Direct methods applied to phase retrieval in high resolution transmission electron microscopy

Usha Bhat and Ranjan Datta
International Center for Materials Science, Chemistry and Physics of Materials Unit, JNCASR, Bangalore 560064, India
E-mail: ranjan@jncasr.ac.in

Keywords: phase retrieval, HRTEM, image reconstruction, fourier transformation

Abstract
An alternative reconstruction method is proposed for retrieving the object exit wave function (OEW) directly from the recorded image intensity pattern in high resolution transmission electron microscopy (HRTEM). The method is based on applying a modified intensity equation representing the HRTEM image. A comparative discussion is provided between the existing methodologies involved in the reconstruction of OEW, off-axis electron holography and the present proposal. Phase shift extracted from the experimental images of MoS2, BN and ZnO are found to be in excellent agreement for most of the atom types investigated with the theoretical reference values. Additionally, it is shown that the Fourier series expansion of diffraction pattern is effective in retrieving the isolated and periodic image functions of certain forms directly. However, for aperiodic object information e.g., defects, dopants, edges etc, the first method works in entirety.

1. Introduction
Phase (φ) is the fundamental quantity in quantitative high-resolution transmission electron interferometry [1–6]. The change in phase (Δφ) of the probe electron wave after interacting with the object potential leads to the formation of specific intensity patterns in the respective image and diffraction planes. The intensity patterns can be recorded through any suitable imaging device, e.g., a charge-coupled device (CCD) camera. In case of transmission electron diffraction, Δφ carry information not only on the crystallographic phase of the scattering potential along high symmetry orientation but also on the electrostatic potential from the atoms and crystals that is essential for the identification and counting of atoms and extracting information on the chemical bonding from the experimental images [7–13]. There are few existing experimental and associated numerical phase retrieval methods in high resolution transmission electron microscopy e.g., through focal image series reconstruction based on HRTEM [2, 14, 15], atomic resolution off-axis electron holography [16, 17], fitting object function directly by intensive computer simulation or so-called direct method [9], transport of intensity equation (TIE) [18], electron channeling [19–21], phase velocity [7], diffraction imaging and ptychography [22–26]. Some of the techniques mentioned above work both at medium and atomic resolutions, and the TIE method was developed for medium resolution applications.

Among various methods, complexities involved in OEW reconstruction based on conventional through focal HRTEM image series are addressed in the present discussion and its analogy and differences with respect to off-axis electron holography are highlighted. Defocus HRTEM image is equivalent to in-line holography. In off-axis electron holography, the wave interference occurs at an angle between the reference and the object waves. The retrieval of OEW is performed at first by Fourier transformation (FT) of the image containing electron interference pattern, then selecting one of the two side bands (SBs) which are complex conjugate (or twin image) to each other followed by inverse-FT [4]. This procedure isolates the CB and the twin image wavefunctions from the recorded image. The phase and amplitude can be evaluated either by the arctan function corresponding to inverse-FT or fitting the inverse-FT pattern with the help of image simulation. As the starting data is the image,
therefore, the FT procedure does not lead to loss of any information in terms of crystallographic phase and inverse FT can return the image intensity pattern. Deconvolution of coherent aberration envelope can be performed posteriori that modifies the aberration figure in the image plane.

In the case of HRTEM, the CB and twin image wave functions overlap in the diffraction plane, and the FT procedure cannot separate them in the frequency space unlike off-axis electron holography [27, 28]. Therefore, almost all the reconstruction methods in HRTEM involve multiplying the image intensity pattern recorded at different focus settings with a complex filter function consisting of a coherent aberration envelope corresponding to each focus and then summing up over all the images to eliminate the unwanted twin image and non-linear components from the wanted OEW function. The complexities involved with reducing the unwanted components from the wanted OEW function led to the development of several reconstruction algorithms [15].

However, in the present manuscript it is demonstrated that by marginally modifying the form of the intensity equation describing the HRTEM image and concomitant justification, it is possible to retrieve the phase information directly from the atomic resolution images. The same equation in an intermediate form based on the wavefunction formalism is used in the existing OEW reconstruction procedures in the case of in-line and off-axis electron holography. Two types of phases are of importance here. One is the crystallographic phase describing the distribution of scattering potential corresponding to the isolated or periodic arrangement of atoms forming specific image or diffraction pattern in the respective planes. The second one is the change in phase of probe electron wave due to strength of atomic potential, which is equivalent to electron density in x-ray crystallography, that determines the intensity of the dots in the image pattern. To understand the workings of the modified equation and underlying complexities involved in the existing methods, at first the flow of phase information from the object to the Gaussian image and diffraction planes are briefly discussed based on the various existing formalism i.e. Fresnel Huygens construction, WPOA, transmitted wave function involving atom scattering factor derived from the Schrödinger integral equation and Fourier method similar to Abbe’s approach of image formation [29–31]. Then, elemental principles behind existing OEW reconstruction methods are described. The change in phase and associated retrieval of OEW function due to object, lens imperfections, and geometry of interference are dependent on the ways the change in phase is incorporated and the form of the equations used in the mathematical formulations. However, it is shown that the phase information both in terms of crystallographic phase at atomic resolution and object potential is not lost in the high-resolution image intensity pattern for both in-line and off-axis electron holography where a reference wave is present. The presence of a reference wave ensures the modulation of the intensity pattern and retaining the phase information in the image plane. The results obtained for MoS2, BN and ZnO are in excellent agreement for most of the atom types with the theoretical reference values within the specified resolution limits [32]. The working of the method is in accordance with the Born rule of probability amplitude and addresses the twin image issue in in-line holography. The effect of thickness and associated multiple scattering phenomena are not considered in the present measurement and analysis. However, both the phase information is completely lost for the intensity pattern recorded in the diffraction plane. But it is shown that it is also possible to retrieve the complete phase information directly from the diffracted intensity alone. This latter method is based on the cosine based Fourier series expansion of the diffraction pattern that is similar to the zero phase retrieval based on Patterson function in x-ray crystallography for small molecular systems [33, 34]. This latter method works exactly for a certain type of functions both in isolated and periodic form. However, for aperiodic object information e.g., defects and dopants the first method works in its entirety.

