Direct structure inversion from exit waves
Part I: Theory and simulations
A. Wang a,*, F.R. Chen b, S. Van Aert a, D. Van Dyck a

a University of Antwerp, Department of Physics, Groenenborgerlaan 171, 2020 Antwerp, Belgium
b National Tsing Hua University, Department of Engineering and System Science, No. 101, Sec. 2, Kuang-Fu Road, 300 Hsinchu, Taiwan

ARTICLE INFO

Keywords:
Argand plot
Electron scattering
Channelling theory
Structure determination
Image simulation

ABSTRACT

In order to interpret the amplitude and phase of the exit wave in terms of mass and position of the atoms, one has to “invert” the dynamic scattering of the electrons in the object so as to obtain a starting structure which can then be used as a “seed” for further quantitative structure refinement. This is especially challenging in case of a zone axis condition when the interaction of the electrons with the atom column is very strong. Based on the channelling theory we will show that the channelling map not only yields a circle on the Argand plot but also a circular “defocus curve” for every column. The former gives the number of atoms in each column, while the latter provides the defocus value for each column, which reveals the surface roughness at the exit plane with single atom sensitivity.

1. Introduction

Recently it has become a matter of routine to reconstruct the exit wave of an object at sub-angstrom resolution either by focal series reconstruction [1], by off-axis holography [2] or in the future possibly by using a phase plate [3]. However, in order to interpret the amplitude and phase of the exit wave in terms of mass and position of the atoms, one has to “invert” the dynamic scattering of the electrons in the object so as to obtain a starting structure which can then be used as a “seed” for further quantitative structure refinement [4]. Since long, plane-wave based methods, such as multislice methods [5] or real space methods [6], successfully describe electron scattering. However, the complexity of the calculation masks an intuitive insight into the underlying physics. For example, it cannot be easily understood why the exit wave function can be locally related to the projected crystal structure even though multislice calculations make such predictions. The physical reason for this “local” dynamic diffraction is the channelling of the electrons along the atom columns parallel to the beam direction. Due to the positive electrostatic potential of the atoms, an atom column acts as a guide or channel for the electron within which the electron can scatter dynamically without leaving the atom column. The channelling theory [7–10] describes this effect well and thus provides physical insight.

If the distance between the atom columns is not too small, a one-to-one correspondence between the exit wave and the columnar structure of the crystal is maintained. Within the atom columns, the electrons oscillate as a function of depth without leaving the atom column. The periodicity is inversely related to the “density” of the atom column. It is important to note that channelling is not a property of a crystal but occurs even in an isolated atom column and is not much affected by neighboring atom columns, provided the atom columns are not too closely spaced. Hence the one-to-one relationship is still present in case of defects such as translation interfaces or dislocations provided they are oriented with the atom columns parallel to the incident beam. Furthermore, it is possible to parameterize the exit wave. According to the channelling theory, the phase of the interaction wave is constant over the atom column and proportional to its “weight”. The amplitude of the column wave oscillates periodically as a function of thickness and the phase increases linearly with thickness. This enables to retrieve the projected structure of the object by matching the parameterized expression with the real exit wave of the columns. A convenient way to visualize this is by plotting for each pixel of the exit wave at the column positions its real and imaginary values as coordinates in a complex plane. Such a plot is known as the Argand plot [9,11,12]. The Argand plot turns out to give the most adequate visualization of the electron-object interaction. Since mass density is a function of thickness one can expect from the channelling theory that the points of columns with different number of atoms, as, for example, in a wedge shape structure, will form a simple continuous curve which is sampled in discrete steps (single atom steps) [13]. By changing mass density, absorption...
and defocus these points will move along specific curves. The images obtained from aberration-corrected microscopes, such as the TEAM0.5 microscope, also exhibit these effects because of the narrow focal depth of the microscope (0.7–1 nm) [14].

In this paper, a brief introduction of the channelling theory will be described in Section 2. In Section 3, the advantages of the representation of the exit wave in an Argand plot are discussed. Moreover, it will be shown how to deduce the number of atoms in the columns. In Section 4, effects of absorption from atoms and from insufficient sampling will be shown. In addition, the advantage to reveal the defocus values for each column will be discussed in Section 5. Finally, in Section 6, conclusions will be drawn for this extended approach on structure determination.

