Electron diffraction simulation

Time complexity estimates

David Waterman (CCP4/STFC) March 2018

# Background

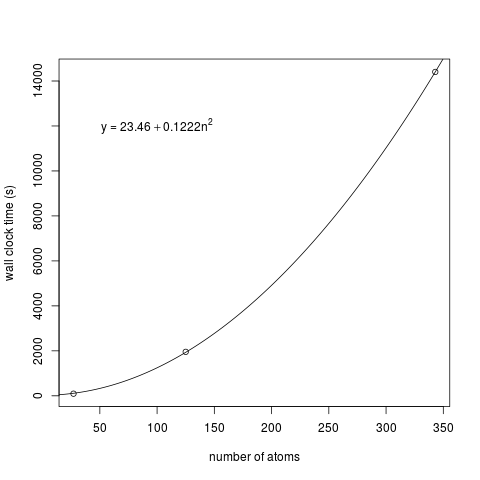
The development of a simulation for electron diffraction proposes to take as a start point the existing simulator nearBragg (<http://bl831.als.lbl.gov/~jamesh/nearBragg/>). This program was written by James Holton for minimal-assumption calculation of total X-ray scattering from samples that may not necessarily be crystalline. We propose to adapt the simple algorithm to the case of electron scattering, with a particular focus on including modelling of absorption and multiple scattering for nanocrystals. The inclusion of multiple scattering incurs the greatest computational cost. Whereas a typical run of the existing nearBragg has an *O*(N) time complexity, where N is the total number of atoms in the sample, to model multiple diffraction each atom will be considered an additional source location, implying that unoptimised single core execution durations will scale with *O*(N2).

# Baseline single image simulation

To get a rough indication of the relationship between the number of atoms N and the execution time we used the unmodified single processor nearBragg code and simply supplied N source positions. This was tested on a Linux desktop computer with an Intel Core i7 CPU 880 @ 3.07GHz processor, using tools downloaded from the nearBragg website to simulate diffraction images from a silicon nanoparticle of various sizes. In the following simple script, LATTICESIZE was set to 1, 2 or 3, producing the results below.

#! /bin/bash  
LATTICESIZE=1  
echo "$LATTICESIZE 5.43071" | makelattice.awk > silicon\_nanoparticle.txt  
rotate.awk -v phix=10 -v phiy=20 -v phiz=30 **\**  
 silicon\_nanoparticle.txt > rotated\_xtal.txt  
awk '{print $0,5}' rotated\_xtal.txt | Bscatter.awk > atoms.txt  
NATOMS=**$(**cat atoms.txt | wc -l**)**  
echo $NATOMS  
time nearBragg -file ./atoms.txt -lambda 1.0 -dispersion 0 -source\_depth 1 **\**  
 -depthsteps $NATOMS -intfile intimage\_round.img -floatfile floatimage\_ideal.bin

|  |  |  |
| --- | --- | --- |
| LATTICESIZE | NATOMS | Runtime (s) |
| 1 | 27 | 94 |
| 2 | 125 | 1955 |
| 3 | 343 | 14404 |



We then use the relationship between number of atoms and overall run time to extrapolate to interesting cases with a larger number of atoms, and datasets with multiple images.

# Extension to realistic sample sizes

The N2 dependence means that it is imperative to choose the smallest possible example case that demonstrates the features we are interested in. The minimum number of unit cells in the simulated sample is set by the limit at which diffraction intensity between Bragg spots can no longer be ignored. When diffraction fringes are significant the accuracy of peak integration methods suffer. This has been discussed, for example in [1]. The minimum number of unit cells at which a sample still reasonably approximates a continuous 3D crystal for the purposes of electron diffraction has yet to be determined. However, the range 10-20 unit cells in each direction may be a fair initial estimate (T. Gruene, personal communication).

