

Electron holography—basics and applications

To cite this article: Hannes Lichte and Michael Lehmann 2008 *Rep. Prog. Phys.* **71** 016102

View the [article online](#) for updates and enhancements.

Related content

- [Progress in electron- and ion-interferometry](#)
Franz Hasselbach
- [The realization of atomic resolution with the electron microscope](#)
David J Smith
- [Imaging of magnetic and electric fields by electron microscopy](#)
Josef Zweck

Recent citations

- [Magnetic domain structures of Sr₃Co₂Z hexaferrite by TEM](#)
William Salazar *et al*
- [Magnetic skyrmions in nanostructures of non-centrosymmetric materials](#)
Nitish Mathur *et al*
- [Rafal E. Dunin-Borkowski *et al*](#)



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Electron holography—basics and applications

Hannes Lichte¹ and Michael Lehmann²

¹ Triebenberg Laboratory, Institute of Structure Physics, Technische Universität Dresden, D 01062 Dresden, Germany

² Institute for Optics and Atom Physics, Technische Universität Berlin, D 10623 Berlin, Germany

Received 16 January 2007

Published 17 December 2007

Online at stacks.iop.org/RoPP/71/016102

Abstract

Despite the huge progress achieved recently by means of the corrector for aberrations, allowing now a true atomic resolution of 0.1 nm, hence making it an unrivalled tool for nanoscience, transmission electron microscopy (TEM) suffers from a severe drawback: in a conventional electron micrograph only a poor phase contrast can be achieved, i.e. phase structures are virtually invisible. Therefore, conventional TEM is nearly blind for electric and magnetic fields, which are pure phase objects. Since such fields provoked by the atomic structure, e.g. of semiconductors and ferroelectrics, largely determine the solid state properties, hence the importance for high technology applications, substantial object information is missing.

Electron holography in TEM offers the solution: by superposition with a coherent reference wave, a hologram is recorded, from which the image wave can be completely reconstructed by amplitude and phase. Now the object is displayed quantitatively in two separate images: one representing the amplitude, the other the phase. From the phase image, electric and magnetic fields can be determined quantitatively in the range from micrometre down to atomic dimensions by all wave optical methods that one can think of, both in real space and in Fourier space.

Electron holography is pure wave optics. Therefore, we discuss the basics of coherence and interference, the implementation into a TEM, the path of rays for recording holograms as well as the limits in lateral and signal resolution. We outline the methods of reconstructing the wave by numerical image processing and procedures for extracting the object properties of interest. Furthermore, we present a broad spectrum of applications both at mesoscopic and atomic dimensions.

This paper gives an overview of the state of the art pointing at the needs for further development. It is also meant as encouragement for those who refrain from holography, thinking that it can only be performed by specialists in highly specialized laboratories. In fact, a modern TEM built for atomic resolution and equipped with a field emitter or a Schottky emitter, well aligned by a skilled operator, can deliver good holograms. Running commercially available image processing software and mathematics programs on a laptop-computer is sufficient for reconstruction of the amplitude and phase images and extracting desirable object information.

(Some figures in this article are in colour only in the electronic version)

This article was invited by Professor J C H Spence.

Contents

1. Introduction	2	2.4. Inelastic interaction	11
2. Electron waves	2	3. Information transfer in a transmission electron microscope (TEM)	12
2.1. Electron interference	3	3.1. Wave optical transfer of information: Abbe theory	12
2.2. Coherence of electrons	5	3.2. Aberrations	15
2.3. Elastic interaction with electric and magnetic fields	7		

4. Electron holography	19	6. Electron holography at atomic resolution	32
4.1. Basic idea	19	6.1. Electron—specimen interaction at atomic dimensions	32
4.2. Electron holography	20	6.2. Benefits of atomic resolution holography	35
4.3. Scheme of image plane off-axis electron holography	20	6.3. Special aspects for acquisition of atomic resolution holograms	36
4.4. Properties of the reconstructed image wave	23	6.4. Correction of coherent wave aberration	38
4.5. Noise properties	24	6.5. Quantitative analysis	39
4.6. Projection problems	24	6.6. Applications	40
5. Electron holography at mesoscopic resolution	25	6.7. Future challenges	42
5.1. Specific aspects at mesoscopic resolution	25	7. Conclusions and outlook	42
5.2. Detection limits	26	Acknowledgments	43
5.3. Applications for the analysis of electric and magnetic fields	27	References	43

1. Introduction

The macroscopic properties of materials, such as conductivity, optical transparency, or mechanical brittleness, are finally determined by the materials structure on an atomic scale. Consequently, for understanding and tailoring materials, solid state physics and materials science aim at setting up structure—properties relations. For this, the structure has to be analyzed with respect to the following questions:

- Where are the atoms?
- Which atoms are where?
- What are the bonds?
- Which mechanic/electric/magnetic fields are around?
- What are the potentials?

There are a variety of methods, which may answer some of these questions:

Scanning tunneling microscopy (STM) and its derivatives, such as atomic force microscopy (AFM), magnetic force microscopy (MFM) and others, allow analysis with respect to nearly all the questions raised above, however, only of the surface structure.

For the interior structure, diffraction methods using x-rays or neutrons, play an essential role; however, in the diffraction pattern one finds the object information averaged over an illuminated area of about $1\ \mu\text{m}$; a true local resolution in the range of $1\ \text{nm}$ or below is not possible, because the phases in Fourier space cannot be determined from a diffraction pattern ('phase problem of diffraction').

The most powerful methods for analyzing interior structures at subnanometre resolution are those of transmission electron microscopy (TEM). In particular since the development of aberration correctors, modern TEM allows access to atomic structures at a resolution of $0.1\ \text{nm}$. However, the intensity recorded in a conventional electron image only represents the squared modulus of the electron image wave; the phase of the image wave is lost. This phase loss means a loss of substantial object information, because most object properties are encoded in the phase of the transmitted electron wave. Loosely speaking, conventional TEM is blind to these object properties, e.g. electric or magnetic fields in the specimen. Since the phase can only be detected by interferometric means, electron

holography has paved the way for a comprehensive analysis of nearly all object properties at medium and at atomic resolution. In this paper, the basics, applications and performance limits of electron holography are outlined.

2. Electron waves

Electrons have particle properties. Taking account of relativistic effects, the momentum of an electron accelerated by a nominal voltage U_a , in space free from electric or magnetic fields, is given by

$$p = \sqrt{2em_0U_a^*}, \quad (2.1)$$

with the rest mass m_0 and the amount of charge e of the electron.

$$U_a^* := U_a \left(1 + \frac{eU_a}{2m_0c^2}\right) \quad (2.2)$$

is the *relativistically modified acceleration voltage* with c the velocity of light [1, 2].

The total energy in an area with electric potential V is given as

$$E = E_{\text{kin}} - eV. \quad (2.3)$$

Please note that for an electron the potential energy is $-eV$, i.e. it is negative in a positive potential V . Therefore, the momentum follows as

$$p = \sqrt{2em_0(U_a + V)^*}. \quad (2.4)$$

Electrons have wave properties. The non-relativistic wave equation for their wave function ψ is the Schrödinger equation

$$\left[\frac{1}{2m_0} (-i\hbar \nabla + e\vec{A})^2 - eV \right] \psi = E\psi \quad (2.5)$$

with electron rest mass m_0 and magnetic vector potential $\vec{A}$; $\hbar$ is Planck's constant.

For relativistic electrons, the wave vector follows from the (relativistic) Klein–Gordon equation as

$$\vec{k} = \frac{\sqrt{2em_0(U_a + V)^*}}{\hbar} \vec{e}_p - \frac{e}{\hbar} \vec{A} \quad (2.6)$$

with $\vec{e}_p$ the unit vector in the momentum direction. In the following, the nominal wavenumber $k_0 = \sqrt{2em_0U_a^*}/h$ is used. The corresponding wavelength for $\vec{A} = \vec{0}$ is the de Broglie-wavelength

$$\lambda = \frac{h}{p}. \quad (2.7)$$

For example, for electrons accelerated by $U_a = 100$ kV, the wavelength amounts to $\lambda = 3.7$ pm, which is much smaller than atomic structures in solids. This is the main reason why electron microscopy is so successful.

Assume for the beginning field free space with $V = 0$ and $\vec{A} = \vec{0}$. Then the most simple solution of the wave equation is a plane wave of amplitude a , described by means of

$$\psi(\vec{r}, t) = a \exp(i(2\pi \vec{k}_0 \vec{r} + \varphi - \omega t)) \quad (2.8)$$

with the wave vector $\vec{k}_0$ of modulus $k_0 = |\vec{k}_0| = 1/\lambda$, pointing in the direction of the electron momentum $\vec{p} = \hbar \vec{k}_0$, and a phase φ . The frequency ω is given by the total energy by means of the Einstein relation $E = \hbar\omega$ with $\hbar = h/2\pi$.

2.1. Electron interference

Interference is the most prominent peculiarity of waves and the very basis of wave optics. Therefore, electron interference is a fundamental phenomenon, e.g. in quantum physics, solid state physics and electron wave optics.

The description of electron interference is virtually the same as of the more familiar light interference. Two coherent waves $\psi_{1,2}(\vec{r}, t) = a_{1,2} \exp(i(2\pi \vec{k}_{1,2} \vec{r} + \varphi_{1,2} - \omega_{1,2}t))$, superimposed on each other, result in the wave $\psi_1 + \psi_2$, which again is a solution of the underlying wave equation. The intensity distribution is

$$I(\vec{r}, t) := (\psi_1 + \psi_2)(\psi_1 + \psi_2)^{cc} \quad (2.9)$$

with ψ^{cc} meaning the complex conjugate of ψ .

The simplest case is the one of two plane waves with $|\vec{k}_1| = |\vec{k}_2| = k_0$ hence $|\omega_1| = |\omega_2|$, superimposed at an angle; further on, $V = 0$ and $\vec{A} = \vec{0}$ is assumed in the following. Introducing a coordinate system such that $\vec{k}_1 = (k_x, 0, k_z)$ and $\vec{k}_2 = (-k_x, 0, k_z)$, the intensity distribution follows as

$$I(x, y) = I(x) = a_1^2 + a_2^2 + 2a_1a_2 \cos(2\pi q_c x + \Delta\varphi). \quad (2.10)$$

In addition to the sum of the single intensities a_1^2 and a_2^2 , it contains cosinoidal interference fringes with the carrier spatial frequency $q_c = 2k_x$, which are shifted aside by the phase difference $\Delta\varphi = \varphi_1 - \varphi_2$. Please note that, because of plane waves with $|\vec{k}_1| = |\vec{k}_2|$ and $|\omega_1| = |\omega_2|$, the intensity is independent from z and t . With respect to the sum of the single intensities, which would result from *incoherent superposition*, interference induces a local redistribution of the intensity. In any case, because $\int_0^{1/q_c} \cos(2\pi q_c x) dx \equiv 0$, the total intensity $a_1^2 + a_2^2$ (i.e. particle number) is conserved.

For detection of the interference pattern, the fringe contrast is much more essential than the intensity: at weak intensities, which often show up at the needed coherent illumination, long exposure times can help much, however, a poor contrast cannot be healed. In electron microscopy, the

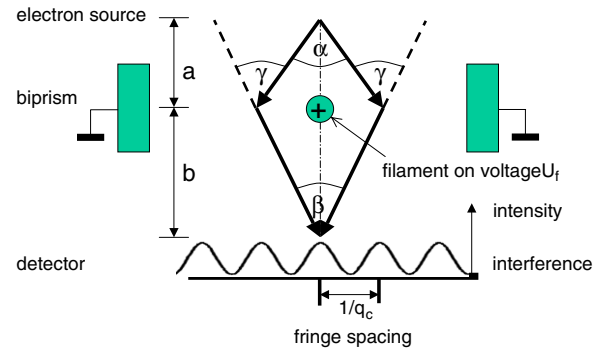


Figure 1. Electron biprism interferometer. The wave front emitted from the source is split by the biprism filament. Due to the positive filament voltage, the two waves passing on the right and the left are deflected by an angle γ and towards each other and hence superimpose in the detector plane. There, a cosinoidal interference pattern arises.

contrast in a point $\vec{r}$ is usually defined as $|I_0 - I(\vec{r})|/I_0$ with I_0 the vacuum intensity. Instead, here we use the definition of contrast

$$C := \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \quad (2.11)$$

with $I_{\max}$ and $I_{\min}$ meaning two adjacent maximum and minimum intensities, respectively. C is normalized, i.e. $0 \leq C \leq 1$. In the above case we obtain $C = (2a_1a_2)/(a_1^2 + a_2^2)$, which is only optimal, i.e. $C = 1$, if the amplitudes are tuned equal as $a_1 = a_2$. Later on it will be shown that also lack of coherence and experimental disturbances reduce the contrast. For $a_1 = a_2$ and with $I_0 = a^2$, we can finally write

$$I(x, y) = 2I_0(1 + \cos(2\pi q_c x + \Delta\varphi)). \quad (2.10a)$$

The Möllenstedt electron biprism. The ingeniously simple device for superposition of electron waves is the electron biprism. Gottfried Möllenstedt and his PhD student Heinrich Düker [3] invented the electron biprism as an electron optical analog of the Fresnel biprism in light optics. The electron biprism consists of a metalized quartz filament of about $0.5 \mu\text{m}$ $\emptyset$, at which a positive voltage U_f of several Volts is applied. It splits the wave front of an incoming electron wave into two coherent partial waves passing the filament on the right and left; in addition, it deflects the split waves by an angle $\gamma = \gamma_0 U_f$ towards each other and hence superimposes them in an area downstream (figure 1). The biprism is an ideal beam-splitter in that the deflection angle γ does not depend on the lateral distance from the filament, meaning that the two wave fronts do not suffer a distortion under deflection.

The resulting interference pattern depends on the simple geometry of the setup: in the detection plane at a distance b downstream, the two waves overlap at an angle $\beta = 2(a/(a+b))\gamma$ easily controllable with γ by means of the biprism voltage U_f . Because γ , and hence β , is very small, only some 10^{-4} rad, the carrier spatial frequency is given by $q_c = 2k_0 \sin(\beta/2) \approx k_0\beta$. Evidently, the fringe spacing

$$s = \frac{1}{q_c} = \frac{a+b}{2k_0a\gamma_0U_f} \quad (2.12)$$

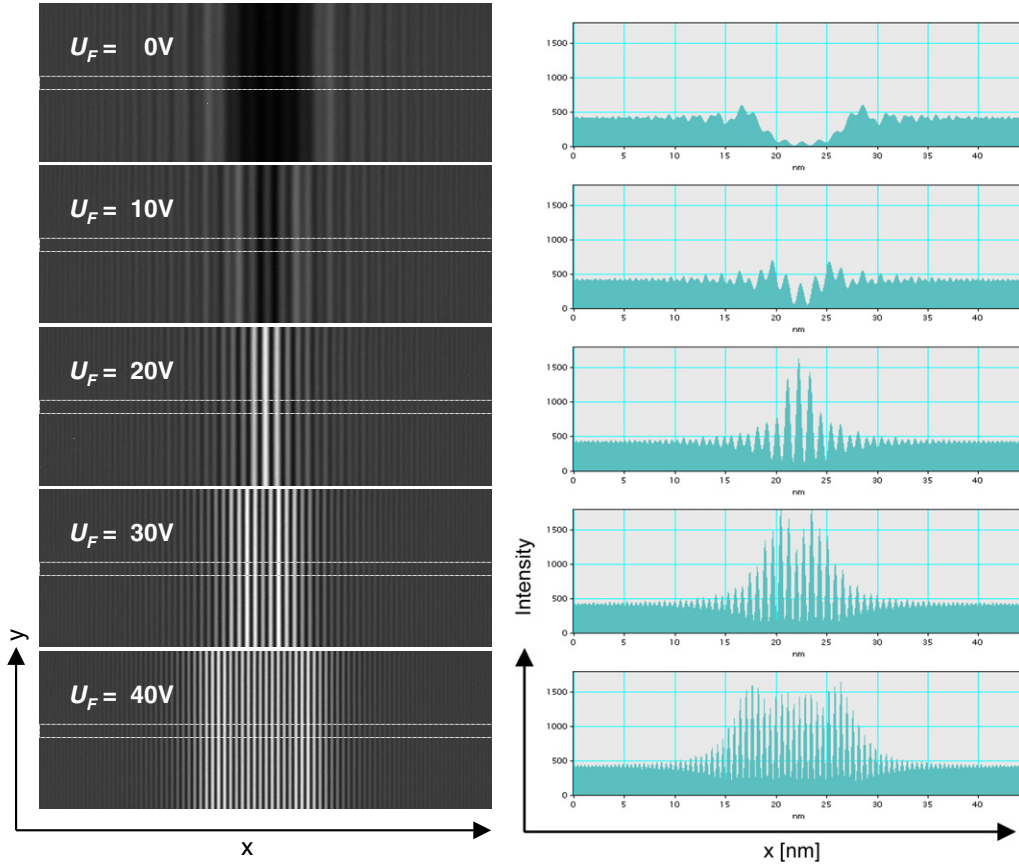


Figure 2. Electron interferograms for increasing biprism voltage. With increasing filament voltage U_f , the two waves overlap at an increasing width, and simultaneously the fringe spacing becomes smaller.

in the detector plane can simply be controlled by the biprism voltage U_f . The upper limit of spacing is given by the width of the shadow thrown by the filament in the observation plane, which requires a certain minimum γ to achieve superposition.

With increasing U_f , the overlapping width also increases according to

$$w = 2b\gamma_0 U_f - 2r_f \frac{a+b}{a} \quad (2.13)$$

with the radius r_f of the biprism filament; the second term is the width of the shadow thrown by the filament in the detector plane. Evidently, with increasing U_f , the fringe spacing shrinks, while the interference field becomes broader. This can be observed in figure 2.