2. Experimental techniques

TEM samples of MoS2 and BN layered materials were prepared by ultrasonicication of respective powders (Sigma Aldrich) for 40 min to exfoliate monolayers and a few layers followed by drop casting on a holey carbon grid [35]. Cross sectional TEM specimen of ZnO epitaxial thin film was prepared by first mechanical thinning and then Ar ion polishing to perforation. HRTEM images were recorded in an aberration corrected FEI TITAN® 80–300 kV transmission electron microscope operating at 300 kV at optimum settings of Cs and defocus [36, 37]. ZnO epitaxial thin film was grown homoeptaxially on a (0001) ZnO substrate by pulsed laser deposition (PLD) following a specific growth procedure as described in [6, 38].

3. Brief overview on various conceptual elements of image formation, flow of phase information and OEW reconstruction

In the following sub-sections, at first, various fundamental concepts and associated mathematical formalism corresponding to the flow of phase information from the object plane to the image and diffraction planes are
described. Subsequently, a discussion is provided on the existing reconstruction methods based on HRTEM through focal image series before proceeding for the alternative proposals on OEW reconstruction in section 4.

3.1. Methods based on Fresnel-Huygens construction and Fourier transformation

A typical plane wave front in 1D with amplitude 1 and initial reference phase $\phi = 0$ is represented by the following complex expression

$$\psi_0(x) = \exp(2\pi ik.x) = \cos(2\pi k.x) + i \sin(2\pi k.x) = Re + Im$$ (1)

Where, $\psi_0$ is the plane wave function, $k = \frac{1}{\lambda}$ is the wave vector, and $x$ is the spatial dimension over which the amplitude of the stationary wave is spread with cosinusoidal oscillation. The wave nature of light is observed e.g., during the passage through the slits under appropriate experimental conditions such as coherent illumination, monochromaticity, relative dimension of illumination wavelength, slit and periodicity. In the case of an electron, which is a matter wave, the wave nature manifests in the form of the diffraction pattern of crystal, Fresnel edge pattern and off-axis electron holography fringes.

The Fraunhofer pattern in the diffraction plane can be calculated within the celebrated Fresnel-Huygens construction. The origin of diffraction phenomena is due to the superposition of various spherical wavelets emanating from object forming envelope waves with modulation in amplitude and phase along different momentum directions. However, the geometry of interference and associated phase correlation between various wave vectors is different for the near-field and the far-field regimes (sections S1.1 and 1.2 (available online at stacks.iop.org/JPCO/6/045007/mmedia)). There exist analytical expressions for various types of aperture geometries through the evaluation of the Fraunhofer integral [29]. The Fraunhofer pattern can also be calculated by absolute Fourier transformation (abs-FT) of the real aperture function [figure 1], but the intensity pattern is given as a function of frequency $k$ of the plane wave basis and needs to be calibrated either with the scattering angle $\theta$ derived from analytical formula or through a calibrated diffraction pattern.

The Fourier based method is analogous to the physical picture of Abbe’s hypothesis of image formation, where the point of interaction between probe and specimen generates outgoing waves with a continuous range of frequencies, waves with higher frequency propagate along larger scattering angles (section S1.3). Whereas, in the Fourier method, the object function is expanded into various frequencies of plane waves with no information on the scattering angle. Both the scattering processes involve the presence of various frequency components of the outgoing waves that are reminiscent of the picture of Bloch wave propagation of probe electrons in a periodic crystal but with a discrete range of frequencies having symmetry with the crystal periodicity [39]. The inverse Fourier transformation returns the aperture function, if the $Re(x, k)$ and $Im(x, k)$ components of the Fourier waves are known. The phase of the wave function that carries the object information, can be calculated by arctan function, i.e., $\tan^{-1}\left(\frac{Im(k \text{ or } x)}{Re(k \text{ or } x)}\right)$ corresponding to diffraction plane and image plane with variables $k$ and $x$, respectively.

For atomic systems, the electrostatic potential replaces the slit object function. The FT of isolated and periodic atomic potential in the diffraction plane is the atom scattering factor $f(k)$ and structure factor $F(g)$, respectively. In this case, the transmitted wave function can be derived from the Schrödinger integral equation involving interference between incident plane wave and outgoing spherical wave with amplitude factor given by

---

**Figure 1.** (a) Example seven slit periodic object with slit width $a$ and slit periodicity $b$. Fraunhofer pattern calculated (b) based on Fourier transformation and (c) by applying analytical formula which is based on the physical picture of constructive and destructive interference between waves along various directions. The phase angle for each Fourier frequency is shown in graph (d).
the atom scattering factor (supp. section 1.4) [31, 40]. The image intensity of the atom can then be calculated following Huygens’s principle in Fraunhofer approximation as given in Equation S16 with the consideration of resolution limiting aberrations acting as coherent envelopes. The phase contrast image calculated varies weakly with atomic number and the peak phase shift \( \varphi_{\text{max}}(\text{rad}) \) follows \( \sim Z^{0.6} \approx Z^{0.7} \), where \( Z \) is the atomic number [32]. This is possible as the formulation of the scattering factor has the information on the atomic potential which corresponds to both crystallographic and object phase information. The strength of the scattering potential or object phase will only modify the magnitude of specific intensity pattern due to crystallographic phase. Generally, this so-called phase information is lost in the experimental diffraction pattern and only the absolute of \( \text{Re} \) and \( \text{Im} \) components or abs-FT are known. However, it is shown in section 4.B that it is possible to retrieve the image intensity function of certain types directly from the diffraction pattern. Once the image function is known, the object phase can be determined according to the standard procedures as described in sec. C or by the alternative method presented in section 4.A.

