On the role of the second-order derivative term in the calculation of convergent beam diffraction patterns

S.C. Hillier [a], E.T. Robertson [b], G.D. Reid [c], R.D. Haynes [a], M.D. Robertson [d],*

[a] Department of Mathematics and Statistics, Memorial University of Newfoundland, St. John's, NL, A1C 5S7, Canada
[b] Department of Mathematics and Statistics, Acadia University, Wolfville, NS, B4P 2R6, Canada
[c] Department of Physics and Astronomy, University of British Columbia, Vancouver BC, V6T 1Z1, Canada
[d] Department of Physics, Acadia University, Wolfville, NS, B4P 2R6, Canada

ARTICLE INFO

ABSTRACT

The simulation of (scanning) transmission electron microscopy images and diffraction patterns is most often performed using the forward-scattering approximation where the second-order derivative term in z is assumed to be small with respect to the first-order derivative term in the modified Schrödinger equation. This assumption is very good at high incident electron energies, but breaks down at low energies. In order to study the differences between first- and second-order methods, convergent beam electron diffraction patterns were simulated for silicon at the [111] zone-axis orientation at 20 keV and compared using electron intensity difference maps and integrated intensity profiles. The geometrical differences in the calculated diffraction patterns could be explained by an Ewald surface analysis. Furthermore, it was found that solutions based on the second-order derivative equation contained small amplitude oscillations that need to be resolved in order to ensure numerical integration stability. This required the use of very small integration steps resulting in significantly increased computation time compared to the first-order differential equation solution. Lastly, the efficiency of the numerical integration technique is discussed.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The quantification of transmission electron microscopy (TEM) and scanning transmission electron microscopy (STEM) images and diffraction patterns continues to be important to solving fundamental problems in material and biological sciences, and has directly followed advances in instrumentation and computing resources. The theory for electron propagation through materials in (S)TEMs is well established [1–5] and many methods for calculating image and diffraction patterns have been developed including Bloch-wave, multislice, matrix, iterative, direct integration, and real-space methods [6–8], as well as references contained therein. In addition, the potentials encountered by incident electrons is a topic of on-going research and most formulations are based on a superposition of isolated neutral atoms and range in complexity from a summation of Gaussian functions to procedures incorporating accurate physical constraints [1,9–13]. The number and sophistication of programs available for (S)TEM simulation is also ever increasing and include: C code [1,14], EMS [15], STEM_CELL [16,17] and MULTEM [13,18] with some incorporating multiple CPU and GPU support for added computational speed.

All of the listed codes incorporate the high-energy approximation in both the calculation of the potentials as well as the form of the differential equations used for the propagation of the electron through the material. Specifically, the second-order differential in z in the Schrödinger equation has been approximated by a first-order differential, since $\partial^2/\partial z^2 \ll 4\pi/\lambda \, \partial/\partial z$, which is appropriate in the high-energy/forward-scattering approximation. In what follows, equations that contain a second-order derivative in z and those that have a first-order derivative in z as the highest-order derivative will be referred to as the second-order and first-order problems, respectively. It is generally assumed that the high-energy approximation is valid for electron energies ≥ 100 keV, however the exact value is somewhat blurry since the electron speed is $0.78c$ at 300 keV, but is still $0.27c$ at a relatively low energy of 20 keV. At lower energies, backscattering of electrons may become important, an effect not included in the high-energy, forward-scattering approximations. Chen and Van Dyck [6] have rigorously formulated a multislice simulation scheme without the use of the high-energy approximation which includes the effects of backscattering. In their correction for the forward-scattering

approximation, the ∇_{xy}^2 and potential terms are present inside a square root operation which must be expanded as a power series in order to perform the calculations. Although more accurate than the conventional forward-scattering approximation, the series expansion is computationally expensive and the Ewald sphere still transforms into a modified parabolic surface, although not to the same extent as the high-energy approximation directly [19]. An accurately shaped Ewald spherical surface is important in the simulation of higher-order Laue zone (HOLZ) diffraction effects.

Ming and Chen [20] have investigated the differences between simulated diffraction patterns of $SmBa_2Cu_3O_{7-x}$[001] using the conventional multislice method (CMS), CMS with a correction factor to more accurately approximate the spherical Ewald surface (PCMS) and the method of Chen and Van Dyck (FCMS) [6]. At an electron energy of 100 keV, they found the three methods are nearly equivalent for zero order Laue zone (ZOLZ) reflections, but the CMS method differed appreciably from the PCMS and FCMS methods for HOLZ reflections. These differences increased with decreasing electron energy and even the PCMS and FCMS methods displayed significant differences at 20 keV. The effects of backscattering were not included in their calculations and the authors acknowledged that backscattering effects would become more important as the electron energy decreased. Recently, Wacker and Schröder [21] have integrated the Schrödinger equation directly using a Runge–Kutta method for the simulation of selected-area diffraction (SAD) patterns in silicon aligned along the [110] zone-axis projection at an incident electron energy of 20 keV neglecting the effects of backscattered electrons and thermal vibration of the atoms. They observed significant differences between the classical multislice, Chen and Van Dyck and direct-integration (Runge–Kutta) solutions. Although the magnitude of the second-order derivative term was about a factor of 10^{-4} less than the first-order derivative term, the second-order effects were cumulative and the differences in the wave functions between the two formalisms were shown to increase with thickness.

In this work, the effects of the second-order derivative in z will be considered, but in the context of convergent beam electron diffraction (CBED) patterns that are the basis of annular bright-field (ABF) and dark-field (ADF) imaging techniques. First, electron propagation across an atom in one dimension (1D) will be considered and the differences between first- and second-order solutions will be investigated. This work will then be extended to the full three-dimensional (3D) system for the calculation of CBED patterns at 20 keV along the [111] zone-axis projection in silicon where thermal effects have been included using the frozen-phonon approximation [22]. An electron energy of 20 keV was chosen since this was the energy where Ming and Chen [20] observed significant differences in simulated diffraction patterns by three multislice methods, as well as the energy used by Wacker and Schröder [21] in some of their simulations, again to highlight differences as a result of simulation methodology. A comparison will be made between the directly integrated first- and second-order solutions, and the implications for ABF and ADF imaging will be discussed. Finally, an observation on the efficiency of the numerical solver will be presented.

