Atomic Resolution Electron Holography

K. Ishizuka
Toyo University, Japan

T. Tanji
Hitachi Ltd., Japan

A. Tonomura
Hitachi Ltd., Japan

Follow this and additional works at: https://digitalcommons.usu.edu/microscopy

Part of the Biology Commons

Recommended Citation
Available at: https://digitalcommons.usu.edu/microscopy/vol1992/iss6/38

This Article is brought to you for free and open access by the Western Dairy Center at DigitalCommons@USU. It has been accepted for inclusion in Scanning Microscopy by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.
ATOMIC RESOLUTION ELECTRON HOLOGRAPHY

K. Ishizuka**, T. Tanji** and A. Tonomura**

Tonomura Electron Wavefront Project, JRDC
c/o Faculty of Engineering, Toyo University, Kawagoe, Saitama 350, Japan
**Advanced Research Laboratory, Hitachi Ltd., Hatoyama, Saitama 350-03 Japan

Abstract

It has been demonstrated that electron holography is a very powerful tool to investigate an electromagnetic potential in medium resolution, since the phase of an electron wave is approximately proportional to the potential. Now, electron holography is at the second stage of development: to establish holography at atomic resolution and further to realize Gabor’s idea to improve the resolution restricted by the spherical aberration of the objective lens. We investigate the possibility of electron holography to get information at atomic resolution by computer simulations as well as by digital processing of electron holograms. We show that the phase distribution has more resemblance to the specimen structure than the amplitude distribution. We also compare electron holography with electron microscopy from an image processing point of view.

Key Words: Electron holography, Electron microscopy, Atomic resolution, High resolution, Phase information, Multislice method, Dynamical scattering, Fast Fourier Transform, Image processing.

Introduction

The resolution of the electron microscope has been restricted by the spherical aberration of an objective lens. According to Scherzer (1949), the resolution is proportional to $(Cs A^3)^{1/4}$, where $Cs$ is the spherical aberration coefficient and $A$ the wavelength of electrons. Therefore, the resolution of the electron microscope has been improved by decreasing the spherical aberration coefficient or the wavelength of electrons. However, this prescription of improvement of resolution has almost reached its limit.

On the other hand, the idea of holography has been proposed by Gabor (1949) in the early days of electron microscopy as a mean to improve resolution of an electron microscope by correcting the spherical aberration of an electron lens. However, since holography requires a coherent source of illumination, there has been little activity to improve the resolution of electron microscopy for a long time.

The holography requires a reference wave which can interfere with an object wave. Gabor has proposed to use an unscattered wave as the reference wave, but this inline hologram has severe restrictions such as conjugate image effects. The idea of holography finds its application in the field of light optics after the development of laser technology. Leith and Upatnieks (1962) have invented so called “off-axis” holography, where the reference wave is tilted with respect to the object wave. Here, the reference wave has been produced by splitting a light wave by a half mirror (amplitude division). The same thing has been tried by splitting an electron wave by using diffraction effect by single crystals with little success. In this case, temporal coherency determines the quality of holograms.
Another method to make the reference wave by splitting a wavefront with a biprism (wavefront division) has been developed by Möllenstedt and his coworkers (e.g. Möllenstedt and Dükér, 1955). Here, spatial coherency realized by a parallel illumination is most important. To obtain parallel illumination, we need an electron source with high brightness. The spatial coherency of the electron source has been improved within these two decades by using a field emission gun first used for STEM by Crewe, Wall and Langmore (1970).

Electron interferometry has been developed by using the electron biprism and the field emission gun. Some recent interesting applications of this technique are the measurement of electrostatic potentials near p-n junctions (Chen et al., 1989) and the demonstration of the Aharonov-Bohm effect (Tonomura et al., 1986).

Electron holography is coupled with reconstruction of an object wave from the electron interferogram and measurement of the reconstructed wave. Thus, the electron interferogram is the same as the electron hologram. However, to reconstruct the wave distribution without interference from the conjugate wave, the fringe spacing of the hologram is usually far less than that of the interferogram. Although this requires an electron gun which can give an illumination with higher spatial coherency, we can obtain more detailed information quantitatively in many cases by using the techniques developed in optical holography.