Interference experiments with single electrons. If we consider a point-like and monochromatic source, subsequently emitted electrons start with random phases; nevertheless, they form spatially well-defined wave-fronts of identical shape. Superimposing the same parts of each wave, the random phases drop out. Therefore, and because of the same geometry for all waves, one obtains the same interference pattern for each electron

$$I(x, y) = 2I_0(1 + \cos(2\pi q_c x)), \quad (2.14)$$

which shows an ideal contrast $C = 1$. Therefore, summing over many electrons yields the same intensity distribution. The hypothesis that we are dealing with waves of single electrons is strongly supported by the experimental evidence for single

particle interference (figure 3). At a very small electron current such that only a few electrons arrive at the detector per exposure time, one finds only well-localized flashes at the impact position of each arriving electron: a single electron does not appear as a wave spread in the whole detector plane. Instead, it behaves like a particle, however different from a classical (deterministic) particle. The associated wave is—according to the Copenhagen-interpretation of quantum mechanics—a probability wave with an intensity describing the probability of finding an electron at a given position. The experiment further shows that interference is not based upon mutual interaction of the electrons. The time of flight of about 10^{-8} s between emitter and detector is much shorter than the average time distance of about 10^{-3} s between two subsequent impacts on the detector. This means that, when one electron hits the screen, the one to come next is still in the conduction band of the emitter about 1 m away. Since all the waves look the same and hence represent the same probability distribution, the electrons fill the common probability distribution with an increasing number of events building up the detectable interference pattern. In this sense, the electrons from a monochromatic point-source are called coherent.

Spin polarization is not considered here. The randomly spin-polarized electrons emitted from the usual source interfere in the sense of single-particle interference always with the same spin-state. Therefore, spin does not affect the appearance of the interference pattern, as long as, on the path from the source

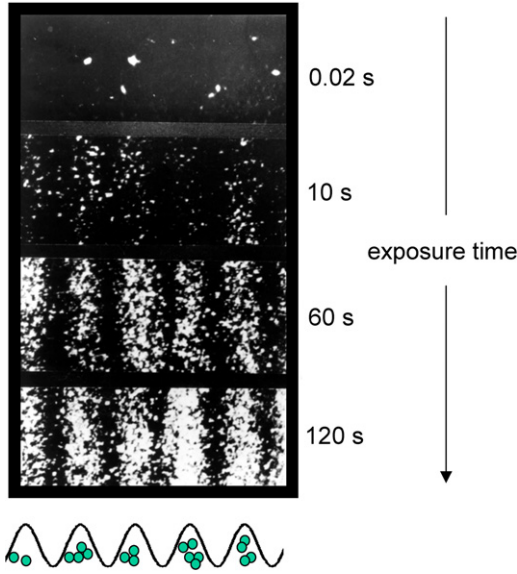


Figure 3. Single electron interference. At very short exposure times, seemingly stochastic impacts of single electrons are found. With increasing exposure time, this shot noise adds up forming a cosinusoidal interference pattern. Since the time of flight of the electrons is five orders of magnitude smaller than the time distance between two impacts, at any one time only a single electron is present in the interferometer.

to the detector, there are no spin-flip processes involved in one wave with respect to the other one. Therefore, spin effects are not yet observed and dedicated experiments changing spin have not yet been performed. However, effects of spin dependence such as those described in [4] promise a very interesting future extension of holography also in this direction.

2.2. Coherence of electrons

In reality, a theoretically possible point-like and monochromatic source would be useless, because it would not emit any electrons; otherwise, the current density at the emitter would be infinite. Consequently, experimentally we always deal with extended polychromatic electron sources. It is generally assumed that an extended and polychromatic electron source is an incoherent source in that electrons emitted from different points or with different energies do not show any detectable phase relation; therefore, they have to be summed up incoherently, i.e. by intensities.

In the following, the description of coherence is outlined at the example of the electron biprism; this description is analogous to the description in light optics for non-laser light, as given in detail, for example, in [5].

Spatial coherence. Spatial coherence of a wave field is determined by the local structure of the issuing source. If, rather than point-like, the source is extended in the ξ -direction parallel to x , and in the η -direction parallel to y , it can be described by the distribution function $i(\xi, \eta)$, which is assumed normalized by

$$\int_{\text{source}} i(\xi, \eta) d\eta d\xi = 1. \quad (2.15)$$

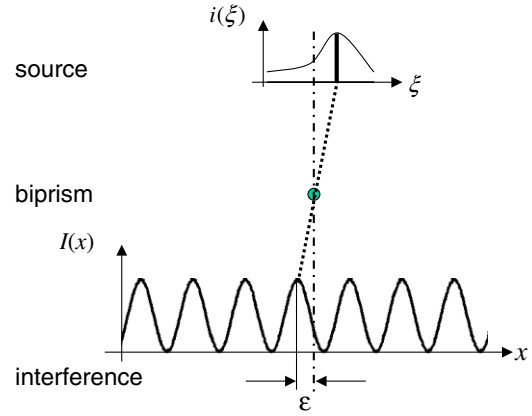


Figure 4. Spatial coherence from an extended source. Each point-like element of the source produces an interference pattern of contrast, $C = 1$, with a corresponding intensity and lateral phase ε . Summing incoherently over all source elements results in an interference pattern with reduced contrast $|\mu^{\text{sc}}|$ and averaged phase $\bar{\varepsilon}^{\text{sc}}$. $|\mu^{\text{sc}}| \exp(i\bar{\varepsilon}^{\text{sc}})$ is the degree of spatial coherence.

In the observation plane, it gives rise to the intensity I_0 .

Each point-like source element $i(\xi, \eta) d\eta d\xi$ contributes with an interference pattern

$$dI(x, y) = 2I_0 i(\xi, \eta) (1 + \cos(2\pi q_c x + \varepsilon(\xi))) d\eta d\xi, \quad (2.16)$$

which again shows the contrast $C = 1$ (figure 4). The lateral phase $\varepsilon(\xi)$ describes a lateral shift of the interference fringes corresponding to the considered source point coordinate ξ , which is assumed parallel to the x -coordinate in the interference pattern; the source coordinate η has no influence, since it is oriented parallel to the fringes. Summing up incoherently, i.e. by intensities, yields the resulting interference pattern

$$I(x, y) = 2I_0 \int_{\text{source}} i(\xi, \eta) (1 + \cos(2\pi q_c x + \varepsilon(\xi))) d\eta d\xi \quad (2.17)$$

and, because of the normalization of $i(\xi, \eta)$, one gets

$$I(x, y) = 2I_0 \left(1 + \int_{\xi} \left(\int_{\eta} i(\xi, \eta) d\eta \right) \cos(2\pi q_c x + \varepsilon(\xi)) d\xi \right). \quad (2.18)$$

This can be written again as a simple cosinusoidal fringe pattern, which, however, is damped in contrast by $|\mu^{\text{sc}}|$ and shifted aside by a phase $\bar{\varepsilon}^{\text{sc}}$

$$I(x, y) = 2I_0 (1 + |\mu^{\text{sc}}| \cos(2\pi q_c x + \bar{\varepsilon}^{\text{sc}})). \quad (2.19)$$

$\mu^{\text{sc}} := |\mu^{\text{sc}}| \exp(i\bar{\varepsilon}^{\text{sc}})$ is called the *degree of spatial coherence*.

By comparison one finds

$$\mu^{\text{sc}} = \int_{\xi} \left(\int_{\eta} i(\xi, \eta) d\eta \right) \exp(i\varepsilon(\xi)) d\xi. \quad (2.20)$$

Inserting $\varepsilon(\xi) = 2\pi q_c (b/a)\xi$ and $q_c = k_0\beta$, one gets

$$\mu^{\text{sc}} = \int_{\xi} \left(\int_{\eta} i(\xi, \eta) d\eta \right) \exp(i2\pi k_0\alpha\xi) d\xi \quad (2.21)$$

with $\alpha = (b/a)\beta$ the angle at the source subtended by two respective waves superimposed subsequently at the angle β (figure 1).

This result gives the degree of spatial coherence between two points with the same y -coordinate in the detector plane, as seen from the incoherent source at an angle α : it is given by the Fourier-transform of the source projected on the ξ - and x -coordinate, i.e. in the direction parallel to the fringes. Because it depends on the angle α , it is also called *angular coherence*.

Up to now, the two points considered have the same y -coordinate, and therefore their mutual coherence is given by the Fourier transform of the source projected on the x -direction. The generalization of two points with arbitrary y -coordinates is straightforward using the 2D-Fourier-transform of the source $i(\xi, \eta)$ and the slice theorem. This theorem says that the correspondingly oriented slice through the origin of Fourier space gives the Fourier-transform of the respective projection in real space.

Therefore

$$\mu^{\text{sc}}(\vec{\alpha}) = \int_{\xi} \int_{\eta} i(\xi, \eta) \exp(i2\pi k_0(\alpha_x \xi + \alpha_y \eta)) d\eta d\xi \quad (2.22)$$

is the general spatial coherence function of a monochromatic wave-field, emitted from the incoherent source $i(\xi, \eta)$. It depends on the angular distance $\vec{\alpha} = (\alpha_x, \alpha_y)$ of the two considered points in a plane, as seen from the source at a large distance (*van Cittert-Zernike theorem*).

The well-known condition for spatial coherence can easily be derived: assuming a slit source of full width ss , the Fourier transform is proportional to $\sin(\pi k_0 \alpha ss)/\pi k_0 \alpha ss$ with α perpendicular to the slit. At the first zero, one finds $\pi k_0 \alpha ss = \pi$, and therefore the coherence requirement reads as $ss \cdot \alpha \ll \lambda$.

Temporal or longitudinal coherence. In a similar way, the effect of the non-monochromaticity of the source can be treated. For this, a point-like, polychromatic source emitting a spectrum $s(k)$ of wave numbers k is considered. Now, the interference patterns obtained for different wave numbers have to be summed incoherently. Thereby, a temporal coherence function

$$\mu^{\text{tc}}(n) = \int_{\text{source}} s(\kappa) \exp(i2\pi \kappa n/k_0) d\kappa \quad (2.23)$$

results, where n is the order of interference; $\kappa = k - k_0$ means the spectrum coordinate relative to the nominal wave number $k_0 = \sqrt{2em_0 U_a^*}/h$, with the assumption $\kappa \ll k_0$ usually satisfied.

Again, the well-known description of longitudinal coherence can easily be derived. Let us assume a rectangular energy distribution with half width of ΔE around the nominal energy eU_a . Then, aside from a relativistic factor in the order of 1, $\Delta k/k_0 = \Delta E/(2eU_a)$ follows as half width of the corresponding distribution of the wave numbers. The Fourier transform results in $\mu^{\text{tc}}(n) \propto \sin(2\pi n \Delta k/k_0)/(2\pi n \Delta k/k_0)$ with the first zero at $n \Delta k/k_0 = 1/2$. Therefore, the total number of visible fringes (i.e. including both sides from zero order) follows as $2n = (2eU_a)/\Delta E$, and the usual coherence length as $L = 2n\lambda$.

The energy width of usual electron beams of only about 1 eV gives a maximum possible order $2n \approx 10^5$ of interference [6]. Therefore, for interference experiments and holography, which use $2n \leq 1000$ at most, temporal coherence does not impose any restriction.

Total degree of coherence. The total degree of coherence $\mu(\vec{\alpha}, n)$ of a wave-field issued from a source given by $i(\xi, \eta; k)$ is difficult to determine in general. If, however, every point of the source emits the same spectrum $s(k)$ of wave-numbers, an interference pattern

$$I(x, y) = 2I_0(1 + |\mu(\vec{\alpha}, n)| \cos(2\pi q_c x + \bar{e})) \quad (2.24)$$

with the factorized degree of coherence

$$\mu(\vec{\alpha}, n) = \mu^{\text{sc}}(\vec{\alpha})\mu^{\text{tc}}(n) \quad (2.25)$$

results. This is assumed to hold for usual electron sources.

Electron current in the coherently illuminated area. The decisive figure of merit for coherent experiments is the electron current available at a certain degree of coherence. Assuming a rotational Gaussian distribution of radius ρ

$$i(\rho) = \frac{1}{\sqrt{\pi}\rho_g} \exp\left(-\left(\frac{\rho}{\rho_g}\right)^2\right) \quad (2.26)$$

the degree of spatial coherence is given by

$$\mu^{\text{sc}}(\alpha) = \exp(-(\pi k_0 \alpha \rho_g)^2) \quad (2.27)$$

This is related to the axial brightness of an electron emitter defined as

$$B := \frac{I}{A\Omega}, \quad (2.28)$$

which is the current per emitting area and solid angle. Setting for a Gaussian source the emitting area $A = \pi \rho_g^2$ and the solid angle $\Omega = \pi \alpha^2$, the brightness can be combined with the respective degree of coherence, giving

$$I_{\text{coh}}(\mu^{\text{sc}}) = -\ln(|\mu^{\text{sc}}|) \frac{B}{k_0^2}. \quad (2.29)$$

Thus the coherent current available at a given degree of spatial coherence is simply given by the ‘reduced brightness’ B/k_0^2 of the electron source. Since the brightness increases linearly with the accelerating voltage U_a^* , and since $k_0^2 = 2em_0 U_a^*/h^2$, the reduced brightness and hence the coherent current are a mere property of the emitter, independent from U_a^* .

Distribution of the coherent current: elliptic illumination. The total coherent current is an optical invariant, which cannot be improved (enhanced) by optical methods such as focussing the beam; the invariance is a consequence of the ‘Abbe sine condition’ saying here that the product $A\Omega$ is invariant under imaging. Nevertheless, the total coherent current can be distributed in the object plane according to the special needs; here, a wide coherence area is needed only in the direction perpendicular to the interference fringes, whereas

in the direction parallel to the fringes it may be significantly smaller. Consequently, the coherently illuminated area in the object need not be circular. It is much more favourable to produce, by means of cylinder optics, a coherent ellipse instead: the long axis agrees with the circle diameter in the x -direction, in the y -direction; however, it is shrunk by a factor ε . Then the current density in the interference plane increases by the factor ε without reduction of interference contrast. Usually, $\varepsilon \approx 10$ brings a huge experimental improvement.

Particles and waves. The first zero of the $\sin(\pi ka ss)/(\pi ka ss)$ -function, which is the Fourier transform of a rectangular source of width ss , is found at $ss \cdot \alpha = \lambda$. From this degree of spatial coherence, as a rule of thumb for experiments, the familiar condition of spatial coherence

$$ss \cdot \alpha \ll \lambda \quad (2.30)$$

can be derived. It means that, in order to see interference phenomena, i.e. to observe electrons as waves, the product of source size ss and angular distance α of subsequently superimposed points must be much smaller than the wavelength λ .

On the other side, the Heisenberg relation, e.g. in the x -direction

$$\Delta p_x \Delta x \geq h/2\pi \quad (2.31)$$

can be translated with $\Delta x = ss/2$ and $\Delta p_x = p\alpha/2$ into

$$ss \cdot \alpha \geq 4\lambda/2\pi \quad (2.32)$$

saying that, in order to predict the position of detection of an electron, the product $ss \cdot \alpha$ has to be larger than about the corresponding wavelength.

Evidently, $ss \cdot \alpha$ gives a figure allowing to decide whether the particle or wave property of electrons can be observed: coherence means non-localizability, whereas localizability in the sense of Heisenberg means incoherence. Consequently, at coherent illumination, one must sum coherently over all paths about which there is uncertainty, i.e. about the corresponding phase space volume.

2.3. Elastic interaction with electric and magnetic fields

From the Schrödinger and the Klein-Gordon equation, respectively, in an area with electric potential V and a magnetic vector potential $\vec{A}$, the wave vector reads as

$$\vec{k} = \frac{\sqrt{2em_0(U_a + V)^*}}{h} \vec{e}_p - \frac{e}{h} \vec{A}. \quad (2.33)$$

Consequently, the eikonal is given as

$$\varphi = 2\pi \int_{\text{path}} \vec{k} \cdot d\vec{s} \quad (2.34)$$

$$= 2\pi \frac{\sqrt{2em_0 U_a^*}}{h} \int \sqrt{\frac{(U_a + V)^*}{U_a^*}} ds - 2\pi \frac{e}{h} \int \vec{A} \cdot d\vec{s} \quad (2.35)$$

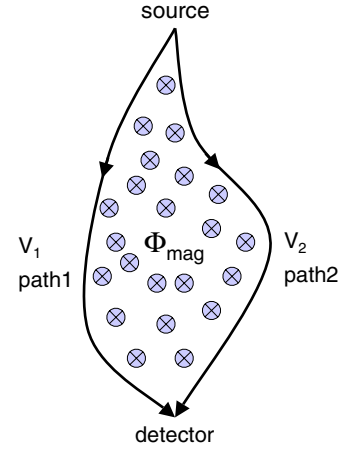


Figure 5. Phase shift of electron waves. On their different paths $path_1$ and $path_2$, the two waves collect an electric phase difference due to the difference of the electric potentials $V_{1,2}$. A magnetic phase difference arises by the magnetic flux Φ_{mag} enclosed by the two paths.

With the wave number

$$k_0 = \frac{\sqrt{2em_0 U_a^*}}{h} \quad (2.36)$$

in field-free space, one finally gets for the phase compared with vacuum

$$\varphi = 2\pi \frac{e}{h\nu} \int V ds - 2\pi \frac{e}{h} \int \vec{A} \cdot d\vec{s}, \quad (2.37)$$

where ν means the electron velocity.

Comparing in an interferometer two different paths through space (figure 5), one finds a phase difference

$$\Delta\varphi = \sigma (V_{\text{proj1}} - V_{\text{proj2}}) - 2\pi \frac{e}{h} \Phi_{\text{mag}} \quad (2.38)$$

with the interaction constant

$$\sigma = 2\pi \frac{e}{h\nu}, \quad (2.39)$$

the ‘projected potential’

$$V_{\text{proj1,2}}(x, y) = \int_{\text{path 1,2}} V(x, y, z) ds \quad (2.40)$$

and

$$\Phi_{\text{mag}} = \int_{\text{path1}} \vec{A} \cdot d\vec{s} - \int_{\text{path2}} \vec{A} \cdot d\vec{s} = \oint_{\text{path1-path2}} \vec{A} \cdot d\vec{s} \quad (2.41)$$

which, according to Stoke’s integral theorem, is the magnetic flux embraced by the closed loop formed by path 1 and path 2.

