Off-axis electron holography in an aberration-corrected transmission electron microscope

BY HANNES LICHTE*, DORIN GEIGER AND MARTIN LINCK

Triebenberg Laboratory, Institute of Structure Physics, Technische Universität Dresden, Germany

Electron holography allows the reconstruction of the complete electron wave, and hence offers the possibility of correcting aberrations. In fact, this was shown by means of the uncorrected CM30 Special Tübingen transmission electron microscope (TEM), revealing, after numerical aberration correction, a resolution of approximately 0.1 nm, both in amplitude and phase. However, it turned out that the results suffer from a comparably poor signal-to-noise ratio. The reason is that the limited coherent electron current, given by gun brightness, has to illuminate a width of at least four times the point-spread function given by the aberrations. As, using the hardware corrector, the point-spread function shrinks considerably, the current density increases and the signal-to-noise ratio improves correspondingly. Furthermore, the phase shift at the atomic dimensions found in the image plane also increases because the collection efficiency of the optics increases with resolution. In total, the signals of atomically fine structures are better defined for quantitative evaluation. In fact, the results achieved by electron holography in a Tecnai F20 Cs-corr TEM confirm this.

Keywords: transmission electron microscopy; electron holography; atomic resolution; noise problems; signal resolution; aberration correction

1. Introduction

Electron holography was proposed by Gabor (1948) as a lensless imaging method avoiding the deleterious effect of the lens aberrations, namely the spherical aberration, which was shown unavoidable by Scherzer (1936). Gabor’s basic idea was to propagate the electron wave issued by the object in free space according to the wave equation. If the propagated wave is detected in an interference pattern (‘hologram’) completely in both amplitude and phase, it can be revived from the hologram and be back propagated to the object exit face (‘reconstruction’), again according to the wave equation; for the reconstruction, every wave (e.g. light or a ‘numerical’ wave) may be used. As no lenses

*Author for correspondence (hannes.lichte@triebenberg.de).

One contribution of 14 to a Discussion Meeting Issue ‘New possibilities with aberration-corrected electron microscopy.’

This journal is © 2009 The Royal Society
are involved in the process, lens aberrations do not play any role. However, a coherent reference wave is needed for the holographic, i.e. interferometric, detection process.

In fact, Haine & Mulvey (1952) were the first to realize the holographic principle with electrons experimentally, yielding a lateral resolution of approximately 1 nm. At the end, they were limited by two reasons: the first is the emergence of two reconstructed images, which are complex conjugates, hence, they are reconstructed at opposite foci; they propagate essentially in the same direction, and hence, unavoidably, intermix with each other. This is the famous ‘twin-image problem’ of in-line holography, arising because the object and reference wave forming the hologram propagate in the same direction in space (‘in-line holography’). The second reason is the limited extension of spatial coherence in the hologram plane, which restricts the maximum diffraction angle that can be recorded coherently; larger angles, corresponding to finer object details, end up outside the hologram, and hence do not contribute to the reconstructed object wave.

The twin-image problem was solved by Leith & Upatnieks (1962), by superimposing the reference wave at an angle to the object wave (‘off-axis’) instead of in-line with the object wave. Consequently, the reconstructed twin images propagate at the mutual double angle, and hence can be separated in Fourier space. The problem of resolution, because of the limited extension of coherence, was overcome by Wahl (1975) in the Möllenstedt Institute in Tübingen, showing the principle of image-plane holography: if the hologram is recorded in an image plane, the limited extension of coherence limits the field of view but not the resolution. Combining the two principles, Wahl realized image-plane off-axis holography in the transmission electron microscope (TEM), using the Möllenstedt electron biprism (Möllenstedt & Düker 1956) as a beam splitter, superimposing the reference wave on the object wave in the image plane at an angle. Since then, off-axis electron holography has been developed and applied, mainly in Tübingen & Dresden (Lichte 1991a), Tokyo (Tonomura 1993), Braunschweig (Hanszen 1982), Bologna (Pozzi 1983), Tempe (McCartney & Smith 2007), Cambridge, Copenhagen (Midgley & Dunin-Borkowski 2009) and Toulouse (Hytch 2008).

2. Principle

The object exit wave leaving the object \( \text{obj}(\mathbf{r}) = a(\mathbf{r}) \exp(i\varphi(\mathbf{r})) \) is imaged in the image plane by the objective lens. An electron biprism inserted between the back focal plane and the image plane superimposes the image wave \( \text{ima}(\mathbf{r}) = A(\mathbf{r}) \exp(i\Phi(\mathbf{r})) \) with a coherent plane reference wave, resulting in the hologram

\[
I_{\text{hol}}(\mathbf{r}) = I_{\text{TEM}}(\mathbf{r}) + I_0 + 2C A(\mathbf{r}) A_0 \cos(2\pi q_0 \mathbf{r} + \Phi(\mathbf{r})),
\]

where \( C \) is the average fringe contrast given by coherence and disturbances, \( I_{\text{TEM}}(\mathbf{r}) \) is the conventional image intensity and \( I_0 = A_0^2 \) is the intensity of the empty reference wave (figure 1). The spacing of interference fringes \( s = 1/|q_0| \) and hologram width \( w \) can be chosen in a wide range by means of excitation of the biprism and the subsequent lenses. These fringes are modulated in contrast.
Electron holography with $C_s$ corrector

Figure 1. Setup for recording image-plane off-axis holograms. The object covering only half the object plane and the adjacent empty area are illuminated coherently. The object plane is imaged in the first image plane. By means of the electron biprism, the image wave and the adjacent empty wave are superimposed at an angle (‘off-axis’). In the image plane, an interference pattern (‘hologram’) arises, consisting of parallel fringes modulated in contrast and fringe spacing by image amplitude and phase, respectively.

