Holography and transmission electron microscopy
G. Matteucci, G. Pozzi, A. Tonomura

To cite this version:
G. Matteucci, G. Pozzi, A. Tonomura. Holography and transmission electron microscopy. Journal de Physique IV Colloque, 1993, 03 (C7), pp.C7-2063-C7-2072. <10.1051/jp4:19937330>. <jpa-00251977>

HAL Id: jpa-00251977
https://hal.archives-ouvertes.fr/jpa-00251977
Submitted on 1 Jan 1993

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Holography and transmission electron microscopy

G. MATTEUCCI, G. POZZI* and A. TONOMURA**

Department of Physics, University of Bologna, via Irnerio 46, 40126 Bologna, Italy
* Department of Materials Science, University of Lecce, via Arnesano, 73100 Lecce, Italy
** Advanced Research Laboratory, Hitachi Ltd., Hatoyama, Saitama 350-03, Japan

Abstract.— The basic principles and methods of off-axis electron holography are presented and illustrated by means of three examples related to its application in high resolution electron microscopy and the investigation of electric and magnetic fields in thin specimens.

1. Introduction.

Gabor devised in-line electron holography [1] in 1948 as a new technique able to circumvent the resolving power of the electron microscope which was at that time about 1.2 nm. The holographic approach foresees firstly the recording of an aberrated object wavefront and then the recovering of amplitude and phase information by optical means. However, electron holography has attained its concrete development only in the last decade, thanks to the introduction in electron microscopy of high brightness sources, such as field emission guns (FEG) [2]. This latter source, combined with a versatile electron interferometer, such as the Möllenstedt-Düker electron biprism [3], allows the recording of high quality off-axis holograms which can be reconstructed and processed by optical and/or digital means. The whole process can be carried out also in line by using the newly developed slow-scan charge coupled device (CCD) cameras [4].

With respect to the standard phase contrast methods in electron microscopy [5,6], holography allows the extraction of quantitative information with increased sensitivity limits owing to the use of phase amplification methods which have no counterpart in electron microscopy.

Aim of this paper is to recall the basic principles of the method and to illustrate its performance with respect to standard transmission electron microscopy techniques, by considering the successes (but also the practical and theoretical limitations) obtained so far in a few typical investigations which range from the correction of spherical aberration and the improvement of high resolution images to the mapping of magnetic and electric fields up to a micrometric scale.

As the researches on electron holography have been mainly carried out so far in few laboratories, the field is covered almost completely in a set of review papers which can serve as guide to the more specialized bibliography. The different realization schemes are illustrated by the papers of Hanszen [7] and Tonomura [8] where the main emphasis is towards theoretical and experimental aspects respectively. The most useful method in electron holography turned out to be image-plane off-axis holography, the relations of which with the parent field of electron interferometry and interference electron microscopy are presented in the early paper by Missiroli et al [9]. The review by Lichte reports on the steps undertaken at Tübingen aimed at the correction of aberrations in high resolution images [10]. Finally, the recent review by Tonomura [11] presents an up-dated account of the latest results and indicates some prospects for future development.
2. General considerations.

It is customary to divide the process of image formation in the electron microscope into three steps: i) interaction of electrons with the specimen; ii) propagation of electrons from the specimen to the final recording plane through the microscope lenses; iii) detection of electrons by means of a photographic plate or an in-line electronic image read-out device. It is also usually assumed that the result of the propagation of electrons from the gun to the specimen is a plane wave. Partial coherence and/or convergent spherical illumination can then be accounted for by a partially coherent superposition of a set of plane waves.