The result can be written as

$$\Delta\varphi = \Delta\varphi_{\text{el}} + \Delta\varphi_{\text{mag}} \quad (2.42)$$

with the electric phase shift

$$\Delta\varphi_{\text{el}} = \sigma (V_{\text{proj1}} - V_{\text{proj2}}) \quad (2.43)$$

and the magnetic phase shift

$$\Delta\varphi_{\text{mag}} = -2\pi \frac{e}{h} \Phi_{\text{mag}}. \quad (2.44)$$

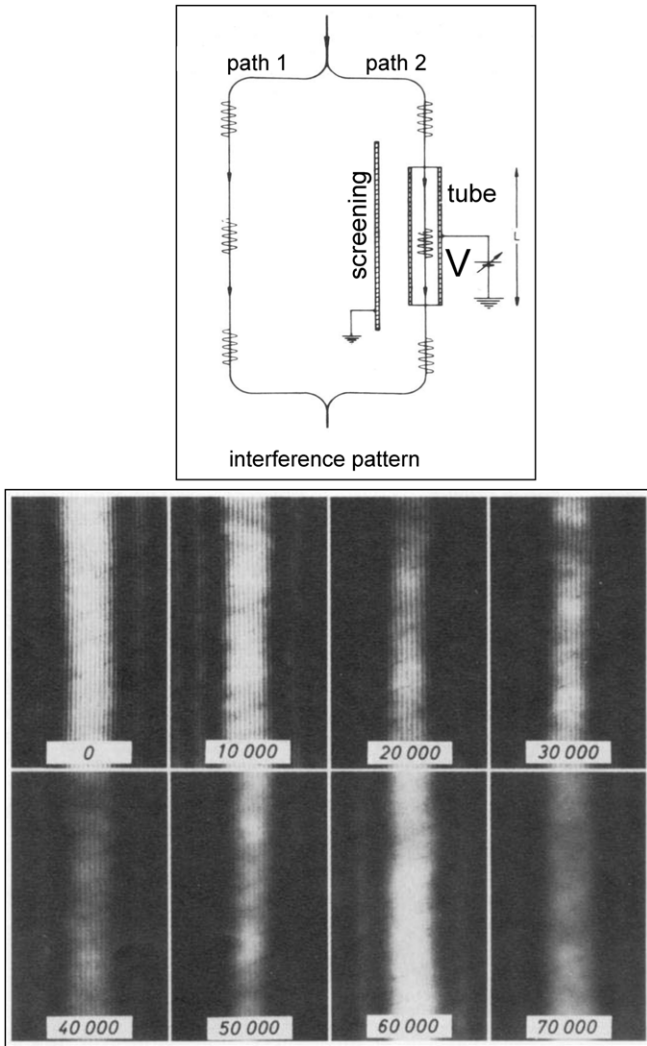


Figure 6. Measurement of electric phase shift and coherence length. One of the split waves is guided through a screened, metallic tube, to which an electric potential V with respect to ground can be applied. Phase shift measured versus V was found in very good agreement with theory. At very large phase shifts, the decay of contrast with the order of interference allows the coherence length to be measured: at a phase shift of $70\,000 \cdot 2\pi$, contrast disappears. This corresponds to a coherence length of 260 nm for 100 keV electrons. From [6].

Electric phase shift. The principle of the electric phase shift was studied in a special experiment [6]. In a biprism interferometer, only one of the two waves was guided through a metallic tube with an electric potential variable with respect to ground (figure 6). The corresponding phase shift, observed by means of the corresponding lateral shift of the interference fringes, was found in excellent agreement with equation (2.43). Furthermore, this experiment was extended for the measurement of the coherence length by producing very high phase shifts and measuring the decay of contrast.

Electric phase shift by solids. The most elementary electric field in solid state physics is the Coulomb field between the nucleus and the electron cloud in an atom. For calculation of the phase shift of the electron wave by a single atom, the atomic potential distribution $V_{\text{atom}}(x, y, z)$ has to be modeled. The

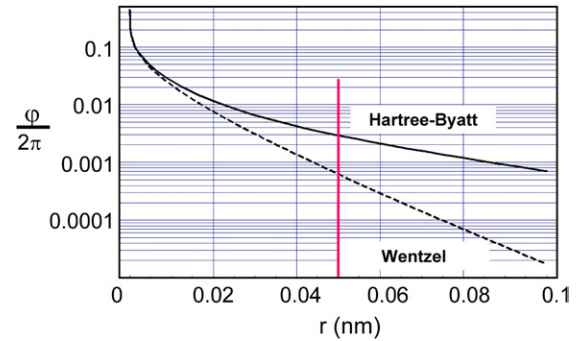


Figure 7. Phase shift by Coulomb potential of atoms. The calculations for a Ge-atom using the Hartree-Byatt model of atomic potentials shows a phase shift, which is peaked in the close vicinity of the nucleus and decays very rapidly with distance. The Wentzel model, only appropriate for very light atoms, gives even smaller phase shifts for Ge.

most simple model is given by the Wentzel distribution, which is valid particularly well for light atoms; the more general Hartree-Byatt model [7] gives a better accuracy especially for heavier atoms. Figure 7 reveals that the phase shift decays very rapidly with the distance from the nucleus such that at the atomic radius of 0.05 nm it is hardly detectable. Indeed, an atom looks like a sharply peaked point object, which, for a resolution worse than 0.1 nm, can roughly be visualized by a delta-function.

Scaling of atomic phase shift with atomic number. In general, the phase shift of atoms increases with the atomic number Z . Unfortunately, there is no simple scaling law because the shape of the electron clouds and hence the shape of the atomic potential changes with Z . By numerical evaluation one finds that there is a coarse dependence for the peak value $\varphi_{\text{atom}} \propto Z^{0.6}$ [8]. A more detailed analysis, however, shows a fine structure in that the atomic phase shift strongly depends also on the electron orbits; this may lead to the effect that sometimes increasing Z means even a reduction of the phase shift [9] (figure 8).

Mean inner potential. In a simple interferometer setup not combined with high-resolution imaging, the phase shift of single atoms cannot be observed, because it is blurred by diffraction in the distant detection plane. Nevertheless, there remains a phase shift detectable from the large area modulation; if the object is amorphous or tilted well away from zone-axis orientation, the phase shift is given by the mean inner potential (MIP) of the material, defined as

$$\text{MIP} := \frac{1}{\text{vol}} \int_{\text{vol}} V_{\text{atom}}(x, y, z) dx dy dz \quad (2.45)$$

integrated over the object volume vol (figure 9). For example, $\text{MIP} \approx 12$ V for silicon. Since the atomic potential $V_{\text{atom}}(x, y, z)$ is always larger than zero in vacuum, the same holds for MIP meaning that electrons are always faster in a material than in vacuum; therefore, the wavelength is shorter and hence the corresponding phase shift is always positive. The interferometric measurement of mean inner potentials was one

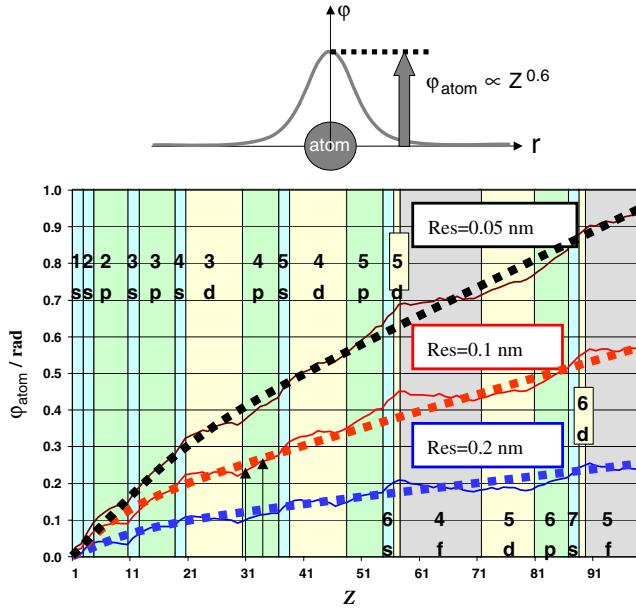


Figure 8. Phase shift of single atoms vs. atomic number Z . Top: schematic image phase distribution produced by an atom at limited resolution. The peak value φ_{atom} is roughly proportional to $Z^{0.6}$ with atomic number Z . bottom: fine structure found by simulation (full lines) in comparison with $Z^{0.6}$ (broken lines). Evidently, the electron orbitals have a strong influence. The signal strongly increases with resolution improving from 0.2 to 0.05 nm. From [9].

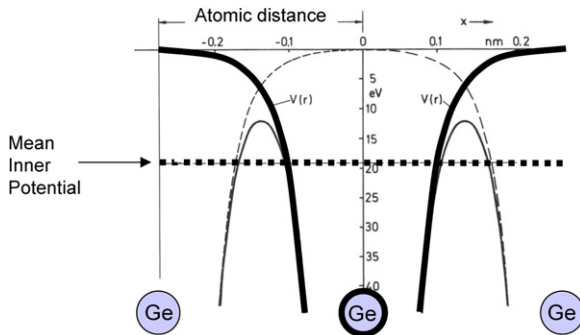


Figure 9. Mean inner potential. When the atomic structure is not resolved in an electron interferometer, one finds a phase shift corresponding to a potential, which is averaged about the potential of many atoms. This is called the mean inner potential. Adapted from [1].

of the first applications of the electron biprism interferometer realized by Möllenstedt and co-workers [10], as shown in figure 10. However, interpretation and calculation of the mean inner potential is a highly sophisticated task, because it is strongly influenced by size effects, bonding between the atomic constituents and valence electron densities in the object [11]. Therefore, accurate holographic measurements of the mean inner potential could also answer these basic questions raised in solid state physics.

Magnetic phase shift. The magnetic phase shift

$$\Delta\varphi_{\text{mag}} = -2\pi \frac{e}{h} \Phi_{\text{mag}} \quad (2.46)$$

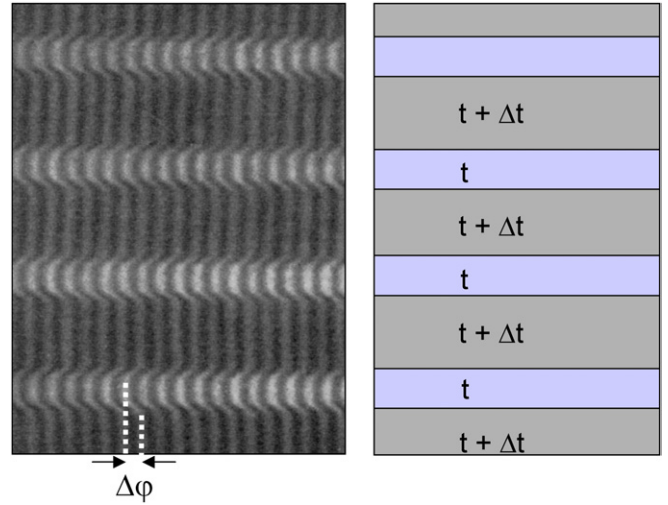


Figure 10. Electron interferometric measurement of the mean inner potential. The object is structured by stripes of different thickness t and $t + \Delta t$, respectively. The measured phase difference allows the mean inner potential to be determined. From [10].

depends solely on the magnetic flux Φ_{mag} enclosed between the two trajectories superimposed in a point of the interference pattern. It is interesting to note that the magnetic phase shift

1. does not offer any means to measure the magnetic vector potential $\vec{A}$. Also here, as in classical electrodynamics, $\vec{A}$ only plays the role of a transient quantity for calculation; it does not show up in any measurable result.
2. does not require that the electrons experience a Lorentz force, i.e. that a $\vec{B}$ -field is present at the trajectories. This is discussed as the Ehrenberg–Siday–Aharonov–Bohm effect in the following section.
3. does not depend on the energy of the electrons, if the trajectories run through field free space. Therefore, in contrast to the electric phase shift, it is not dispersive and does not increase the order of interference and hence does not allow to measure the coherence length of the electron wave to be measured by large magnetic phase shifts.

Ehrenberg–Siday–Aharonov–Bohm effect. Phase shift of electrons by the enclosed magnetic flux Φ_{mag} , without classical interaction with a $\vec{B}$ -field, was predicted by Ehrenberg and Siday [12] as well as by Aharonov and Bohm [13]. Therefore, we name it the ESAB-effect. It was experimentally verified by Chambers [14], inserting a thin magnetic iron whisker between the two waves. An even more striking experiment was performed by Möllenstedt and Bayh [15], who, by means of three biprisms, guided the two electron waves at the comparably large mutual distance of about $60 \mu\text{m}$; hereby they attained sufficient space between the waves to insert a micro-coil (figure 11). Changing the current through the coil, they found the predicted phase shift in the interference pattern of the two waves enclosing the flux. From their record in figure 11 one also finds the proof that the electrons did not experience any magnetic force in that the Fresnel fringes stay unchanged; since the Fresnel fringes occur from interference of waves diffracted at the first biprism, propagating to the detector along

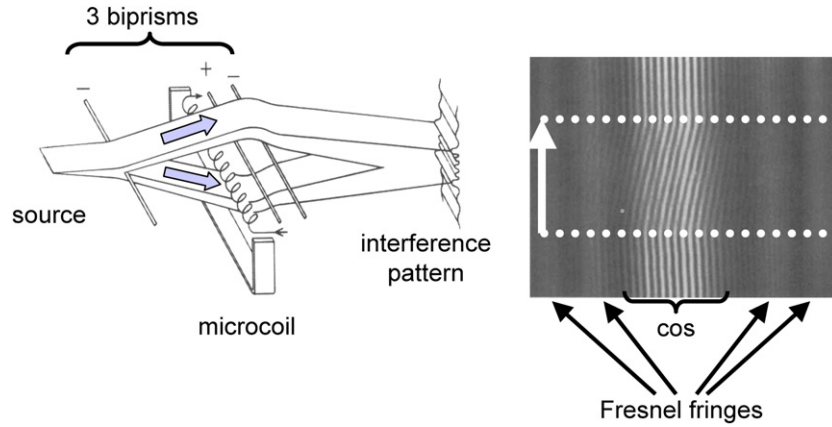


Figure 11. Magnetic phase shift due to an enclosed magnetic flux (ESAB-effect). Left: by means of 3 biprisms, the two coherent electron waves are guided around a microcoil at some distance. To prevent magnetic stray-fields from leaking out to the waves, a yoke short-circuits the magnetic flux. Right: the interference pattern is dynamically recorded, i.e. the white arrow shows time. The cos-fringe position (centre) is shifted aside with increase of current through the microcoil showing the magnetic phase shift of the two waves. The electrons do not experience any stray-field hence no Lorentz-force, otherwise also the Fresnel fringes would be displaced. From [15].

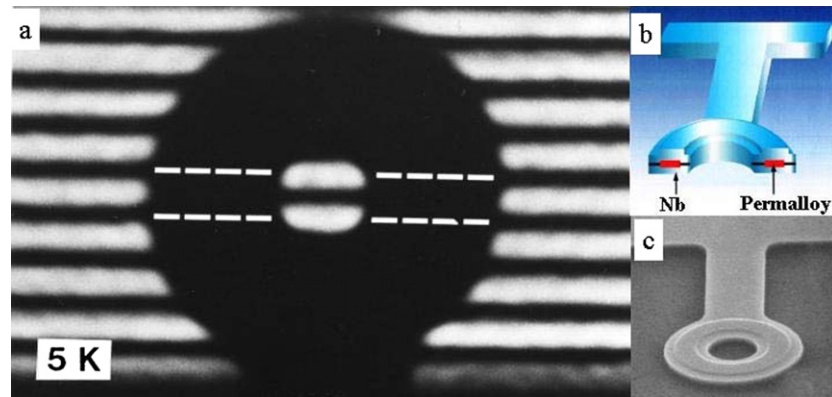


Figure 12. Magnetic phase shift due to enclosed magnetic flux (ESAB effect). (a) Interference pattern; (b) schematic setup; (c) SEM image. Doughnut-shaped (toroidal) ferromagnets six micrometres in diameter are covered with niobium superconductors to completely confine the magnetic field within the doughnuts to remove any flux leakage due to the Meissner effect. With the magnet maintained at 5 K, a phase difference of π is detected between one electron beam passing through the hole in the doughnut and the other passing on the outside of the doughnut. The phase quantization in π units indicating the flux quantization assures no flux leakage. From [16].

the same side of the coil, the straight Fresnel fringes prove that there was no stray $\vec{B}$ -field at the electrons trajectory. The ESAB-effect was later also verified by Schmid [6] with the micro-coil encapsulated in a glass tube evaporated with gold, to prevent electrons from entering the interior flux. Tonomura *et al* [16] used a magnetic ring, which was embedded in superconducting material to shield the magnetic flux perfectly from leaking outside (figure 12). In both cases the ESAB-effect was measured in excellent agreement with theory.

Also, if the electrons run through the $\vec{B}$ -field and hence experience a Lorentz force, the phase shift is always given by the enclosed magnetic flux. In general, the bending of the trajectories has to be taken into account for evaluation of the integral. With thin, electron-transparent magnetic films, however, one can mostly neglect the bending and integrate simply over z (figure 13).

Comparison of phase shift and trajectory displacement. A tilt of the trajectory of the electron particle can be considered equivalent to the phase gradient of the corresponding electron

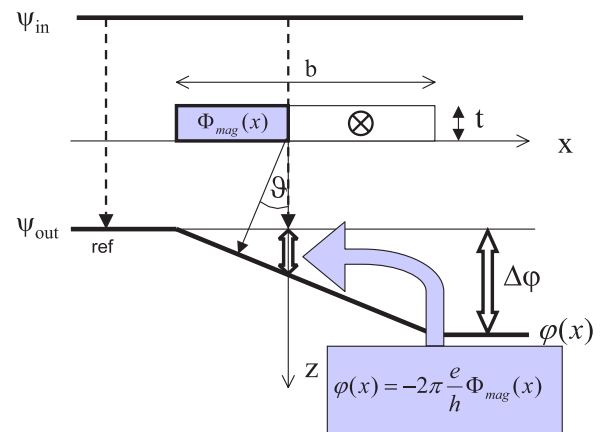


Figure 13. Phase shift from a magnetized thin film. With respect to the indicated reference point, the phase of the object exit wave ψ_{out} is shifted proportional to the respectively enclosed flux $\Phi_{\text{mag}}(x)$. For large x -values beyond the field region, the phase remains at the shifted level according to the ESAB-effect.

wave. Since in field-free space the wave front is the surface orthogonal to the bundle of all trajectories, the relation

$$\nabla\varphi = 2\pi \vec{k} = \frac{2\pi}{h} \vec{p} \quad (2.47)$$

holds. Consequently, wave and trajectories contain equivalent information about fields and potentials and hence both can be used for measuring them. Nevertheless, wave and particle are not completely equivalent as to measurement accuracy.