2. Channelling theory

Basically, the multislice theory expands the electron wave function inside the crystal in terms of plane waves which provide a complete basis for any wave function. It handles thousands of simultaneous plane waves to compute the exit wave whereas the channelling theory uses a more appropriate basis of eigenstates of the projected atom columns which leads to a simple closed-form expression for the exit wave. For that reason, the multislice theory is mostly used as a black box whereas the channelling theory provides more intuitive physical insight into the interaction between the electron wave and the object. The details of the channelling theory may be found in [7–10]. Here the basic concept will be described. Due to the strong electrostatic potential of the atoms, an atom column acts as a channel for the incoming electrons in which the electrons scatter dynamically without leaving it. This is schematically shown in Fig. 1. A simple intuitive way to understand this channelling theory is to consider each atom as a thin lens [15]. By passing through successive lenses, the electron wave is focused at periodic distances. This oscillation length or extinction distance is a function of the average mass density of the column which is related to the atomic number and the repeated distance between successive atoms along the column. Therefore, the extinction distance depends on chemical composition and zone axis. Because of the one-to-one correspondence between exit wave and structure, it is suitable to invert the multiple electron scattering and to retrieve structure information, mainly the number of atoms, along the beam direction. In what follows, the channelling theory will be described mathematically.

In quantum mechanics, the motion of the electron passing through a crystal can be derived by solving the time-dependent Schrödinger equation assuming the speed of the electrons to be constant so that the depth is proportional to the time. A more rigorous approach starting from the stationary Schrödinger equation with boundary conditions at entrance and exit face yields similar results [16]. For a crystal in zone-axis orientation, the exit wave for an isolated atom column can be expressed mainly by the lowest bound state, the s-state (denoted by subscript s in the following equations), as

$$\psi(r, z) = 1 + c_s \phi_s(r - \beta) \exp \left( -i\frac{E_s}{E_0} k z \right)$$

(1)

assuming the entrance wave to be equal to 1 (plane wave), with r the two-dimensional vector in the plane perpendicular to the incident beam direction, \( \beta \) the column position, \( E_0 \) the incident beam energy, k the reciprocal wavelength and z the thickness. \( E_s \) is the eigenenergy of the s-state function \( \phi_s(r - \beta) \). The exact shape of the s-state is easily corrupted due to unavoidable experimental factors such as atom vibrations. It has been shown in [8] that the s-state function may be approximated by a quadratically normalized and parameterized Gaussian function with column dependent width \( a \):

$$\phi_s(r) = \frac{1}{a \sqrt{2\pi}} \exp \left( -\frac{r^2}{4a^2} \right),$$

(2)

where \( r = |r| \). The excitation coefficient \( c_s \) is given by

$$c_s = 2\sqrt{2} a.$$

(3)

For most cases, this expression is sufficiently accurate except for thick objects containing heavy atoms, where other higher order states will become more important when the distance between adjacent atom columns decreases. It turns out to be more convenient to subtract the entrance wave from the exit wave. We will call this the interaction wave. From Eq. (1) the amplitude and phase of the interaction wave are [8]

$$\text{abs}(\psi(r, z) - 1) = 2c_s \sin \left( -\pi \frac{E_s}{E_0} k z \right) \phi_s(r),$$

(4)

and

$$\arg(\psi(r, z) - 1) = -\pi \frac{E_s k}{E_0 Z} z - 1.$$

(5)

From Eq. (4), it follows that the amplitude is peaked at the atom column position and that it varies periodically with depth. From Eq. (5), it follows that the phase, which is a constant over the column, is proportional to the average mass density of the column. The phase linearly increases with depth. The amplitude can be used to determine the positions of the atom columns and the phase can be used to determine the composition of the atom columns.