By considering only elastic scattering events, the interaction of the specimen with the electron beam can be described through a complex transmission function (object wavefunction) \( O(\mathbf{r}) \) which represents the ratio between the amplitudes of the out-going and the in-going electron wavefunctions; \( \mathbf{r} = (x,y) \) is a bidimensional vector perpendicular to the optic axis \( z \) which is parallel and in the same direction as the electron beam. In the standard phase-object approximation

\[
O(\mathbf{r}) = \exp(i\phi(\mathbf{r}))
\]  

where the phase term \( \phi(\mathbf{r}) \) is given by

\[
\phi(\mathbf{r}) = \frac{\pi}{\lambda E} \int V(x,y,z)dz - \frac{e}{\hbar} \int A_z(x,y,z)dz
\]  

The integral is taken along a trajectory \( l \) parallel to the optical axis \( z \) inside and outside the specimen to include stray fields, \( V(x,y,z) \) and \( A_z(x,y,z) \) are the electrostatic potential and the \( z \) component of the magnetic vector potential \( \mathbf{A}(x,y,z) \) respectively and \( E \) is the accelerating voltage in the non-relativistic approximation. \( e, \lambda \) and \( \hbar \) are the absolute value of the electron charge, the electron wavelength and the reduced Planck constant respectively.

As the information is mainly encoded through the phase (the observable quantity being the phase difference), it is important to recall that, contrary to the optical case where three-dimensional effects are strikingly impressive, only essentially two-dimensional information is available in transmission electron microscopy. In fact in the electric case \( \phi \) is proportional to the potential averaged along the electron path whereas, in the magnetic case, the maximum of information encoded in the beam corresponds to the magnetic flux enclosed between two trajectories. Therefore, once the three-dimensional electromagnetic field is known, it is possible to determine the two-dimensional phase difference, but not the reverse.

The electrostatic potential includes the contributions both of external applied fields on a mesoscopic scale, as is the case of p-n junctions, and of the atoms building up the specimen at a microscopic scale, so that also the structural information is encoded in the object phase. However, it turns out that the phase object approximation holds only for very thin specimen, but can be implemented to cover also the case of thicker specimens by means of the multislice method [12]. By this approach the specimen is divided into thin slices, for each of which the phase object approximation holds, and the propagation between slices is treated by means of the Kirchhoff-Fresnel theory. In this way the exit wavefunction is calculated by a series of interaction-propagation steps over the slices. Today most of the computer programs used for the interpretation of high resolution images are based on this algorithm which has been shown to be able to treat, in addition to the interaction problem, also the propagation of electrons through the microscope lenses [13]. Therefore, if \( O(\mathbf{r}) \) is the exit wavefunction, the image wavefunction \( \Psi(\mathbf{r}) \) taking into account the microscope aberrations is given by:

\[
\Psi(\mathbf{r}) = O(\mathbf{r}) \otimes FT^{-1}\{ B(\mathbf{u}) \exp(i\chi(\mathbf{u})) \} = C(\mathbf{r}) \exp(i\varphi(\mathbf{r}))
\]  

where \( \otimes \) is the convolution, \( FT \) is the Fourier Transform, \( \chi \) is the isoplanatic wave aberration function of the objective lens, including the effects of the defocus \( \Delta z \) and of the spherical aberration...
$C_S = (u_x, u_y)$ is the spatial frequency vector and $B(u)$ takes into account the effect of the objective aperture. $C(r)$ and $\varphi(r)$ are the amplitude and the phase respectively of the image wavefunction.

The distortions in the aberrated image can be better understood if a weak sinusoidal phase grating is considered \cite{14}, described by the object wave:

$$O(r) = 1 + i \phi_o \cos(2\pi u_o x)$$

(2.4)

Then the image wavefunction is given by

$$\Psi(r) = 1 - \phi_o \cos(2\pi u_o x) \sin(\chi) + i \phi_o \cos(2\pi u_o x) \cos(\chi)$$

(2.5)

and is modulated both in amplitude and phase by the transfer functions $\sin(\chi)$ and $\cos(\chi)$.