To illustrate the relation between wave and particle aspect, assume the situation sketched in figure 13: an electron beam is transmitted through a thin film of width b and thickness t , which carries a B -field oriented in the y -direction perpendicular to the drawing plane. On one side, the trajectories are deflected by the Lorentzian force resulting in the angle

$$\vartheta = \frac{p_x}{p_z} \approx \frac{eBt}{p}. \quad (2.48)$$

On the other side, here the phase gradient of the electron wave is given by

$$\frac{\partial\varphi}{\partial x} = 2\pi \frac{e}{h} \frac{\partial\Phi_{\text{mag}}}{\partial x} = 2\pi \frac{e}{h} B t = 2\pi \frac{\vartheta p}{h}. \quad (2.49)$$

Consequently, formally, the phase shift $\Delta\varphi$ can be mathematically determined by measuring $\vartheta(x)$ and integrating over x

$$\Delta\varphi(x) = \int_0^x \frac{\partial\varphi}{\partial x} dx = \frac{2\pi}{h} \int_0^x \vartheta p dx. \quad (2.50)$$

However, because of the Heisenberg relation for particles

$$\Delta p_x \Delta x \geq h/2\pi, \quad (2.51)$$

which, with $\Delta x = b/2$, reads here

$$p\vartheta \cdot b \geq h/\pi \quad (2.52)$$

it follows a classical limit

$$\Delta\varphi_{\text{class}}(x) \geq 2. \quad (2.53)$$

Consequently, from the distribution of deflection angles of particle trajectories at incoherent illumination, e.g. obtained by *Lorentz microscopy*, only structures can be determined significantly, which, in the wave image, would produce a phase shift of at least 2 rad. This was already pointed out by Wohlleben [17] and Cohen [18]. Using their Heisenberg-relation $\Delta p \cdot b \geq h$, for evaluating trajectories, one finds the classical limit $\Delta\varphi_{\text{class}}(x) \geq 2\pi$, which is more intuitive. In any case, by interferometric phase measurements, however, small fractions of 2π can easily be measured; the Heisenberg-limit does not exist for waves. The measurement accuracy of waves is only limited by the signal/noise properties of the interferogram, as later shown in section 4.

Phase shift: résumé. Interpretation of the measured phase shift in terms of an object is often very tedious. For example, at an object consisting of both electric and magnetic phase shifting components, there is the need for distinguishing the two. This can be achieved by taking two holograms at different accelerating voltages, because the electric phase shift is sensitive against the electron velocity whereas the magnetic one is not. Another possibility [19] needs to take a second hologram with an object flipped over. Then the sum of the two measured phase distributions reveals the electric component whereas the difference reveals the magnetic one.

2.4. Inelastic interaction

Decoherence by energy transfer: coherence at inelastic interaction. The interaction of electrons with electric and magnetic fields is perfectly elastic, i.e. it does not give rise to any energy transfer, as long as the fields are constant in time. If, however, e.g. the electric potential changes while an electron is in the field, energy is transferred from the field to the electron or vice versa ('inelastic'). If this happens only in one of the two waves, there arises an energy difference δE between the interfering waves $\psi_{1,2}(\vec{r}; t) = a_{1,2} \exp(i2\pi \vec{k}_{1,2}\vec{r} - \omega_{1,2}t)$. For small energy differences such as ≤ 1 eV, the resulting spatial frequency of the fringes may be written as $q_c = (k_1 + k_2)\beta/2$, which is virtually unchanged. However, the energy difference gives rise to a time-dependent interference pattern ('beat')

$$I(x, y; t) = 2I_0(1 + \cos(2\pi q_c x + 2\pi \nu t)) \quad (2.54)$$

with $\nu = \delta E/h$. This beat-phenomenon is extremely sensitive. For example, a beat frequency of $\nu = 1$ cps results for $\delta E = 4.135 \times 10^{-15}$ eV. It depends on the time resolution τ of the detector, whether one can still detect the interference fringes in the recorded intensity

$$I_{\text{rec}}(x, y) = 2I_0 \left(1 + \frac{1}{\tau} \int_0^\tau \cos(2\pi q_c x + 2\pi \nu t) dt \right), \quad (2.55)$$

which can be written as

$$I_{\text{rec}}(x, y) = 2I_0 (1 + |\mu^{\delta E}(\tau)| \cos(2\pi q_c x + 2\pi \nu \tau/2)) \quad (2.56)$$

with

$$\mu^{\delta E}(\tau) = |\mu^{\delta E}(\tau)| \exp(i2\pi \nu \tau/2) = \frac{1}{\tau} \int_0^\tau \exp(i2\pi \nu t) dt. \quad (2.57)$$

hence with

$$|\mu^{\delta E}(\tau)| = \frac{\sin(\pi \nu \tau)}{\pi \nu \tau} \quad (2.58)$$

with a first zero at

$$\tau = \frac{1}{\nu}. \quad (2.59)$$

Consequently, it strongly depends on exposure time τ , whether coherence is found or not. For frequencies $\nu > 1$ cps, i.e. $\delta E > 4.135 \times 10^{-15}$ eV, one would need such short exposure times that the number of collected electrons is very low, hence the signal disappears in quantum noise. Compared

with 4.135×10^{-15} eV, the energy transfer to a wave, usually occurring at inelastic interaction, is huge: excitation of phonons with several meV, of plasmons at several eV or of inner shell excitations at several 10 eV destroys coherence with respect to an elastically scattered wave, or to an unscattered one.

Contrast reduction by inelastic interaction. We start again with two coherent waves $\psi_{1,2}$ producing the interference pattern

$$I(x, y) = a_1^2 + a_2^2 + 2a_1a_2 \cos(2\pi q_c x). \quad (2.10)$$

Assume that the wave ψ_2 is suffering an inelastic interaction $\gg 10^{-15}$ eV with a probability P_{inel} that an electron suffers at least one inelastic event. Then the fraction $P_{\text{inel}}a_2^2$ of the intensity of ψ_2 will be incoherent with ψ_1 , whereas $(1 - P_{\text{inel}})a_2^2$ will remain coherent with it. Therefore, the interference pattern results

$$I(x, y) = a_1^2 + a_2^2 + 2a_1a_2\sqrt{1 - P_{\text{inel}}} \cos(2\pi q_c x). \quad (2.60)$$

This means that the contrast is reduced by the factor

$$C_{\text{inel}} = \sqrt{1 - P_{\text{inel}}}. \quad (2.61)$$

Including coherence effects, the interference pattern reads

$$I(x, y) = a_1^2 + a_2^2 + 2|\mu| C_{\text{inel}} a_1 a_2 \cos(2\pi q_c x) \quad (2.62)$$

with the contrast $|\mu| C_{\text{inel}} \frac{2a_1 a_2}{a_1^2 + a_2^2}$.

For the usual assumption

$$P_{\text{inel}} = 1 - \exp(-d/\lambda_{\text{inel}}) \quad (2.63)$$

with object thickness d and mean free path for inelastic interaction λ_{inel} , one obtains

$$C_{\text{inel}} = \exp(-d/2\lambda_{\text{inel}}). \quad (2.64)$$

Inelastic coherence. The interesting question remains, whether there is spatial coherence, and how wide is the coherence area, within an inelastically scattered wavefield of a well-defined energy loss. To answer this question, the interferometer has to be combined with an energy filter. A first experiment showed coherence [20]; in a second one, the area of coherence was estimated larger than 10 nm [21]. A more systematic investigation measured the decay of the degree of coherence of plasmon-scattered electrons with mutual distance, and it revealed that coherence reaches out as far as 30 nm [22–24]. Interestingly, the visibility of interference fringes proves that the two superimposed waves, inelastically scattered at $\delta E \approx 15$ eV, have an energy difference smaller than about 10^{-15} eV.

Please note that in these experiments the interferograms were recorded in the image plane, i.e. in the plane optically conjugate to the inelastic process. The finding of coherence in the inelastic wave fields is very interesting insofar as, at first sight, an inelastic interaction may be considered as a measurement process collapsing the wave into a point, hence showing no spatial coherence in the object plane. A

basic understanding may be established by the following possibly over-simplified reasoning: as shown by Howie [25], an energy transfer from E to $E \pm \delta E$ is always accompanied by a deflection with the angular spread $\vartheta_{\delta E} \approx \delta E/(2E)$. Therefore, the momentum of the electrons is correspondingly undefined, which gives rise to a Heisenberg-Uncertainty (equation (2.31)) for the position of the inelastic process, by the diameter

$$D = \lambda/(\pi \vartheta_E); \quad (2.65)$$

this is usually called *delocalization of inelastic interaction*. Because of the equivalence with the spatial coherence condition, the delocalization disc produced in the object plane agrees approximately with the extension of spatial coherence in the inelastic wave field issued from and referred back to the object. A more general and sophisticated description based on Rose and Kohl [26] is given by Verbeeck *et al* [27]

3. Information transfer in a transmission electron microscope (TEM)

An electron wave transmitted through an object may be modulated in amplitude and phase according to the interaction processes with the object structure. To reveal the object structure, the wave has to be imaged at a sufficient magnification using a microscope. The imaging process can be analyzed as follows:

Basically, an optical imaging system is a transfer system for waves. The input-signal is the object exit wave in the object exit plane with $\vec{r} = (x, y)$

$$\text{obj}(\vec{r}) = a(\vec{r}) \exp(i\varphi(\vec{r})), \quad (3.1)$$

which is transferred into the output-signal, i.e. the image wave found in the detector plane

$$\text{ima}(\vec{r}) = A(\vec{r}) \exp(i\phi(\vec{r})). \quad (3.2)$$

The coordinate $\vec{r}$ in the image wave is related to the object exit plane. The image wave is analyzed in order to yield data sets for the comprehensive characterization of the object structure. For interpretation of the image wave in terms of the object properties, one has to understand the relation between amplitudes $a(\vec{r})$ and $A(\vec{r})$, as well as between the phases $\varphi(\vec{r})$ and $\phi(\vec{r})$ of the two waves. This is established by means of the transfer theory sketched in the following.

3.1. Wave optical transfer of information: Abbe theory

According to Abbe, microscopic imaging is described in two steps, i.e. diffraction and interference of waves (figure 14). In the first step, the object exit wave

$$\text{obj}(\vec{r}) = a(\vec{r}) \exp(i\varphi(\vec{r})) \quad (3.3)$$

modulated by the object structure is propagating in free space. In correspondence with the Huygens principle, the propagating wave is described by the Kirchhoff diffraction integral. In the near field of the object, this is well approximated by a Fresnel Integral ('Fresnel diffraction'), e.g. giving rise to

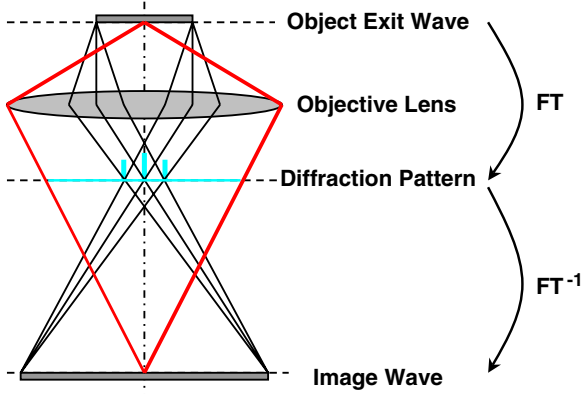


Figure 14. Abbe theory of imaging. The objective lens forms a pattern of diffracted waves of the object exit wave in the back focal plane; this is far-field diffraction described as a Fourier-transform. The mutually coherent diffracted waves interfere in the image plane building up the image wave, described as an inverse Fourier-transform. The achievable resolution is given by the opening angle of the objective lens, which can also be seen in the extent of the diffraction pattern.

Fresnel fringes; in the far field, the solution is given by a Fourier integral ('Fraunhofer diffraction'), which, in the case of periodic patterns in the object, shows well-separated reflections. It is called the Fourier spectrum of the object exit wave

$$\text{spec}(\vec{q}) := \int_{-\infty}^{\infty} \text{obj}(\vec{r}) \exp(-i2\pi \vec{q} \cdot \vec{r}) d\vec{r} \quad (3.4)$$

in short

$$\text{spec}(\vec{q}) = \text{FT}[\text{obj}(\vec{r})], \quad (3.5)$$

which again consists of amplitude and phase

$$\text{spec}(\vec{q}) = |\text{spec}(\vec{q})| \exp(i\rho(\vec{q})). \quad (3.6)$$

The spatial frequency $\vec{q} = k\vec{\alpha}$ is the variable in Fourier space, where $\vec{\alpha}$ is the *diffraction angle* as seen from the object, and k is the wave number.

Since the back focal plane of a lens is conjugate to infinity, one finds there roughly the Fourier-transform of the object exit wave. 'Roughly' means that the amplitude $|\text{spec}(\vec{q})|$ of the Fourier spectrum is correct. The phase found in the back focal plane, however, agrees with the phase $\rho(\vec{q})$ of the Fourier transform only, if the object exit wave is positioned in the front focal plane of the lens. Strictly speaking, Abbe theory correctly describes the imaging procedure by means of a $4f$ -setup of two lenses with two focal points coinciding. Nevertheless, at microscopic imaging using one lens at large magnification, the error is negligible.

In the second imaging step of Abbe, the Fourier spectrum is considered as a 2D-array of mutually coherent point sources at positions $\vec{q}$ emitting spherical waves weighted by $\text{spec}(\vec{q}) = |\text{spec}(\vec{q})| \exp(i\rho(\vec{q}))$. The waves propagate into the image plane and interfere. Since, at high magnification, the image plane is in the far-field of Fourier-space, the resulting wave is obtained again by a Fourier transformation, now leading from Fourier space back into real space by means of

$$\text{ima}(\vec{r}) = \int_{-\infty}^{\infty} \text{spec}(\vec{q}) \exp(+i2\pi \vec{q} \cdot \vec{r}) d\vec{q} \quad (3.7)$$

in short

$$\text{ima}(\vec{r}) = \text{FT}^{-1}[\text{spec}(\vec{q})]. \quad (3.8)$$

Inserting $\text{spec}(\vec{q}) = \text{FT}[\text{obj}(\vec{r})]$, one finds $\text{ima}(\vec{r}) \equiv \text{obj}(\vec{r})$, hence $A(\vec{r}) \equiv a(\vec{r})$ and $\phi(\vec{r}) \equiv \varphi(\vec{r})$, i.e. amplitudes and phases of image and object wave agree perfectly. This is called ideal imaging.

Interventions in the back focal plane. The Fourier spectrum $\text{spec}(\vec{q}) = |\text{spec}(\vec{q})| \exp(i\rho(\vec{q}))$ represents the object information sorted by spatial frequencies.

Interventions can alter the modulus $|\text{spec}(\vec{q})|$ by an aperture $B(\vec{q})$ damping or masking out diffracted waves from contributing, and the phase $\rho(\vec{q})$ can be changed by means of a phase plate $\chi(\vec{q})$. In general, these are accounted for by the *wave transfer function*

$$\text{WTF}(\vec{q}) = B(\vec{q}) \exp(-i\chi(\vec{q})). \quad (3.9)$$

Then the Fourier spectrum building up the image wave is given as

$$\text{spec}_{\text{ima}}(\vec{q}) = \text{spec}_{\text{obj}}(\vec{q}) \cdot \text{WTF}(\vec{q}) \quad (3.10)$$

and the image wave follows as

$$\text{ima}(\vec{r}) = \text{FT}^{-1}[\text{spec}_{\text{obj}}(\vec{q}) \text{WTF}(\vec{q})]. \quad (3.11)$$

With the convolution theorem of Fourier transformations

$$\text{FT}[f \cdot g] = \text{FT}[f] \otimes \text{FT}[g], \quad (3.12)$$

the image wave can also be written as

$$\text{ima}(\vec{r}) = \text{obj}(\vec{r}) \otimes \text{PSF}(\vec{r}), \quad (3.13)$$

i.e. by convolution ($\otimes$) of the object wave with the *point spread function*

$$\text{PSF}(\vec{r}) = \text{FT}^{-1}[\text{WTF}(\vec{q})]. \quad (3.14)$$

In signal transfer theory, PSF is called *impulsive response*.

Diffraction limited imaging. Due to the finite extension of the lens, the maximum acceptable diffraction angle α_{max} limits the Fourier spectrum by an equivalent aperture

$$B(q) = \begin{cases} 1 & \text{for } |q| \leq |q_{\text{max}}| = k\alpha_{\text{max}}, \\ 0 & \text{for } |q| > |q_{\text{max}}| = k\alpha_{\text{max}}. \end{cases} \quad (3.15)$$

Therefore the point spread Function is rotational symmetric, given as

$$\text{PSF}(r) = \frac{J_1[2\pi q_{\text{max}} r]}{2\pi q_{\text{max}} r} \quad (3.16)$$

with the Bessel-function of first order $J_1(z)$. For the first zero of $|\text{PSF}(r)|$, one finds

$$q_{\text{res}} = 1.6 k\alpha_{\text{max}} \quad (3.17)$$

as resolution. For the small angles occurring in a TEM, this is equivalent to the famous Abbe-formula for the minimum resolved distance

$$\delta := \frac{1}{q_{\text{res}}} = 0.61 \frac{\lambda}{\alpha_{\text{max}}}. \quad (3.18)$$

Interpretation. Although the resulting expression $\text{ima}(\vec{r}) = \text{obj}(\vec{r}) \otimes \text{PSF}(\vec{r})$ is very clearly interpreted visually as ‘smearing out’, the determination of the image wave is in general very difficult to perform analytically. Therefore, transfer is formulated in Fourier space. Without much loss of generality, one can restrict to weak objects, which frequently occur in electron microscopy. They have an amplitude $a(\vec{r}) = 1 - t(\vec{r})$ with the amplitude modulation $t(\vec{r}) \ll 1$, and a phase $\varphi(\vec{r}) \leq \frac{2\pi}{10}$. Then the object wave $\text{obj}(\vec{r}) = a(\vec{r}) \exp(i\varphi)$ can be expanded as

$$\text{obj}(\vec{r}) = (1 - t(\vec{r}))(1 + i\vec{\varphi}(\vec{r})) \quad (3.19)$$

and, omitting higher than linear terms,

$$\text{obj}(\vec{r}) = 1 - t(\vec{r}) + i\vec{\varphi}(\vec{r}) \quad (3.20)$$

remains.