...and position by the amplitude and phase of the image wave, respectively. It turns out that at least three fringes have to sample one period to be reconstructed. Consequently, the fringe spacing finally limits the lateral resolution obtainable in the reconstructed wave. For the object, fringe spacings of only a few 0.01 nm can be obtained, allowing the grasping of all information needed for a lateral resolution of approximately 0.1 nm.

From such a hologram, the image wave is reconstructed in the form of two images, an amplitude and a phase image. Numerical image processing performs the reconstruction, rendering all data quantitatively for further evaluation in terms of the object. The basics are outlined in detail in Lichte & Lehmann (2008).

3. Benefits of electron holography

Despite the problems, which are still under solution, electron holography nowadays offers substantial advantages over conventional TEM imaging. First of all, the data representing the object structure are complete in that one obtains

(i) amplitude and phase cleanly separated (figure 2),
(ii) quantitative data, which are gauged by means of an empty reference hologram additionally recorded without a specimen using the same parameters,
(iii) linear transfer of amplitude and phase from the object into the reconstructed wave according to the wave equation, and
(iv) pure elastic data due to perfect zero-loss filtering (less than $10^{-15}$ eV).

Phil. Trans. R. Soc. A (2009)
The complex object wave is retrieved by cutting out one sideband of the hologram spectrum and applying an inverse Fourier transform (FT$^{-1}$). Afterwards, the reconstructed complex wave can be evaluated in amplitude and phase separately and quantitatively.

Moreover, comprehensive wave optical image processing is applicable for evaluation of all kinds of information encoded in the object exit wave because the whole wave is available (Rau & Lichte 1998). This means

(i) \textit{a posteriori} numerical correction of all coherent aberrations, including higher order aberrations showing up in the future,
(ii) no crosstalk, i.e. the amplitude and phase can be interpreted in terms of physics and materials properties separately,
(iii) no delocalization, which presently gives rise to misinterpretation,
(iv) evaluation of amplitude and phase up to the respective noise limit, and
(v) full wave optical evaluation also in Fourier space, for example, by single-reflection analysis and nanodiffraction (figure 3).

Examples of a variety of applications elaborated in the Triebenberg laboratory are summarized in Lichte et al. (2007).

4. Role of aberrations in the transmission electron microscope

It is true that, in contrast to Gabor’s idea for avoiding lenses, for image-plane holography, lenses are needed. Therefore, the role of lens aberrations also has to be considered. Following the linear transfer theory, the image wave

\[ \text{ima}(r) = \text{obj}(r) \otimes \text{PSF}(r) \]

is given by the convolution of the object wave $\text{obj}(r) = a(r) \exp(i\varphi(r))$ with the complex point-spread function $\text{PSF}(r)$ arising in the image plane from lens aberrations; $r$ is the two-dimensional coordinate in the image plane and the
Electron holography with C₃ corrector

Figure 3. Holographic nanodiffraction. Like electron diffraction patterns, (b) the nanodiffraction patterns obtained by Fourier transforms of the respective areas of (a) the complex wave show the variation of excitation of the different opposite reflections. Therefore, thickness and tilt variations of the object across the field of view can be analysed from a holographic wave. This is not possible from a conventional image. Object: ZnTe in [110] orientation. Hologram recorded in the Philips CM30 Special Tübingen TEM.

Object plane. By the convolution, the image wave \( \text{ima}(r) = A(r) \exp(i \Phi(r)) \) is altered in amplitude and phase with respect to the object wave. Details can also be found in, for example, Lichte & Lehmann (2008).

The point-spread function describes the effect of the aberrations of the objective lens in real space as

\[
\text{PSF}(r) = \text{FT}^{-1}[\text{WTF}(q)],
\]

i.e. the inverse Fourier transform of the wave-transfer function

\[
\text{WTF}(q) = E^{sc}(q) E^{tc}(q) \exp(-i \chi(q)),
\]

defined in Fourier space with coordinate \( q \). It contributes with a damping of the Fourier components owing to the envelope functions \( E^{sc}(q) \) and \( E^{tc}(q) \) given by the deficiencies of spatial and temporal coherence, respectively. They damp the contributions of the respective spatial frequencies and hence destroy information, in particular, of the high spatial frequencies. The arising limit is called the information limit (figure 4).