Further degradation of the information encoded in the electron beam happens in the recording step since the signal is proportional to the square modulus of the image wavefunction i.e.

$$I(r) = |\Psi(r)|^2 = 1 + 2\phi_o \cos(2\pi u_o x) \sin(\chi(u_o))$$

(2.6)

This equation shows that only limited information is preserved. In particular, depending on the spatial frequency $u_o$, no information is transferred at all at the zeroes of the phase contrast function $\sin(\chi)$. The loss of information is even more serious when the phase object approximation holds and for ideal imaging: in that case the phase information is completely lost in the Gaussian image of the object and special methods, so-called phase-contrast \cite{5,6}, should be employed in order to partly recover this information. For the high resolution case, the phase-contrast effects are automatically introduced owing to the combined effect of defocus and spherical aberration which give rise to a heavily distorted image of the structure.

Contrary to standard phase-contrast techniques, electron holography is the unique method by means of which it is possible to obtain the complete retrieval of the two-dimensional information stored in the electron beam after its interaction with the specimen (that is, the object wavefunction unperturbed by the aberrations) and to display and evaluate that information in a vivid and, more important, a quantitative way.

3. Basic principles for recording and processing electron holograms.

In brief, holography is a two step imaging process, by means of which it is possible to record and subsequently recover both the amplitude and the phase of the object wavefunction. This is accomplished in the off-axis image scheme by superimposing in the first step, within the electron microscope, a tilted coherent plane reference wave $R = \exp(2\pi ixu_c)$ with carrier spatial frequency $u_c$ on an image $\Psi(r)$ of the object wavefunction.

Both wave front and amplitude beam splitting devices can be used for realizing this task \cite{9}. However, the most diffused and versatile type of electron interferometer is the Möllenstedt-Düker electron biprism \cite{3} which belongs to the class of wavefront division interferometers. The biprism is realized by placing a thin charged wire $W$ between two earthed plates $P$, Figure 1 (a) and is located under the back focal plane of the objective lens $Ob$. The wire splits the wavefront of the incoming electron beam and its electrostatic field produces a deflection and a subsequent overlapping in the observation plane $OP$ below the wire. As the width of the overlapping region and the interference fringe spacing can be controlled by varying the charge on the wire by means of an external voltage supply, it is possible to obtain electron interferograms and holograms. In particular, if the interference region contains a large number of fringes (above 100-200) superimposed on a focused image of the specimen, then the interferogram is called image off-axis electron hologram. The preferred arrangement is to use only one half of the object plane for the specimen $S$, illuminated by the coherent beam $EB$, and the other for the reference beam, as shown in Figure 1 (a). If the specimen occupies the whole plane, the interpretation of the pattern is more difficult because the information...
available concerns the phase difference between points in the specimen. Finally, it should be noted that the whole set-up comprises additional lenses in the condenser stage for reduction of the source to an angular size smaller than that of the interference fringes as seen from the biprism wire, in order to ensure that they are of high contrast. Moreover, the projector stage PL should magnify the fringe system which forms in the observation plane so that the fringe spacing matches with the resolution of the recording medium in the hologram plane H.

Interference fringes of spacing as small as possible are necessary because, in order to obtain the complete separation of the reconstructed wave from the twin wave and from the straight through beam, the carrier fringe spacing \( s = 1/u_c \) should be at least three times finer than the finest detail to be resolved in the image. This is the main reason why these experiments are difficult to realize, unless the electron microscope is equipped with a FEG.

The intensity recorded in the interferogram, henceforth called hologram, is given by

\[
H = |R + \Psi|^2 = |R|^2 + |\Psi|^2 + \Psi^* R + \Psi R^* = 1 + C(r)^2 + 2C(r) \cos(2\pi u_c x + \varphi(r))
\]  

(3.1)

showing that both amplitude \( C \) and phase \( \varphi \) of the image wavefunction are encoded in the hologram, contrary to the conventional intensity \( I = C(r)^2 \).

The idea of Gabor was to recover the information contained in the hologram by optical means. This is the second step, reconstruction and processing, of the holographic method.