3.2. Zernike phase object and WPOA

The method described here yields an image intensity pattern at or near the Gaussian image plane. The definition of weak phase object is introduced here and its implication on the recorded intensity pattern. For pure phase or weakly scattering object, the one-dimensional object function is mathematically represented by a complex function of the form \( F(x) = e^{i\phi(x)} \) or \( \sim 1 + i\phi(x) \) for small \( \phi \), where \( \phi \) is a real function corresponding to the discrete or periodic transparent object and is known as WPOA. The effect of the Zernike phase plate modifies the intensity of the object wave that depends linearly on the phase change due to object according to \( I(x) = 1 \pm 2\phi(x) \). This is the foundation of the Zernike’s phase contrast theory [29]. In fact, the object function is real, it is the replica of the object carried by the probe plane wave that is a complex wave function (similar to the Fourier waves propagation with the information of the object function) and absolute of this function after the interaction is imaged on the recording device. Therefore, in the recorded image pattern the information on the object function or crystallographic phase is preserved. The phase corresponding to OEW may be determined from this pattern using the arctan function if the corresponding object wave is known.

However, a similar approach for the atomic system incorporated the change in phase in the plane wave illumination in terms of change in magnitude of the wave vector \( \Delta k = \frac{1}{\Delta \lambda} \) [30, 31]. This brings about transmission function and under axial imaging condition is further convoluted with the point spread function \((psf)\) in the image plane or multiplication with the phase contrast transfer function (PCTF) function which acts as a coherent envelope in the diffraction plane incorporating the effect of aberration phase shift to form the final image [27, 30]. The intensity depends linearly on the phase change due to object potential after ignoring the higher order interaction terms (equation S21) that is similar to Zernike like transfer if the complete envelope function is set to 1. Considering that the spatial resolution criteria of the instrument is being met, again, the information on the crystallographic phase is not lost in the image intensity pattern. However, if the WPOA approximation is not considered, then the information related to object function \( \phi(x) \) and complete phase information is lost [section S1.4.2]. Kindly note that, the phase shift due to aberration cannot be added in the trigonometric function in the diffraction plane as that shifts the wave, rather it is used as coherent envelope or a frequency filter function. In real space, the psf due to PCTF gives weight to the intensity (e.g. Scherzer transfer will have maximum weight for the optimum value of spherical aberration and defocus) that depends on the integration value of PCTF over the limits of reciprocal vector acting as a frequency filter and other coherent aberration e.g., aperture function determines the broadening or full width at half maximum (FWHM) of the potential function [27, 41]. As the convolution procedure changes the magnitude of the resultant function significantly, a flux balance approach is helpful to observe the decrease in intensity and increase in FWHM due to aberration in comparison to the ideal image free from any aberration (figure S7) [41].

However, the above description of WPOA does not draw any analogy in terms of interference geometry between Gabor’s in-line holography, Fresnel diffraction geometry and defocus HRTEM image except transmission function between in-line holography and transmission function derived based on the change in magnitude of electron wavelength (compare equation (S24) and (S25)). However, instead of considering the change in momentum vector direction due to interaction with the object potential provides insights in all three pictures in terms of the geometry of interference (sections S1.1 and S1.2). An account of comparative image simulation based on various methods described above can be found in [41]. In brief, due to various ways of considering the flow of phase information by different methods, the simulated image patterns do not correspond to each other for the same atom type.

3.3. Various existing procedures on HRTEM focal series reconstruction

Reconstructions methods associated with the HRTEM through focal image series aim to retrieve the unknown phase \( \phi \) of the OEW function of the form \( \psi_1 = A(x,y)e^{i\phi(x,y)} \). The OEW function can be used to interpret the
object structure \( f(x, y, z) \) based on the model methods as described in the previous sections. Generally, the illumination is considered to be monochromatic and for semi-monochromatic waves, partial coherence theory appears into the stage. While the theoretical model behind the off-axis electron holography technique to retrieve the OEW function is straightforward (section S2.2), the methods corresponding to inline holography using through focal HRTEM image series are much more elaborate involving intensive data refinement and fitting procedures \( [17] \). This is due to the nature of the fitting equation considered to retrieve the \( \psi_i \) that attempts to reduce the effect of complex conjugate \( \psi_i^* \) and non-linear image components \( \psi_i \psi_i^* \) from the recorded image set [equations (5) and (7)].

There exist a few different focus variation methods e.g., Wiener formulation by Schiske, 3D paraboloid method (PM), maximum likelihood method (ML) and various numerical schemes associated with them \([2, 3, 42, 43]\). Saxton showed the equivalence between the different reconstruction techniques in terms of equivalence in the form of restoring filter applied to the intensity expression in Fourier space to retrieve the wave function. The simplified form of the restoring filter is given as

\[
t_\gamma(k) = \frac{1}{N} \exp\{i\gamma(k)\}
\]

and the associated aberration function

\[
\gamma_n(k) = \pi C_3 \lambda k^4 - \pi \lambda z_n k^2
\]

finally, the restored wave function is written as

\[
\psi(k) = \sum_n I_n(k) t_\gamma(k)
\]

Where, \( C_3 \) is the third order spherical aberration, \( \lambda \) is the wavelength, \( z_n \) is the defocus corresponding to the \( n \)th image, and \( k \) is the spatial frequency. The application of restoring filter is similar to deconvolving the effect of aberration. The inverse-FT of equation (4) will return the wave function in the image plane. This is only possible if one starts with the intensity pattern recorded in the image plane with sufficient spatial resolution and during inverse-FT, spatial information on the crystallographic phase is preserved through \( R(x, k) \) and \( Im(x, k) \) of Fourier waves. The pattern of the wave function in terms of phase \( \phi \) and amplitude \( A \) thus obtained required to be fitted with the model calculation for further interpretation \([9]\).

However, Saxton derived the final form of the wave function without considering any restoration filter from a different start in the form of the intensity expression as given in equation (5). The approach provides insight into the effect of restoration on dispersing the effect of the conjugate wave function and non-linear image component. Under perfect coherence, the image intensity at some focus \( z \) near the Gaussian image plane is written as

\[
i(x, z) = |1 + \psi|_1^2 = 1 + \psi_i(x, z) + \psi_i^*(x, z) + h(x, z)
\]

Where, \( \psi_i(x, z) \) is the wanted part or the OEW, \( h(x, z) = \psi_i(x, z) \psi_i^*(x, z) \) and \( \psi_i^*(x, z) \) are the unwanted non-linear term and complex conjugate, respectively. \( \psi_i^*(x, z) \) is the twin image of \( \psi_i(x, z) \). Kindly note that the reference wave \( \psi_i \) is set to 1 by considering axial illumination (\( k_0 = 1 \)) in the above expression.