We have chosen two examples to represent larger simulations of the sort required to demonstrate the effects we are interested in. One is a larger version of the silicon nanoparticle example used to determine the baseline result above. The second is the structure of a SAPO-34 zeotype that has recently been studied experimentally by electron diffraction [2]. We anticipate that the results of simulations will be easier to interpret when more unit cells are included in the model for the crystal. A large number of unit cells will minimise fringes between diffraction spots and will allow standard crystallographic data processing software to be used, which ignore intensity outside of Bragg peaks. Improvements to simulation duration obtained by optimisation of the code will be best utilised by allowing access to larger sample sizes and more valuable cases (such as SAPO-34 rather than the Si nanoparticle), rather than reducing the overall computational cost.

|  |  |  |
| --- | --- | --- |
| Example | Si nanoparticle | SAPO-34 |
| No. unit cells | 20×20×20 | 10×10×10 |
| No. atoms | 8000 | 108'000 |
| Full still image simulation, 1024×1024 pixels (CPU hours) | 2200 | 400'000 |

# Regions of interest

The table above indicates that generation of datasets consisting of many full still images for a standard detector size of 1024×1024 pixels is impracticable. However, we can make use of a trivial optimisation afforded to us by the crystalline nature of our simulated samples. Significant intensity is allowed only in the directions close to the Bragg spot peaks. We need therefore only simulate a small fraction of the total number of pixels. This fraction is sample-dependent. Samples with larger unit cells produce more reflections within a particular scattering angle, while larger number of unit cells in the crystal reduce the Bragg spot size to impinge upon fewer crystals.

To estimate the ROI size for each case we make some simple assumptions. We assume that a Bragg spot on a single still image will be contained within a 20×20 pixel box. We also assume that each reflection will be present for 1° of a rotation dataset. We choose to divide the full sphere of 360° into 1800×0.2° rotation images. Each rotation image will be formed as the sum of 20 still images related by 0.01° rotations. That implies that a full sphere rotation dataset would consist of 36'000 separate still image simulations, and each reflection will simulated within a 3D ROI *shoebox* of 20×20×(5×20)=40'000 pixels

|  |  |  |
| --- | --- | --- |
| Example | Si nanoparticle | SAPO-34 |
| No. reflections in 360° | 690 to 1 Å resolution | 448 to 2 Å resolution |
| No. pixels in ROIs in 360° | 690×40'000=27.6M | 448×40'000=17.92M |
| No. pixels in full images in 360° | 36'000×10242=37.75G | |
| Av. no. ROI pixels per image | 27.6M/36'000=767 | 17.92M/36'000=498 |
| ROI factor | 1368 | 2107 |
| ROI still image simulation, (CPU hours) | 1.61 | 190 |

The table shows how the expected simulation duration for single still images has decreased by more than three orders of magnitude by considering regions of interest in each image.

# Rotation datasets

In order to investigate the features of interest we need rotation datasets rather than individual still images. Rotation datasets will consist of rotation images, approximated by the sum of a sufficient number of still images, as discussed above. The accuracy of a rotation image will improve the more still sub-images it is divided into. As we are interested in the total intensity of Bragg spots rather than the shape of reflection profiles it is likely that a fairly modest number of sub-images will suffice, such as 20. The total angular width of rotation required to investigate improvements to scaling algorithms is somewhat case-dependent. A high symmetry crystal implies a smaller rotation dataset is required, whilst investigation of absorption correction is dependent on crystal shape and sets a lower limit of perhaps 90° of data.

|  |  |  |
| --- | --- | --- |
| Example | Si nanoparticle | SAPO-34 |
| ROI still image simulation, (CPU hours) | 1.61 | 190 |
| Simulation of 0.2° rotation image in 20 steps (CPU hours) | 32.2 | 3800 |
| Simulation of 1° dataset (CPU hours) | 161 | 19'000 |
| Simulation of 90° dataset (CPU hours) | 14490 | 1.7M |

This analysis indicates that basic modifications to the current code makes the simulation of 90° datasets for a Si nanoparticle crystal feasible with a sufficient number of independent parallel processes. The larger case of SAPO-34 remains difficult, but the current analysis is based on unoptimised code, with a single gross algorithmic rearrangement to consider regions of interest. It is expected that further savings will be identified during the conversion and optimisation of the nearBragg algorithm for this case.

# References

1. Williams, S.R.; Dilanian, R.A.; Quiney, H.M.; Martin, A.V. Analysis of Diffracted Intensities from Finite Protein Crystals with Incomplete Unit Cells. Crystals 2017, 7, 220.
2. Tinti, G., Frojdh, E., van Genderen, E., Gruene, T., Schmitt, B., de Winter, D. A., Weckhuysen, B. M. & Abrahams, J. P. Electron crystallography with the EIGER detector (2018). IUCrJ 5.