In this case the electron hologram \( H \), registered on a photographic emulsion is first developed so that its amplitude transmittance is a linear function of the intensity. The plate is then inserted in an optical bench and illuminated by a coherent plane laser wave, Figure 1 (b). The lens L performs, in the back focal plane, the Fourier transform of the intensity distribution of the hologram H. In addition to the straight through beam, two sidebands are displayed which represent the Fourier
spectrum of the image wave that is (1) and its conjugate (−1). In the off-axis scheme, if the carrier spatial frequency is three times larger than the maximum spatial frequency of the object, the sidebands do not overlap with the zero beam (autocorrelation). Therefore, the (1) sideband can be isolated by means of a spatial filter F inserted in the back focal plane of the reconstruction lens L. Allowing a free propagation of the wave from the back focal plane downwards, the inverse Fourier transform is performed on the plane RI where an exact optical replica of the object wave function is found. A detector, placed in the plane RI, will record however only the intensity of the object wavefunction, while the phase is again lost. To display also the phase of the image wave, a number of techniques can be applied which have no counterpart in the electron microscope. The capabilities include bright and dark field, phase contrast imaging, optical interferometry, diffraction image filtering, low angle diffraction, phase detection and its amplification.

For the reconstruction of a high resolution image wave, the compensation of the aberrations could be performed using a suitable phase plate in the back focal plane of the lens L, Figure 1 (b). This is a very difficult task because of the inherent difficulty for preparing a suitable correction plate matching precisely all the parameters related to the wave aberrations. The first successful attempt to compensate the spherical aberration made use of an optical convex lens, in this way half-spacing fringes in the image of an evaporated gold particle were observed [15].

Today, however, image reconstruction by numerical methods presents considerable advantages with respect to the light optical technique, and the effects of the correction of aberrations at high resolution have been demonstrated by numerical image processing of a digitized hologram of a carbon foil which provides all relevant spatial frequencies [16]. In this case the Fourier Transform of the hologram H is taken and the sideband given by \( \tilde{O}(u + v \text{i}) \exp[i\chi(u + v \text{i})] \), where \( \tilde{O}(u) \) is the Fourier Transform of the object wave, is chosen. Shifting the origin in the Fourier space by \( u \text{i} \), which is equivalent to tilting the illumination as in Figure 1 (b), we obtain the wavefunction \( \tilde{O}(u) \exp[i\chi(u)] \), from which the lens aberration term can be eliminated in principle by multiplication with the inverse of the aberration phase factor \( \exp[-i\chi(u)] \). The inverse Fourier Transform gives finally the amplitude and phase of the true object wave \( O(r) \).

4. High resolution observation of InP atomic structure.

In this section recent results of the application of high resolution electron holography to the investigation of InP crystals [17], a material which is now widely used in the semiconductor industry, are presented and discussed. Previous investigations of this material identified the In and P atom columns by comparing multislice calculations with through-focus micrographs [18]. The same result can be achieved by means of high resolution electron holography using a single electron hologram. Figure 2 shows a schematic view of a InP crystal, which has a zinc-blende structure (a=0.587 nm), with In and P columns separated along the (001) direction. Figure 3 (a) shows a high-resolution electron hologram of InP, taken with a 350 kV field-emission electron microscope operating at 300 kV with an underfocus of about 20 nm. The carrier fringe spacing is 0.058 nm. The Fourier spectrum of the hologram is shown in Figure 3 (b), displaying the required separation between the two sidebands and the central autocorrelation spectrum. The upper sideband has been selected for the reconstruction, using an aperture of 6.7 nm−1 diameter, which includes the (200) reflections, essential for identifying In and P atom columns, because the structure factor \( F_{200} = 4|f_{In} - f_{P}| \) is useful for differentiating atom species.