Because of the involved Fourier transforms, it is very helpful to analyze the transfer for one spatial frequency q_0 , and to generalize the result correspondingly. Therefore, in the following, the object wave

$$\text{obj}(x) = 1 - (t_0 - i\varphi_0) \cos(2\pi q_0 x + \varepsilon) \quad (3.21)$$

is considered, which represents an elementary mixed amplitude/phase grating with a position relative to the optic axis given by the lateral phase ε .

The Fourier spectrum

$$\begin{aligned} \text{spec}_{\text{obj}}(q) = & \delta(q) + \frac{1}{2} \sqrt{t_0^2 + \varphi_0^2} \exp(i\vartheta) \{ \exp(i\varepsilon) \delta(q - q_0) \\ & + \exp(-i\varepsilon) \delta(q + q_0) \} \end{aligned} \quad (3.22)$$

consists of three reflections. The phase of the two off-axis reflections is given by $\vartheta = \arctan \left[\frac{\varphi_0}{t_0} \right]$ symmetrical in q , and by the lateral phase ε , which is anti-symmetrical in q . Symmetry and anti-symmetry of the phases decide whether they contain the amplitude/phase property of the grating such as ϑ , or the lateral position such as ε .

Under ideal imaging, the image wave perfectly agrees with the object wave and hence the image intensity reads as

$$\begin{aligned} I_{\text{ima}}(x) = \text{ima} \cdot \text{ima}^{\text{cc}} = & 1 - 2t_0 \cos(2\pi q_0 x + \varepsilon) \\ & + (t_0^2 + \varphi_0^2) \cos^2(2\pi q_0 x + \varepsilon). \end{aligned} \quad (3.23)$$

Whereas the wave is always transferred linearly, in intensity only the first two terms are linearly related to object properties. The third one is quadratic in t_0 and φ_0 , which, because of $\cos^2(z) = \frac{1}{2}(1 + \cos(2z))$, represents the double spatial frequency $2q_0$ (‘half-spacings’). This is an artifact of the intensity not present in the object, because the linear intensity contributions display the true object structure. Additionally, the contrast $C = 2t_0$ only shows the amplitude modulation (‘amplitude contrast’); phase contrast, i.e. contrast induced by the object phase φ_0 , is missing. (figure 15)

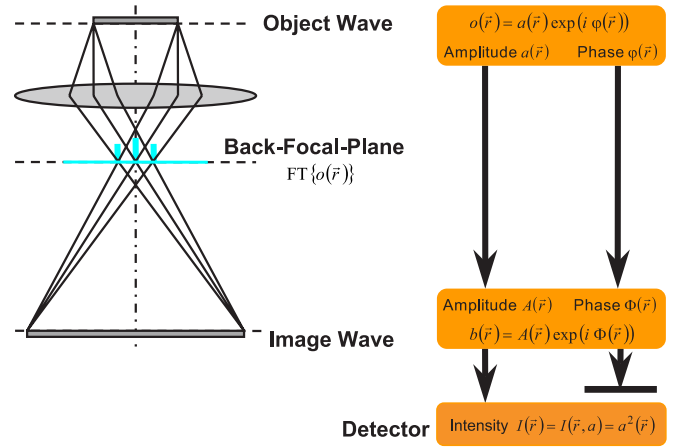


Figure 15. Wave transfer by an ideal objective lens. Amplitude and phase of the object wave are uniquely transferred into amplitude and phase of the image wave. In the image intensity one finds the amplitude squared, whereas the phase is not displayed. The phase channel is blocked. The scheme is valid for a weak object.

Zernike phase contrast by a ‘ $\lambda/4$ -phase plate’. Zernike [28] recognized that the mathematical difference between amplitude and phase is that the phase is multiplied by the imaginary unit $i = \exp(i\pi/2)$. He showed the way to optimum phase contrast by inserting into Fourier space a ‘ $\lambda/4$ -phase plate’, which introduces an additional phase shift $\chi(q) = \pm\pi/2$ between the zero beam ($q = 0$) and the diffracted beams ($q \neq 0$). This phase plate changes the object spectrum to the image spectrum

$$\begin{aligned} \text{spec}_{\text{ima}}(q) = & \delta(q) + \frac{1}{2} \sqrt{t_0^2 + \varphi_0^2} \exp \left(i \left(\vartheta \pm \frac{\pi}{2} \right) \right) \\ & \times \{ \exp(i\varepsilon) \delta(q - q_0) + \exp(-i\varepsilon) \delta(q + q_0) \}, \end{aligned} \quad (3.24)$$

which results in an image wave

$$\text{ima}(x) = 1 - (\pm i t_0 \mp \varphi_0) \cos(2\pi q_0 x + \varepsilon). \quad (3.25)$$

Now, amplitude and phase in the image wave are exchanged with respect to the object wave, hence the linear intensity component

$$I_{\text{ima}}(x) = 1 \mp 2\varphi_0 \cos(2\pi q_0 x + \varepsilon) \quad (3.26)$$

shows perfect Zernike phase contrast $C = 2\varphi_0$, whereas the amplitude is invisible (figure 16).

Effect of a more general phase plate. If the phase plate shifts the phase of the Fourier components by an arbitrary value χ , the image spectrum reads as

$$\begin{aligned} \text{spec}_{\text{ima}}(q) = & \delta(q) + \frac{1}{2} \sqrt{t_0^2 + \varphi_0^2} \exp(i(\vartheta + \chi)) \\ & \times \{ \exp(i\varepsilon) \delta(q - q_0) + \exp(-i\varepsilon) \delta(q + q_0) \}, \end{aligned} \quad (3.27)$$

which produces an image wave

$$\text{ima} = 1 - (T_0 + i\phi_0) \cos(2\pi q_0 x + \varepsilon) \quad (3.28)$$

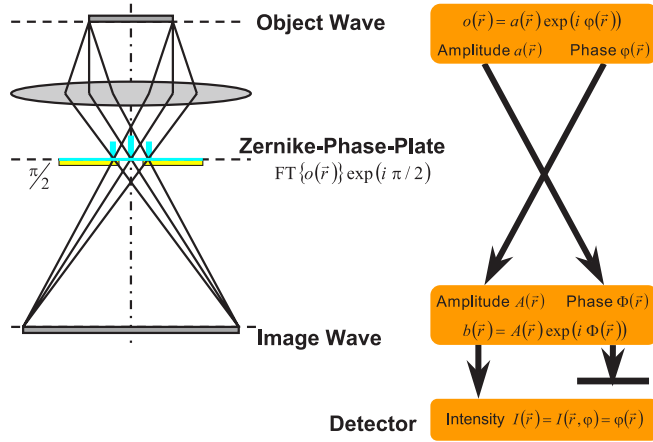


Figure 16. Wave transfer with Zernike phase plate. Inserting a $\pi/2$ -plate shifting the phases of the diffracted waves (or of the undiffracted wave) exchanges the amplitude and the phase channel between object and image wave. Now in the image intensity the object phase is visible. The image phase containing object amplitude is blocked. The scheme is valid for a weak object.

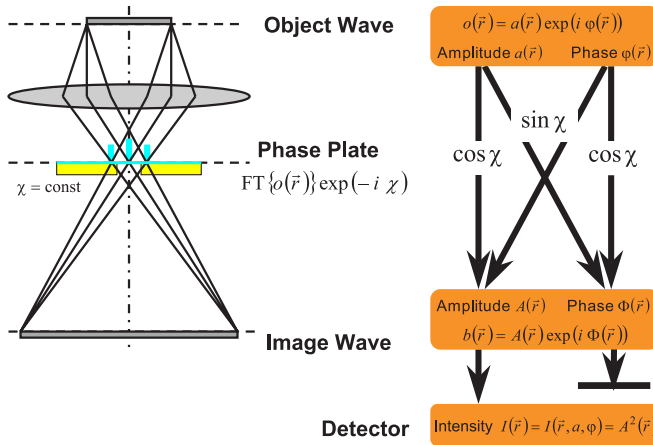


Figure 17. Wave transfer with arbitrary phase plate. An arbitrary phase shift χ of the diffracted waves (or of the undiffracted wave) produces an intermediate of figures 15 and 16. Partly the object amplitude and phase are distributed amongst the amplitude and phase of the image wave, straight and across according to the $\cos \chi$ and $\sin \chi$ -values. In the image intensity, one finds a corresponding mixture of the object amplitude and phase. The image phase channel with the complementary object information is blocked. The scheme is valid for a weak object.

with the amplitude modulation

$$T_0 = (t_0 \cos \chi + \varphi_0 \sin \chi) \quad (3.29)$$

and the phase

$$\phi_0 = (t_0 \sin \chi - \varphi_0 \cos \chi). \quad (3.30)$$

Consequently, both amplitude and phase of the object wave contribute to both amplitude as well as phase of the image wave with the respective weighting factors $\cos \chi$ and $\sin \chi$. Therefore, in the image wave, amplitude and phase of the object wave are mixed up according to the scheme shown in figure 17.

Since the image intensity is given by

$$I_{\text{ima}}(x) = 1 - 2T_0 \cos(2\pi q_0 x + \varepsilon) \quad (3.31)$$

the weighting factors

$$\text{PCTF} := \sin \chi \quad (3.32)$$

and

$$\text{ACTF} := \cos \chi \quad (3.33)$$

control phase contrast and amplitude contrast and are hence called *phase contrast transfer factor* and *amplitude contrast transfer factor*, respectively.

3.2. Aberrations

Up to now, aberration-free imaging was considered. This is realized in light optics, since one has learned already in the 19th century, to build lenses free from aberrations. For example, spherical aberration was overcome by combining lenses with corresponding positive and negative spherical aberration coefficients. In electron optics, however, Scherzer showed that round lenses, which are free from space charges and have a focusing power constant in time, always have a positive spherical aberration coefficient [29]. Since the resolution of an electron microscope is predominantly determined by the objective lens, only the aberrations of the objective lens are considered in the following.

Coherent aberrations. *Coherent aberrations* means in short that these aberrations are effective also at perfectly coherent illumination of the object, both spatial and temporal. They can be described by means of a phase plate $\exp(-i\chi(\vec{q}))$, with the *wave aberration* $\chi(\vec{q}) = 2\pi k \sum_i \Delta_i$, which distorts the phases in Fourier space. For the transfer of the wave up to a resolution of $q_{\text{res}} = 10 \text{ nm}^{-1}$, the following contributions Δ_i have to be taken into account:

- Defocus $\frac{1}{2} D z \left(\frac{q}{k}\right)^2$ (3.34)

- Spherical aberration of 3rd order $\frac{1}{4} C_s \left(\frac{q}{k}\right)^4$ (3.35)

- Twofold astigmatism $\frac{1}{2} A_2 \left(\frac{q}{k}\right)^2 \cos(2(\alpha - \alpha_{A2}))$ (3.36)

- Threefold astigmatism $\frac{1}{3} A_3 \left(\frac{q}{k}\right)^3 \cos(3(\alpha - \alpha_{A3}))$ (3.37)

- Axial coma $\frac{1}{3} B \left(\frac{q}{k}\right)^3 \cos(\alpha - \alpha_B)$ (3.38)

with the azimuth α of $\vec{q}$. The phase plate $\exp(-i\chi(\vec{q}))$ represents the phase factor of the wave transfer function (WTF).

Since symmetry of the phases in Fourier space plays an important role, please note that the first three are point symmetrical about the optic axis; they change the Zernike phase ϑ and hence influence the amplitude/phase character of the object structure in the image wave. The last two are anti-symmetrical about the optic axis and hence change the lateral position ε .

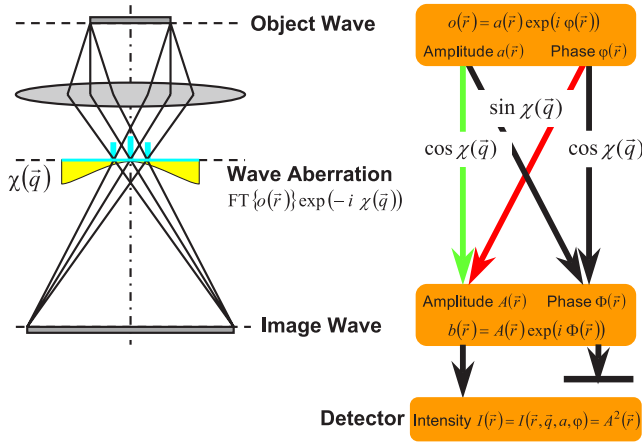


Figure 18. Wave transfer with coherent aberrations. Coherent aberrations form a phase plate $\chi(\vec{q})$. Again, the object wave is transferred along the four channels, however, this is more complicated in that the weighting factors are now dependent on spatial frequency $\vec{q}$. In the image intensity one finds a complex mixture of object amplitude and phase. A considerable amount of object information is lost in the blocked image phase channel. Scheme is valid for a weak object.

Optimization of coherent aberrations. Axial coma can be avoided by very careful alignment of the illumination direction on the optic axis of the objective lens. The two astigmatisms can be corrected by means of respective stigmators. Focus can easily be adjusted at any value. Until recently, spherical aberration could not be corrected. In the following, the route to atomic resolution with spherical aberration is sketched, which also shows a way around the missing phase contrast problem pointed out by Scherzer [30]. Scherzer showed that spherical aberration and defocus have to be combined in such a way that optimum phase contrast transfer results, because the most interesting components of the object structure, e.g. the atoms, are predominantly phase objects.

For an arbitrary object we have to consider all spatial frequencies. Therefore the previously introduced contrast transfer factors have to be generalized as functions, i.e. as *phase contrast transfer function*

$$\text{PCTF} := \sin \chi(\vec{q}) \quad (3.39)$$

and

amplitude contrast transfer function

$$\text{ACTF} := \cos \chi(\vec{q}). \quad (3.40)$$

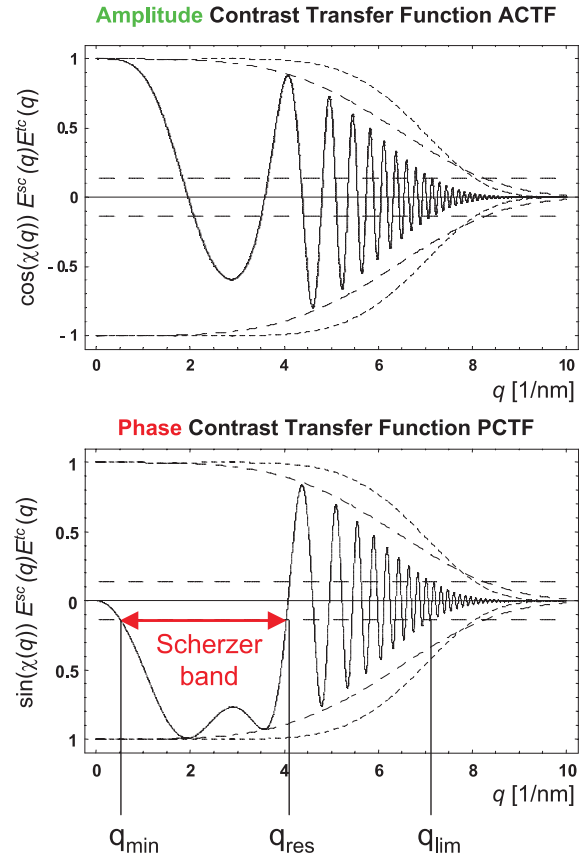
The corresponding flux diagram is shown in figure 18.

For optimum phase contrast, $\text{PCTF} = \pm 1$ and $\text{ACTF} = 0$ is needed for all spatial frequencies. Therefore, the residual wave aberration

$$\chi(q) = 2\pi k \left(\frac{1}{4} C_s \left(\frac{q}{k} \right)^4 + \frac{1}{2} D_z \left(\frac{q}{k} \right)^2 \right) \quad (3.41)$$

has to approach the Zernike phase value $\pi/2$ for a spatial frequency range of maximal width. This is achieved at the *Scherzer-focus* given by

$$D_{z,\text{sch}} = -1.2 \sqrt{\frac{C_s}{k}}. \quad (3.42)$$



Philips CM200FEG ST:

$C_s = 1.3$ mm, $D_{z,\text{sch}} = -68$ nm,

$C_c = 1.5$ mm, $\Theta_c = 0.1$ mrad, $\Delta E = 1$ eV

Figure 19. Contrast transfer functions at Scherzer focus. The phase contrast transfer exhibits a broad Scherzer band reaching from $q_{\min}$ to q_{res} , where all spatial frequencies are transferred at the same sign. This is, however, not true for the amplitude contrast transfer showing oscillations for these spatial frequencies. Beyond the information transfer limit q_{lim} the information is damped below e^{-1} and hence assumed lost in noise. Damping is due to the damping functions $E^{\text{sc}}(q)$ and $E^{\text{tc}}(q)$ from spatial and temporal coherence, respectively.