3. Argand plot

A convenient way to visualize the effect of electrons passing through a column is by plotting each pixel of the complex exit wave, which is located at a projected atom column position, in an Argand plot [9,11,12]. This is a representation in which each pixel is plotted as a point in a complex plane with its x-coordinate corresponding to the real pixel value and the y-coordinate corresponding to the imaginary pixel value. The position of the pixel at the center of the column is most sensitive to the scattering factors, the Debye–Waller factor and absorption. It turns out to be more convenient to shift the exit wave over a vector corresponding to the entrance wave so as to represent the interaction wave. From Eq. (1), it follows that the real and imaginary parts of the interaction wave are given by

$$\text{Re}\psi(r, z) - 1 = -2c_s \sin^2 \left( -\pi \frac{E_s k}{E_0 Z} \right) \phi_s(r),$$

(6)
has a circular shape in the Argand plot. Fig. 3 shows the Argand plot of the pixel values of the atom columns are radially symmetric and electron channelling. The absorption results from inelastic interactions such as phonon scattering, plasmon scattering and single electron excitations from inner atom shells [20–22]. Since the cross section from phonon scattering is also sharply peaked at the atom cores, we believe that this effect on the channelling is the strongest one [23]. We will use the model of Humphries and Hirsch [22], in which the electron–phonon interaction is described by introducing a complex potential where the imaginary part is proportional to the real part,

\[
V = V_k + iV_l = V_k + i\alpha V_k,
\]

where \(\alpha\) represents the proportionality constant. After replacing the new potential into the Schrödinger equation, the exit wave then becomes

\[
\psi(r, z) = 1 + c_s \phi_s \left( \frac{r}{\sqrt{1 + i \alpha \lambda}} \right) \exp \left( -\frac{i \pi E_r k z}{E_0} \right) \exp \left( -\alpha \pi \frac{E_r k z}{E_0} \right) - 1. \]

4. Absorption

It is well known that absorption has a strong effect on the electron channelling. The absorption results from inelastic interactions such as phonon scattering, plasmon scattering and single electron excitations from inner atom shells [20–22]. Since the cross section from phonon scattering is also sharply peaked at the atom cores, we believe that this effect on the channelling is the strongest one [23]. We will use the model of Humphries and Hirsch [22], in which the electron–phonon interaction is described by introducing a complex potential where the imaginary part is proportional to the real part,

\[
V = V_k + iV_l = V_k + i\alpha V_k,
\]

where \(\alpha\) represents the proportionality constant. After replacing the new potential into the Schrödinger equation, the exit wave then becomes

\[
\psi(r, z) = 1 + c_s \phi_s \left( \frac{r}{\sqrt{1 + i \alpha \lambda}} \right) \exp \left( -\frac{i \pi E_r k z}{E_0} \right) \exp \left( -\alpha \pi \frac{E_r k z}{E_0} \right) - 1. \]
thickness. From the slope of this inclined line, the ratio between the inelastic and elastic potential can be determined. However, a kind of pseudo-absorption can also occur when multislice calculations are done with an insufficient number of beams [24]. A small number of beams in the calculations neglects high spatial frequencies which contain more detailed structure information. If the number of beams is insufficient, the intensity in the calculations leaks out so that the circles become spirals as shown in Fig. 6. In the major simulation packages used today, such as MacTempas [19] (used in this paper) or EMS [25], this pseudo-absorption can be easily controlled through the total wave intensity. For example, one can input different maximum frequency values ($g_{\text{max}}$) in the parameters box in MacTempas. Fig. 3 shows Argand plots for different atom column types where the required spatial frequency is indicated. The heavier the atom column the larger the number of beams required. Including frequencies larger than those indicated does not significantly affect the circles in the Argand plot. Thus, from the Argand plot, one can also visualize the accuracy of multislice calculations. Fig. 7 shows the Argand plot of simulated exit waves based on the multislice theory of a Au[1 0 0] wedge-shape structure with the presence of absorption and different spatial frequency ranges. It is shown that both an insufficient spatial frequency range as well as absorption turns the Argand plot from a circle into a spiral. In other words, calculations with insufficient spatial frequencies may be interpreted as artificial absorption. Therefore, special caution has to be taken to ensure the accuracy of the calculations. A similar spiral when the absorption is due to the effect of the objective aperture is also shown in [9].