2. Theoretical background

Fundamentally, the Dirac equation governs the interaction of relativistic electrons with atoms. This equation requires the solution of four coupled, second-order differential equations yielding four solutions containing matter and anti-matter, as well as spin up and down, effects. There is a very accurate approximation that simplifies this problem for high-energy electrons; Fujiwara [23,24], and more recently Rother and Scheerschmidt [25], have shown that the Dirac equation can be replaced with the non-relativistic Schrödinger equation, where the electron mass, m, and wavelength, λ, have been relativistically corrected. Thus, the problem is reduced to solving a single, second-order differential equation with a single solution. Using the notation of Kirkland [1], the Schrödinger equation can be expressed as,

$$\left[-\frac{\hbar^2}{2m}\nabla^2 - eV(x,y,z)\right]\Psi_f(x,y,z) = E\Psi_f(x,y,z) \tag{1}$$

where $\hbar$ is Planck's constant, e the charge of an electron, $V(x, y, z)$ is the three-dimensional (3D) atomic potential interacting with the incident electron, and E is the kinetic energy of the incident electron,

$$E = \frac{2\pi^2\hbar^2}{m\lambda^2} \tag{2}$$

Unless otherwise specified, it will be assumed the wave function and potential are both functions of x, y and z. The direct integration of Eq. (1) is usually not performed as the solutions are highly oscillatory with a period close to the electron wavelength. Hence, a rapidly varying plane-wave term is factored out of the solution. Making the variable substitution, $\Psi_f(x,y,z) = e^{2\pi iz/\lambda}\Psi(x,y,z)$ and using Eq. (2) for the energy leads to the differential equation [1],

$$\left[\nabla_{xy}^2 + \frac{\partial^2}{\partial z^2} + \frac{4\pi i}{\lambda}\frac{\partial}{\partial z} + \frac{2me}{\hbar^2}V\right]\Psi = 0 \tag{3}$$

For high-energy electrons, it is assumed that $\partial^2/\partial z^2 \ll 4\pi\,\partial/\partial z$ and the differential equation becomes first order in z,

$$\frac{\partial\Psi}{\partial z} = \left[\frac{i\lambda}{4\pi}\nabla_{xy}^2 + \frac{2mei\lambda}{4\pi\hbar^2}V\right]\Psi \tag{4}$$

In both Eqs. (3) and (4), the Laplacian operator was calculated using the Fourier transform method as given by Eq. (19) [1,21].

2.1. One-dimensional solution

The solution to the full 3D differential equation is complex and non-analytical, and in order to study some of the properties of the first- and second-order problems, a simplification to a 1D problem in z will be made. Thus, setting $\Psi(x,y,z) \rightarrow \Psi(z)$, $V(x,y,z) \rightarrow V(z)$ and $\nabla_{xy}^2\Psi(z) = 0$, Eqs. (3) and (4) become,

$$\left[\frac{d^2}{dz^2} + \frac{4\pi i}{\lambda}\frac{d}{dz} + \frac{2meV(z)}{\hbar^2}\right]\Psi(z) = 0 \tag{5}$$

and

$$\frac{d\Psi(z)}{dz} = \frac{me\lambda iV(z)}{2\pi\hbar^2}\Psi(z) \tag{6}$$

respectively.

Many of the features of the solutions to the 1D problem can be illustrated by studying the case of constant potential with $V(z) = V$. In the case of the first-order problem the solution is,

$$\Psi^{1st}(z) = \exp\left(i\frac{me\lambda V}{2\pi\hbar^2}z\right) \tag{7}$$

with a real component,

$$\Psi_R^{1st}(z) = \cos\left(\frac{me\lambda V}{2\pi\hbar^2}z\right) \tag{8}$$

Similarly, the solution for the second-order problem for a constant potential is given by,

$$\Psi^{2nd}(z) = e^{-2\pi iz/\lambda}\left[\cos(kz) + \frac{2i\pi}{\lambda k}\sin(kz)\right] \tag{9}$$

where,

$$k = \sqrt{\frac{2meV}{\hbar^2} + \frac{4\pi^2}{\lambda^2}} \tag{10}$$

The real component of Eq. (9) for the special case of a weak potential can be put in the form,

$$\Psi_R(z) = \cos\left(\frac{me\lambda V}{2\pi\hbar^2}\right) - \frac{me\lambda^2 V}{8\pi^2\hbar^2}\left[1 - \cos\left(\frac{4\pi}{\lambda}z\right)\right] \tag{11}$$

$$= \Psi_R^{1st}(z) - \frac{me\lambda^2 V}{8\pi^2\hbar^2}\left[1 - \cos\left(\frac{4\pi}{\lambda}z\right)\right] \tag{12}$$

See the Appendix for the full derivation. The first term inside the square brackets of Eq. (12) represents a small negative shift of the value of the second-order solution with respect to the first-order solution and the second term is a small amplitude oscillation with half the period of the incident electron wavelength.

The second case to be studied in one dimension is for a soft-core atomic potential defined as [26],

$$V(z) = \frac{eZ}{4\pi\epsilon_0}\frac{1}{\sqrt{a^2 + z^2}} \tag{13}$$

where Z is the atomic number, ϵ_0 is the vacuum permittivity and a is a constant to avoid the singularity at the nucleus of the atom located at $z = 0$. For $Z = 14$ and $a = 0.03$ Å, this form of the potential approximates that given by Peng et al. for silicon [27]. Apart from avoiding the singularity in a Coulomb potential, the use of this potential results in an analytic solution to the first-order differential equation,

$$\Psi(z) = \left(\frac{z + \sqrt{a^2 + z^2}}{z_0 + \sqrt{a^2 + z_0^2}}\right)^{ik'} \tag{14}$$

where the boundary condition $\Psi(z_0) = 1$ and

$$k' = \frac{me^2\lambda}{8\epsilon_0\pi^2\hbar^2} \tag{15}$$

were used.

3. Results and discussion

The numerical results presented will focus on a relatively low electron energy, specifically 20 keV, since the applicability of Eq. (4) has been firmly established for high energy electrons (i.e., 100+ keV). See, for example, Kirkland [1] and the references cited therein. This was also our experience and we did not find any appreciable differences between the first- and second-order solutions at high electron energies that would justify not using Eq. (4). As in the numerical simulations of Wacker and Schröder, the effects of backscattered electrons will not be considered [21].

3.1. One-dimensional solutions

All of the simulations in this section were performed at an incident electron energy of 20 keV, $Z = 14$ and $a = 0.03$ Å. The starting point of the numerical integration was $z_0 = -3.0$ Å and proceeded until $z = 3.0$ Å as these were regions of negligible atomic potential. The differential equations were numerically integrated in MATLAB^{TM} using the ode45 solver with maximum absolute and relative per-step tolerances of 10^{-8}. As a check of the accuracy of the numerical solver, the differences between the analytical and numerical solutions at $z = 3$ Å for the soft-core potential were 6.2×10^{-9} and 1.2×10^{-8} for the real and complex components of the wave function, respectively.