The PCTF at Scherzer-focus is plotted in figure 19. It shows that a rather broad Scherzer-band is transferred satisfactorily. For spatial frequencies beyond the first zero ('point resolution') at

$$q_{\text{sch}} = 1.5 \sqrt[4]{\frac{k^3}{C_s}}, \quad (3.43)$$

the PCTF increasingly oscillates, which gives rise to severe falsifications by contrast reversal and information loss at the subsequent zeros below noise. Therefore, q_{sch} is called point resolution limit, beyond which the interpretation is very difficult. Furthermore, there is no phase contrast for spatial frequencies smaller than

$$q_{\min} = 0.385 \sqrt[4]{\frac{k^3}{C_s}} \approx q_{\text{sch}}/4, \quad (3.44)$$

meaning that large area phase structures are missing in the intensity image. In any case, coherent aberrations distribute the

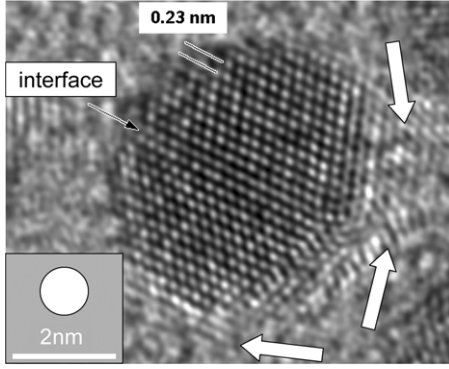


Figure 20. Gold particle in a TEM without Cs-corrector. The dots seem to represent the real atoms. However, information is delocalized by the point spread function with a diameter shown by the white circle above the marker. Therefore, the dots represent averaged atomic information, which may be displaced outside the particle (white arrows). Quantitative interpretation is very difficult locally, if possible at all.

object information into amplitude and phase of the image wave, however, they virtually do not destroy information. Summing up, a TEM is an imaging device with pass-band characteristics given by the Scherzer-band.

Delocalization of information. In the intensity image one finds atomic details with the resolution q_{schertz} , and even beyond, because of nonlinearities. Of course, the appearance of bright or dark dots on an atomic scale is very suggestive and leads one to believe that one sees the real atoms. However, this is only partially true, because the atomic dots may sit in wrong positions and have wrong intensities due to ‘delocalization of information’. Delocalization smears the information into a disc of confusion; it is described by the point spread function (PSF) mentioned earlier. Under incoherent illumination it would be a blurred disc without fine structure, but under coherent illumination it is filled with atomic dots, however averaged.

Of course, delocalization is a consequence of aberrations. A suitable measure for delocalization is the diameter of the PSF given as

$$\text{psf} = \text{grad}_{\max} \chi(\vec{q}) / \pi, \quad (3.45)$$

i.e. the maximum of the gradient of the wave aberration in the range of spatial frequencies admitted to the image [31]. Very interesting is the situation at Scherzer-focus. Inserting the corresponding wave aberration, one finds that

$$\text{psf} = 4.75 / q_{\text{schertz}} \quad (3.46)$$

is simply given by the point resolution. This means that we have a comparably strong delocalization. For example, at $q_{\text{schertz}} = 5 \text{ nm}^{-1}$, the atoms are found in a position, which may be wrong by $\text{psf} \approx 1 \text{ nm}$. Inside of large, perfectly periodic crystals, this just changes the contrast according to the transfer functions. At discontinuities like defects, interfaces and surfaces, however, the findings are not interpretable in terms of the object, neither qualitatively nor quantitatively (figure 20). A remedy against delocalization

is only the correction of aberrations, including spherical aberration.

Incoherent aberrations. In addition to the coherent ones, *incoherent aberrations* come up because of the limited degree of coherence of electron beams. In combination with the coherent aberrations, the spread of illumination direction (spatial coherence) and of wave-numbers (temporal coherence) gives rise to damping functions, which act like apertures. Spatial coherence results in the *spatial coherence damping function*, temporal coherence induces the *temporal coherence damping function*.

Spatial coherence damping function. At a tilt of illumination, the imaging process changes in two different aspects: first, the interaction with the object changes in that the projection direction and the diffraction at crystalline structures and hence excitation of reflection change in amplitude and phase; at small tilt angles, these effects can be neglected. The second aspect, however, has a much stronger influence in that tilting by an angle $\vec{\alpha}_{\text{ill}}$ shifts the Fourier spectrum of the object by the spatial frequency $\vec{Q} = k\vec{\alpha}_{\text{ill}}$. Therefore, it experiences a wave aberration $\chi(\vec{q} + \vec{Q})$, which can be expanded as

$$\chi(\vec{q} + \vec{Q}) \approx \chi(\vec{q}) + \vec{Q} \cdot \text{grad} \chi(\vec{q}) \quad (3.47)$$

for the usual small tilt angles. If the wave aberration $\chi(\vec{q}) = \chi(-\vec{q})$ is symmetrical, $\text{grad} \chi(\vec{q})$ is antisymmetric in $\vec{q}$, and hence the image wave of an elementary grating is laterally shifted as

$$\text{ima}_{\vec{Q}}(\vec{r}) = 1 + (T_0 + i\phi_0) \cos(2\pi q_0 x + \varepsilon + \vec{Q} \cdot \text{grad} \chi(\vec{q}_0)), \quad (3.48)$$

and the intensity reads as

$$I_{\text{ima}, \vec{Q}}(\vec{r}) = 1 - 2T_0 \cos(2\pi q_0 x + \varepsilon + \vec{Q} \cdot \text{grad} \chi(\vec{q}_0)). \quad (3.49)$$

Since the waves for different $\vec{Q}$ are incoherent, the intensity resulting under illumination from a normalized extended source $i(\vec{Q})$ is given by the sum of the respectively weighted intensities

$$I_{\text{ima}}(\vec{r}) = \int_{\text{source}} i(\vec{Q}) I_{\text{ima}, \vec{Q}}(\vec{r}) d\vec{Q}, \quad (3.50)$$

which results in

$$I_{\text{ima}}(\vec{r}) = 1 - 2E^{\text{sc}}(q_0)T_0 \cos(2\pi q_0 x + \varepsilon + \rho(q_0)) \quad (3.51)$$

with

$$E^{\text{sc}}(\vec{q}) \exp(i\rho(\vec{q})) = \int_{\text{source}} i(\vec{Q}) \exp(i2\pi \text{grad} \chi(\vec{q}) \cdot \vec{Q}) d\vec{Q} \quad (3.52)$$

given by the Fourier-transform of the source function $i(\vec{Q})$. Since it is related to the degree of spatial coherence, $E^{\text{sc}}(\vec{q})$ is called the damping function of spatial coherence.

In a well-aligned microscope with a rotational-symmetric Gaussian source with

$$i(\vec{Q}) = \frac{1}{\pi Q_0^2} \exp(-Q^2 / Q_0^2), \quad (3.53)$$

i.e. an illumination aperture (half width at half maximum) of

$$\alpha_{\text{ill}} = Q_0/k \quad (3.54)$$

and wave aberration given by spherical aberration and defocus only, one finally obtains

$$E^{\text{sc}}(\vec{q}) = \exp \left(-\pi^2 \frac{k^2 \alpha_{\text{ill}}^2}{\ln 2} \left(C_s \left(\frac{q}{k} \right)^3 + Dz \left(\frac{q}{k} \right) \right)^2 \right). \quad (3.55)$$

Temporal coherence damping. An analogous procedure is used to grasp the influence of the spectrum $s(k)$ of wave numbers emitted from the electron source. Of course, by means of $k = \sqrt{2m_0 E}/h$, $s(k)$ is related to the energy spectrum $G(E)$ occurring at the emission process, which is usually given as source-characteristics. With a FWHM-width of $\Delta E \approx 1 \text{ eV}$ at $U_a = 300 \text{ kV}$ one can estimate the width $\Delta k/k \approx \Delta E/(2E) \approx 10^{-6}$ of the wave number distribution to be very small.

Similarly to the tilt of illumination, a variation of the wave number introduces changes at the interaction with the object, e.g. diffraction angles, and the interaction coefficient σ . Again, these effects can be neglected, because the width of $s(k)$ is very small.

However, the influence on the imaging process has to be considered, since the wave aberration function $\chi(\vec{q})$ depends on k and hence on E . The main contribution stems from the defocus-term $\frac{1}{2} Dz \left(\frac{q}{k} \right)^2$ (3.34), because the focal power depends on electron energy and is normally given for the nominal energy eU_a^* . At a deviation δE , Dz has to be replaced by $Dz + C_c \frac{\delta E}{eU_a^*}$ depending on energy via C_c the coefficient of chromatic aberration of the objective lens. Therefore, the intensity of the weak phase grating reads as

$$I_{\text{ima}, \delta E}(\vec{r}) = 1 - 2T_0(\delta E) \cos(2\pi q_0 x + \varepsilon). \quad (3.56)$$

Integration over all these intensities weighted with the (normalized) energy spectrum $G(\delta E)$

$$I_{\text{ima}}(\vec{r}) = \int_{\text{source}} G(\delta E) I_{\text{ima}, E}(\vec{r}) d\delta E \quad (3.57)$$

delivers

$$I_{\text{ima}, E}(\vec{r}) = 1 - 2 |E^{\text{tc}}(q)| T_0 \cos(2\pi q_0 x + \varepsilon). \quad (3.58)$$

The function

$$E^{\text{tc}}(q) = \int G(\delta E) \exp \left(i\pi k_0 C_c \frac{\delta E}{eU_a^*} \left(\frac{q}{k_0} \right)^2 \right) d\delta E \quad (3.59)$$

is complex, if $G(\delta E)$ is not symmetric; then the phase describes an additional mean defocus. If $G(\delta E)$ is symmetric, $E^{\text{tc}}(q)$ is a real function (*temporal coherence damping function*) damping the resulting contrast of the image intensity of the grating with spatial frequency q . If, for example, $G(\delta E)$ is Gaussian with standard deviation std_E , it is given as

$$E^{\text{tc}}(q) = \exp \left(-\frac{\pi^2}{2} k_0^2 \Delta^2 \left(\frac{q}{k_0} \right)^4 \right) \quad (3.60)$$

with

$$\Delta = C_c \sqrt{\frac{\text{std}_E^2}{(eU_a^*)^2}} \quad (3.61)$$

the *focus spread*.

Usually, Δ is generalized by the contributions from variations of U_a and lens current I_{lens} , as

$$\Delta = C_c \sqrt{\frac{\text{std}_E^2}{(eU_a^*)^2} + \frac{\text{std}_{U_a}^2}{U_a^{*2}} + 4 \frac{\text{std}_{I_{\text{lens}}}^2}{I_{\text{lens}}^2}}. \quad (3.62)$$

In the case of conventional imaging of strong objects the situation is much more difficult. With the appropriate distribution functions for illumination aperture $F(\vec{\alpha}) \hat{=} i(\vec{Q})$ and for energy-related defocus $H(Dz) \hat{=} G(\delta E)$, at partial coherence the image intensity is found as

$$I_{\text{ima}}(\vec{r}) = \iint b(r) b^*(\vec{r}) F(\vec{\alpha}) H(Dz) d\vec{\alpha} dDz. \quad (3.63)$$

In Fourier space, this corresponds to

$$S_{I_{\text{ima}}}(\vec{q}) = \int_{-\infty}^{+\infty} S_{\text{obj}}(\vec{q} + \vec{h}) S_{\text{obj}}^*(\vec{h}) TCC(\vec{q} + \vec{h}, \vec{h}) d\vec{h} \quad (3.64)$$

with the transmission cross coefficient

$$TCC(\vec{q}', \vec{q}'') = \iint H(Dz) F(\vec{q}_{\text{ill}}) \exp(-i(\chi(\vec{q}' + \vec{q}_{\text{ill}}, z + Dz) - \chi(\vec{q}'' + \vec{q}_{\text{ill}}, z + Dz))) d\vec{q}_{\text{ill}} dDz \quad (3.65)$$

and the illumination direction given by $\vec{q}_{\text{ill}} = k\vec{\alpha}$ with wave number k . Evidently, in general $I_{\text{ima}}(\vec{r}) \neq A^2(\vec{r})$ holds.

Summary of aberrations. Coherent aberrations lead to a mixing of the amplitude and phase components of the object in the image. Therefore, interpretable resolution is limited to Scherzer-resolution. Mixing destroys object information only in the image intensity, whereas it does not destroy information in the image wave. Therefore, the full object information can be recovered by all methods analyzing the complete image wave, i.e. all sorts of holography.

Incoherent aberrations damp the contributions to the image intensity as well as to the image wave. They destroy information in both, if damped below noise. The resulting damping envelope function is usually taken as

$$E(\vec{q}) := E^{\text{sc}}(\vec{q}) \cdot E^{\text{tc}}(\vec{q}) \quad (3.66)$$

which defines the information limit q_{lim} , where damping leads below an assumed noise level.

Interpretation of findings. The transfer scheme (figure 18) together with the wave transfer function

$$\text{WTF}(\vec{q}) = E(\vec{q}) \cdot \exp(-i\chi(\vec{q})) \quad (3.67)$$

shows the problems of conventional imaging in a TEM:

1. Mixing of amplitude $a(\vec{r})$ and phase $\varphi(\vec{r})$ of the object wave in amplitude $A(\vec{r})$ and phase $\phi(\vec{r})$ of the image wave.

2. Phase contrast in the image intensity $A^2(\vec{r})$ only interpretable within the spatial frequency band reaching from $q_{\min}$ to q_{schertz} even at Scherzer focus for optimum phase contrast.
3. Spatial frequencies beyond q_{schertz} up to the information limit q_{lim} are transferred, however damped and with oscillating phase contrast transfer function hence difficult to interpret.
4. Spatial frequencies below about $q_{\min} \approx q_{\text{schertz}}/4$ invisible.
5. Delocalization of the ‘resolved’ structures by coherent blurring (point spread function) of the image wave prohibits analyzing the true structures at the interfaces and defects. Even at Scherzer focus, the delocalisation amounts to about $4.75/q_{\text{schertz}}$.
6. Nonlinear information (higher harmonics and mixing of different reflexions) shows up as artifacts in the image intensity, but not in the image wave.
7. Loss of the image phase $\phi(\vec{r})$ in the recorded image intensity.

In spite of the fact that these problems are theoretically precisely understood, the obtained images cannot be corrected from the aberrations, because the needed phase ϕ of the image wave is missing.

Cs-corrector for spherical aberration correction. Of course, right from the early days, electron microscopists were searching for a possibility of correcting spherical aberration. However, as already mentioned, Scherzer had already shown in 1936 [29] that the coefficient C_s of spherical aberration is always positive for round, time-independent lens fields, which are free from space charges. Therefore, the method of light-optics, i.e. combining usual lenses for compensating negative and positive spherical aberration, could not be realized in electron optics.

After attempts of many paving the way, Harald Rose in 1990 designed a system of (un-round) multipole lenses, which effectively shows imaging properties of a round system but a negative C_s -value [32] and hence is able to compensate the positive C_s of the usual objective lens. This *Cs-corrector* was successfully built and implemented into a TEM [33]. Meanwhile, a variety of TEMs with *Cs-corrector* are commercially available, where all aberrations of up to 3rd order can be corrected. This means a huge step forward in that ideal imaging up to a true resolution of $\approx 10 \text{ nm}^{-1}$ is presently reachable also in electron microscopy (e.g. [34, 35]). The information limit is finally determined only by the temporal damping function, i.e. by chromatic aberration.

The main benefits of *Cs-correction* are illustrated by means of figure 21. The delocalization has shrunk to the apparent resolution, the atomic dots are interpretable as atoms both in position and intensity. The consequence is that small details are clearly observable, such as small deviations from perfect lattice planes. In fact, often lattice planes are no planes at all.

However, the phase contrast problem is even worse with the *Cs-corrector*: corresponding to figure 15, the object phase is completely invisible in the intensity image. Here, only holographic methods can help.

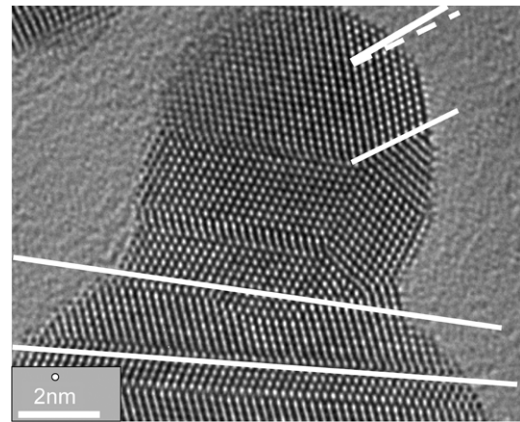


Figure 21. Gold particle in a TEM with *Cs-corrector*. The dot above the marker showing the diameter of point spread function has the same size as the atomic dots. These may now be interpreted as atoms. This image reveals that lattice planes are not planes but are actually curved due to the strain in the particle. In the top right, one finds a rotation of lattice planes indicating larger atomic distances at the surface than inside. Now quantitative interpretation is possible on the atomic scale. Recorded with a FEI Tecnai F20/*Cs-corr* electron microscope.

4. Electron holography

4.1. Basic idea

The Hungarian physicist Dennis Gabor realized the problems occurring with electron microscopy from the unavoidable spherical aberration of the objective lens. Therefore, he developed the idea of lens-less imaging by means of the following method [36, 37]: let the object wave propagate in space according to the well-known wave equation. If one succeeds in recording the complete wave with amplitude and phase by means of a detector at some distance, the wave can be back-propagated according to the same wave equation. The only critical point is that the detector must record the propagated wave completely, i.e. including both amplitude and phase. This complete detection was achieved by Gabor by interfering the propagated wave with a known reference wave. The arising interference fringes are modulated in contrast and position by amplitude and phase of the wave, respectively. This means that the wave is indeed recorded completely in an interference pattern, which he therefore named ‘hologram’ (from Greek *ὅλος* meaning *complete*).

Furthermore, Gabor recognized that the hologram can be considered a highly complicated diffraction grating. This grating has the astonishing property that, illuminated with a copy of the reference wave previously used for interference, it issues—amongst others—a diffracted wave, which is a revival of the propagated object wave. This one has to be isolated from the others and back-propagated to the object exit plane, to recover the object exit wave for further analysis. Interestingly, this ‘reconstruction’ can be accomplished by means of any wave following the wave equation, be it electromagnetic, mechanic, quantum or numeric, if the corresponding constraints (fringe distance, etc) are met. The most severe condition is that coherence is indispensable for both recording the hologram and reconstructing the wave.

In light optics, nowadays, with the LASER as a nearly perfectly coherent light source, holography is widespread in a variety of holographic techniques. The hologram may be recorded in the near-field (Fresnel holograms), in the far-field (Fraunhofer hologram), in the Fourier-spectrum (Fourier hologram) etc. In any case, the reconstructed object wave, accordingly back-propagated to the object exit plane, should always be the same.

