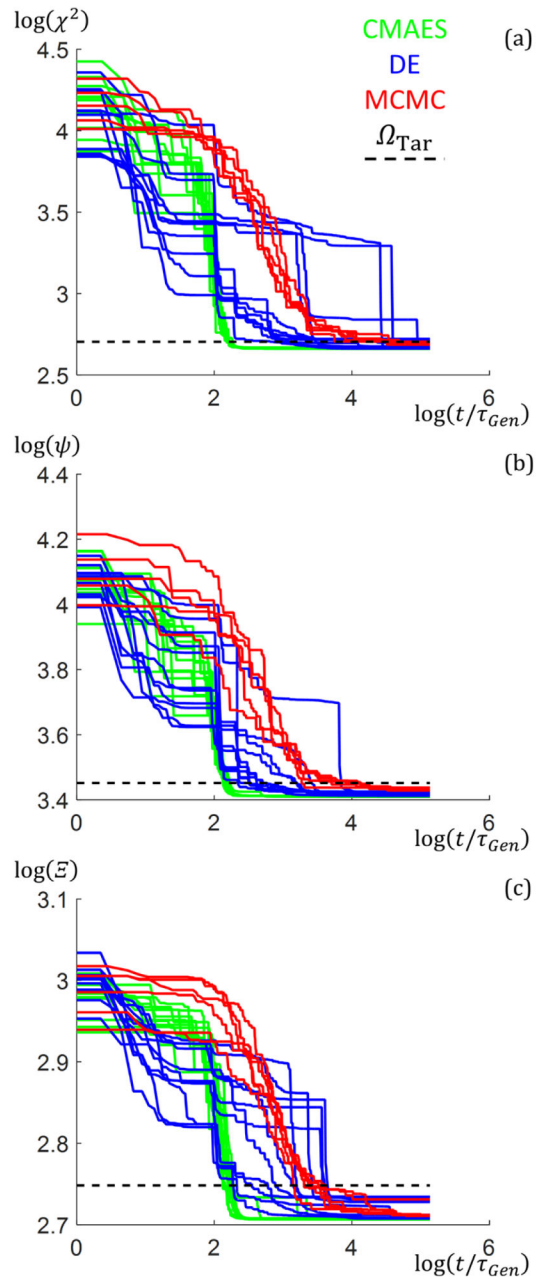


**Fig. 15.** Plots of the goodness of fit values for all three algorithms versus the normalized time  $t/\tau_{Gen}$  for the SC  $M=1$  structure. Plots are on log base 10 scale. Inset is a key for which algorithm the colored curves correspond and a schematic of the target structure. The plots have black dashed lines showing the  $\Omega_{Tar}$  values. Plots have all  $N_{Runs}$  for the three algorithms of the goodness of fit values versus  $t/\tau_{Gen}$  superimposed with different shades of the algorithm distinguishing colors for better clarity. Note there was one case for the DE in (c) that reached  $\Omega_{Tar}$  faster than the CMAES, but the final converged value of  $\mathcal{E}$  was slightly higher. (a)  $\Omega = \chi^2$ . (b)  $\Omega = \psi$ . (c)  $\Omega = \mathcal{E}$ .



**Fig. 16.**

Plots of the goodness of fit values for all three algorithms with  $\Omega = \mathcal{E}$  versus the normalized time  $t/\tau_{Gen}$  for the multiple trapezoid target structures. Plots are on log base 10 scale. Inset is a key for which algorithm the colored curves correspond and a schematic of the target structure is inset in each corresponding plot. The plots have black dashed lines showing the  $\Omega_{Tar}$  values. Plots have all  $N_{Runs}$  for the three algorithms of the goodness of fit values versus  $t/\tau_{Gen}$  superimposed. Note that fewer  $N_{Runs}$  were performed for the higher  $N_{Param}$  structures for the MCMC and DE based on just doing enough to find examples that converged within  $\Omega_{Tar}$  as those runs started to take much longer than  $t/\tau_{Gen} = 10^5$  to converge on average (corresponding to a few weeks of simulation time). (a) SC  $M = 2$ . (b) SC  $M = 3$ . (c) SC  $M = 4$ . (d) DC  $M = 1$ . (e) DC  $M = 2$ .



**Fig. 17.**

Plots of the goodness of fit values for all three algorithms versus the normalized time  $t/\tau_{Gen}$  for the  $M=3$  experimental data set. Plots are on log base 10 scale. Inset is a key for which algorithm the colored curves correspond. The black dashed lines indicates the  $\Omega_{Tar}$  values. Plots have all  $N_{Runs}$  for the three algorithms of the goodness of fit values versus  $t/\tau_{Gen}$  superimposed. (a)  $\Omega = \chi^2$ . (b)  $\Omega = \psi$ . (c)  $\Omega = \Xi$ .