5. Defocus

Neighboring atom columns may have different numbers of atoms so that they protrude from the exit surface with different “heights”. Therefore, the defocus value of the exit face to the image may vary from column to column. Even if the sample is very thin, this effect may be seen in the Argand plot. Moreover, even if the number of atoms in a column is the same, the vertical positions of the columns can be different. An example of such a structure is schematically shown in Fig. 8. Ideally, the exit wave corresponds to the plane where the electrons just leave the atom columns. For the structure in Fig. 8, however, there is no flat plane where the exit wave can be reconstructed. Therefore, exit waves reconstructed with conventional methods such as focal-series reconstruction [26–28] have to be used with caution since they only reconstruct the exit wave at a hypothetical plane close to the exit surface. As such, there may be a distance between the plane of reconstruction and the true exit. This will affect the wave function. In order to avoid that the defocus effect would be
confused with the number of atoms in a column, we have considered this effect in the channelling very carefully.

The effect of defocus on the exit wave for a particular atom column can mathematically be described as a convolution product of Eq. (1) and the defocus propagator. The defocus propagator is given by the inverse Fourier transform of

\[ p(g) = \exp(-i\pi \alpha g^2), \]

(10)

where \( e \) is the defocus distance, \( g \) the two-dimensional reciprocal wave vector and \( g = |g| \). From the convolution theorem, it can be shown that the defocused exit wave of one column is given by

\[ \psi(r,z) = 1 + c_e \frac{2\sqrt{2\pi}a}{4\pi a^2 + i\alpha} \left\{ \exp\left(-\frac{R_0^2}{4a^2 + i\alpha}\right) \right\} \left[ \exp\left(-i\frac{E}{E_0}kz\right) - 1 \right]. \]

(11)

Using this equation, an exit wave of a Au[100] atom column containing 5 atoms has been simulated for different defocus values ranging from -150 to +150 Å with a defocus step of 1. The complex values at the atom column of each defocused exit wave are shown in Fig. 9(a). This curve is called a defocus circle.

In order to make the calculations more robust, we will average neighboring pixels around the column position, or in other words, all pixels in a disk around the center of the column will be included. Since the transformation to the Argand plot is a linear operation it commutes with the averaging of the pixel values and with the linear effect of the defocus propagator, which makes the measurement more robust. Then the averaged defocused exit wave becomes

\[ \langle \psi(r,z) \rangle = 1 + c_e \frac{2\sqrt{2\pi}a}{\pi R_0} \left\{ 1 - \exp\left(-\frac{R_0^2}{4\alpha}\right) \right\} \left[ \exp\left(-i\frac{E}{E_0}kz\right) - 1 \right]. \]

(12)

with \( R_0 \) the radius of the averaged circular area. For small defocus values, \( \alpha \ll 4a^2 \), the averaged defocused exit wave is approximately equal to

\[ \langle \psi(r,z) \rangle = 1 + c_e \frac{2\sqrt{2\pi}a}{\pi R_0} \left\{ 1 - \exp\left(-\frac{R_0^2}{4\alpha}\right) \right\} \left[ \exp\left(-i\frac{E}{E_0}kz\right) - 1 \right] \]

(13)
and for large defocus values, $\frac{\varepsilon \lambda}{\pi} \gg 4\sigma^2$, it approaches

$$\langle \psi(r, z) \rangle = 1 - i e^{-2\sqrt{2} \pi a \frac{\varepsilon \lambda}{\pi}} \exp \left(-i \frac{E_0}{E} k z \right) - 1. \quad (14)$$