Up to present, electron holography achieved a great success to measure the electromagnetic potentials in medium resolution. Although some works in high resolution have been reported (e.g. Tonomura, Matsuda and Endo, 1979; Lichte, 1986), we have to say that the quality of the reconstructed image from the electron hologram is at present not so good as that of the corresponding electron micrograph taken by the same microscope. The next step of electron holography is thus to establish the technique to obtain a high quality reconstructed wave at atomic resolution and to realize Gabor's idea to break through the resolution limit achieved by present day electron microscopes.

**Computer simulation**

At atomic resolution, the specimen behaves as a phase object for incident electrons. Therefore, phase information is expected to give a more direct description of the specimen structure. By electron holography, we can obtain a phase distribution as well as an amplitude distribution from a reconstructed wave. In order to see what we can expect from the phase distribution, we calculate wave functions at some resolutions. Here, we use a model specimen of TiO₂: Tetragonal, \( a=b=0.45937, c=0.29581 \) nm. Fig. 1(a) and (b) show a potential distribution projected along c axis and its reciprocal lattice, respectively. Odd index reflections along both axes are absent, and titanium atoms only contribute to the reflections marked by small circles due to the specimen symmetry. Wave functions at several specimen thicknesses are calculated by the FFT multislice method (Ishizuka and Uyeda, 1977) taking account of dynamical scattering effect. An accelerating voltage of 200 kV is assumed. We also assume the Scherzer resolution limit of 4.2 nm⁻¹, which corresponds to the spherical aberration coefficient of 1.0 mm for this accelerating voltage.

Fig. 2 shows the amplitude and phase distributions of diffraction limited wave functions at the Scherzer resolution and at two other resolutions of 6.0 and 8.4 nm⁻¹ corresponding to \( \sqrt{2} \) and 2 times of the Scherzer resolution, respectively. These resolution limits are shown by the circles 1, 2 and 3 in Fig. 1 (b). The specimen thickness assumed is 10 unit cell, i.e. 2.96 nm. The phase distribution has...
more resemblance to the specimen structure than the amplitude distribution at 6.0 or 8.4 nm\(^{-1}\) resolution. The phase change between TiO\(_2\) clusters is almost zero in each figure, and shows dark contrast. The highest phase values, which appears at titanium positions in every image, are 0.33\(\pi\), 0.42\(\pi\) and 0.73\(\pi\) rad in Fig. 2 (a), (b) and (c), respectively. Oxygen atoms can only be observed from the phase distribution at 8.4 nm\(^{-1}\) resolution. The phase distribution calculated at one unit cell thickness for each resolution is approximately one tenth of the phase distribution at 10 unit cells. However, the amplitude distributions at one unit cell only show less than 1% change from unity for every resolution. This small decrease in the amplitude results from excluding high frequency component by the resolution limiting aperture. The images at higher resolution are expected from an ideal treatment of electron hologram at lower resolution by correcting the wave aberration. To do this, however, we have to estimate a spherical aberration coefficient.

<table>
<thead>
<tr>
<th>Spherical aberration coefficient</th>
<th>Resolution</th>
<th>Scherzer Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 mm</td>
<td>4.2 nm(^{-1})</td>
<td>55 nm</td>
</tr>
<tr>
<td>0.25</td>
<td>6.0</td>
<td>27.5</td>
</tr>
<tr>
<td>0.0625</td>
<td>8.4</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 1 Resolution and Scherzer focus at 200 kV

In electron microscopy, the resolution will be improved by reducing the spherical aberration coefficient. Since the resolution will be improved as \(Cs^{1/2}\), the resolutions of 6.0 and 8.4 nm\(^{-1}\) will respectively be obtained with one fourth and one sixteenth of the original spherical aberration coefficient (see Table 1). Fig.3 shows the best focus image (electron micrograph at Scherzer focus) for each of
Figure 4: (a) and (b): amplitude and phase distribution of the resolution limited wave function. Accelerating voltage: 200 kV, specimen thickness: 5.9 nm, resolution 8.4 nm$^{-1}$. (c) and (d): simulated electron holograms with two different carrier directions. The carrier frequencies are respectively 26.2 and 27.7 nm$^{-1}$. Note that an abrupt phase change indicated by the arrow in (b) results in irregular fringe patterns.