Presented in Fig. 1 are the real and imaginary components of the wave function for both first- and second-order differential
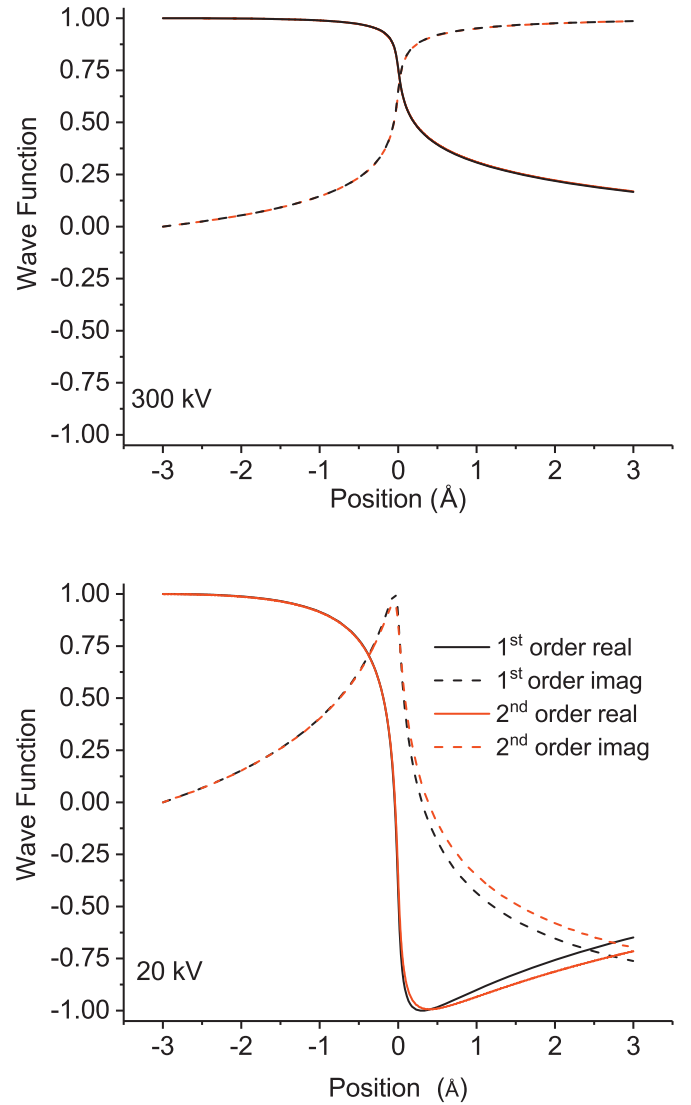

Fig. 1. The real and imaginary components of the first- and second-order solutions using the soft-core potential for electron energies of (top) 300 keV and (bottom) 20 keV for an atom located at $z = 0$. Differences between the first- and second-order solutions provides a measure of the quality of the forward scattering approximation. At an electron energy of 300 keV it is observed that the forward scattering approximation holds very well, whereas at 20 keV this approximation appears to break down. The legend applies to both figures.

equations calculated for electron energies of 300 keV and 20 keV. Although the 300 keV results did not show any appreciable difference between the first- and second-order solutions, it is observed that the first- and second-order wave function solutions diverge significantly at 20 keV at the $z = 3$ Å integration limit. While it is recognized that it is the intensity ($\Psi^*\Psi$) that is the physically important quantity, it does not make sense to plot this quantity in the 1D case since it would equal very close to 1 over the entire integration interval. However, the presence of a difference in the wave functions may be indicative of differences occurring in the full 3D calculation.

The second-order calculations took 84 times longer to compute than the first-order calculations based on an average timing from 100 runs. Naively, it is expected that the second-order problem should take twice as long to compute since it can be written equivalently as two coupled first-order differential equations [28]. In order to better understand the increase in computation time, displayed in Fig. 2 is the same plot as the 20 keV solution in Fig. 1 but the region near $z = -3.0$ Å has been highlighted. Here
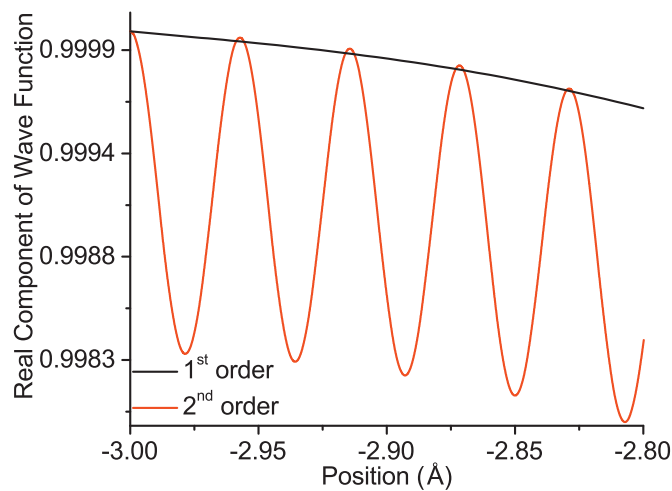
Fig. 2. A comparison of the real component of the wave functions evaluated using the first- and second-order calculations. The wave functions calculated using the second-order equations contain small oscillations that need to be resolved by the numerical integrator in order to ensure a stable and accurate calculation. This necessitates the use of much smaller step sizes than required for the first-order calculation resulting in increased computation time.
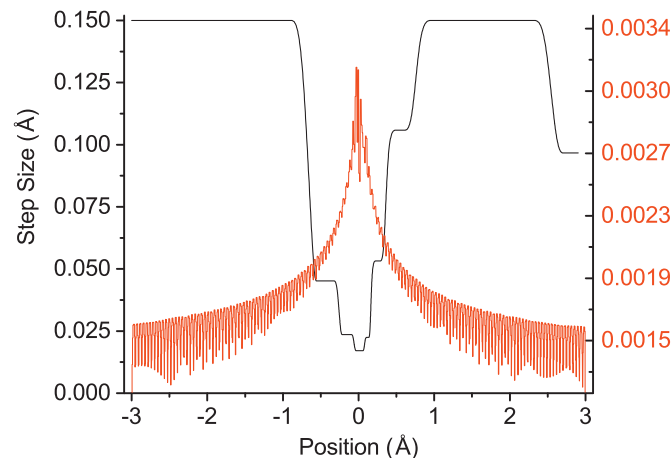

Fig. 3. The numerical integration step size as a function of position, where the results from the first- and second-order calculations are shown in black and red, respectively. The reduced step size of the second-order calculations with respect to the first-order equations is a result of the need to resolve the small amplitude oscillations present in the wave function of the second-order solution. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

it is observed that the average value of the second-order solution has decreased compared to the first-order solution and contains small amplitude oscillations, as expected from Eq. (12). The period of the oscillations is very close to half of electron wavelength of 0.086 Å at 20 keV. This is an important result since the numerical solver must use a much smaller step size in the second-order calculation as compared to the first-order calculation in order to resolve the oscillations. The numerical solution becomes unstable when too large of a step size is selected, as observed by Wacker and Schröder [21].