Most essential, however, is the way in which the reference wave is superimposed on the object wave for recording the hologram. Initially, Gabor proposed *in-line holography*, where the reference wave propagates in the same direction as the object wave. He found that, under reconstruction, two conjugate waves (*twin-waves*) arise, which—in the in-line case—overlap coherently and hence cannot be separated from each other. This *twin-image problem* was solved by Leith and Upatnieks [38], who proposed to superimpose reference and object wave at an angle β . Then the reconstructed twin-waves are separated angularly by 2β and hence can be isolated from each other in Fourier space.

4.2. Electron holography

Historically, Gabor developed holography aiming at electron holography, to overcome the problems of electron microscopy. In fact, in 1951 Haine and Mulvey were the first to record electron holograms, namely Fresnel-in line holograms [39]. In the reconstructed wave they showed details of about 1 nm, limited by the twin-image problem. In those days, further progress was inhibited by the comparably poor stability of electron microscopes and the lack of more coherent electron sources.

In contrast to light holography using the LASER, electron holography started flourishing much later, since the needed degree of coherence is much more difficult to achieve with electrons. Therefore, coherent electron optics was performed only in a few especially experienced laboratories. For example in Tuebingen, off-axis electron interferometry was developed and well understood by Möllenstedt and co-workers from 1954. In 1968, Möllenstedt and Wahl, for holographic purposes, recorded the first *lens-less* (no objective lens) off-axis Fresnel hologram and successfully reconstructed the electron wave with LASER light [40]. However, thinking about the limits of this method, Wahl recognized that lens-less imaging is not promising for electron holography, because the achievable resolution is limited by the restricted degree of spatial coherence of electrons. The reason is that the waves diffracted at the object must be detected well inside the restricted coherently illuminated patch to be caught in the hologram, together with the coherent reference wave. Consequently, the farther away from the object the hologram is recorded, the more waves diffracted at large angles are missing in the hologram and hence also in the reconstructed wave. Only if the hologram is recorded in the object plane (or a conjugate image plane), all diffracted waves are collected; hence, not resolution, but field of view is limited by lack of coherence. Therefore, from light optics Wahl adopted the method of *image plane off axis holography* into electron microscopy and

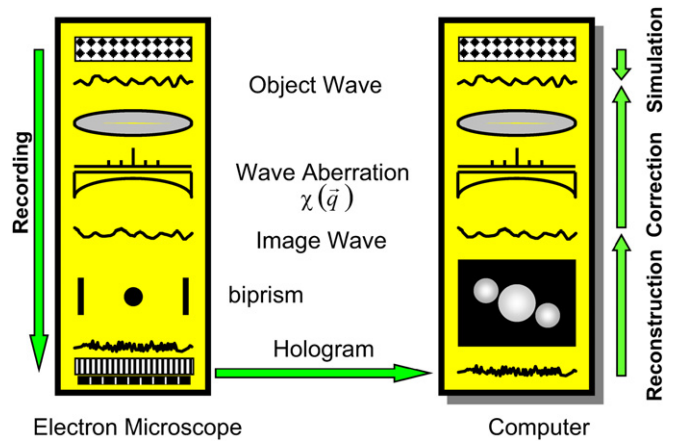


Figure 22. Scheme of image plane off-axis electron holography. Left: the object wave is imaged into the image wave by the conventional optics of the TEM. Superposition of the image wave with a reference wave by means of the electron biprism produces the hologram in the final image plane. The hologram is recorded by means of a CCD-camera. Right: the hologram is fed to the computer. After numerical reconstruction, the image wave is back-propagated through an equivalent virtual microscope, including aberration correction, if needed. The resulting object wave may be analyzed quantitatively by any wave optical tool.

developed *image plane off axis electron holography* [41]. To date, this is the most successful and widespread holographic method applied in electron microscopy. Therefore, other schemes for electron holography [42], which have been attempted without comparable success, are not discussed in the following.

4.3. Scheme of image plane off-axis electron holography

The basic scheme is shown in figure 22: in a TEM equipped with a highly-coherent field emission electron source and an electron biprism, the hologram is produced at adequate magnification in the final image plane. It is recorded by means of a CCD-camera and transferred to a computer, where the image wave is reconstructed with the help of numerical Fourier-optics. This means that the numerical image wave is back-propagated through a virtual microscope modelled according to the one used for recording the hologram; thereby, aberrations are corrected. At the object exit face, one finds the complete object exit wave, which can be evaluated according to all needs one can think about.

Recording the hologram. For recording a hologram, the TEM has to be modified only slightly in that an electron biprism is inserted such that superposition with a plane reference wave is achieved in the image plane. Usually, the biprism is inserted slightly above the first intermediate image produced by the objective lens (figure 23), because in any electron microscope there is available a port for the selected area aperture, which may also be used for a biprism holder. Unfortunately, this is a non-optimum compromise, because at a fixed geometry the overlapping width (giving field of view) and fringe spacing (giving resolution) cannot be selected independently. Optimum positions of the electron biprism in the path of rays

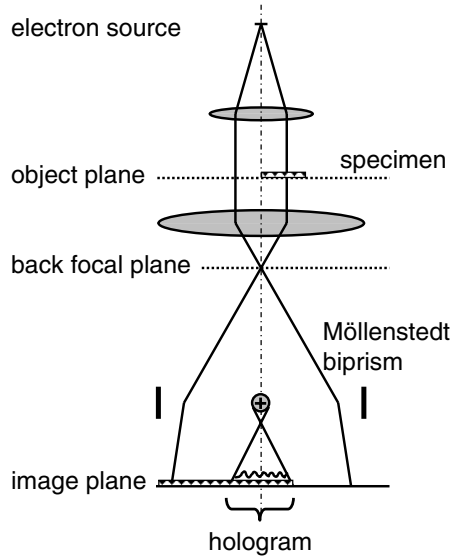


Figure 23. Setup for recording off-axis image plane holograms. The specimen on the right side of the optic axis and the adjacent reference wave on the left are imaged into the image plane. The biprism superimposes them in the image plane. The arising interference pattern is the hologram, which is magnified by the subsequent lenses according to the needs of recording by the CCD-camera.

for the different applications of electron holography, e.g. at medium or at highest resolution, are discussed in [43]. Elliptic illumination for increasing the coherent current density in the object is achieved by means of the condenser stigmator.

In any case, assume the image wave $\text{ima}(\vec{r}) = A(\vec{r}) \exp(i\phi(\vec{r}))$ transferred by the objective lens out to a maximum spatial frequency q_{lim} . Superimposed with a plane reference wave of modulus 1 at an angle β in the x -direction, it produces a hologram with the intensity distribution

$$I_{\text{hol}}(\vec{r}) = 1 + I_{\text{el}}(\vec{r}) + I_{\text{inel}}(\vec{r}) + 2C \cdot A(\vec{r}) \cos(2\pi q_c x + \phi(\vec{r})). \quad (4.1)$$

$q_c := k\beta$ is the carrier spatial frequency, and C again means the contrast of the hologram fringes. Here and in the following, all variables describing the interference such as q_c and β are referred back to the object plane by division with magnification of the objective lens.

$I_{\text{inel}}(\vec{r})$ represents all inelastic interaction larger than about 10^{-15} eV. It is interesting to remember that inelastically scattered electrons do not contribute to the cosinoidal interference term.

Properties of detector. As detectors, CCD-cameras are used. By means of a scintillator, the electron image is converted into a photon image, which is transferred via fiber optics to the CCD-chip. The obtained signal is corrected for variations in gain and dark-current. For adaptation to the scattering in the scintillator, the pixels on the CCD-chip are comparably large, e.g. $27 \mu\text{m}$. Therefore, a CCD-chip with 1024×1024 pixels measures about 30 mm. Due to imperfections of the fiber optics, the output image suffers from slight geometrical distortions, which fortunately can be corrected afterwards.

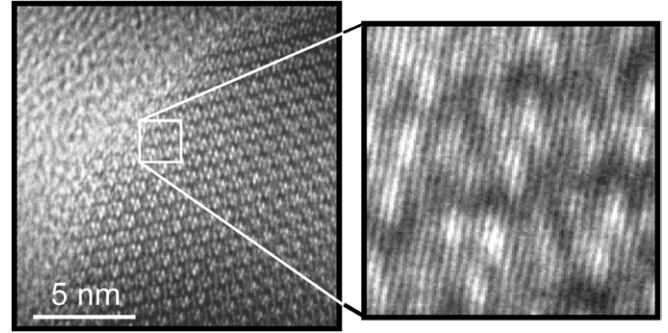


Figure 24. Electron hologram. In the highly magnified part, the interference fringes are discernible. Their modulation both in contrast and position gives rise to amplitude and phase modulations in the reconstructed image wave. Object: Si_3N_4 .

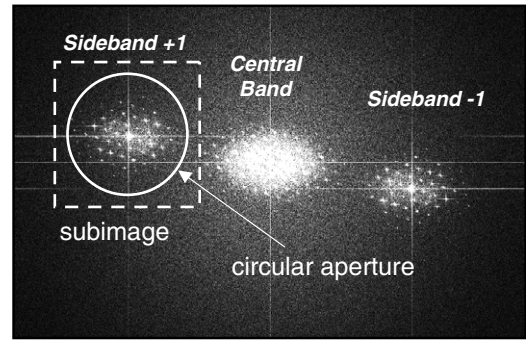


Figure 25. Reconstruction of the image wave. The Fourier-transform of the hologram exhibits three bands. The sideband +1, representing the Fourier-transform of the image wave, is masked out by a circular aperture and centered around the origin of Fourier space. The smaller sub-image is inverse Fourier-transformed into real space to reveal the image wave.

Figures of merit are the pixel number $n_{\text{pix}} = 1024$ and 2048 in a 1 k- and 2 k-camera, respectively. Furthermore, the camera is characterized by the modulation transfer function (MTF) out to Nyquist frequency sampled just by 2 pixels. Also important is the detection quantum efficiency $\text{DQE} = \frac{(\text{Signal/Noise})_{\text{out}}^2}{(\text{Signal/Noise})_{\text{in}}^2}$. DQE is a measure for the noise added by the camera to the noise of the incoming signal [44]. An example of a hologram is shown in figure 24.

Reconstruction of the complete electron wave from the hologram. The Fourier spectrum (figure 25) of the hologram

$$\text{spec}(\vec{q}) =$$

$$\begin{aligned} & \text{FT} [1 + I_{\text{el}}(\vec{r}) + I_{\text{inel}}(\vec{r})] && \text{centerband} \\ & + C \cdot \text{FT} [A(\vec{r}) \exp(i\phi(\vec{r}))] \otimes \delta(\vec{q} - \vec{q}_c) && + 1 \text{ sideband} \\ & + C \cdot \text{FT} [A(\vec{r}) \exp(-i\phi(\vec{r}))] \otimes \delta(\vec{q} + \vec{q}_c) && - 1 \text{ sideband} \end{aligned} \quad (4.2)$$

reveals 3 bands:

- the center band represents essentially the conventional image; it contains both elastically and inelastically scattered electrons and both linear and nonlinear terms. It does not contain the image phase $\phi(\vec{r})$ and hence is not of further interest here.

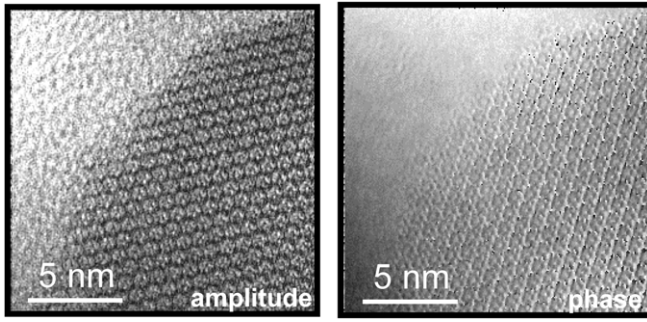


Figure 26. Reconstructed image wave. The reconstructed image wave shows a strong modulation both in amplitude and phase. Whereas a conventional image would essentially contain the amplitude only, here also the phase is accessible. This wave showing atomic details cannot yet be interpreted in terms of the object, because there is still a strong mixing between object amplitude and phase in the image wave. This will be healed by corrections of aberrations, as described in section 6.

- the ± 1 -sidebands contain the Fourier-spectrum of the complete image wave, damped by the factor C and convoluted around $\vec{q} = \pm \vec{q}_c$, respectively.

The two sidebands offer the big advantages that they only contain the elastically scattered electrons, and that amplitude and phase are linearly related to the object properties. The two sidebands are conjugate, i.e. equivalent and redundant in that already one sideband contains all image wave information, including quantum noise. Therefore, one sideband—usually the $+1$ -sideband—is masked out by an aperture, and centered around $\vec{q} = \vec{0}$ in an appropriate sub-image. Finally, by inverse Fourier-transformation, the reconstructed image wave

$$\text{ima}_{\text{rec}}(\vec{r}) = A_{\text{rec}}(\vec{r}) \exp(i\phi_{\text{rec}}(\vec{r})) \quad (4.3)$$

with

$$A_{\text{rec}}(\vec{r}) = C A(\vec{r})$$

is found in real space (figure 26).

Rectification by means of an empty reference hologram. Geometric distortions, e.g. from the electron lenses and the CCD-camera, as well as from charging effects and dirt particles at the biprism, produce bending of the hologram fringes and hence artificial phase structures, which may conceal the reconstructed phase image $\phi_{\text{rec}}(\vec{r})$. Additionally, mistilt of the reconstructed wave may be introduced by improper centering of the sideband. Therefore, for each hologram, an empty reference hologram is recorded at the same optical parameters, but without object. The reconstructed ‘empty wave’ is used for correction of artifacts and mistilt, and furthermore, for normalization of the amplitude $A_{\text{rec}}(\vec{r})$.

Visual inspection. The wave is displayed by amplitude and phase. Mostly the phase is the most interesting part. Since the phases are extracted from complex numbers by means of the arctan-function, they are displayed in the range $[0, 2\pi]$, mod 2π (‘phase wrapping’). Therefore, if at strong phase objects the phase dynamics exceeds 2π , phase jumps $0 \leftrightarrow 2\pi$ occur, which show up as equiphase lines at maximum contrast

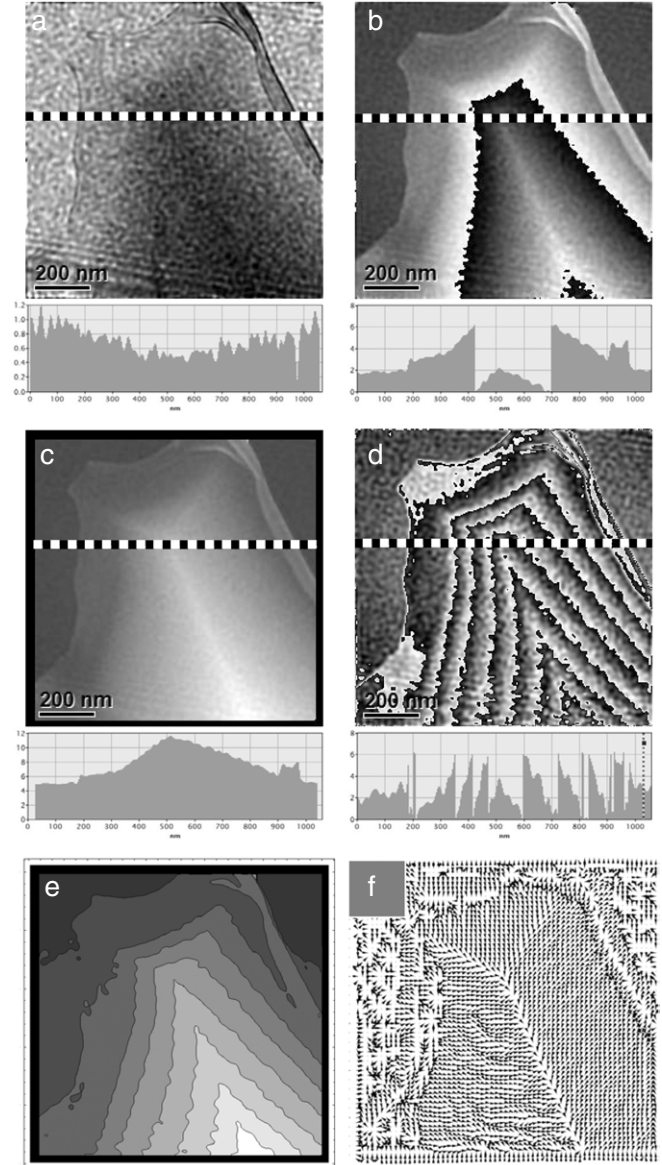


Figure 27. Evaluation of reconstructed wave from a PZT-crystal. (a) Amplitude and (b) phase as reconstructed. The gray scale increases from black to bright with increasing phase values. The lines with strong black/white contrast are equal-phase lines. They show the transition $0 \leftrightarrow 2\pi$ arising from ‘phase wrapping’ under evaluation of the arctan-function. As equiphase lines, they give an intuitive idea about the phase distribution. Note from linescans along dotted lines that signal/noise is much better in the phase than in the amplitude. Furthermore the phase can directly be interpreted as projected mean inner potential. Differently processed phase images based on the same data set: (c) unwrapped, (d) $4\times$ amplified, (e) contours and (f) arrow-plot of phase gradient. The four images display the same phase distribution of the etched PZT-crystal looking like a bird-eye view of a hipped roof. All phase images very clearly reveal the etching structure of the crystal. Assuming constant inner potential, the phase represents the thickness distribution.

(figure 27). These lines are very helpful in that they give a very instructive overview over the phase distribution and allow a first intuitive interpretation. If, on the contrary, the phase structure of interest is very weak, these equiphase lines can be selected at small fractions of 2π , to enable visual interpretation (‘phase amplification’). In any case, before a detailed quantitative

analysis, the phase jumps have to be removed by means of a phase-unwrapping procedure, which adds or subtracts 2π at the phase wrapping lines accordingly.