Furthermore, the wave-transfer function introduces the phase shift \( \chi(q) \) (‘wave aberration’), distorting the wave front according to the ‘coherent’ aberrations, i.e. defocus, spherical aberrations, astigmatisms, comae etc; these are called coherent aberrations because they are effective, even under perfectly coherent illumination. Taking account of defocus \( C_1 \) and spherical aberration \( C_3 \) of the third order only,

\[
\chi(q) = 2\pi k \left( \left( \frac{C_1}{2} \right) \left( \frac{q}{k} \right)^2 + \left( \frac{C_3}{4} \right) \left( \frac{q}{k} \right)^4 \right)
\]

results in wavenumber \( k \). In the image plane, it gives rise to crosstalk between the amplitude and phase of the object wave, as well as to contrast reversal. Antisymmetric aberrations with \( \chi(q) = -\chi(-q) \), for example, from three-fold

\[\text{Phil. Trans. R. Soc. A} (2009)\]
astigmatism or axial coma, additionally would introduce a corresponding lateral shift of the spatial frequency components in the image wave. In neither of these cases, however, does the wave aberration destroy object information, as the above envelope functions do. This means that, in contrast to the envelope functions, the coherent aberrations can be corrected \textit{a posteriori}.

The complete description of aberrations by means of the point-spread function $PSF(r) = FT^{-1}[WTF(q)]$ mentioned above is, in general, difficult to handle explicitly in real space. This is because the Fourier transform in most cases—containing more than just defocusing resulting in the Fresnel propagator—cannot be performed analytically, and hence cannot be evaluated easily. Nevertheless, the amount of the corresponding delocalization, i.e. the radius of the point-spread function, can be estimated by making use of the fact that trajectories run orthogonally to the wave front. Therefore, in Fourier space, the local distortion of the wave front leads to a respective trajectory mistilt that is proportional to $\text{grad} \chi(q)$. Consequently, in real space, the delocalization of information can be estimated from the maximum trajectory mistilt found in the whole Fourier spectrum, extrapolated into the corresponding displacement in the image plane (figure 5). A detailed calculation (Lichte 1991b) results in a radius of the point-spread function given by $\text{psf} = \text{grad}_{\text{max}} \chi(q) / 2\pi$. To give an idea, the point-spread function of an uncorrected TEM at Scherzer conditions has a diameter of about five times the Scherzer resolution; for example, it amounts to 1 nm in a TEM with 0.2 nm point resolution.

Often, TEM users are not aware about delocalization because the appearance of seemingly highly resolved details, such as ‘atoms’, within the point-spread function is very seductive. These details appear because the coherently illuminated patch stretches further out than the psf, and hence, interference producing the fine details is visible; at incoherent illumination of the point-spread

![Figure 4. Phase-contrast transfer function at Scherzer’s defocus.](image)

The imaginary part of the wave-transfer function (WTF) basically characterizes contrast transfer from object phase to the image intensity. The oscillations, which are only present in the uncorrected TEM, restrict the interpretable resolution (Scherzer resolution) to below the highest spatial frequency $q_{\text{max}}$ transferred. $q_{\text{max}}$ is called the information limit given by the envelope functions $E^\text{sc}$ and $E^\text{tc}$ of the restricted spatial and temporal coherence, respectively. Please note that, for small spatial frequencies, no phase contrast is found, and hence large-area phase structures, such as extended electric or magnetic fields, are invisible.
Figure 5. Radius psf of the point-spread function PSF. The aberrations of the objective lens introduce an additional phase shift $\chi(q)$ to the ideal spherical wave. According to the trajectories, which are perpendicular to the aberrated wave front, the size of the PSF can be derived from the maximum trajectory mistilt $\theta_{\text{max}}$ corresponding to the maximum of the aberration gradient.

function, by appropriate blurring within the point-spread function, they would smear out and the atoms would not be resolved. In any case, the atoms are averaged, in both position and contrast, about the real atoms in the area of the point-spread function. Therefore, such an image is neither directly nor quantitatively interpretable, in particular not at interfaces and defects.

5. Holography in an aberrated transmission electron microscope

The aberrated image wave is reconstructed from the hologram completely by amplitude $A(r)$ and phase $\Phi(r)$, which, however, are scrambled with respect to the original object wave, and, in particular, averaged over the point-spread function. However, as the complete wave is at hand, it can be corrected for the coherent aberrations. These possibilities were realized for defocus by Wahl (1975) using light optical reconstruction. Nowadays, holographic aberration correction is carried out using wave-optical image processing in the computer. Putting aside, in this paper, all the problems related with the precise determination of the aberration parameters, as well as sampling problems in Fourier space and
avoiding artefacts, a *posteriori* aberration correction (assumed to be isoplanatic) can, in principle, be accomplished in a comparably simple way (Lichte 2008). The Fourier spectrum of the reconstructed image wave has to be divided by a numerical phase plate exp\((-i\chi_{\text{num}}(q))\), which agrees, as closely as possible, with the coherent part exp\((-i\chi(q))\) of the wave-transfer function active in the TEM. Within the information limit, the resulting holographic object wave

\[
\text{obj}_{\text{hol}}(r) = \text{FT}^{-1} \left[ \frac{\text{FT}[\text{ima}_{\text{hol}}(r)]}{\exp(-i\chi_{\text{num}}(q))} \right]
\]

agrees with the initial object exit wave if |\(\chi(q) - \chi_{\text{num}}(q)\)| ≤ π/6; therefore, a highly accurate knowledge of the aberration parameters is indispensable (Lichte et al. 1996). Here, because of possible holographic modifications, the reconstructed image wave ima_{hol}(r) is distinguished from the image wave ima(r) present in the image plane of the TEM. An example of a *posteriori* aberration correction is shown in figure 6.