A typical plot of the integration step sizes as a function of position for the soft-core potential is shown in Fig. 3. It is observed that the step sizes for the first-order calculation are larger away from the core of the atom and get smaller near the core where both the slopes of the atomic potential and wave functions are greatest. Perhaps somewhat counter intuitively, the step sizes for the second-order calculation display opposite behaviour; they are larger nearer to the core. The reason for this behaviour is that
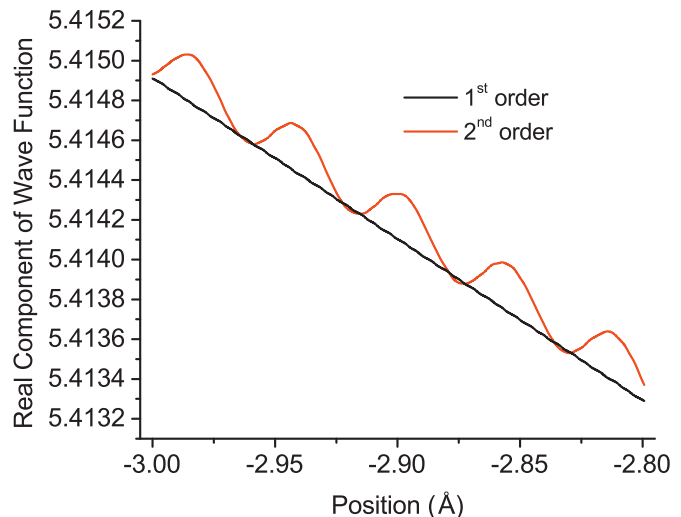

Fig. 4. The real components of the first- and second-order calculated wave functions obtained from the centre of the (000) diffraction disk. Note that the plot starts 3 Å above the top surface of the super cell where the first plane of atoms is located at $z = 0$. As in the 1D case, the increase in computational time is largely due to resolving small amplitude oscillations caused by the second-order derivative term.

when the slope of the wave function is large, the oscillations have less of an effect on the shape of the wave function locally and the numerical solver can take larger steps while maintaining the same per-step error tolerance. Even the largest step in the second-order calculation is about an order of magnitude smaller than the smallest step in the first-order calculation. Therefore, the main cause of the increased time of computation required for the second-order equation is the reduced step size necessary to resolve the small amplitude oscillations.

3.2. Three-dimensional CBED simulations

CBED patterns were simulated for silicon at the [111] zone-axis projection by direct integration of Eqs. (3) and (4). The specimen consisted of a super cell of dimension 38.4 Å × 39.9 Å × 498.5 Å and the full 3D potential of Kirkland [1] was used. The effects of thermal vibrations were included in the frozen-phonon approximation [22] using a root-mean-square deviation of 0.078 Å at a temperature of 300 K. Sixty-four thermal vibration configurations were run and averaged incoherently to ensure convergence of the solution. The microscope parameters were: an electron energy of 20 keV, an objective aperture semiangle of 16.0 mrads, a defocus of 983 Å and a spherical aberration coefficient of 2.0 mm. The equation given by Spence [29, p.272] to obtain the narrowest electron probe was used for the defocus, but the objective aperture semiangle was increased from 10.3 mrads to 16.0 mrads in order to resolve some of the fine structure within the CBED disks. Numerically, the size of the wave function was set to 512× 512 pixels and the bandwidth was limited to 2/3 of its maximum Nyquist value to avoid aliasing effects during the Fourier transform calculations. These parameters resulted in a maximum radial distance in reciprocal space of 367.3 mrads. A constant slice thickness of 0.00192 Å was used to ensure stable simulation calculations. Eqs. (3) and (4) were integrated using a Dormand-Prince 4(5) method [28] which is the same method implemented in the MATLAB^{TM} ode45 routine. The integration started 3 Å above and extended 3 Å below the top and bottom planes of the super cell, respectively.

The second-order 3D calculation ran much slower than the first-order calculation and the slow down was again mainly due to having to resolve the small oscillations in the second-order wave function. Presented in Fig. 4 is a plot of the real component of
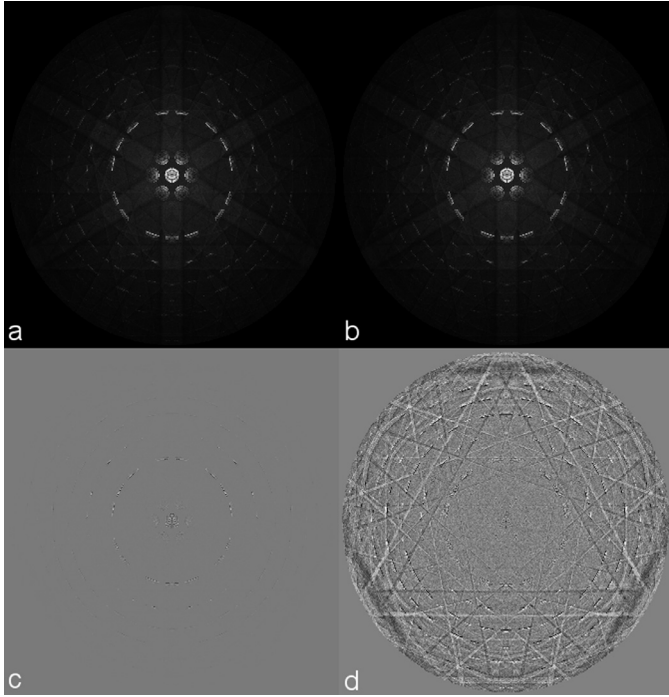
Fig. 5. Si [111] zone-axis orientation CBED patterns using (a) first-order and (b) second-order equations. The square root of the intensities of images (a) and (b) were plotted in order to display weak intensities. (c) Difference map of (b)-(a) where black represents a difference of -5.6×10^7 and white a difference of 6.1×10^7. (d) Percent difference map between (b) and (a) as calculated using $100*[(b)-(a)]/[(b)+(a)]/2$ where black represents -193% and white 188%. An incident electron energy of 20 keV and specimen thickness of 498.5 Å were used in the calculations.

the electron wave function using both the first- and second-order equations along the z direction through the centre of the (000) diffraction disk. The period of oscillation of the wave function at this location is half of the incident electron wave length, as expected when the atomic potential is weak.