Figure 4 shows the aberration-corrected amplitude (a) and phase (b), with a point resolution of 0.17 nm, i.e. 3 times the carrier fringe spacing. For aberration correction the required accuracy of the spherical aberration coefficient \( C_S \) and the defocus \( \Delta z \) are estimated to be 0.04 mm and 2.5 nm respectively, at 300 kV. This accuracy is much lower than the error associated to the experimental determination of those parameters, given by \( C_S = 1.0 \text{ mm} \) and \( \Delta z \) about 20 nm underfocus. Therefore, the reported images have been selected from a set of about 100 reconstructions made with values of \( C_S \) ranging between 0.9 mm and 1.1 mm with a 0.02 mm interval and of \( \Delta z \) ranging between 15 nm and 25 nm (underfocus) at 1 nm interval. The InP structure is well recognized in
Fig. 2 — Schematic view of the InP crystal (viewing direction (001)).

Fig. 3 — (a) High-resolution electron hologram of InP (carrier fringe spacing = 0.058 nm; bar represents 0.5 nm). (b) Fourier spectrum of (a). (Courtesy Physical Review Letters)

Fig. 4 — Aberration corrected reconstructed images of Figure 3 (a) (bar represents 0.5 nm). A unit cell is drawn in. (a) Amplitude. (b) Phase. (Courtesy Physical Review Letters)
the reconstructed images, In atom columns are seen as black dots in the amplitude image, Figure 4 (a), and as large white dots and small white dots, respectively, in the phase image, Figure 4 (b). It should be noted that the phase image is clearer than the amplitude one, since the thin sample affects the phase more than the amplitude. The phase shifts due to In and P columns are 0.61π and 0.23π respectively, which corresponds to a specimen thickness of 4.4 nm.

These results demonstrate that, whereas the aberration problem is solved in principle, its practical solution depends on a more accurate knowledge of the electron optical parameters. A careful theoretical analysis of this issue made by Lichte [19] (who also examined the related problem of the optimum focus for taking a hologram [20]) shows that the required precision in the determination of the spherical aberration and defocus for improving the resolution up to the information limit of 0.1 nm is of about 1μm and 1 nm respectively. Thus the major difficulty for attaining the highest resolution is the determination of the wave aberration of the electron microscope. However, it should be pointed out that this difficulty is not specific of electron holography, but is inherent for all the electron microscopy techniques.

5. Mapping of electric and magnetic fields

Whereas the application of electron holography in the high resolution realm has not completely fulfilled Gabor’s dream, the main obstacle being the imprecise knowledge about the aberrations of the microscope, in the medium resolution the problem of phase contrast has been completely solved.

The information contained in the phase term, equation (2.2), about electric and/or magnetic fields of the specimen can be released from the hologram by optical or digital interferometric techniques. Usually a plane wave is superimposed parallelly to the reconstructed object wave. A new interference pattern, called contour map, is obtained where the fringes display the loci of points with equal phase difference. According to (2.2) the contour fringes map either the lines of force of the magnetic field or the projected potential distribution of the electric one.

The improvement of electron holography versus Lorentz microscopy methods can be appreciated by the following two examples concerning observations carried out on reverse-biased p-n junctions and micromagnetic domains.

An array of parallel ion implanted diodes has been realized on a wafer, which was chemically thinned until a hole was formed with a thin area around it [21,22]. The specimen was then mounted on a special holder equipped with electrical contacts for biasing the p-n junctions. Figure 5 (a) shows an out-of-focus image of the selected region: the reverse bias applied to the junction is 4V and the defocus distance −1.25 mm. The micrograph shows sharp contrast lines (see arrows) with a fine wave-optical structure which arises in correspondence with the regions having the highest phase gradient, i.e. the highest electric field.

Let us compare this result with those obtained by means of electron holography. Figure 5 (b) shows an interferogram obtained by the simultaneous reconstruction of one hologram at 2V reverse bias and another without specimen. The optical fringes propagate in the whole interference field and fan out through the hole connecting neighbouring junctions, whereas within the specimen they are influenced also by its thickness variations and topographical features. A particularly interesting aspect is represented by the fact that the projected potential lines between the junctions are closed. Thanks to the finding of a model for the field associated to an array of p-n junctions [23], it has been possible to calculate the projected potential and to establish that this effect is not due to the field itself, but is an artefact due to the perturbation of the reference wave by the long range tail of the electrostatic field.