Figure 5: Phase diagrams illustrating phase discontinuity. Solid and dotted curves correspond to regularly spaced vertical and horizontal parallel lines of the complex amplitude distribution, where the amplitude approaches zero. Here, some lines show a phase change of 2\(\pi\), when the curves cross the negative side of the x axis.

Thus, phase information is expected to be more valuable than amplitude or intensity information even at atomic dimension. However, the phases become indefinite when the amplitudes at these points decrease to zero at certain specimen thicknesses. This may happen when dynamic scattering becomes important.
respect to others. Then, some lines have a continuous phase, while others which cross the negative side of the x axis suffer a phase change of \(2\pi\). In the case where the amplitude exactly becomes zero at a certain point, we have a phase change of \(\pi\) by crossing the origin of the phase diagram. Thus, the phase distribution may have a discontinuity within a very small region, when the amplitude approaches to zero. This phase discontinuity can not be eliminated by adding or subtracting \(2\pi\), or rotating the axis of phase diagram. It should be noted that the wave function is continuous in complex plane, and both the real and imaginary parts of the wave function have no discontinuity even in this case. Thus, the phase distribution does not always have a physical meaning.

There are simple relations between the phase change and electromagnetic potentials (e.g. Feynman, Leighton and Sands, 1964). Here, the specimen is considered as a single layer phase object, and the exit wave function has the amplitude of unity everywhere. However, the phase discontinuity in the reconstructed image from the real electron holograms may happen even for such a specimen, since the wave aberration may give the image of small amplitude at some points. It should be noted again that the phase discontinuity may occur at the places where the amplitude approaches zero due not to the the computation error for phase calculation or the measurement problem from the bad signal-to-noise ratio (S/N) of the hologram, but to dynamical scattering and/or the lens aberration.

**Holography vs Microscopy**

We discuss here electron microscopy and electron holography from an image processing point of view. There are many algorithms for processing electron micrographs, most of which assume a linear relation between an image intensity \(l(r)\) and a complex object amplitude \(\phi(r)\) suffered from an aberration:

\[
l(r) = |1 + \phi(r)|^2 = 1 + 2\text{Re}[\phi(r)]
\]

This linearity assumption will hold when an amplitude of the wave function after suffering aberration does not deviate so much from unity. Thus, the linearity in the image processing depends on the resolution as well as aberration. The processing algorithms usually assume the weak phase object approximation (WPOA). This approximation depends on acceleration voltage and scattering power of materials, and will only be valid for very thin specimens. It should be noted that the amplitude at the dark region of Fig. 3 (b) and (c) is less than 0.5. Thus, the linearity assumption becomes questionable even at 3 nm thickness.

The "off-axis image hologram" \(H(r)\) will be recorded as an intensity distribution of a complex image amplitude \(\psi(r)\) plus a reference wave \(\exp(-i\mathbf{k} \cdot \mathbf{r})\):

\[
H(r) = |\exp(-i\mathbf{k} \cdot \mathbf{r}) + \psi(r)|^2
\]

\[
= 1 + |\psi(r)|^2 + \psi(r) \exp(+i\mathbf{k} \cdot \mathbf{r}) + \psi(r)^* \exp(-i\mathbf{k} \cdot \mathbf{r})
\]

center-band

Here, the complex image amplitude is the complex object amplitude plus an incident wave: \(\psi(r) = 1 + \phi(r)\). By Fourier transforming the hologram, the second and third terms are shifted by a carrier frequency \(k\) in Fourier space. Thus, we can select one of the side-bands, (+) or (-), if the carrier frequency is high enough to separate these three bands. Therefore, information obtained from a hologram always has a linear relation with the complex amplitude. It is not necessary to assume the weak phase object approximation: the scattering may be dynamical.