Visually, the CBED patterns simulated using the first- and second-order equations appeared similar and presented in Fig. 5(a) and (b) are the CBED patterns using the first- and second-order equations, respectively. In order to quantify the differences between these two simulations, second-order - first-order difference and percent difference maps were prepared and are displayed in Fig. 5(c) and (d), respectively. In the difference map of Fig. 5(c), the largest differences occurred within the (000) diffraction disk

and along the Laue-zone rings. However, when percent differences are plotted, the differences are close to a minimum at the centre of the pattern and increase in a systematic way with increasing scattering angle.

This increasing difference with scattering angle can be explained using the Ewald surface construction shown in Fig. 6. Here $K = 1/\lambda$ is the magnitude of the incident electron wave vector, and g_z and g_R are the distances to a point on the Ewald surface in the vertical and radial directions, respectively. The form of the Ewald surface for the second-order equation is a sphere defined by [6,19,30],

$$g_R^2 = 2Kg_z - g_z^2 \tag{16}$$

and is a parabola for the first-order equation,

$$g_R^2 = 2Kg_z \tag{17}$$

The difference between these two surfaces is g_z^2 and for a given g_R increases with decreasing accelerating potential. For example, the radius of the Ewald sphere, K, increases by a factor of 4.3 when the potential is increased from 20 kV to 300 kV. Hence, the difference between the two surfaces becomes negligible for high energy electrons even at high scattering angles.

Another way of visualizing the differences in the CBED simulations is to look at the radially integrated electron intensity. This quantity is important since it is required to quantitatively model ABF and ADF image contrast. Displayed in Fig. 7 are the integrated intensity profiles calculated using the first- and second-order equations. There is very little separating the three curves at low-to-medium scattering angles (0–200 mrads) and only modest differences for higher scattering angles. It is observed that the peaks from HOLZ scattering are shifted to higher scattering angles for the first-order method as compared to the second-order method, as expected from Ewald surface considerations. The intensities of the thermal diffuse background remain nearly constant for each method over the range of scattering angles plotted. This would imply under these conditions ABF and ADF images formed should be similar regardless of the method chosen.

However, it is expected that the fine structure within the CBED disks located in the HOLZs could change significantly between the first- and second- order methods at low incident electron energies. Furthermore, slight differences in shape of the Ewald surfaces can have a large impact on the structure of selected area diffraction (SAD) patterns as demonstrated by Wacker and Schröder [21]. In both of these cases, a full second-order calculation may be required to ensure accurate matches between simulation and experiment depending on the particular (S)TEM parameters and material system under investigation.
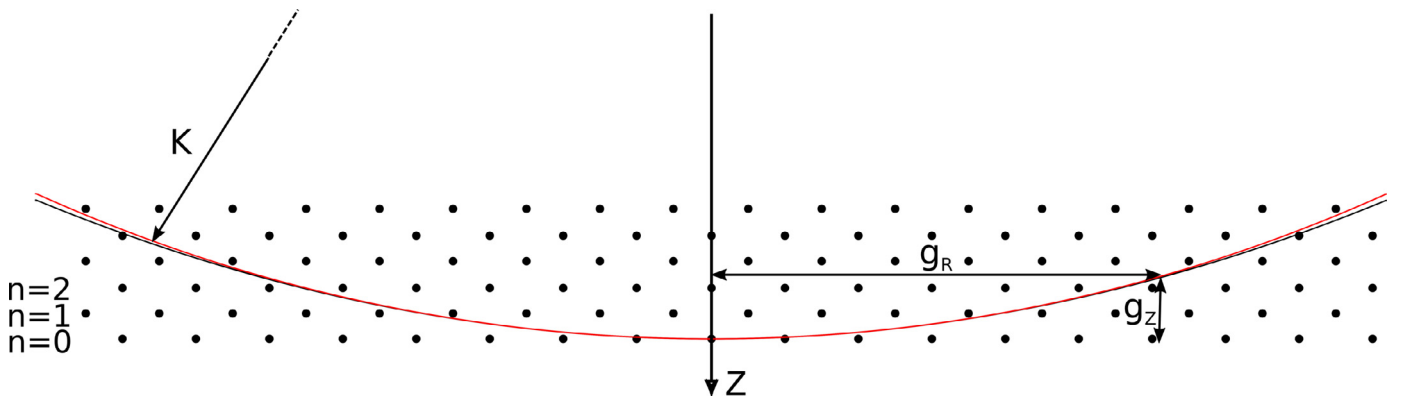

Fig. 6. Ewald surfaces with the radius drawn to scale based on (black) first-order and (red) second-order equations. The bottom row of dots corresponds to the zeroth-order Laue zone, the second row the first-order Laue zone, etc. The difference between the two curves scales as g_z^2 and increases with decreasing electron energy. Scale: horizontal dot spacing = 5.2 Å^{-1} = 44.7 mrads. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
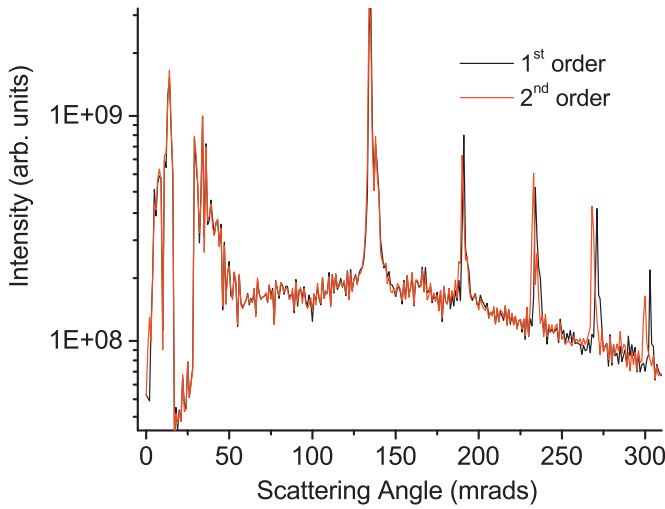
Fig. 7. The radially integrated intensity of the CBED patterns using the first- and second-order differential equations. The main difference between these two curves is that the first-order results are shifted to higher scattering angles as compared to the second-order calculation, where the magnitude of the shift increases with scattering angle. This is a result of the differences in Ewald surface shape between the two methods.
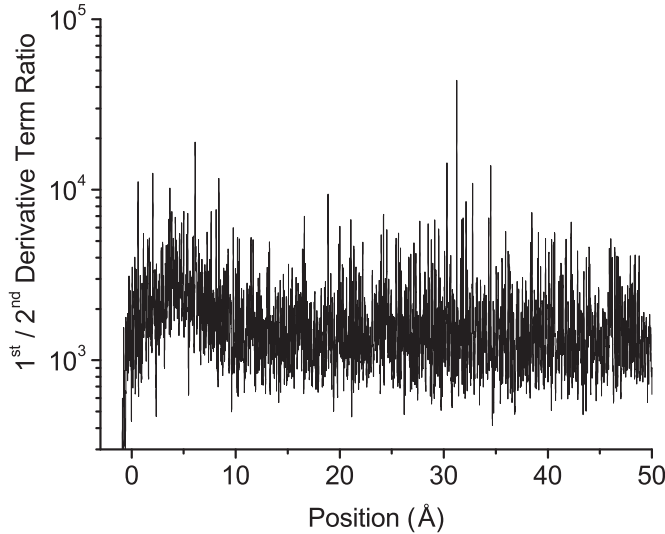