The 2D Fourier transformation of equation (5) and then explicitly adding the dependence on defocus results in the following expression.

\[
I(k, z) = \delta(k) + \psi(k) \exp(\pi i \lambda z k^2) + \psi^*(-k) \exp(-\pi i \lambda z k^2) + H(k, z)
\]

Where, \( H(k, z) \) is the non-linear component that describes the autocorrelation between the two linear terms in the reciprocal space. The paraboloid method as proposed by van Dyck can be derived from the above expression by taking 3D-FT with respect to defocus where the wave function and its complex conjugate follow the reflected parabola from the reference diffraction plane \([2]\).

Now, within the coherent detection i.e., multiplying by corresponding phase conjugate and summing over \( N \) images after assuming \( k \) is non-zero which allows to omit the delta function and for a constant \( k \), the image intensity becomes

\[
\sum_n I_n \exp(-\pi i \lambda z_n k^2) = \psi \sum_n 1 + \psi^* \sum_n \exp(-2\pi i \lambda z_n k^2) + \sum_n H_n \exp(-\pi i \lambda z_n k^2)
\]

From the above expression, it was concluded that the wave functions \( \psi \) and \( \psi^* \) accumulate to \( N \) and \( \sqrt{N} \) times to its original value, respectively \([3]\). The non-linear term \( H_n \) behave randomly with the defocus. Thus, by dividing the above equation by \( N \), the wanted OEW function may be recovered. Equation (7) is widely considered in almost all the through focus image series OEW reconstruction. And all the efforts on developing image reconstruction codes primarily deal with eliminating the non-linear and complex conjugate terms and finding the best fit with the model calculation. However, the presence of complex conjugate and non-linear terms will always be present depending on the extent their weights are subdued.
Nonetheless, there exists a contention, after complete evaluation of equation (5), yields the following final form
\[
I_{\text{in line}} = |\psi_0 + \psi_i|^2 = A_0^2 + A_i^2 + 2A_0A_i\cos(\phi_i - \phi_0) \tag{8}
\]
Where, \( \psi \) is replaced with the wave function of the form \( A(x, y) e^{i\phi(\alpha,\beta)} \), describing the image intensity pattern based on self-interference between reference incident and scattered waves within the picture of single electron wave interference phenomena [Equation S49 & S50]. The expression has the similarity with the wave interference between the reference and object waves in off-axis electron holography (section S2.2). However, in off-axis geometry, there is an additional phase term \( Qx \) due to wave interference at an angle that gives rise to spatial modulation in the interference field [5]. Moreover, equation (5) is in an intermediate state, which is used to eliminate the effect of twin image and non-linear terms by working in the diffraction plane [section S2.1]. However, the existence of the final form of the expression implies that fitting the intensity equation alone and evaluating the phase term should, in principle, allow extracting the relative phase change from the image plane. For more details on associated twin image wave functions and applicability of equation (8), see section 4. The results based on the existing schemes can be found in [6, 17] and do not show any systematic trends with the sample thickness.

Finally, the approach based on partial coherence theory considers the effect of finite source size, chromatic defocus spread, current voltage fluctuation of the instrument, objective aperture size and wave aberration function [1, 44, 45]. The reconstruction method based on partial coherence theory is known as an iterative linear restoration that addresses the residual non-linear term. Repeatedly applying a linear restoring filter from the subtraction of the calculated non-linear term improves the initially guessed wave function. Various derivations available based on partial coherence theory are given in section S2.4. One can notice that the coherence and interference phase shift in the formalism has the origin in convolution procedure in real space and cross correlation in Fourier space.

4. Alternative methods on the phase retrieval

In this section, alternative proposals are introduced for the retrieval of the \( \phi \) and \( A \) of OEW function from the image and diffraction planes, respectively. The methods are straightforward and do not require through focal image series acquisition, similar to off-axis electron holography, where only a single image embedded with the hologram is sufficient. Addressing the effect of large defocus together with various coherent aberration envelopes, and thickness on the rich variations in phase and image intensity pattern by the alternative procedure will be part of a future discussion as this requires a different viewpoint from the traditional approach due to unique experimental observation overlooked in the past.

4.1. Recovering phase from the HRTEM image intensity pattern

The method described here works on atomic resolution HRTEM image recorded under suitable imaging conditions i.e., with a particular combination of spherical aberration coefficient \( C_s \) and defocus \( \Delta f \) that sets the optimum contrast and resolution.

The object phase can be recovered by applying equation (9) as given below, which is a modified form of equation (8) describing the image intensity pattern in HRTEM within few nanometers [section S2.2.1] from the Gaussian image plane.
\[
I_{\text{in line}}(x, y) = |\psi_0 + \psi_i|^2 = A_0^2 + A_i^2 + 2\alpha A_0A_i \sin \{\phi_i(x, y)\} \\
= \alpha_0 + \beta_0 + 2\sqrt{\alpha_0 \times \beta_0} \sin \{\phi_i(x, y)\} \tag{9}
\]
Where, \( I_0 \) is the mean vacuum intensity. The factors \( \alpha \) and \( \beta \) represent the fractions of direct and scattered part of the intensity and can be determined by analyzing the image pattern, where \( \alpha + \beta = 1 \). Typical values of \( \alpha \) and \( \beta \) are found to be \( \sim 0.88/0.95 \) and \( 0.12/0.05 \), respectively from the experimental images of MoS\(_2\)/BN [section S5].