For small defocus values, the averaged defocused exit wave contains two terms: the first term is the wave function without defocus and the second term incorporates the effect of defocus. As shown in Fig. 9(a), the defocus points lie on a circular curve crossing the zero-defocus point ($e = 0$) on the Argand circle and the origin (entrance wave) for large defocus values. This procedure can be repeated for every atom column. In Fig. 9(b), it is shown that the defocus curve of each thickness gives an extra circle on the Argand plot. From this figure, it is clear that the defocus circles for atom columns containing different number of atoms pass through a common point, which is the entrance wave. If the common reference point is known, a straight line from this point through the center of the defocus circle intercepts the curve at zero defocus. This procedure can be used to determine the exact defocus values of the exit face of each atom column by calculating the distance between the point corresponding to the column exit wave and this zero defocus point. As such, a defocus map of the sample can be obtained providing the surface roughness of the exit face at atomic scale. It should be mentioned that other techniques exist. For example, in [29], the interfacial roughness of a Si quantum well is mapped using the QUANTITEM technique, which is cross-correlating the entire image with a unit cell template obtained from the image intensities. In our proposed method, the surface roughness is analyzed from the defocus values atom column by atom column.

In order to test this effect, we have done multislice simulations using a Au wedge structure shown in Fig. 10(a). The structure is arranged as a fcc structure with lattice parameter 4 Å. The atoms of the $a$ columns are set at corner positions and top face-center positions, whereas the atoms of the $b$ columns are at the centers of the side faces. Atoms in the $a$ and $b$ columns are vertically displaced by half a lattice parameter. The atom columns contain 1–5 atoms. The electron beam direction is along [001]. The exit wave of this structure is taken at the plane where the electron wave leaves the atom columns containing 5 atoms. As a result, atom columns containing less than 5 atoms are defocused. The calculation of the defocus circles for every column is shown in Fig. 10(b). Each column is defocused from $-15$ to $+15$ Å, with focus step 1 Å. It is clearly shown that the defocus circles for the $a$ and $b$ columns are separated and slightly rotated. This is due to the difference of the entrance wave for each column type. The entrance wave for the $b$ columns is not a perfect plane wave but contains a “shadow” wave from the neighboring protruding $a$ columns. The Argand plot thus also contains information about the surface roughness of the entrance face. Moreover, according to the theory, all defocus circles still cross a common point. The centers of these defocus circles are shown in Fig. 10(c). A defocus map of the structure can be obtained Fig. 10(d), which directly reveals the vertical surface structure at the exit face, by applying the procedure described above.

Fig. 10. (a) A simulated wedge-shape fcc structure using Au atoms. The $a$ and $b$ columns containing 1–5 atoms are half a lattice parameter in difference. (b) The defocus circles for every atom column. Each atom column is defocused from $-15$ to $+15$ Å, with focus step 1 Å. (c) The centers of the defocus circles. (d) The defocus map of the wedge-shape structure.
The zero-defocus point of the defocus circle is not easily determined for experimental data since it is sensitive to the exact value of the entrance wave. Therefore, in experiments, the defocus circles for columns with different number of atoms may not cross a single common point and, furthermore, for objects that have only one type of column with the same number of atoms. Therefore, in the following, we describe a method in which we determine the zero-defocus point from the positions in the Argand plot of an equidistant focal series for the same column. For simplicity, only the first part related to defocus of the interaction wave in Eq. (11) is considered at the column position \( r = 0 \):

\[
\frac{2\sqrt{2}a}{4\pi a^2 + i\lambda}.
\]

Substituting Eq. (3) into the above equation, we have

\[
2 \frac{1}{1 + \frac{\epsilon \lambda}{\pi 4 a^2}} = 2 \frac{1}{1 + \chi}
\]

where \( \chi = \frac{\epsilon \lambda}{\pi 4 a^2} \). In what follows, the pre-factor 2 is excluded for simplicity since it is only a scaling factor for the radius of the circle. A simulated defocus curve according to Eq. (16) is shown in Fig. 11(a). The angle \( \theta \), the angle between two neighboring defocus points drawn from the center of the defocus circle, is calculated from

\[
\tan \theta = \frac{y}{x - \frac{1}{2}} = \frac{2\chi}{1 - \chi^2},
\]

where \( x \) and \( y \) are the real and imaginary parts from Eq. (16), respectively, and

\[
\chi = \tan \frac{\theta}{2}.
\]