Fig. 8. The ratio of the first-order derivative term $\left(\frac{4\pi}{\lambda}\frac{\partial}{\partial z}\right)$ with respect to the second-order derivative term $\frac{\partial^2}{\partial z^2}$ along the z direction located at the centre of the (000) CBED disk for the first 50 Å of the integration.

Apart from Ewald surface effects, there is little to differentiate the CBED intensities as illustrated in Fig. 5(c) and (d). This is a result of the second-order derivative term being a factor of 10^3 smaller than the first-order derivative term as shown in Fig. 8 for the first 50 Å of a simulation run. However, the second-order derivative effects can compound with specimen thickness as shown by Wacker and Schröder [21], particularly above a thickness of 500 Å for silicon at 20 keV.

It was previously mentioned that the small amplitude oscillations present in the wave functions computed with the second-order method led to increased computation time and instability when the integration step was set too large to resolve the oscillations. However, there is another important factor which limits the numerical efficiency of the calculations; the rate at which the numerical solver converges to a given global error tolerance. The global error was defined as the maximum difference between pixel intensity values when comparing pixels at the same locations in
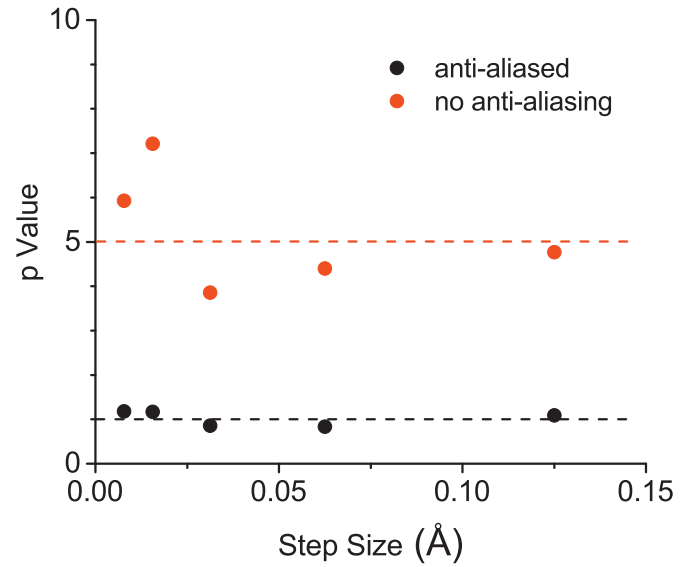

Fig. 9. The rate of the error reduction for a given numerical integration step size for the full 3D model using the first-order equation (black) with and (red) without anti-aliasing of the FFTs. A p value of 5 is expected for a Dormand-Prince 4(5) numerical integration scheme applied to smooth functions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the numerical and exact solutions. In the following text, all references to errors implies global errors. The error between the numerical and exact solutions for a given step size should follow a power-law dependence on the step size, $e_h \approx \eta h^p$, where h is the step size, η is an unknown constant and p is the rate. Thus, the errors should decrease with the pth power of the step size provided that constant step sizes are used during the integration. For example, the simple Euler method should yield $p = 1$ and the Dormand-Prince 4(5) solver used in this work should give $p = 5$. It is expected that the errors associated with the Dormand-Prince method should decrease much faster at a given step size than the Euler method, allowing for larger step sizes to be taken and a computationally faster integration. Taking the ratio of the error with step size h to the error with step size $h/2$ leads to the following expression for the rate,

$$p = \log_2\left(\frac{e_h}{e_{h/2}}\right) \tag{18}$$

Hence, by calculating the errors at step sizes h and $h/2$ the rate can be determined. In the case of the 1D numerical simulations using the soft-core potential, the exact solution was known and a value of p very close to 5 was obtained. In the full 3D simulation, the $e \approx \eta h^p$ scaling is expected to hold where e is the difference between the numerical solution and the exact solution of the system of the ODEs that results upon using Eq. (19) for the Laplacian in Eq. (4). Since the exact solution of this system of ODEs is not known analytically, we provide a numerical surrogate computed with a very small z step size. In this work, the first-order equations were used for the determination of the p value and a step size of 0.000244 Å was selected as representative of the exact solution. The step sizes were halved starting from 0.125 Å down to 0.0078125 Å since step sizes larger than 0.125 Å led to unstable solutions. Presented in Fig. 9 are the results of this experiment for two cases: (1) the FFTs were anti-aliased as in a standard calculation and (2) the FFTs were not anti-aliased. Incorporation of anti-aliasing of the FFTs is important to avoid the introduction of artifacts into the simulation because of wrap-around effects at the boundaries in reciprocal space. See chapter 6 of Kirkland [1] for a thorough description of this

topic. When anti-aliasing is included in the calculation a p value close to one was obtained, indicative of an order one numerical integration technique, such as forward Euler [27], and not what is expected for a Dormand-Prince solver. But when anti-aliasing is turned off, the expected $p = 5$ behaviour is obtained. This same behaviour was observed regardless of whether a steep filter function was used for bandwidth limiting or a smoothed Fermi function applied as per Wacker and Schröder [21]. The explanation lies in performing the forward and inverse FTs in the calculation of the ∇_{xy}^2 terms in Eq. (4) [1],