**Table 1**

Comparison of the normalized time to converge  $t_{\text{Conv}}/\tau_{\text{Gen}}$  within the given  $\mathcal{Q}_{\text{far}}$  value and the best  $\mathcal{Q}_{\text{Best}}$  saturated goodness of fit values for the various algorithm and objective function combinations. Values of  $t_{\text{Conv}}/\tau_{\text{Gen}}$  are reported with standard uncertainty and averaged from all runs.  $\mathcal{Q}_{\text{far}}$  and  $\mathcal{Q}_{\text{Conv}}$  are listed for each objective function on the right for reference.

	CMAES	DE	MCMC	$\mathcal{Q}_{\text{far}}$	$\mathcal{Q}_{\text{Conv}}$
$\chi^2$	$t_{\text{Conv}}/\tau_{\text{Gen}}$	$230 \pm 20$	$650 \pm 70$	463.5	421.3
	$\mathcal{Q}_{\text{Best}}$	412.1	412.9		
$\Psi$	$t_{\text{Conv}}/\tau_{\text{Gen}}$	$160 \pm 30$	$510 \pm 40$	2307	2097
	$\mathcal{Q}_{\text{Best}}$	2097	2098		
$\mathcal{E}$	$t_{\text{Conv}}/\tau_{\text{Gen}}$	$80 \pm 10$	$520 \pm 40$	514.0	467.2
	$\mathcal{Q}_{\text{Best}}$	465.6	465.7		

**Table 2**

Comparison of various parameters found for symmetric periodic silicon grating line structure found using experimental data. CMAES with  $\mathcal{E}$  was used (normalized by  $N_q - 1$  here). The parameters presented are the  $\mathcal{Q}_{\text{Best}}$  values. For the cases of  $M = 1, 2, 4,$  and  $5$  only the total height  $H_{\text{Tot}}$  and full width at half max height  $W_{\text{FWHM}}$  were compared since the individual trapezoid components making up the structure in those cases are not comparable.

Parameter	$M = 1$	$M = 2$	$M = 3$	$M = 4$	$M = 5$	Previous Study $M = 3$
$H_{\text{Tot}}$ [nm]	25.66	26.65	26.55	26.97	26.73	26.24
$W_{\text{FWHM}}$ [nm]	13.55	13.85	13.87	13.92	13.97	14.19
$DW$ [nm]	1.51	1.29	1.20	1.20	1.25	1.26
$I_{\text{Exp}}$	-2.33	-2.34	-2.36	-2.36	-2.33	-2.22
$I_{\text{Bk}}$	0.76	0.65	0.64	0.64	0.65	0.49
$\mathcal{Q}_{\text{Best}} \rightarrow \mathcal{E}$	0.236	0.199	0.186	0.185	0.187	0.195



**Table 3**

The relative fractional percentage each slice contributed to the total goodness of fit for each  $\Omega$  in its own basis using the CMAES algorithm. Values are calculated using the objective function noted at the top of table.

	$\chi^2$	$\psi$	$\mathcal{E}$
% Slice $q_j$ Contributes to Total $\Omega$			
$q_1$ ( $q_x=0.223 \text{ nm}^{-1}$ )	99.5 %	91.0 %	48.3 %
$q_2$ ( $q_x=0.451 \text{ nm}^{-1}$ )	0.3 %	4.5 %	42.2 %
$q_3$ ( $q_x=0.677 \text{ nm}^{-1}$ )	0.1 %	2.9 %	4.5 %
$q_4$ ( $q_x=0.905 \text{ nm}^{-1}$ )	0.1 %	1.6 %	5.0 %

**Table 4**

Target parameters with bounds used for SC  $M=1$  trapezoid target structure. Units for  $W_f$ ,  $H_f$ , and  $DW$  are all in nm.

SC $M=1$	$W_B$	$W_T$	$H$	$DW$	$I_{\text{exp}}$	$I_{\text{blk}}$
Parameter Value	30	10	15	0.5	-2	0.5
Lower Bound	1	1	1	0	-5	0.01
Upper Bound	40	40	40	2	0	1

**Table 5**

Target parameters with bounds used for SC  $M=2$  trapezoids target structure. Units for  $W_j$ ,  $H_j$ , and  $DW$  are all in nm.

SC $M=2$	$W_1$	$W_2$	$W_3$	$H_1$	$H_2$	$DW$	$I_{Exp}$	$I_{Bk}$
Parameter Value	30	20	7	5	15	0.5	-2	0.5
Lower Bound	1	1	1	1	1	0	-5	0.01
Upper Bound	40	40	40	40	40	2	0	1

**Table 6**

Target parameters with bounds used for SC  $M=3$  trapezoids target structure. Units for  $W_j$ ,  $H_j$ , and  $DW$  are all in nm.