6. Limitations of electron holography

The performance limits are given by the incoherent aberrations, i.e. information limit, as well as by the signal detection limits, which are determined by the brightness of the electron source, and instabilities.

(a) Lateral resolution

As mentioned above, lateral resolution is first of all determined by the fringe spacing 1/q0 sampling the object wave. After many years of development, this is presently sufficient for object details of 0.1 nm. However, the lateral resolution
Figure 7. Information content in a hologram. The quality of an electron hologram is given by the field of view, lateral resolution and signal resolution. As a figure of merit, \( \text{InfoCont} = n_{\text{rec}} n_{\text{sig}} \) is given by the number of pixels reconstructed across the field of view and the numbers of phase values discernible in the interval \([0, 2\pi]\).

in the amplitude or phase of the corrected wave cannot exceed the information limit of the TEM, even if the fineness of the hologram fringes would allow the recording of finer details. Therefore, the improvement of the information limit is also indispensable for extending resolution in holography.

(b) Signal resolution

For evaluation of an imaging method, one has to account, not only for lateral resolution, but also for signal resolution. Signal resolution means the smallest signal difference detectable in the final image at a given confidence level. In Lichte (2008), it is shown that, in the holographically reconstructed phase image, lateral and signal resolution are connected, and hence cannot be considered separately. It turns out that a corresponding measure can be defined, which we call the holographic information content,

\[
\text{InfoCont} = n_{\text{rec}} n_{\text{sig}}.
\]

It is the product of the number \( n_{\text{rec}} \) of pixels reconstructed along one edge of the field of view and the number \( n_{\text{sig}} \) of phase values discernible in the range \([0, 2\pi]\) (figure 7). For example, a reconstructed field of view with \(128 \times 128\) pixels with a phase difference of \(2\pi/50\) discernible between two pixels would result in \( n_{\text{rec}} = 128 \) and \( n_{\text{sig}} = 50 \); such a hologram would represent an information content of \( \text{InfoCont} = 6400 \). If, from the same hologram, only \(64\) pixels were reconstructed at a corresponding sacrifice of lateral resolution, the phase resolution would be improved to \(2\pi/100\).

It is worthwhile emphasizing that, in \( \text{InfoCont} \), the number of reconstructed pixels shows up, and not the lateral resolution or the field of view separately. This means that, from the standpoint of information content, a large field of view at a correspondingly poor lateral resolution is equivalent to a small field of view at an excellent lateral resolution. Correspondingly, one has to check, before taking a hologram, what will be needed for the specific purpose. For example, recording a hologram with a given field of view, lateral resolution and signal resolution have to be traded off against each other. In the same way, aiming at atomic resolution, the field of view and signal resolution are coupled.

In any case, the information content should be as large as possible. However, it is restricted by experimental shortcomings when recording the hologram. The
detailed analysis shows that the upper limit may be estimated as

\[ \text{InfoCont} = \frac{2\pi}{\text{NoiseFigure} \times \text{snr}} \times C_{\text{inel}}, \]

where \( \text{snr} \) means the desirable signal-to-noise ratio in a reconstructed phase image, giving the confidence level for a phase measurement; usually, \( \text{snr} = 3 \) is chosen. \( C_{\text{inel}} \) is the fringe contrast damping by inelastic interaction, which is a property of the object, and is discussed in more detail subsequently.

\[ \text{NoiseFigure} = \sqrt{\pi} \frac{\mu^{sc} C_{\text{inst}} C_{\text{MTF}}}{\log(|\mu^{sc}|)(B/k^2 e)\tau DQE} \]

is a figure of merit for the holographic quality of the whole recording system, the smaller the better. It explicitly depends on

(i) \( \mu^{sc} \), the degree of spatial coherence at an ellipticity \( \varepsilon \) of illumination; temporal coherence does not play any role here,
(ii) \( C_{\text{inst}} \), the fringe contrast damping by instabilities, such as of an electron biprism, an object and a microscope,
(iii) \( C_{\text{MTF}} \), the fringe contrast damping by the modulation transfer function of a charge-coupled device (CCD) camera,
(iv) \( \text{DQE} \), the detective quantum efficiency of a CCD camera,
(v) \( B \), the axial brightness of illumination at wavenumber \( k \),
(vi) \( e \), the elementary charge, and
(vii) \( \tau \), the exposure time.

For example, inserting the numbers of the Philips CM200-TEM in our Triebenberg laboratory, we find \( \text{InfoCont} \approx 7000 \) for a pure elastic interaction with the object \( (C_{\text{inel}} = 1) \).