By using electron holography, we can automatically obtain an energy filtered image. The center-band corresponds to an electron micrograph \(|\psi(r)|^2\) plus a constant distribution due to the reference wave, and does not have an energy filtering effect. However, the side-bands are energy filtered due to interference between the reference wave and the image wave. The energy width of filtering is equal to the energy spread of incident electrons. Thus, information obtained from a side-band should have high contrast without any background due to inelastic scattering usually observed in an electron micrograph. An inelastic scattering event which will be filtered out is mainly plasmon loss. The mean free path for the plasmon loss is usually of the order of 100 nm (Egerton, 1986). Thus, for a thick specimen this energy filtering reduces the signal intensity of the side-bands, and results in the reconstructed image with a lower S/N compared with the electron micrograph.
Another disadvantage of electron holography is its high requirement for the spatial coherence of electrons. The wavefront division holography usually requires very high spatial coherency for the reference wave to interfere with the object wave at the observation plane. Furthermore, Fresnel diffraction effect from a biprism filament results in non-equidistant interference fringes, and makes processing difficult. To avoid this difficulty due to Fresnel diffraction, the interference region should be wider than a few tens of nm. Thus, a field emission gun is indispensable for electron holography. In the case of ordinary electron microscopy (CTEM), the requirement for spatial coherency is not so stringent, and a conventional thermionic gun has been commonly used for high resolution microscopy (Ishizuka, 1986).
Atomic resolution electron holography

Figure 7: Filter processing of MgO crystal. (a) and (b): original hologram and its Fourier transform. (c) and (d): amplitude and phase distributions of the reconstructed complex amplitude. Note that the side-bands shown by the small circle are totally inside the centerband indicated by the large circle. The scale bars in real and reciprocal spaces are 1 nm and 5 nm\(^{-1}\), respectively. The arrows in (b) indicate the reflections used in reconstruction.

Digital Processing

The electron holograms were taken with Hitachi HF-2000 microscope operated at 200 kV with a cold field emission gun. A negative of the hologram was digitized by a CCD camera into 512 by 512 pixels of 256 gray levels, and processed by a personal computer (an IBM PC/AT clone with Intel 80386 cpu). The main software we are using is SEMPER (Saxton, Pitt and Horner, 1979), and we have added some routines for processing the electron hologram. Fast Fourier transforms (FFT) were mainly performed by using an array processor of 20 MFLOPS installed into the computer.

Observation of crystal surface

The surface of the small magnesium oxide (MgO) crystal was observed by high resolution electron holography like a profile image technique (Tanji and Cowley, 1985). Fig. 6 (a) shows an electron hologram taken from MgO cube of approximately 100 nm. Here, [011] axis is parallel to the electron beam. Thus, the specimen is a wedge-shaped, and the thickness from right to left of this image is from 8 to 17 nm. The bi-prism is approximately parallel to (100) edge of the crystal. Fig. 6 (b) shows the amplitude of the spectrum of the hologram digitally calculated by FFT.

The electron micrograph outside the interference band indicates that the hologram was
K. Ishizuka, T. Tanji and A. Tonomura

taken at a little bit over-focus. Thus, we calculate the wave functions at several under-focuses by propagating the reconstructed wave within the computer. The wave function obtained by underfocusing 40 nm after the image reconstruction gives the most clear phase image. The amplitude and phase distributions at this defocusing are shown in Fig. 6 (c) and (d), respectively. Here, no attempt was made to correct the spherical aberration. The phase distribution shows an asymmetry especially at thick region. This probably results from small misalignment of the crystal relative to the electron beam, as demonstrated by the asymmetric spectrum distribution of the side-band shown in (b), and probably from an asymmetric wave aberration due to small misalignment of the optic axis. Fig. 6 (e) is a line profile of the phase distribution along the surface as shown in (d). We can clearly observe the surface with much less effect from Fresnel diffraction due to the specimen surface (Tanji et al., 1993).

Abrupt contrast changes from black to white are observed within the crystal region of the phase distribution shown in Fig. 6 (d), especially at the thick region (left part). These contrast changes correspond to phase change from \(-\pi\) to \(+\pi\). As explained above, we can find the points where the amplitude approaches to zero. Thus, we can not make this phase distribution continuous. This is partly due to an asymmetry of the reconstructed phase. However, the specimen is rather thick, and we can expect the phase discontinuity even from the hologram taken at the ideal condition as demonstrated above.

