GPU-Based Defect Image Simulations using the Scattering Matrix Formalism

Saransh Singh[1], Marc De Graef[1]

[1] Dept. of Materials Science and Engineering, Carnegie Mellon Univ., Pittsburgh PA 15213, USA

Systems of first order differential equations frequently occur in disparate fields spanning all of the sciences. Examples include the population evolution of interacting species in an ecosystem, the concentrations of molecules in a chemical reaction, the production of goods, the price of supplies in economic processes and so on. These equations also describe the dynamical scattering of electrons in a periodic lattice potential, in which case they are known as the Darwin-Howie-Whelan (DHW) equations. The DHWs describe how the probability amplitude of an electron wave changes as a function of depth in a crystal. This change depends on the difraction geometry and on the mutual interaction strength of different diffracted beams [1]. The equations are often written as:

$$\frac{d\psi_{\mathbf{g}}}{dz} - 2\pi i s_{\mathbf{g}}\psi_{\mathbf{g}} = i\pi \sum_{\mathbf{g}'} \frac{e^{i\theta_{\mathbf{g}-\mathbf{g}'}}}{q_{\mathbf{g}-\mathbf{g}'}}\psi_{\mathbf{g}'},$$

where $\psi_{\mathbf{g}}$ is the amplitude of the beam diffracted by the lattice planes corresponding to the reciprocal lattice vector $\mathbf{g}$, $s_{\mathbf{g}}$ represents how well the Bragg condition is satisfied for the planes $\mathbf{g}$, and $q_{\mathbf{g}-\mathbf{g}'}$ describes the strength of interaction between the beams $\mathbf{g}$ and $\mathbf{g}'$. The equations are commonly written in the form of a matrix equation, given by:

$$\frac{d\mathbf{S}}{dz} = i\mathcal{A}\mathbf{S}, \qquad \mathcal{A}_{\mathbf{gg}'} = \begin{cases} 2\pi s_{\mathbf{g}} & \mathbf{g} = \mathbf{g}' \\ \frac{\pi}{q_{\mathbf{g}-\mathbf{g}'}} & \mathbf{g} \neq \mathbf{g}' \end{cases}$$

where $\mathbf{S}$ is the column vector of unknown diffracted amplitudes and $\mathcal{A}$ is the structure matrix; the diagonal entries of $\mathcal{A}$ contain the diffraction geometry via the excitation errors $s_{\mathbf{g}}$, and the off-diagonal elements represent the beam interactions. The equations have a well-known solution of the form $\mathbf{S}(z_0) = e^{i\mathcal{A}z_0}\mathbf{S}(0) = \mathcal{S}(z_0)\mathbf{S}(0)$, relating the amplitude at depth z_0 to the initial amplitude, which involves computation of the exponential of the structure matrix, resulting in the scattering matrix, $\mathcal{S}(z_0)$.

The presence of a lattice defect usually implies a broken translational crystal symmetry, resulting in atoms being displaced from their regular positions. Thus, a defect is usually quantified by a displacement field, $\mathbf{R}(\mathbf{r})$, that may vary continuously (e.g., misfitting inclusions and dislocations) or discontinuously (e.g., stacking faults) from one position to another. While there are other approaches to compute the amplitude changes of electrons traversing a periodic lattice (e.g., Bloch waves), the scattering matrix approach is particularly well suited for defect simulations as the defect contributions appear as phase factors in the off-diagonal terms only, while the overall structure of the equation remains intact [1]. The modified DHW equations, in the presence of one or more defects, take on the form:

$$\frac{d\psi_{\mathbf{g}}}{dz} - 2\pi i s_{\mathbf{g}}\psi_{\mathbf{g}} = i\pi \sum_{\mathbf{g}'} \frac{e^{i\theta_{\mathbf{g}-\mathbf{g}'}}}{q_{\mathbf{g}-\mathbf{g}'}} e^{-i\alpha_{\mathbf{g}-\mathbf{g}'}}\psi_{\mathbf{g}'},$$

where $\alpha_{\mathbf{g}} \equiv 2\pi\mathbf{g} \cdot \mathbf{R}(\mathbf{r})$, with $\mathbf{R}(\mathbf{r})$ the total displacement at position $\mathbf{r}$ due to the superposition of all individual defect displacement fields. It should be noted that the defect phase shifts can be written in matrix form as $\mathcal{P}_{\mathbf{gg}'}(\mathbf{r}) = e^{-i\alpha_{\mathbf{g}-\mathbf{g}'}}$, where the diagonal of the matrix contains 1s. Hence, the structure matrix in the presence of a defect is modified as $\mathcal{A}^d_{\mathbf{gg}'}(\mathbf{r}) = \mathcal{A}_{\mathbf{gg}'}\mathcal{P}_{\mathbf{gg}'}(\mathbf{r})$, where the matrix product is performed element by element instead of as a regular matrix product. This decomposition into a perfect crystal factor $\mathcal{A}$ and a defect factor $\mathcal{P}$ has important consequences for the efficient numerical implementation of defect image simulations.