And the intensity expression corresponding to the off-axis electron hologram is given by
\[
I_{\text{off axis}} = 1 + a^2(x, y) + 2a(x, y) \cos \{2\pi Qx + \phi(x, y)\} \tag{10}
\]

The appearance of sinusoidal function in equation (9) in contrast to equation (8) is due to the OEW phase term having a relative phase term with respect to vacuum phase \( \phi_0 \) in addition to \( \phi_i \) due to object potential in the case of HRTEM. This phase difference is absent in off-axis electron holography [compare equations (9) and (10)] and is generally considered to be \( \pi / 2 \) i.e., \( \Delta \phi = \pi / 2 - \phi_0 \) that sets the vacuum phase value to zero. In case of off-axis geometry, the two halves of the wave on either side of the bi-prism carry the same vacuum reference phase term and gets eliminated in the final intensity except for part of the wave that carries the object phase.
information [section S2.2]. For an intuitive physical picture on the origin of such a \( \pi/2 \) phase shift between diffracted and primary incident waves based on interference geometry is provided in section S 2.2.1 [figure S11]. The modifications incorporated in equation (9) are essential and yield excellent results from various experimental images with different types of atoms present for a given crystal. For more discussion on the factors \( \alpha \) and \( \beta \) and the particular use of amplitude terms and various forms of equation (10) and their relationship with experimental image see section S5.

Now, both the equations are founded on the physical picture of self-interference of single electron wave function involving interaction in the form of phase change with the object potential [28]. The interaction is mostly elastic due to the fast probe electron, and the small probability of inelastic interaction is useful for the analytical techniques. The primary difference between equations (8) and (10) is that the additional phase term of \( 2\pi Qx \) that appears in the trigonometric function in off-axis electron holography due to wave interference at an angle (section S2.2). This particular phase term oscillates with the spatial coordinate \( x \) over the field of the electron hologram. The phase term \( \phi(x, y) \) corresponding to OEW phase is acquired by another half of the wave while passing through the sample. This OEW phase manifests as holographic fringe bending in the image plane, e.g., relative fringe bending between vacuum and MgO crystal and differently striped thickness of object structure, that allows to determine the mean inner potential (MIP) at medium resolution [4, 46].

Another difference between the two holography techniques is the twin image. It is well known that for in-line holography, separation of twin images is an issue, whereas, for off-axis geometry, they get separated at \( \pm Q \) in the frequency space. Ptychography has also been used to eliminate the twin image issue. However, in the image plane, they superimpose on top of each other in both the techniques. In the standard practice of OEW reconstruction, equations (9) and (10) are generally written in an intermediate state (see equation (5) for in-line and Equation S54 for off-axis) having both OEW function \( \psi \) and its conjugate \( \psi^* \) in consistent with the holographic principle of image formation [29]. The motivation behind this was based on the concept that the twin image components can be separated from the direct component (DC) if allowed to propagate along the scattering direction i.e., from the image plane to the diffraction plane and emphasizing the direct retrieval of the wave function. However, it is argued here that this does not pose an issue if one wishes to work in the image plane of the intensity pattern to retrieve directly both the phase and amplitude of OEW directly. The choice and limitations between the diffraction plane and image plane can be understood by Heisenberg’s uncertainty principle, i.e., only momentum and position variables of the OEW are known in the diffraction plane and image plane at a time, respectively.

Now, from equations (9) and (10), it is evident that the phase information is preserved in the final form of the intensity equation that is responsible for the intensity modulation as described by the trigonometric function. At the same time, no information on the wave functions or twin images and DC component are available. The application of the final form of intensity expression instead of wave function based intermediate state of the equation can be guided by the Born rule in Quantum Mechanics. Born rule states that it is the probability density \( \psi_0 \) given by the square of the probability amplitude of particle’s wave function \( \psi \) that is real and observable quantity during the measurement and not the associated complex wave function which acts as a state vector [47, 48]. There exists controversies in the literature on the measurement of such state vector which is a complex quantity but can be constructed based on the information of probability density [48–50]. Born rule is an important link between the abstract mathematical formalism of quantum theory based on the complex wave function and the experimental measurement and constitutes an essential part of the Copenhagen interpretation of quantum mechanics [51]. Therefore, the final form of the intensity equation is equivalent to measuring the total probability density due to contribution from all the three components of wave functions, i.e., DC part, wanted wave function \( \psi \) and its complex conjugate \( \psi^* \). And in the process of recording, the information on the wave functions is lost, recovering of which is emphasized in all the reconstruction approaches through intermediate state of the equations. Retaining the phase term is inherent to self-interference and resulting modulation in intensity pattern and similar to the off-axis electron holography fringe bending due to object potential in the image plane. The predicament of evaluating the local phase information from real space image in terms of fringe bending primarily comes from the restriction on spatial resolution and will be addressed in the forthcoming discussion along with additional aspects already mentioned earlier. In standard practice, the image is Fourier transformed and then a digital aperture function is used in that plane that takes care of the resolution. The use of intermediate state of the equation to disperse the twin images and DC part in the diffraction plane made the earlier methods elaborate and complex for HRTEM reconstruction where twin images overlap on the same reciprocal space and required to enhance the weight of \( \psi \) over the \( \psi^* \) and \( \psi_0^* \) terms through summation over many images. Though the information on the wave functions is lost, however, the phase term is preserved in the final state of the equation which are the same for both \( \psi \) and \( \psi^* \) but having opposite sign.

Now, the twin images are the exact copy of each other and superimposes at the Gaussian image plane and do not cause any loss in information in terms of phase and amplitude. This can be understood with the help of off-axis electron holography, where the inverse FT corresponding to each side band preserves the same intensity.
pattern. In inline holography, the twin images propagate along the opposite direction with defocus similar to the parabola picture of OEW reconstruction where complex conjugate pairs lie on a reflected parabola with respect to the diffraction plane and focus variation. This picture is reminiscent of the transactional interpretation of quantum mechanics where the wave function and its complex conjugate experience phase change in the opposite directions equivalent to the forward and backward propagation in time [52–54].