The required signal resolution has to be adapted to the object under investigation. The phase shift produced in the image wave by single atoms of different species was described as \( \phi \propto Z^{0.6} \) by Kirkland (1998). For our purpose, it has been studied in detail in Linck et al. (2006) by means of multi-slice calculations (Stadelmann 1987; Weickenmeier & Kohl 1991; Rother et al. 2009), evaluating to numbers between, for example, \( 2\pi/50 \) for oxygen and \( 2\pi/12 \) for gold at a lateral resolution of \( 10 \text{nm}^{-1} \); the thermal motion of atoms was accomplished by an appropriate Debye–Waller factor. For a single elementary charge, a phase shift of about \( 2\pi/30 \) is found. These values increase with lateral resolution. For the analysis of a magnetic specimen, it is disappointing that a single Bohr magneton produces a phase shift of only about \( 2\pi/100000 \), and hence several thousand atoms are needed in a particle to detect magnetism; this is in agreement with the results shown in Dunin-Borkowski et al. (2004).

(c) The role of inelastic interaction

Generally, increasing the thickness of a specimen increases the phase signal, but, alas, also the probability for inelastic interaction; the latter is known to destroy the coherence of the inelastic electrons with elastic ones (Van Dyck et al. 2000), and hence induces background noise in the hologram. Optimization of the signal-to-noise ratio leads to an optimum thickness of twice the mean free path for

Phil. Trans. R. Soc. A (2009)
inelastic interaction, resulting in $C_{\text{inel}} = e^{-1}$. At atomic resolution, for reasons of interpretability, specimen thickness is usually chosen so small that $C_{\text{inel}} \approx 1$ holds. Only the elastically scattered electrons contribute to the holographic sideband, and hence to the reconstructed wave. This means that holography intrinsically offers nearly perfect energy filtering for energy transfers larger than $h/\tau = 4.135 \times 10^{-15}$ eV, where $h$ is Planck’s constant and the exposure time $\tau = 1$ s. However, the fraction of noise created by inelastic electrons, which leaks from the centre band into the sideband area, contributes to the noise in the reconstructed wave. This would improve slightly with an energy filter according to its dispersion; presently, it could help for energy transfers larger than approximately 1 eV, but not for phonons or thermal diffuse scattering at some meV. A monochromator would not help to improve the holographic fringe contrast because longitudinal coherence is not critical; in fact, it would also help holography by improving the chromatic information limit, but at the cost of the correspondingly reduced brightness, and hence increased quantum noise. Consequently, for this purpose, a chromatic corrector would be much better because it preserves the brightness.

(i) Inelastic coherence

At energy transfers larger than $4.135 \times 10^{-15}$ eV, the coherence with elastically scattered or unscattered electrons collapses. Therefore, even a chromatic corrector could not add inelastic object information to a usual hologram taken with a vacuum reference wave. However, it was found that coherent patches, some 10 nm wide, exist inside the inelastic wave fields, for example, generated by plasmons (Lichte & Freitag 2000; Verbeeck et al. 2005). This opens interesting perspectives for investigating inelastic processes by electron holography if the reference wave is also taken from the coherent patch in the inelastic wave field. For such experiments, an energy filter is indispensable.

7. Benefits from the aberration corrector for holography

An aberration corrector (Haider et al. 1998) makes the wave aberration $\chi(q)$ disappear, and hence also $\text{grad} \chi(q)$. This brings the following benefits.

(a) Improvement of information limit

The lateral resolution of holography is finally determined by the information limit. As shown above, it cannot be corrected by $a \text{ posteriori}$ image processing. Therefore, $a \text{ priori}$ improvement of the TEM is indispensable. In an aberration-corrected microscope, i.e. $\chi(q) \equiv 0$, the envelope function of spatial coherence, given by

$$E^{\text{sc}}(q) = \exp \left( -\frac{\pi^2 k^2 \alpha_{\text{ill}}^2}{\ln 2} (\text{grad} \chi(q))^2 \right),$$

where $\alpha_{\text{ill}}$ is the angular size of the electron source, results in $E^{\text{sc}}(q) = 1$ because $\text{grad} \chi(q)$ vanishes for all $q$. For example, at 300 kV accelerating voltage, the residual information limit is enhanced to approximately 12 nm$^{-1}$, corresponding to details of 0.08 nm. This is only determined by the envelope function of temporal
coherence (energy spread in combination with chromatic aberration)

\[ E^{tc}(q) = \exp \left( -\frac{\pi^2}{2} k^2 \Delta^2 \left( \frac{q}{k} \right)^4 \right), \]

with the defocus spread

\[ \Delta = C_c \sqrt{\frac{\text{std}_E^2}{(eU_a^*)^2} + \frac{\text{std}_{U_a}^2}{U_a^2} + 4 \frac{\text{std}_{I_{\text{lens}}}^2}{I_{\text{lens}}^2}}, \]

where \( C_c \) is the coefficient of the chromatic aberration of the objective lens, \( U_a \) is the accelerating voltage and std means the respective standard deviation of the energy spread, high tension and lens currents. As shown recently by the CEOS company (Hartel et al. 2008), \( C_c \) also is correctable, with the consequence that the information limit in a \( C_s/C_c \)-corrected TEM will reach beyond 20 nm\(^{-1}\); consequently, the holographic lateral resolution will also improve correspondingly.