Recently, Graphics Processing Units (GPUs) have become very popular for high performance computing applications, allowing for heterogeneous combinations of GPUs and CPUs to carry out computationally intensive tasks in a massively parallel way. The GPU has a very different memory structure and programming model from traditional serial programs; while the performance gain may be significant, the development effort and time involved in GPU programming is usually rather substantial. In this contribution, we present a new GPU-based algorithm for defect image simulations using the scattering matrix approach. In a typical CPU implementation, the intensity of each image pixel would be calculated in a sequential fashion; incorporation of multi-core techniques, for instance through the use of OpenMP constructs, can accelerate the computation, but even in the best case scenario, the speed-up factor is usually less than the number of available cores. The GPU-based algorithm subdivides the image into smaller blocks and performs the computation of all pixels in a sub-block simultaneously. The perfect crystal structure matrix $\mathcal{A}$ is stored in global GPU memory, whereas the series of defect phase factor matrices $\mathcal{P}_j$ for each of the slices j along the integration column is stored in the local work item memory. The exponential of the defect structure matrix $\mathcal{A}^d$ is then calculated for each slice using scaling and squaring combined with the optimized Taylor expansion, considering the first nine terms in the expansion [2]. Multiplication of the first slice scattering matrix $\mathcal{S}_1$ with the initial probability amplitudes at the entrance surface, followed by repeated matrix-vector multiplications then completes the integration along the column. Such an approach minimizes the amount of data transfer between the CPU and the GPU, which would otherwise dominate the overall computation time.

The scattering matrix formalism, as implemented on a GPU platform, provides a unified approach for defect image simulations for both the Transmission Electron Microscope (TEM) forescatter geometry and the Scanning Electron Microscope (SEM) backscatter geometry. The intensity of an image pixel in a TEM image is essentially given by the squared modulus of the relevant electron wave function component at the sample exit plane, i.e. $|\psi_{\mathbf{g}}(z_0)|^2$, where $\mathbf{g} = \mathbf{0}$ for bright field images, and $\mathbf{g} \neq 0$ for dark field mode. STEM-mode diffraction contrast simulations can also be implemented by properly integrating the intensities over those diffracted beams that reach the detector [3]. For SEM image modes, the detected signal is typically given by a depth-integrated intensity. The observed backscattered intensity in an SEM can be described as [4]:

$$\mathcal{P}(\mathbf{k}) = \sum_{i \in S} \frac{Z_i^2 DW_i}{z_0} \int_0^{z_0} \lambda(\mathbf{z}) |\Psi(\mathbf{r}_i)|^2 d\mathbf{z}$$

where, Z_i and DW_i are the atomic number and Debye-Waller factor of the atom at the ith lattice site, $\lambda(z)$ is a weight factor to account for the variation in the number of backscattered electrons as a function of depth and z_0 is the maximum depth from which the electrons are backscattered; both parameters are computed via Monte Carlo trajectory simulations. The GPU-based scattering matrix approach provides a fast and efficient algorithm for defect image computations for both TEM and SEM mode; examples of simulation results will be provided, as well as an analysis of the speed-up achieved by moving from CPU to GPU-based computations.

References

[1] Marc De Graef in "Introduction to Conventional Transmission Electron Microscopy", (2003, Cambridge University Press, New York) p. 310, p. 462.
[2] R.S. Pennington, F. Wang and C.T. Koch, Ultramicroscopy, **141**, 32–37 (2014).
[3] P.J. Phillips, M.J. Mills, and M. De Graef, Ultramicroscopy, **111**, 1483–1487 (2011).
[4] P.G. Callahan and M. De Graef, Microscopy and Microanalysis, **19**, 1255–1265 (2013).
[5] Research supported by the Air Force Office of Scientific Research, MURI contract # FA9550-12-1-0458.