This result emphasizes an important point. In the applications of electron holography it has always been taken for granted that the reference beam in the first recording step is an unperturbed plane wave, an assumption which is violated in the above case. The difficulties encountered in such a case have been thoroughly investigated [24] and the results of this analysis can be summarized as follows, (1) the most reliable contour map can only be obtained by recording a double-exposure
Fig. 5 — (a). Out-of-focus image of a region where two p-n junctions are present. Arrows mark contrast lines associated with depletion layers. Reverse bias 4V.
(b) Optical interferogram of the specimen at 2V reverse bias. (Courtesy Ultramicroscopy)

Fig. 6. — Recorded magnetization pattern in cobalt magnetic tape: (a) schematic of the recording method; (b) interference micrograph. (Courtesy Journal of Applied Physics)
hologram directly in the electron microscope, (2) the information stored in the resulting contour map is that of a fictitious object whose phase is given by the difference between the object phase and that of the modulated reference beam, (3) to extract the maximum information from these holograms prior knowledge of the field is necessary in order to achieve the best fit between theoretical and experimental results.

When thin ferromagnetic films are observed by electron holography, the contour lines which appear in the optical interference image indicate projected magnetic lines of force in $h/e$ units [25]. This method has demonstrated to be effective in high-resolution observation of magnetic domain structures [26], which has become an essential part of the arsenal of modern technologies in the field of magnetic recording and memories.

In addition to solving practical problems, electron holography also provides information about fundamental physics, such as the Aharonov-Bohm effect [27] and magnetic flux quantization [28].

The last example reported here shows an application of electron holography to magnetic recording [26]. As a magnetic head is moved along an evaporated thin cobalt film, a bit pattern can be recorded as magnetization in alternate directions, as shown in Figure 6 (a). An interference micrograph of the recorded cobalt film is shown in Figure 6 (b). The film covers only the upper part of the micrograph and is observed from above. The lower part is free space. Arrows in the micrograph indicate the directions of the recorded magnetization. Recorded magnetic lines of force can be observed as contour fringes along the arrows. At the boundaries between two opposite magnetizations, magnetic lines of force behave like vortices. As they approach the film edge they meander and then leak outside into the free space. The leakage flux can be quantitatively measured in the micrograph.

In order to achieve higher-density recording, the width of the transition region has first to be reduced. This width depends greatly on the magnetic characteristics of the film material, and also on the recording method. With observation by sensitivity enhanced electron holography, a bit length of 85 nm has been demonstrated [29].

5. Conclusions.

Since research in electron holography has been carried out so far only in few laboratories, application fields are still limited and undeveloped. Further usage is expected to become possible in a wide range of fields from basic physics to technological frontiers not yet explored. First steps towards the application of high resolution electron holography in material science are being taken [30], with quantitative evaluation of the experimental data made possible by the combination of CCD image read-out devices and modern digital image-processing technology [30,32].

Finally the long-standing problem of flux line observation in thin films has been solved by applying both standard [33] and holographic [34] methods to superconducting niobium. It is expected that electron holography will be of help to clarify the flux pinning mechanism both for conventional and high $T_c$ superconductivity and to answer more fundamental questions, like the value of the magnetic flux for each fluxon whether it is always $h/e$ or if it varies for new kinds of superconductors.

Acknowledgements.

One of the authors (G.P.) thanks Professor H. Lichte and Dr T. Kawasaki for enlightening discussions about high resolution electron holography. Mr S.Patuelli skilful technical assistance is also acknowledged.

This work has been done in a collaboration scheme of one of the authors (G.P.) with the Hitachi Advanced Research Laboratory, Hatoyama, Japan. Financial support from MURST coordinated by Consorzio INFM and CNR-GNSM is acknowledged.
References