Note that in the derivation of Eq. (18), the negative sign resulting from the imaginary part of Eq. (17) does not change the value of the equation but only the quadrant of the angle. As shown in the figure, the distance between two neighboring defocus points is smaller as the defocus values are further from the zero-defocus point. In other words, the distance between two neighboring defocus points is the largest at zero defocus. Since the focus step is equidistant the derivative of Eq. (18) equals to 1:

\[
\frac{d \tan \frac{\theta}{2}}{d \chi} = 1.
\]
The left-hand side of Eq. (19) can be rewritten as

\[ \frac{1}{\cos^2 \theta} \frac{d\theta}{dz} \]  

(20)

Thus,

\[ \frac{d\theta}{dz} = -2 \cos \frac{\theta}{2} \cos \theta + 1. \]  

(21)

As shown in Fig. 11(b), a simple cosine curve can be drawn from the defocus points and the zero-defocus point can be obtained by estimating the maximum point of the curve. Since the focus step is very small (\(~1 \text{Å}\)),

\[ \tan \theta = \theta - AB, \]  

(22)

where \(AB\) is the distance between any two neighboring defocus points on the defocus circle (an example is indicated in Fig. 11(a)). This cosine curve can be obtained in an easier way by plotting a distance curve between every two neighboring defocus points. Fig. 12(a) is a simulated defocus curve according to Eq. (11) for an Au atom column containing 3 atoms. As shown in Fig. 12(b), the zero-defocus point can also be obtained from the maximum peak of the distance curve. In a forthcoming paper, this method will be used to find the zero-defocus point from experimental data in order to obtain the actual defocus value for every atom column.

6. Conclusions

In this paper, the channelling theory has been used to derive an extended method to determine structure information, the number of atoms along the beam direction and the surface roughness, from an exit wave using an Argand plot. If the mass density of a column varies, the curve is a circular shape on the Argand plot with a radius which is independent of the atom column type. The angular increment, corresponding to an extra number of atoms along the beam direction and the surface roughness, is very small.

\[ \cos \theta = \frac{1}{H + d 

Thus, $d\theta/dz = 2\cos^2 \theta/2 - \cos \theta + 1$. As shown in Fig. 11(b), a simple cosine curve can be drawn from the defocus points and the zero-defocus point can be obtained by estimating the maximum point of the curve. Since the focus step is very small (~1 Å),

\[ \tan \theta = \theta - AB, \]  

(22)

where \(AB\) is the distance between any two neighboring defocus points on the defocus circle (an example is indicated in Fig. 11(a)). This cosine curve can be obtained in an easier way by plotting a distance curve between every two neighboring defocus points. Fig. 12(a) is a simulated defocus curve according to Eq. (11) for an Au atom column containing 3 atoms. As shown in Fig. 12(b), the zero-defocus point can also be obtained from the maximum peak of the distance curve. In a forthcoming paper, this method will be used to find the zero-defocus point from experimental data in order to obtain the actual defocus value for every atom column.

6. Conclusions

In this paper, the channelling theory has been used to derive an extended method to determine structure information, the number of atoms along the beam direction and the surface roughness, from an exit wave using an Argand plot. If the mass density of a column varies, the curve is a circular shape on the Argand plot with a radius which is independent of the atom column type. The angular increment, corresponding to an extra number of atoms along the beam direction and the surface roughness, is very small (~1 Å),

\[ \tan \theta = \theta - AB, \]  

(22)

where \(AB\) is the distance between any two neighboring defocus points on the defocus circle (an example is indicated in Fig. 11(a)). This cosine curve can be obtained in an easier way by plotting a distance curve between every two neighboring defocus points. Fig. 12(a) is a simulated defocus curve according to Eq. (11) for an Au atom column containing 3 atoms. As shown in Fig. 12(b), the zero-defocus point can also be obtained from the maximum peak of the distance curve. In a forthcoming paper, this method will be used to find the zero-defocus point from experimental data in order to obtain the actual defocus value for every atom column.

References


[14] Private discussion with Dr. C. Kisielowski from NCEM in Berkeley, USA.


Acknowledgments

A. Wang and S. Van Aert are grateful to the Fund for Scientific Research - Flanders (FWO) (Project no. G.0188.08). A. Wang and F.R. Chen would like to give thanks to the National Science Council of Taiwan (Project no. 96–2628–E–007–017–MY3).