(b) Enhancement of signal

Owing to the opening of the imaging aperture, electrons scattered into larger angles also contribute to the image wave. First of all, this improves the lateral resolution. A likewise important benefit is the increase of signals such as atomic phase shifts in the image wave. From the simulations in Linck et al. (2006), one can derive the gain in the phase shift of single atoms with improving lateral resolution, as plotted in figure 8. It turns out that the phase shift is enhanced by a factor of 2 if the lateral resolution is improved from 5 to 10 nm\(^{-1}\), as well as from 10 to 20 nm\(^{-1}\) by roughly another factor 2. Note that the phase shift does not increase indefinitely, but approaches an upper limit, which is terminated by the nature of the scattering process given, among others, by the scattering amplitude and thermal motion of the atoms.

Phil. Trans. R. Soc. A (2009)
Electron holography with Cs corrector

Figure 9. Required hologram width in an (a) uncorrected and (b) a corrected TEM. Whereas in the uncorrected microscope, the hologram has to be wider than $4\varnothing$PSF, in the corrected microscope, it just has to cover the field of view (fov) of interest. Therefore, in the latter case, the coherent electron current is focused to a higher density and much more electrons are collected in the fov of interest. Consequently, quantum noise is strongly reduced.

(c) Enhancement of signal resolution

For the signal resolution, the width $\text{psf} = \text{grad}_{\text{max}} x(q)/2\pi$ of the point-spread function plays a crucial role because the width $w$ of a hologram not only has to cover the field of view (fov) of interest. Instead, it has to meet the condition $w \geq \text{fov} + 4\varnothing$PSF to collect all the delocalized information (Lichte 1991b). This means that, in the case of a non-corrected microscope, the number of reconstructed pixels $n_{\text{rec}}$ has to be very high to achieve atomic resolution. Consequently, the signal resolution $n_{\text{sig}} = \text{InfoCont}/n_{\text{rec}}$ is comparably poor. Therefore, another main benefit of the aberration corrector is the fact that, with $\text{grad} x(q) \downarrow 0$, the point-spread function also shrinks to atomic dimensions. The hologram width $w$ can be selected to be nearly as small as the area field of view of interest, for example, about a grain boundary and its close neighbourhood. In such a narrow hologram, the coherent current density increases and hence the dose recorded at the same exposure time. Alternatively, exposure time can be chosen to be smaller, which reduces the damping of fringe contrast by instabilities (figure 9; Geiger et al. 2008). Both measures will decrease the noise figure, and hence increase the information content. From the findings with our aberration-corrected Tecnai F20 Cs-corr TEM, we conclude that $\text{InfoCont} \approx 14\,000$, which is significantly higher compared with that found in a conventional microscope. As an example, figure 10 shows two holograms of the same specimen recorded at the conventional Philips CM30 TEM and the Tecnai F20 Cs-corr TEM.

(d) More degrees of freedom in the optical setup

The aberration corrector increases the number of degrees of freedom for changing optical parameters such as magnification, effective focal lengths and so on. Therefore, the flexibility for realizing better or new paths of rays for holography also improves. At medium resolution, the transfer lens of the corrector

Phil. Trans. R. Soc. A (2009)
Figure 10. Improvement of holograms by aberration corrector. Comparison of holograms recorded in (a) the uncorrected CM30 Special TEM and (b) the corrected Tecnai F20 TEM convincingly shows the progress achieved by an aberration-corrected TEM for holography: the fringe contrast and definition by the higher dose are much better. Object: GaAs/AlAs-multilayer, provided by Dr Kisielowski (NCEM, Berkeley). Scale bar (a) 5 nm (inset 0.5 nm); (b) 2 nm (inset 0.5 nm).

For atomic resolution holography, an additional improvement has turned out to be the higher magnification from the object plane to the first image plane. In fact, even though the interference pattern is generated in the first image plane below the biprism, the fringe spacing and the hologram width are only of interest with respect to the object plane, i.e. divided by the magnification in the first image plane. Therefore, for achieving the same hologram fringe spacing in an aberration-corrected microscope, a significantly smaller biprism voltage is required with respect to the conventional TEM. A reduction of biprism voltage to approximately 85 per cent has been reported in Geiger et al. (2008). This considerably improves the stability of the experimental setup. Consequently, the hologram fringe contrast is higher, and hence the information content improves. Note that this benefit is not earned from the correction of aberrations, but merely from simple geometric reasons.