Now proceeding for the experimental reconstruction of the OEW function, at first, the phase can be determined directly by applying equation (9). If the quantity inside the arcsine function is more than 1, then this should be divided by 1 and the computer program should read the quotient as phase jump by the amount (quotient × π/2) plus the phase corresponding to the remainder and then calculate the total phase change and corresponding counts of the atoms accordingly. The π/2 phase jump does not appear for one-layer thick MoS2 and BN but is encountered in 3D crystal like ZnO where thickness includes several atoms in the column. Next, the amplitude corresponding to ψ needs to be evaluated from the image intensity. The total amplitude of the recorded image is simply the square root of it but not equal to the amplitude of the single OEW function. As already mentioned that the total amplitude in the image intensity pattern has three components, DC, and two twin image components for off-axis and DC and only diffracted components in case of in-line holography. In the case of in-line holography, the twin image components overlap with the DC part. The DC component has approximately one order of magnitude higher total intensity compared to individual twin images from SBs and diffracted FT spots as measured directly from the FT of the images by placing a region of interest around various spots and SBs and evaluating the sum. Thus, the intensity of the OEW function is approximately given by the factor β for HRTEM, and this will be further divided into halves between two twin image components in case of off-axis holography. The amplitude of ψ_HRTEM = ψ0 + ψ1 is given by the square root of I_{in\, line}(x, y) and only amplitude component A1 associated with trigonometric function is shown in reconstructed wavefunctions, where A1 and A0 are the magnitudes of individualdiffracted and reference complex wave functions, respectively. Details on the evaluation on the amplitude from the intensity pattern, discussion on the interpretation of amplitude of isolated wavefunction, and on the appropriate form of conventional intensity expression of atomic resolution off-axis hologram and HRTEM are given in section S2.2. Kindly note that the subtraction procedure to adjust the intensity value is automatically taken care of by equation (9) while calculating the phase. Therefore, it is possible to reconstruct the OEW function from the information obtained from the measurement as described above i.e., phase and amplitude.

We have analyzed selected experimental images of MoS2, BN and ZnO using equation (9) and the reconstructed phase and amplitude images are presented in figure 2 and 3 for MoS2 and BN, respectively. One can see from the phase images that the vacuum phase is zero and phase values are high and low depending on the brightness in the dots which represents the periodic arrangement of atoms in the lattice. Theoretical peak phase shift values extracted from [32] for the resolution range 1–0.5 Å are given in table 1 for the atoms of interest in the present investigation. According to this table, the peak phase shift ϕ_{max} (rad) follows \sim Z^{0.6} \sim Z^{0.7}, where Z is the atomic number [32]. The exponent of Z increases with increasing spatial resolution resulting in a higher peak phase shift value for a given atomic number. Though the trend can be complicated depending on the valence electron filling and for specific atoms with higher Z can have smaller contrast compared to atoms with lower Znext to each other in the periodic table. From the present reconstruction method, a peak phase shift value of \sim 0.56 and 0.45 rad are obtained from the Mo and S atom positions, respectively throughout the image, which are little higher than the theoretical values. Kindly note that for MoS2 there are two S atoms along the [0001] projection. Therefore, the phase shift value is almost double compared to the single S theoretical phase shift value. The values are 0.06 and 0.21 rad, for B and N atoms, respectively which are close match with the
theoretical estimation. However, the value for N atom appear to be almost double than the theoretical reference value, this might because of the strong dependence of $\alpha$ and $\beta$ values on the local intensity maximum, sample contamination and resolution dependence of theoretical reference values. Results on ZnO and bi-layer MoS$_2$ are given in figure S13 and S14. For ZnO thin film, as the sample thickness increases from the vacuum edge of the specimen, a systematic increase in phase and atoms numbers can be observed. As an example the phase value corresponding to Zn column changes from 1.23 to 2.3 rad from the edge to the interior of the specimen. This corresponds to a change in atom number/thickness from 4 (1.3 nm) 8 (2.6 nm), respectively. For the MoS$_2$ bilayer sample, a AB type of stacking is observed. The weak spot in between the strong hexagonal white dots are from S atoms. The intensity of dots corresponding to joint Mo and S atom columns are stronger and intensity of dots corresponding to S columns are weaker. The phase value corresponding to weaker dots $\sim 0$ and from bright dots phase values of 0.67 and 0.5 rad are obtained, does not help to identify the atoms. This is because of image intensity depends on defocus [41] for more details and the intensity has a strong role to play to obtain the correct phase number according to equation (9). This will be part of a future communication on the reconstruction based on thicker crystal. The uncertainty in phase and amplitude values from the experimental measurement is $\sim 0.006$.

Table 1. Theoretical peak phase shift extracted for resolution in the range 1 - 0.5 Å [35] along with experimentally reconstructed values.

| Atoms (Z) | Theoretical phase shift (rad) | Reconstructed phase shift (rad) |
|----------|-------------------------------|-------------------------------|
| B (5)    | 0.076–0.078                   | 0.06                          |
| N (7)    | 0.100–0.102                   | 0.21                          |
| O (8)    | 0.106–0.108                   | —                             |
| S (16)   | 0.164–0.23                    | 0.2                           |
| Zn (30)  | 0.242–0.41                    | —                             |
| Mo (42)  | 0.300–0.51                    | 0.56                          |

The reconstructed phase numbers calculated by equation (9) for Zn and O atoms in ZnO are in good agreement with the earlier report using atomic resolution off-axis electron holography [6]. However, the reconstruction from the HRTEM through focus series using the Gerchberg-Saxton scheme as implemented in MacTempas package did not yield expected variation in phase with sample thickness unlike the application of equation (9) yield systematic variation in phase with sample thickness [6, 17]. For thicker sample, channeling theory and diffraction imaging based reconstructions are being utilized at present [20, 21, 27]. However, more robust and straight forward reconstruction scheme considering relatively thicker specimen and large defocus will be part of a separate communication as this requires a different fundamental start than traditional approaches [41].