$$\nabla_{xy}^2 \Psi = FT^{-1}\left[-4\pi^2 k^2 FT[\Psi]\right] \qquad (19)$$

An FT is required to go from real to reciprocal space so that a scaling factor can be applied as well as the anti-aliasing operation. Anti-aliasing limits the bandwidth to 2/3 of the Nyquist limited value and results in close to 2/3 of the pixels in the diffraction pattern being set to zero. Thus, the reciprocal space wave function is truncated, a large amount of phase information is lost, and a small amount of electron intensity is removed from the diffraction pattern. An inverse FT is then performed returning to real space. Since some data has been removed between integration steps, the wave functions are no longer smooth along the electron propagation direction and there are small discontinuities between adjacent real-space wave functions. These discontinuities reduce the maximum order of continuous derivatives in the wave functions leading to reduced p behaviour. More generally, discontinuities occurring between integration steps result in a reduction to order one behaviour [28]. Therefore, the desired advantage of faster computation with higher rate solvers has been negated and a lower rate solver with less computations required per step may lead to lower computation times for the same specified error tolerance. This behaviour remains to be investigated.

4. Conclusions

It has been shown that the source of increased computation time and instability in the solution of the second-order differential equation for electron scattering can be attributed to having to resolve small amplitude oscillations that are present in the second-order problem but not in the first-order problem. The percentage differences in the intensities of the CBED patterns between the first- and second-order solutions were a minimum at low scattering angles and increased radially with scattering angle. This could be explained using an Ewald surface construction where the Ewald surface is a sphere for the second-order differential equation but becomes a parabola in the first-order approximation. The surface of the Ewald parabola is located at greater reciprocal distance for non-zero scattering angle compared to the surface of the Ewald sphere. Hence, for a given set of conditions, this results in a shift of the HOLZ diffraction rings towards higher scattering angles. In contrast, the intensity of the thermal diffuse background remained nearly the same between the two methods as a function of scattering angle. Furthermore, at least within the scope of these simulations, minimal differences in ABF and ADF simulated images would be expected between the solutions based on first- and second-order equations.

Numerically, the rate of the Dormand-Prince 4(5) solver was found to be 1, a factor of 5 less than expected, and resulted from including anti-aliasing of the FFTs in the calculations. Since the anti-aliasing step is required to avoid artifacts from being introduced in the simulations, there may be opportunities for higher computational efficiency by implementation of lower rate solvers. In addition, significant improvements in computational speed could be realized in calculating the second-order solutions if a method could be found where numerically resolving the small amplitude oscillations could be avoided.

Appendix

Eq. (5) can be written as,

$$\left[\frac{d^2}{dz^2} + \alpha\frac{d}{dz} + \beta\right]\Psi(z) = 0 \qquad (20)$$

where,

$$\alpha = \frac{4\pi i}{\lambda} \text{ and } \beta = \frac{2meV}{\hbar^2} \qquad (21)$$

This is a linear, homogeneous differential equation with analytical solution,

$$\Psi(z) = c_1 \exp\left[-\frac{z}{2}\left(\sqrt{\alpha^2 - 4\beta} + \alpha\right)\right]$$
$$+ c_2 \exp\left[\frac{z}{2}\left(\sqrt{\alpha^2 - 4\beta} - \alpha\right)\right] \qquad (22)$$

where c_1 and c_2 are determined from the boundary conditions,

$$\Psi(0) = 1 \text{ and } \Psi'(0) = 0 \qquad (23)$$

Applying these two boundary conditions leads to,

$$\Psi(z) = e^{-\alpha z/2}\left\{\frac{1}{2}\left[\exp\left(\frac{z}{2}\sqrt{\alpha^2 - 4\beta}\right) + \exp\left(-\frac{z}{2}\sqrt{\alpha^2 - 4\beta}\right)\right]\right.$$
$$+ \frac{\alpha}{2\sqrt{\alpha^2 - 4\beta}}\left[\exp\left(\frac{z}{2}\sqrt{\alpha^2 - 4\beta}\right)\right.$$
$$\left.\left. - \exp\left(-\frac{z}{2}\sqrt{\alpha^2 - 4\beta}\right)\right]\right\} \qquad (24)$$

Substituting for α and β,

$$\Psi(z) = e^{-2\pi i z/\lambda}\left\{\frac{1}{2}[\exp(ikz) + \exp(-ikz)]\right.$$
$$\left. + \frac{\pi}{\lambda k}[\exp(ikz) - \exp(-ikz)]\right\} \qquad (25)$$

and using the identities $2\cos x = e^{ix} + e^{-ix}$ and $2i\sin x = e^{ix} - e^{-ix}$ gives Eq. (9).

From Eq. (10), in the weak potential approximation, using $(1 + x)^{1/2} \approx 1 + x/2$ leads to,

$$k \approx \frac{2\pi}{\lambda}\left(1 + \frac{me\lambda^2 V}{4\pi^2\hbar^2}\right) \qquad (26)$$

and

$$\frac{2\pi}{\lambda} - k \approx -\frac{me\lambda^2 V}{4\pi^2\hbar^2}; \quad \frac{2\pi}{\lambda k} \approx 1 - \frac{me\lambda^2 V}{4\pi^2\hbar^2} \qquad (27)$$

Expanding the exponential term in Eq. (9) using $e^{-ix} = \cos(x) - i\sin(x)$ and multiplying through gives,

$$\Psi(z) = \cos\left(\frac{2\pi}{\lambda}z\right)\cos(kz) + i\frac{2\pi}{\lambda k}\cos\left(\frac{2\pi}{\lambda}z\right)\sin(kz)$$
$$- i\sin\left(\frac{2\pi}{\lambda}z\right)\cos(kz) + \frac{2\pi}{\lambda k}\sin\left(\frac{2\pi}{\lambda}z\right)\sin(kz) \qquad (28)$$

Rewriting the cosine-cosine and sine-sine terms using the trigonometric identities $\cos(x)\cos(y) = \cos(x - y) - \sin(x)\sin(y)$ and $\sin(x)\cos(y) = \sin(x - y) - \cos(x)\sin(y)$ we obtain,

$$\Psi(z) = \cos\left[\left(\frac{2\pi}{\lambda} - k\right)z\right] - \sin\left(\frac{2\pi}{\lambda}z\right)\sin(kz)$$
$$+ \frac{2\pi}{\lambda k}\sin\left(\frac{2\pi}{\lambda}z\right)\sin(kz) - \frac{2\pi}{\lambda k}\cos\left(\frac{2\pi}{\lambda}z\right)\sin(kz)$$
$$- i\left\{\sin\left[\left(\frac{2\pi}{\lambda} - k\right)z\right] + \cos\left(\frac{2\pi}{\lambda}z\right)\sin(kz)\right\} \quad (29)$$