(e) Replacement of crystal tilt by beam tilt

The better the lateral resolution, the more critical is the orientation of the object with respect to the electron beam. However, the usual mechanical specimen tilt is only as precise as the often-insufficient precision of the goniometer. Unfortunately, beam tilt cannot be simply used for compensating specimen tilt because beam tilt would also affect the aberration parameters. When applying beam tilt in an aberrated TEM, a lot of additional aberrations are introduced, which boost the aberration gradient to that even holography cannot correct for. This is much better in the C₄-corrected microscope since the aberrations are minor and the resulting aberrations owing to tilted illumination stay small. Therefore, it becomes feasible to use the beam tilt for the final alignment of mutual beam–specimen orientation. In particular, off-axis electron holography
Electron holography with $C_s$ corrector

3787
tilt axis
sideband
SB+
FT
CB
10 nm$^{-1}$
10 nm$^{-1}$
tilt by 5 mrad
full symmetric sideband
SB–

Figure 11. (a) Beam tilt for fine-tuning object tilt. In the sideband, i.e. the Fourier transform of the image wave, a clear asymmetry owing to object mistilt is found (b), which disappears by a beam tilt of 5 mrad (c). This fine-tuning can be performed interactively by observing the real-time fast Fourier transform of the hologram; such a procedure is not viable under conventional imaging. Object: centrosymmetric crystal of SrRuO$_3$, provided by Dr Honda, Fujitsu. These experiments were carried out in the FEI Tecnai F20 Cs-corr TEM.

with real-time reconstruction facilities can make use of this feature. Immediately after hologram acquisition, a slight specimen misalignment can be recognized in the Fourier transform of the hologram intensity. Asymmetric excitation of Bragg spots in the hologram sideband directly reveals a crystal misorientation, which allows interactive tilt corrections by means of beam tilt to achieve a perfect crystal orientation on the zone axis, as shown in figure 11. This is not possible on the basis of diffractograms from usual intensities because these always reveal point symmetry. A small lateral shift of the hologram with respect to the object arising by beam tilt is easily compensated by shifting the biprism accordingly.

(f) Improvement of isoplanacy

To meet the requirements of high spatial coherence while collecting a high number of electrons, the use of an elliptic illumination for hologram acquisition is essential. In fact, this violates isoplanacy because a plane wave is not used for the illumination of the object. Instead, the incident wave is of cylindrical shape, and hence exhibits a local beam tilt. Although it is a small beam tilt, which hardly affects the interaction with the object, a position-dependent axial coma is introduced in the imaging process. It has been shown that the corresponding violation of isoplanacy can be taken into account by adding a numerical phase curvature before a posteriori aberration correction and removing it afterwards (Lehmann 2004). As already shown above, a beam tilt in the $C_s$-corrected microscope hardly introduces additional aberrations.

Phil. Trans. R. Soc. A (2009)
Figure 12. Phase plates for holographic correction of aberrations. In the case of (a) an uncorrected microscope, the strong increase in the wave aberration gives rise to undersampling in Fourier space, as indicated by the appearing Moiré patterns; this limits the correctable range of spatial frequencies, and hence lateral resolution in the reconstructed wave. (b) In an aberration-corrected TEM, the residual wave aberration varies so slowly that no undersampling shows up. Scale bar, 10 nm$^{-1}$.

Therefore, the residual coherent aberrations can be approved as quasi-isoplanatic, and hence no phase curvature has to be taken into account, although it is still present.

(g) More accurate a posteriori aberration correction

The small aberration gradient $\nabla \chi(q)$ in the aberration-corrected microscope also brings improvements for the numerical back-propagation process from the image plane to the object plane. The numerically generated phase plate $\exp(-i\chi_{\text{num}}(q))$ not only has to describe very accurately the aberrations, which were present in the TEM during the recording of the hologram, but, at the same time, the gradient of the aberration function must fulfil the condition $|\nabla \chi_{\text{num}}(q) \cdot \Delta q| \leq \pi/6$ for a correct sampling in Fourier space, otherwise massive artefacts would arise from the numerical correction procedure; $\Delta q$ denotes the width of a Fourier space pixel. Figure 12 compares the numerical phase plates of the uncorrected Philips CM30 UT/Special Tübingen TEM and our aberration corrected Tecnai F20 TEM. It is obvious that far less numerical artefacts are produced for high-frequency object details in the latter case.

(h) Benefits from holography for aberration-corrected transmission electron microscopy

In addition to the general benefits for the thorough wave-optical analysis, holography offers the following advantages.

Although conventional TEM images and holograms are recorded under preferably aberration-free conditions, residual defocus and the remaining higher order aberrations are difficult to control. Here, holography offers a posteriori fine-tuning by image processing. Because residual aberrations can be removed...
Figure 13. Reconstructed object wave of the hologram in figure 10b. The object is a cross-sectioned specimen of a GaAs/AlAs multi-layer system. It consists of [110]-oriented GaAs layers of constant width and AlAs layers increasing in width from bottom to top by one monolayer each (indicated by dashed lines). (a) Amplitude and (b) phase are corrected for the residual aberrations of the Tecnai F20 Cs-corr. Owing to improved signal-to-noise properties based on the $C_s$-corrected holographic setup, the reconstructed data are only little affected by noise, and hence are much more reliable.