The method described above using equation (9), in fact translates the intensity information into a phase value and an excellent trend is obtained for images with low and high Z compounds giving lower and higher phase shift values, respectively without any additional data refinement procedures. Reference vacuum is not required as one can approximate the mean of atomic resolution image intensity as mean vacuum intensity. Intensity pattern if at all present other than OEW, e.g., Fresnel fringe, electron interference fringe in off-axis electron holography should be removed from the image data before applying equation (9). These unwanted patterns are generally removed from the image before the final analysis in case of off-axis electron holography.
4.2. Recovering image function from diffracted intensity by Fourier series expansion

In this sub-section, the recovery of image function containing information on both crystallographic and object phase is described if the information is available only in the diffraction plane. The real image function \( f(x) \) of both symmetric non-periodic and periodic in particular forms e.g., Gaussian, reciprocal etc. can be retrieved by cosine Fourier series expansion of the absolute FFT of such functions followed by summation over all frequencies. As already mentioned, that the absolute FFT and Fraunhofer pattern are equivalent to each other. The procedure is similar to the zero phase reconstruction using the Patterson function in x-ray crystallography applicable to smaller size molecules \([33, 34]\). Result is presented for the example Gaussian function in 1D both in isolated and periodic form (figure 4). Illustrations of other type of functions and 2D forms are given in section S6.

The unique feature underlying the workings of this reconstruction method is that for most functions the integration of imaginary part over spatial parameter or abs-Im component is negligible compared to real part or abs-Re \([\text{figure S9}]\). Therefore, the original function \( f(x) \) can be retrieved completely by the following real Fourier series alone. If \( F(k) \) is the absolute FFT of \( f(x) \), then \( f(x) \) can be retrieved from the \( F(k) \) directly via following real Fourier series expansion

\[
f(x) = \text{absolute} \left( \sum_{n=-k}^{n=k} \frac{1}{n} C_{n} \cos(2\pi nx) \right)
\]

(11)

Where, \( n = \) number of data points in the frequency axis, \( k \) is the frequency, \( C_{k} = \) absolute of Fourier transformations or diffracted intensity at some frequency \( k \), and \( x \) is the 1D spatial coordinate over which real image function will be defined.

For periodic function in 1D, \( f(x) = f(x \pm px) \), where \( p = 0, 1, 2, \ldots \), one need to ensure that the range of \( n \) should be \( \pm \frac{2p}{M} \), where \( M = \) number of data points both in real and diffraction space. In the case of a 2D isolated function, equation (11) will have another summation over \( n_{y} \) for the second orthogonal axis and the modified Fourier series equation for isolated and periodic 2D function are given in supp. Section 7 \([\text{equation S85 and S86}]\), where the functional form of cosine functions changes depending on the isolated 2D and periodic 2D functions in order to get an exact fit.

The method described above works well for both isolated and periodic functions. From the image intensity function, one can then apply the first method to retrieve the OEW and associated object phase related to potential. However, if there is any inhomogeneity in the distribution of periodic function e.g., dopants, edges, interfaces etc then the accompanying information will be required from the image intensity pattern and the method described in the previous section works suitably.
5. Conclusions

In conclusion, alternative reconstruction methods to retrieve the phase and amplitude of OEW in HRTEM imaging are introduced. The first method is based on directly applying a modified HRTEM intensity equation to retrieve the phase of OEW from the image. The reconstruction results in terms of peak phase shift values are in excellent agreement for most of the atom types with the theoretical estimation. The second method described is applicable for retrieving the image intensity function from the information available in the diffraction plane for both isolated and periodic functions.

Acknowledgments

The authors at JNCASR sincerely acknowledge ICMS, JNCASR for the financial support. Ranjan Datta would like to give a special thanks to Deutsche Forschungsgemeinschaft (DFG) for providing funding for an illuminating and exceptional research visit in TU-Berlin in the year 2016.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