Applying the two weak potential approximations for $2\pi/\lambda - k$ and $2\pi/\lambda k$, and noting that $\sin(kz) \approx \sin(2\pi z/\lambda)$ gives,

$$\Psi(z) = \left[\cos\left(\frac{me\lambda V}{2\pi\hbar^2}\right) + i\sin\left(\frac{me\lambda V}{2\pi\hbar^2}\right)\right]$$
$$- \frac{me\lambda^2 V}{4\pi^2\hbar^2}\sin\left(\frac{2\pi}{\lambda}z\right)\left[\sin\left(\frac{2\pi}{\lambda}z\right) + i\cos\left(\frac{2\pi}{\lambda}z\right)\right] \quad (30)$$

with a real component of,

$$\Psi_R(z) = \cos\left(\frac{me\lambda V}{2\pi\hbar^2}\right) - \frac{me\lambda^2 V}{4\pi^2\hbar^2}\sin^2\left(\frac{2\pi}{\lambda}z\right) \quad (31)$$

$$= \cos\left(\frac{me\lambda V}{2\pi\hbar^2}\right) - \frac{me\lambda^2 V}{8\pi^2\hbar^2}\left[1 - \cos\left(\frac{4\pi}{\lambda}z\right)\right] \quad (32)$$

where $\sin^2(x) = [1 - \cos(2x)]/2$ has been used.

References

[1] E. Kirkland, Advanced Computing in Electron Microscopy, second ed., Springer, New York, 2010.
[2] D. Williams, C. Carter, Transmission Electron Microscopy, Springer, New York, 1996.
[3] C. Humphreys, The scattering of fast electrons by crystals, Rep. Prog. Phys. 42 (1979) 1825–1887.
[4] M.D. Graef, Introduction to Conventional Electron Microscopy, Cambridge University Press, Cambridge, 2003.
[5] L. Reimer, Transmission Electron Microscopy - Physics of Image Formation and Microanalysis, third ed., Springer-Verlag, New York, 1993.
[6] J. Chen, D. Van Dyck, Accurate multislice theory for elastic electron scattering in transmission electron microscopy, Ultramicroscopy 70 (1997) 29–44.
[7] D. Van Dyck, Image calculations in high-resolution electron microscopy: problems, progress, and prospects, Adv. Elec. Elec. Phys. 65 (1985) 295–355.
[8] I. Lobato, D. Van Dyck, Improved multislice calculations for including higher-order Laue zones effects, Ultramicroscopy 119 (2012) 63–71.
[9] L.-M. Peng, G. Ren, S. Dudarev, M. Whelan, Robust parameterization of elastic and absorptive electron atomic scattering factors, Acta Cryst. A52 (1996) 257–276.
[10] L.-M. Peng, Electron atomic scattering factors, Debye–Waller factors and the optical potential for high-energy electron diffraction, J. Elec. Micros. 54 (2005) 199–207.
[11] A. Weickenmeier, H. Kohl, Computation of absorptive form factors for high-energy electron diffraction, Acta Cryst. A47 (1991) 590–597.
[12] I. Lobato, D. Van Dyck, An accurate parametization for scattering factors, electron densities and electrostatic potentials for neutral atoms that obey all physical constraints, Acta Cryst. A 70 (2014) 636–649.
[13] I. Lobato, D. Van Dyck, MULTEM: a new multislice program to perform accurate and fast electron diffraction and imaging simulations using graphics processing units with CUDA, Ultramicroscopy 156 (2015) 9–17.
[14] C. Dwyer, Simulation of scanning transmission electron microscope images on desktop computers, Ultramicroscopy 110 (2010) 195–198.
[15] P. Stadelmann, EMS - a software package for electron diffraction analysis and HREM image simulation in materials science, Ultramicroscopy 21 (1987) 131–146.
[16] V. Grillo, E. Rotunno, STEM_cell: a software tool for electron microscopy: part I–simulations, Ultramicroscopy 125 (2013) 97–111.
[17] V. Grillo, F. Rossi, STEM_cell: a software tool for electron microscopy. Part 2 analysis of crystalline materials, Ultramicroscopy 125 (2013) 112–129.
[18] I. Lobato, S. Van Aert, J. Verbeeck, Progress and new advances in simulating electron microscopy datasets using MULTEM, Ultramicroscopy 168 (2016) 17–27.
[19] C. LV, Q. Liu, C. Cai, J. Huang, G. Zhou, Y. Wang, Dynamical electron diffraction simulation for non–orthogonal crystal system by a revised real space method, J. Micros. 261 (2016) 105–114.
[20] W. Ming, J. Chen, Validities of three multislice algorithms for quantitative low-energy transmission electron microscopy, Ultramicroscopy 134 (2013) 135–143.
[21] C. Wacker, R. Schröder, Multislice algorithms revisited: solving the Schrödinger equation numerically for imaging with electrons, Ultramicroscopy 151 (2015) 211–223.
[22] R. Loane, P. Xu, J. Silcox, Thermal vibrations in convergent-beam electron diffraction, Acta Cryst. A47 (1991) 267–278.
[23] K. Fujiwara, Relativistic dynamical theory of electron diffraction, J. Phys. Soc. Jpn. 16 (1961). 2262–2238
[24] K. Fujiwara, Relativistic dynamical theory of electron diffraction, J. Phys. Soc. Jpn. 17 (Suppl. B-II) (1962) 118–123.
[25] A. Rother, K. Scheerschmidt, Relativistic effects in elastic scattering of electrons in TEM, Ultramicroscopy 109 (2009) 154–160.
[26] T. Dziubak, J. Matulewski, Stabilization of one–dimensional soft–core and singular model atoms, Eur. Phys. J. D 59 (2010) 321–327.
[27] L.-M. Peng, G. Ren, S. Dudarev, M. Whelan, Robust parameterization of elastic and absorptive electron atomic scattering factors, Acta Cryst. A52 (1996) 257–276.
[28] U. Ascher, L. Petzold, Computer Methods for Ordinary Differential Equations and Differential–Algebraic Equations, Society for Industrial and Applied Mathematics, Philedelphia, 1998.
[29] J. Spence, High-Resolution Electron Microscopy, third ed., Oxford University Press, New York, 2003.
[30] D. Bird, Theory of zone axis electron diffraction, J. Elect. Micros. Tech. 13 (1989) 77–97.