Completely afterwards, no compromise on defocus and spherical aberration or other aberrations is necessary, as needed for the usual phase-contrast imaging. Instead, residual aberrations can be selected freely and independently, as long as the aberration gradient remains small. This may offer free parameters for taking optimized holograms, for example, with respect to the local signal-to-noise properties (Linck 2007). As an example, figure 13 shows the reconstructed and aberration-corrected object wave of a GaAs/AlAs multi-layer. Object amplitude and phase are barely suffering from artefacts. Note that, although no averaging or Fourier filtering was applied to the reconstructed wave, the dataset has a very smooth background compared with hitherto reconstructed waves. Furthermore, figure 14 demonstrates the excellent holographic performance of the $C_s$-corrected TEM at atomic resolution.

A special aspect is the exact focusing of the reconstructed wave into details of the object. In principle, one can change focus by as little as 0.1 nm reproducibly in the computer. However, this does not mean that we can look into the different planes of the object. We cannot see, for example, an illuminating plane wave at strong underfocus, or the modulation by single object slices at a weaker underfocus and a complete object wave in focus. This can only be achieved by tomographic holography (Twitchett-Harrison et al. 2008; Wolf et al. in preparation). In two-dimensional holography, we always deal only with the complete object exit wave leaving the object, which we can propagate backwards and forwards by defocus. With increasing lateral resolution, the question ‘which focus is the best?’ to find the true object information in both amplitude and phase has to be discussed, together with ‘what are the guidelines?’ to find it. This needs highest precision simulations and comparison with experimental findings. Holography offers these means.

Phil. Trans. R. Soc. A (2009)
Figure 14. Atomic resolution details of the indicated subareas in figure 13. Whereas the different atom species can hardly be distinguished in (a) the amplitude, the AlAs and GaAs layers can be directly distinguished in (b) the phase image owing to phase jumps in both atomic columns of the GaAs layers; AlAs does not show this effect in the Al columns because of the significantly smaller atomic number. From the atomic phase shift, a specimen thickness of approximately 9 nm can be derived by comparison with (c) the EMS simulations. It is evident that the phase signal provides much more reliable information about atomic species than the amplitude, although atomic resolution is available in both.

8. Summary and outlook

By the development of aberration correction of the TEM, all related methods have made huge progress: TEM and scanning transmission electron microscopy imaging, energy filtering transmission electron microscopy and electron energy loss spectroscopy as well as electron holography. These methods deliver a respective variety of signals, which, because of their complementarity, allow the comprehensive characterization of the objects to an atomic scale. It would be a very fruitful task to compare and combine these methods for the final goal of understanding materials.

Electron holography is a unique method for comprehensive wave-optical analysis of structures and fields in materials. Its uniqueness lies in the availability of both amplitude and phases down to an atomic scale, allowing access, for example, to electric and magnetic fields. In the comparison of off-axis holography with focal series holography, also reconstructing amplitude and phase, there are still some open questions (Lehmann et al. 2002). In particular, the phases resulting from the two methods are significantly different; this will be a topic for future investigations.

In this article, the improvement of off-axis electron holography by using the benefits obtained from an aberration-corrected TEM is described, mainly with respect to atomic resolution. In spite of the fact that the coherent aberrations can be corrected by \textit{a posteriori} image processing of the holographically reconstructed image wave, an aberration corrector in the TEM used for recording the hologram substantially improves the performance of electron holography. The reason is that, in addition to the lateral resolution, a signal resolution is also needed, which is capable of measuring tiny phase structures such
as single atoms. The aberration corrector significantly improves both the
obtainable signal resolution and the signal delivered by the object so that the
coherent electrons are used much more efficiently. The corrector for chromatic
aberration, recently successfully tested at the CEOS company, will further
boost this.

Presently, the performance described by the information content is limited by
the brightness of the electron source, by the not-yet-ideal path of rays obtainable
with the electron biprism in the usual TEM and by the deficiencies of the detector.
However, presently, there are exciting improvements under development, i.e. the
XFEG (Freitag et al. 2008), which provides a brightness better by roughly one
order of magnitude, by realizing the optimum position of the biprism (Lichte
1996), relaxing the needs for extreme stability and by the future generation of
detectors (Moldovan et al. 2008) with improved modulation transfer and detective
efficiency. Using these achievements, the information content of a hologram may
be expected to improve by another factor of three to five.

We thank Harald Rose and Max Haider for designing the Cs-corrector and making it available. The joint efforts with CEOS and FEI in combining off-axis holography and the Cs-corrector is indispensable for this development. Also, the creative discussion within the Triebenberg laboratory and with Michael Lehmann (now at TU Berlin) was very fruitful. The financial support obtained for various related projects from DFG and EU (Framework 6 Integrated Infrastructure, reference 026019 ESTEEM) enabled us to realize our research.

References

Geiger, D., Lichte, H., Linck, M. & Lehmann, M. 2008 Electron holography with a Cs-
Haine, M. E. & Mulvey, T. 1952 Formation of the diffraction image with electrons in the Gabor
Germany: Springer Verlag.
Lehmann, M. 2004 Influence of the elliptical illumination on acquisition and correction of
coherent aberrations in high-resolution electron holography. Ultramicroscopy 100, 9–23.
(doi:10.1016/j.ultramic.2004.01.005)

Phil. Trans. R. Soc. A (2009)