Filter processing of MgO crystal

Usually to separate the reconstructed wave from the other two terms in the case of a strong scattering object, the frequency of the interference fringe (carrier frequency) should be larger than three times of the resolution you want, as shown in Fig. 6 (b). However, if the specimen is a crystal, strong Bragg reflections will occur. Then, information of the center-band and two side-bands will be localized at lattice points. The lattices of the side-bands are identical to the center-band lattice at the origin, and shifted from the origin to the side-band center. They do not coincide each other provided that the direction or the periodicity of the carrier fringes is not commensurate with the crystal lattice. Thus, we can filter the side-band information from the center-band in the case of a crystal specimen even when the carrier frequency is equal to the required resolution.

Fig. 7 (a) and (b) show an electron hologram and its Fourier transform, respectively. Here, the carrier frequency is about the same as the resolution limited by the aperture indicated by the small circle in (b). Thus, the carrier frequency does not satisfy the ordinary requirement to separate the side-band from the center-band. However, the lattice of the center-band does not coincide with the lattice of the side-band, and we can easily filter out unwanted information. The amplitude and phase distributions of the reconstructed wave respectively shown in Fig. 7 (c) and (d) were obtained by filtering with masks of diameter of 0.95 nm\(^{-1}\) at each lattice point indicated by arrows in Fig. 7 (b). We have used Hanning window to reduce the effect of "leakage" (Brigham, 1974), which will become serious for a small mask in filtering.

Now, we can produce a fine interference fringe of less than 0.03 nm, which will be sufficient to reconstruct the wave up to 0.1 nm resolution (Tanji, Urata and Ishizuka, 1991). However, this filtering technique will find its application, because we can take the hologram at a lower magnification. Since a required magnification to take a hologram is determined by the highest spacing to be recorded, the magnification for this technique is one half of the ordinary treatment where the side-band is separated from the center-band.

Conclusions

We investigate the possibility of electron holography at atomic resolution by computer simulations as well as by digital processing of electron holograms. It has been shown that the phase distribution has more resemblance to the specimen structure than the amplitude distribution. However, it should be noted that the phase becomes indefinite when the amplitude decreases to zero due to dynamical scattering. In terms of image processing, electron holography is more general than electron microscopy. However, the requirement for high spatial coherency makes high resolution electron holography difficult.
Atomic resolution electron holography

References

Mollenstedt, G. and Düker, H. (1955) Fresnelscher Interferenzversuch mit einem Biprisma für Elektronenwellen (Fresnel interference studies with a biprism for electron waves), Naturwissenschaften 42, 41.

Ultramicroscopy 49, 259-264

Discussion with Reviewers

H. Lichte: What is the width of the hologram used to reconstruct the results given in Fig. 6? What is the maximum spatial frequency contributing to the reconstructed wave?
Authors: The magnification used to take the hologram is 900 k. The fringe spacing and the width of the interference band are 0.034 nm and 4.5 nm, respectively, on the specimen. Although the objective aperture used is 0.105 nm, the maximum spatial frequency contributing to the side-band is 0.127 nm (311 reflections).

Reviewer: The phase discontinuity may be overcome by avoiding zero value of wave function. In numerical simulation, some methods, e.g. mask can have an amplitude not equal to zero. But does it have physical meaning?
Authors: The phase discontinuity results from a basic characteristic of the wave when we describe the wave with the amplitude and phase. If the amplitude becomes to zero, the phase has no physical meaning. However, the fact that the amplitude approaches zero has the physical meaning. Therefore, even if we can make the amplitude not equal to zero by some methods, the resulting amplitude does not have the physical meaning.

H. Lichte and D. Van Dyck: The authors said that the specimen behaves as a phase object for incident electrons at atomic resolution. This is not necessarily true. For instance, the case of Au columns, the phase object approximation breaks down even at a few nm thickness and a noticeable amplitude modulation of the electron object wave is produced.
Authors: In the phase object approximation, every part of the specimen modulates only the phase of the electron wave. However, the propagation of the modulated electron wave through the specimen results in the amplitude
change. This is the heart of the physical optics of Cowley and Moodie (Acta Cryst. (1957) 10, 609). For instance, an optical lens considered as a phase object to the light produces the intensity modulation. Thus, the amplitude of the wave function may deviate from unity even for a phase object. Here, the single-phase-object approximation breaks down but not the phase object approximation itself.