SC $M=3$	$W_1$	$W_2$	$W_3$	$W_4$	$H_1$	$H_2$	$H_3$	$DW$	$I_{\text{exp}}$	$I_{\text{BK}}$
Parameter Value	35	25	15	5	7	25	4	0.5	-2	0.5
Lower Bound	1	1	1	1	1	1	1	0	-5	0.01
Upper Bound	40	40	40	40	40	40	40	2	0	1

**Table 7**

Target parameters with bounds used for SC  $M = 4$  trapezoids target structure. Units for  $W_j$ ,  $H_j$ , and  $DW$  are all in nm.

SC $M = 4$	$W_1$	$W_2$	$W_3$	$W_4$	$W_5$	$H_1$	$H_2$	$H_3$	$H_4$	$DW$	$I_{\text{Exp}}$	$I_{\text{Bk}}$
Parameter Value	35	25	20	15	10	5	25	30	15	0.5	-2	0.5
Lower Bound	1	1	1	1	1	1	1	1	1	0	-5	0.01
Upper Bound	40	40	40	40	40	40	40	40	40	2	0	1

**Table 8**

Parameters with bounds used for DC  $M=1$  trapezoid per column target structure. Units for  $W_j$ ,  $H_j$ , and  $DW$  are all in nm.

DC $M=1$	$W_{1B}$	$W_{1T}$	$H_1$	$W_{2B}$	$W_{2T}$	$H_2$	$DW$	$I_{Exp}$	$I_{Bk}$
Parameter Value	30	15	25	10	20	15	0.5	-2	0.5
Lower Bound	1	1	1	1	1	1	0	-5	0.01
Upper Bound	20	20	40	20	20	40	2	0	1

**Table 9**

Parameters with bounds used for DC  $M = 2$  trapezoids per column target structure. Units for  $W_j$ ,  $H_j$ , and  $DW$  are all in nm.

DC $M = 2$	$W_{11}$	$W_{12}$	$W_{13}$	$H_{11}$	$H_{12}$	$W_{21}$	$W_{22}$	$W_{23}$	$H_{21}$	$H_{22}$	$DW$	$I_{\text{Exp}}$	$I_{\text{Bk}}$
Parameter Value	15	7.5	10	12.5	7.5	5	10	2.5	7.5	10	0.5	-2	0.5
Lower Bound	1	1	1	1	1	1	1	1	1	1	0	-5	0.01
Upper Bound	40	40	40	40	40	40	40	40	40	40	2	0	1

**Table 10**

Parameters with bounds used for SC  $M=3$  trapezoids experimental data structure. Units for  $W_j$ ,  $H_j$ , and  $DW$  are all in nm.

Experiment	$W_1$	$W_2$	$W_3$	$W_4$	$H_1$	$H_2$	$H_3$	$DW$	$I_{\text{exp}}$	$I_{\text{Bk}}$
Parameter Value	21.83	15.13	12.89	1.92	5.74	17.57	2.93	1.26	-2.22	0.49
Lower Bound	1	1	1	1	0.01	0.01	0.01	0	-5	0.01
Upper Bound	40	40	40	40	40	40	40	2	0	10



Table 11

Comparison of various parameters found for symmetric periodic silicon grating line structure found using experimental data. CMAES with  $\mathcal{E}$  was used. The parameters presented are the best fit values (corresponding to the lowest goodness of fit values). For the cases of  $M = 1, 2, 4$ , and 5 only the total height  $H_{\text{Tot}}$  and full width at half max height  $W_{\text{FWHM}}$  were compared since the individual trapezoid components making up the structure in those cases are not comparable (NC).

Parameter	$M = 1$	$M = 2$	$M = 3$	$M = 4$	$M = 5$	Previous Study $M = 3$
$H_{\text{Tot}}$ [nm]	25.66	26.65	26.55	26.97	26.73	26.24
$H_1$ [nm]	NC	NC	7.80	NC	NC	5.74
$H_2$ [nm]	NC	NC	16.15	NC	NC	17.57
$H_3$ [nm]	NC	NC	2.59	NC	NC	2.93
$W_{\text{FWHM}}$ [nm]	13.55	13.85	13.87	13.92	13.97	14.19
$W_1$ [nm]	NC	NC	21.36	NC	NC	21.83
$W_2$ [nm]	NC	NC	14.67	NC	NC	15.13
$W_3$ [nm]	NC	NC	12.30	NC	NC	12.89
$W_4$ [nm]	NC	NC	1.84	NC	NC	1.92
$DW$ [nm]	1.51	1.29	1.20	1.20	1.25	1.26
$f_{\text{Exp}}$	-2.33	-2.34	-2.36	-2.36	-2.33	-2.22
$f_{\text{BK}}$	0.76	0.65	0.64	0.64	0.65	0.49
$\mathcal{E}$	0.236	0.199	0.186	0.185	0.187	0.195