ORCID iDs

Ranjan Datta © https://orcid.org/0000-0002-8553-3724

References

[1] Coene W, Jansen G, Op De Beeck M and Van Dyck D 1992 Phase retrieval through focus variation for ultra-resolution in field-emission transmission electron microscopy Phys. Rev. Lett. 69 3743
[2] Van Dyck D and Op De Beeck M 1990 New direct methods for phase and structure retrieval in HREM Proc. of the Xlth Int. Congress on Electron Microscopy 26–7
[3] Saxton W O 1994 What is the focus variation method? is it new? is it direct? Ultramicroscopy 55 171–81
[4] Lichte H and Lehmann M 2008 Electron holography - basics and applications reports prog Phys. 71 016102
[5] Geiger D, Lichte H, Linck M and Lehmann M 2008 Electron holography with a cs-corrected transmission electron microscope Microsc. Microanal. 14 68
[6] Bhat U and Datta R 2019 Quantitative counting of Zn and O atoms by atomic resolution off-axis and in-line electron holography J. Appl. Phys. 125 154902
[7] Van Dyck D, Jinschek J R and Chen F R 2012 Big bang tomography as a new route to atomic-resolution electron tomography Nature 486 243–6
[8] Jin C, Lin F, Suenaga K and Iijima S 2009 Fabrication of a freestanding boron nitride single layer and its defect assignments Phys. Rev. Lett. 102 3
[9] Jia C L, Mi S B, Barthel J, Wang D W, Dunin-Borkowski R E, Urban K W and Thust A 2014 Determination of the 3D shape of a nanoscale crystal with atomic resolution from a single image Nat. Mater. 13 1044
[10] Jia C L, Lentzen M and Urban K 2003 Atomic-resolution imaging of oxygen in perovskite ceramics Science 299 870
[11] Meyer J C et al 2011 Experimental analysis of charge redistribution due to chemical bonding for high-resolution transmission electron microscopy Nat. Mater. 10 209
[12] Jia C L, Mi S B, Urban K, Vrejoiu I, Alexe M and Hesse D 2008 Atomic-scale study of electric dipoles near charged and uncharged domain walls in ferroelectric films Nat. Mater. 7 57
[13] Midgley P A and Dunin-Borkowski R E 2009 Electron tomography and holography in materials science Nat. Mater. 8 271
[14] Meyer R R, Kirkland A I and Saxton W O 2002 A new method for the determination of the wave aberration function for high resolution TEM 1 Measurement of the Symmetric Aberrations Ultramicroscopy 92 89–109
[15] Carter C B and Williams D B 2016 Transmission Electron Microscopy: Diffraction, Imaging, and Spectrometry (Berlin: Springer)
[16] Genz F, Niermann T, Buijsse B, Freitag B and Lehmann M 2014 Advanced double-biprism holography with atomic resolution Ultramicroscopy 147 33
[17] Niermann T and Lehmann M 2016 Holographic focal series: differences between inline and off-axis electron holography at atomic resolution J. Phys. D: Appl. Phys. 49 194002
[18] Ishizuka K and Allman B 2005 Phase measurement of atomic resolution image using transport of intensity equation J. Electron Microsc. 54 191
[19] Chen F R, Van Dyck D and Kiesielowski C 2016 In-line three-dimensional holography of nanocrystalline objects at atomic resolution Nat. Commun. 7 1
[20] Wang A, Chen F R, Van Aert S and Van Dyck D 2010 Direct structure inversion from exit waves: I. Theory and simulations Ultramicroscopy 110 527
[21] Wang A, Chen F R, Van Aert S and Van Dyck D 2012 Direct structure inversion from exit waves: II. A practical example Ultramicroscopy 116 77
[22] Hope V W 1969 Beugung im inhomogenen primärstrahlwellenfeld. I. prinzip einer phasenmessung von elektronenbeugungsinterferenzen Acta Cryst. A25 495
[23] Scott M C, Chen C C, Mecklenburg M, Zhu C, Xu R, Ecius P, Dahmen U, Regan B C and Miao J 2012 Electron Tomography at 2.4-angstrom resolution Nature 483 444
[24] Putkunz C T et al 2012 Atom-scale pychographic electron diffractive imaging of boron nitride cones Physical Review Letters 108 073901
[25] Nellist P D, McCallum B C and Rodenberg J M 1995 Resolution beyond the ‘information limit’ in transmission electron microscopy Nature 374 630
[26] Hie F, Rodenberg J M, Maiden A M, Sweeney F and Midgley P A 2010 Wave-front phase retrieval in transmission electron microscopy via pychography Physical Review B 82 121415
[27] Spence J C H 2003 High-Resolution Electron Microscopy 3rd Edition (Oxford Science Publications)
[28] Fuhr B and Howe J M 2008 Transmission Electron Microscopy and Diffractometry of Materials (Berlin: Springer)
[29] Born M and Wolf E 1999 Principles of Optics 7th ed (Cambridge: Cambridge University Press)
[30] Williams D B and Carter C B 2009 Transmission Electron Microscopy: A Textbook for Materials Science 2nd ed. (Berlin: Springer)
[31] Kirkland E J 2010 Advanced Computing in Electron Microscopy: Second Edition (Berlin: Springer)
[32] Linck M, Lichte H and Lehmann M 2006 Off-axis electron holography: materials analysis at atomic resolution Int. J. Mat. Res. 97 890
[33] Patterson A I 1934 A Fourier series method for the determination of the components of interatomic distances in crystals Phys. Rev. 46 572
[34] Rius J 2014 Application of Patterson-function direct methods to materials characterization IUCr 1 291
[35] Sahu R et al 2017 Nature of low dimensional structural modulations and relative phase stability in Re5Mo(W)32S2 transition metal dichalcogenide alloys J. Appl. Phys. 121 105101
[36] Dileep K, Sahu R, Sarkar S, Peter S C and Datta R 2016 Layer specific optical band gap measurement at nanoscale in MoS2 and ReS2 van Der Waals compounds by high resolution electron energy loss spectroscopy J. Appl. Phys. 119 114309
[37] Datta R, Kanuri S, Karthik S V, Mazumdar D, Ma J X and Gupta A 2010 Formation of antiphase domains in NiFe2O4 thin films deposited on different substrates Appl. Phys. Lett. 97 071907
[38] Loukya B, Sowjanya P, Dileep K, Shripa R, Kanuri S, Panchakarla I S and Datta R 2011 Controlling structural quality of ZnO thin film on C-plane sapphire during pulsed laser deposition J. Cryst. Growth 329 20
[39] Loukya B, Negi D S, Dileep K, Pachauri N, Gupta A and Datta R 2015 Effect of Bloch wave electron propagation and momentum-resolution signal detection on the quantitative and site-specific electron magnetic chiral dichroism of magnetic spinel oxide thin films Phys. Rev. B 91 134412
[40] Scherzer O 1949 The theoretical resolution limit of the electron microscope J. Appl. Phys. 20 20
[41] Bhat U and Datta R 2021 Image simulation in high resolution transmission electron microscopy considering atom as an electrostatic interferometer J. Phys. Commun. 5 085004
[42] Schiske F 2002 Image reconstruction by means of focus series J. Microsc. 207 154
[43] Van Dyck D and Op De Beeck M 1990 New Direct Methods for Phase and Structure Revival in HREM Proc. of the XIIth Int. Congress for Electron Microscopy 26–7
[44] Frank J 1973 The envelope of electron microscopic transfer functions for partially coherent illumination Optik 38 519
[45] Ishizuka K 1980 Contrast Transfer Ultramicroscopy 5 55
[46] Linck M, Freitag B, Kuijawa S, Lehmann M and Niermann T 2012 State of the art in atomic resolution off-axis electron holography Ultramicroscopy 116 13
[47] Born M 1926 Zur quantenmechanik der Stoßvorgänge Zeitschrift Für Phys. 37 863
[48] Born M The Statistical Interpretation of Quantum Mechanics, Nobel Lecture (https://nobelprize.org/uploads/2018/06/born-lecture.pdf)
[49] Zuo J M, Kim M, O’Keefe M and Spence J C H 1999 Direct observation of d-orbital holes and Cu–Cu bonding in Cu2O Nature 401 49
[50] Humphreys C J 1999 Physical chemistry: electrons seen in orbit Nature 401 21
[51] Pham B Q and Gordon M S 2017 Can orbitals really be observed in scanning tunneling microscopy experiments? J. Phys. Chem. A 121 4851
[52] Siddiqui S and Singh C 2017 How diverse are physics instructors’ attitudes and approaches to teaching undergraduate level quantum mechanics? Eur. J. Phys. 38 035703
[53] Kastner R E 2012 The New Transactional Interpretation of Quantum Mechanics: The Reality of Possibility (Cambridge: Cambridge University Press)
[54] Cramer J G 1997 Quantum nonlocality and the possibility of superluminal effects NASA Breakthr. Propuls. Phys. Work. Proc. 1