Quantitative wave optical analysis by numeric image processing. The reconstructed image wave represents the object exit wave, however, still distorted by the aberrations of the objective lens. Nevertheless, the big advantage at this stage of the comparably complicated holographic process over conventional imaging is the availability of both amplitude and phase. These are intrinsically quantitative data, and hence all delineated information is quantitative. The reconstructed image wave is the starting point of a complete wave optical analysis by numerical image processing for the extraction of the object data, which cannot be obtained by conventional TEM. Like in a ‘virtual microscope’, the wave can be defocused, it can be corrected from aberrations, and it can be analyzed quantitatively in all ways one can think about; for example, it can be displayed in bright-field mode as perfect amplitude or phase image, or in dark-field mode masking out or admitting any reflection in Fourier space. Furthermore, one can determine amplitude and phase of any reflection in Fourier space, determine the reflections contributing to any detail of the object (‘holographic nanodiffraction’), or the contribution of any reflection to amplitude and phase of the field of view (‘holographic reflection analysis’). Also in Fourier-space everything is quantitatively accessible by amplitude and phase. Examples are given in the section about high resolution holography.

4.4. Properties of the reconstructed image wave

The reliability of the object information determined by wave optical analysis depends on the degree of fidelity of the reconstructed wave with the ‘true’ image wave in the electron microscope. Ideally, $\text{ima}_{\text{rec}}(\vec{r}) \equiv \text{ima}(\vec{r})$ holds, i.e. $C = 1$, $A_{\text{rec}}(\vec{r}) \equiv A(\vec{r})$, and $\phi_{\text{rec}}(\vec{r}) \equiv \phi(\vec{r})$.

However, there are some discrepancies:

1. The contrast C of the hologram fringes damps the reconstructed wave, e.g. by the restricted spatial coherence, by instabilities, by the modulation transfer function of the CCD-camera, and by inelastic interaction, i.e. loss of coherent electrons.
2. The reconstructed image wave represents a time-averaged wave, which possibly never existed as such in the electron microscope. Averaging comes about from collecting a large number of electrons in the hologram, where each may have ‘seen’ a slightly different object. The coherence conditions make sure that fringes show up in the hologram; but the slight differences, e.g. due to an object changing the phase shift in time, remain. Therefore, with the time dependent amplitude $A(\vec{r}, t)$ and phase $\phi(\vec{r}, t)$, the reconstructed wave results as

$$\text{ima}_{\text{rec}}(\vec{r}) = \frac{1}{\tau} \int_0^\tau A(\vec{r}, t) \exp(i\phi(\vec{r}, t)) dt, \quad (4.4)$$

where τ again means exposure time of the hologram.

This averaged wave may never exist as such. To give an example, assume an object with a pure phase detail, which changes phase shift from $\phi = \phi_0(\vec{r})$ to $\phi = 0$ halfway during exposure time; for example, this could happen by an atom hopping away under electron irradiation. The reconstructed phase will be $\phi_{\text{rec}}(\vec{r}) = \phi_0(\vec{r})/2$, and the reconstructed amplitude $A_{\text{rec}}(\vec{r}) = \cos(\phi_0(\vec{r})/2)$.

3. The radius q_{rec} of the aperture, used to mask out the sideband in Fourier space, determines the resolution in analogy to Abbe theory of imaging; for the highest resolution, it should be as wide as possible. The maximum sensible radius $q_{\text{rec,max}}$, however, is given by the distance q_c of the sideband from the centerband. Since, in general, the radius of the centerband ($\text{FT}[1 + I_{\text{el}}(\vec{r}) + I_{\text{inel}}(\vec{r})]$) measures up to twice the one of the sideband ($\text{FT}[A \exp(i\phi)]$),

$$q_{\text{rec,max}} \leq q_c/3 \quad (4.5)$$

has to be met. For a weak pure phase object, i.e. $A = 1$ hence $\text{FT}[A^2] = \delta(\vec{q})$, this relaxes to

$$q_{\text{rec,max}} \leq q_c. \quad (4.6)$$

Usually, a value in between these two limits is applicable. In any case, the used part of the sideband must not overlap with the centerband, to avoid disturbance from information of the centerband. Therefore, if for any reason q_c is limited, the Fourier spectrum of the image wave has to be restricted such that the relations (4.5) and (4.6) hold; this must be done under recording the hologram by means of an appropriate aperture in the back focal plane of the objective lens. In any case, only the part not overlapping with the centerband may be used for reconstruction. On the other hand, it is disadvantageous to use larger carrier frequencies than q_c given above, because then the very narrow fringes are more sensitive against instabilities; furthermore, one has to spend correspondingly more CCD-pixels, to cover the same field of view.

4. The reconstructed image wave is sampled by a comparably small number n_{rec} of reconstructed pixels. n_{rec} can be derived for recording the hologram by means of a CCD-camera with n_{pix} camera pixels as follows: at least 4 pixels have to sample one hologram fringe [45], to avoid falsification of amplitude and phase. Since 3 fringes make up one reconstructed pixel, one finds $n_{\text{rec}} \leq n_{\text{pix}}/12$ for a general object, and $n_{\text{rec}} \leq n_{\text{pix}}/4$ for a pure weak phase object.
5. Signal/Noise-properties. The hologram is built up by a finite number of electrons and hence always shows quantum noise. Therefore, contrast and position of the hologram fringes are not arbitrarily sharply defined; the same is true for the derived amplitude and phase of the reconstructed wave. The error in phase may be estimated by means of the *phase detection limit*

$$\delta\varphi_{\text{lim}} = \frac{\sqrt{2} \text{snr}}{C\sqrt{N}} \quad (4.7)$$

for an intended signal/noise ratio (snr), where C is the fringe contrast of the hologram, and N the number of electrons collected per reconstructed pixel [45–47]. The phase detection limit is examined in more detail in the following section.

4.5. Noise properties

As mentioned above, the phase detection limit $\delta\varphi_{\text{lim}} = \frac{\sqrt{2} \text{snr}}{C\sqrt{N}}$ answers the question, *which is the smallest phase difference detectable at a given snr?* This is elaborated in the following. The contrast of the recorded hologram fringes

$$C = |\mu| C_{\text{inel}} C_{\text{inst}} C_{\text{MTF}} \quad (4.8)$$

is given by the degree of coherence μ of illumination in the object entrance plane, practically given by spatial coherence μ^{sc} ; longitudinal = temporal coherence need not be considered here, because, in usual holograms with less than 1000 fringes, it is always close to unity. C_{inel} is the loss of coherence due to inelastic interaction, C_{inst} the reduction of contrast stemming from instabilities of the object, the microscope and disturbances from the environment, e.g. AC-stray fields; C_{MTF} describes contrast damping by the modulation transfer function (MTF) of the CCD-camera.

The average number N of electrons collected in a reconstructed pixel can be elaborated in the following way: the total coherent current available at the degree of spatial coherence μ^{sc} is given as

$$I_{\text{coh}} = -\ln(|\mu^{\text{sc}}|) \frac{B}{k^2} \quad (4.29)$$

with brightness B and wave number k of illumination. With the hologram width w and the pixel size $1/(2q_{\text{max}})$ at resolution q_{max} , the number of electrons per reconstructed pixel, reduced by the detection quantum efficiency (DQE) of the CCD-camera, reads as

$$N = -\ln(|\mu^{\text{sc}}|) \frac{B}{ek^2} \frac{1}{(2q_{\text{max}}w)^2} \varepsilon \tau \text{ DQE}. \quad (4.9)$$

ε is ellipticity of illumination and τ exposure time.

Finally,

$$\delta\varphi_{\text{lim}} = \frac{\sqrt{2\pi} \text{snr } n_{\text{rec}}}{|\mu^{\text{sc}}| C_{\text{inel}} C_{\text{inst}} C_{\text{MTF}} \sqrt{-\ln(|\mu^{\text{sc}}|) \frac{B}{ek^2} \varepsilon \tau \text{ DQE}}} \quad (4.10)$$

results with the number $n_{\text{rec}} = 2q_{\text{max}}w$ of reconstructed pixels. In effect, the degree of spatial coherence controls the term

$$\frac{1}{|\mu^{\text{sc}}| \sqrt{-\ln(|\mu^{\text{sc}}|)}}, \quad (4.11)$$

which renders a minimal $\delta\varphi_{\text{lim}}$ for $|\mu^{\text{sc, opt}}| = 0.61$.

Sorting out the parameters given by the microscopic setup and by the disturbance level of the lab, one can define a figure of merit for the quality of a holographic system as

$$\text{Noise figure} = \frac{\sqrt{2\pi}}{|\mu^{\text{sc}}| C_{\text{inst}} C_{\text{MTF}} \sqrt{-\ln(|\mu^{\text{sc}}|) \frac{B}{ek^2} \varepsilon \tau \text{ DQE}}}, \quad (4.12)$$

which may depend on the fringe spacing resulting in different values for C_{inst} at the same disturbance level. This definition allows the abbreviated form

$$\delta\varphi_{\text{lim}} = \text{Noise figure snr} \frac{n_{\text{rec}}}{C_{\text{inel}}}, \quad (4.13)$$

which can be applied most easily once the Noise figure of a holographic system is determined. In a good laboratory, Noise figure $< 10^{-3}$ can be reached.

Amount of information in a hologram. A suitable measure for the total phase information volume ascertainable from a hologram can be defined as

$$\text{PIV} := \frac{n_{\text{rec}}}{\delta\varphi_{\text{lim}}/2\pi} = \frac{2\pi C_{\text{inel}}}{\text{Noise figure snr}}, \quad (4.14)$$

which means the number of reconstructed pixels multiplied by the number of phase values distinguishable in the phase range $[0, 2\pi]$ at a selected snr; evidently, PIV is a constant for a given microscopic setup used for recording any hologram. The only parameter depending on the object is the contrast reduction by inelastic interaction, which is related to object thickness d and mean free path for inelastic interaction λ_{inel} by means of $C_{\text{inel}} = \exp(-d/(2\lambda_{\text{inel}}))$ (equation (2.64)).

Evidently, at constant PIV, $\delta\varphi_{\text{lim}}$ can be adapted to the needs of measurement by means of the field of view w and resolution q_{max} . For a given hologram width, there is still the possibility of improving $\delta\varphi_{\text{lim}}$ by selecting a smaller reconstruction aperture, which increases the size of the reconstructed pixels and hence reduces n_{rec} , alas, at the cost of lateral resolution q_{max} [48]. Specific details are discussed in the following sections for medium and high resolution.

4.6. Projection problems

As in conventional electron microscopy, the object exit wave is a 2D-projection of a 3D-object. Therefore, interpretation of a phase image in terms of the object is in general not unique and difficult to perform. In the following, the problems are discussed in more detail for electric objects.

The electric phase shift $\varphi(x, y) = \sigma V_{\text{proj}}(x, y)$ (equation (2.43)) with $V_{\text{proj}}(x, y) = \int_{\text{object}} V(x, y, z) dz$ reveals only the projected structure of the object. Even in the simplest case with constant potential $V(x, y, z) = V_0$ and constant thickness $d = d_0$, the measured phase gives no more than the product $V_0 d_0$. To delineate the potential V_0 distinctively, one has to measure the thickness d_0 by another way. A very elegant way is to measure it from the amplitude, which is damped according to $\exp(-d/(2\lambda_{\text{inel}}))$, reconstructed from the same hologram [49]; however, this method depends on the accuracy to which the mean free path λ_{inel} is known. For crystalline specimens, convergent electron beam diffraction is an accurate method for specimen thickness determination; it is independent from the t/λ_{inel} -method and therefore a combination of both would be optimal.

In any case, the phase can only be interpreted in a simple way, if the potential $V(x, y, z)$ does not change in the z -direction, and if the object thickness is constant. To display

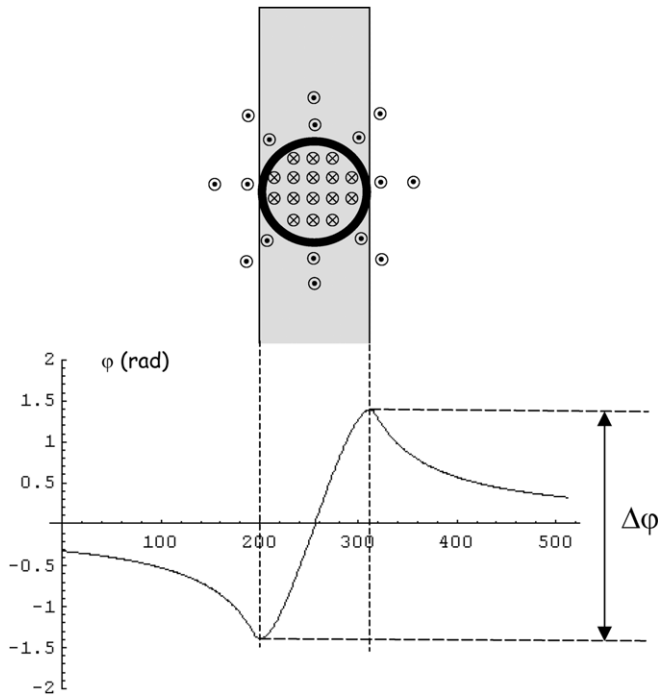


Figure 28. Role of stray fields around an object for the measured phase shift. Top: meridian section through a sphere (circle) homogeneously magnetized in the viewing direction produces a stray field in the opposite direction decaying with distance. Bottom: The phase of a traversing electron wave is given by the respective enclosed flux. Assuming the reference point on the far left, the enclosed flux first increases in the negative direction. At the surface of the sphere, it reaches a minimum. Then the positive flux in the sphere shifts the phase to a maximum at the right surface. With increasing distance, the phase approaches zero again. The phase difference $\Delta\phi$ is given by the flux inside the sphere reduced by the stray flux in the shaded area. Because of rotational symmetry, the measured phase difference is smaller by a factor $2/\pi$ compared with the sphere alone. Consequently, to determine the magnetization in the sphere, one has to evaluate $\frac{\pi}{2}\Delta\phi$. For a more general geometry, the stray fields present are much more complicated to analyze.

from a phase-image at least the potential distribution exactly, the thickness has to be correspondingly homogeneous. For example, if an accuracy of 1% is needed, the thickness variation also has to be smaller than 1%. This poses severe problems on the preparation methods of the objects [50].

Furthermore, electric and magnetic stray fields may expand around the specimen (figure 28). Above and below the specimen in the z -direction, they contribute to the measured phase values accordingly. In the case of simple geometries, e.g. spheres, cylinders, etc, this can be accounted for under evaluation of the phase image. The effect poses problems in particular at medium resolution, because the corresponding fields reach far out. At high resolution, the high spatial frequencies delivered by the object are damped very fast exponentially so that they do not contribute much.

Plane reference wave. Fields expanding in the (x, y) -direction sometimes produce problems in that they also affect the reference wave [51]. This means that the measured phase distribution is in fact the phase difference distribution of the

object wave with the reference wave also modulated by the object.

Summary: performance of electron holography. Despite the problems, to which solutions are still to be found, electron holography offers substantial advantages over conventional TEM imaging. First of all, the data representing the object structure are complete in that one obtains

- amplitude and phase cleanly separated
- quantitative data gauged by means of an empty reference hologram
- linear transfer of amplitude and phase from the object into the reconstructed wave
- pure elastic data due to perfect zero loss filtering ($<10^{-15}$ eV).

Moreover, comprehensive wave optical image processing is applicable for evaluation of all kinds of information encoded in the object exit wave, since the whole wave is available. This means

- *a posteriori* numerical correction of all coherent aberrations
- no crosstalk, i.e. amplitude and phase can be interpreted in terms of physics and materials properties separately
- no delocalization leading to misinterpretation
- evaluation of amplitude and phase up to the respective noise limit
- full wave-optical evaluation also in Fourier space, e.g. by single reflection analysis and nanodiffraction.

At the end, the performance limits are definitely given by the incoherent aberrations, i.e. information limit, as well as by the signal/noise limits determined by brightness of the electron source, and instabilities.

5. Electron holography at mesoscopic resolution

According to the phase contrast transfer function $PCTF = \sin \chi(\vec{q})$ discussed in section 3, a conventional TEM is virtually blind for large area phase objects: these objects contain mainly spatial frequencies much smaller than the lower limit $q_{\min} \approx q_{\text{Scherz}}/4$ of the Scherzer-band. For these and smaller ones, the wave aberration $\chi(\vec{q})$ is not effective and hence $PCTF \approx 0$ and $ACTF \approx 1$. This is true, in particular, at Gabor-focus $D_{z_{\text{gabor}}} = 0.56 D_{z_{\text{Scherz}}}$ introduced specially for holography at mesoscopic resolution [31]. Consequently, the image wave agrees with the object wave, and hence the holographically reconstructed wave can be interpreted directly, i.e. without correction of aberrations, in terms of the object. Meanwhile, holography turned out unrivalled for the analysis of electric and magnetic microfields.

5.1. Specific aspects at mesoscopic resolution

Field of view and resolution. For large area structures, e.g. magnetic or electric fields on the micrometre scale, one needs large fields of view w , i.e. wide holograms. In principle, every field of view can be realized, if the optics is set up accordingly; in a usual TEM, of course, the free parameters

for arranging lenses and biprisms according to the special needs are very restricted. In any case, constraints hold in that the hologram has to be grasped by the CCD-chip with given dimensions and pixel number n_{pix} . At the minimum sampling of 4 CCD-pixels per fringe, one obtains $n_{\text{rec}} \leq n_{\text{pix}}/12$ for a general object and $n_{\text{rec}} \leq n_{\text{pix}}/4$ for a pure phase object. Since q_{max} has to be sampled by ≥ 2 pixels, a resolution $q_{\text{max}} = (n_{\text{rec}}/2w)$ can be reached. For a usual CCD-camera with $n_{\text{pix}} = 1024$, one gets $n_{\text{rec}} \approx 90$ and $n_{\text{rec}} \approx 270$, respectively.

Examples for taking mesoscopic holograms. With the usual position of the biprism in the selected area aperture holder, the following values can be realized:

- $w \approx 10 \mu\text{m}$ and $q_{\text{max}} \approx 1/220 \text{ nm}^{-1}$ ($q_{\text{max}} \approx 1/74 \text{ nm}^{-1}$) with the objective lens nearly switched off ('low magnification mode') and the diffraction lens working as first imaging lens.
- $w \approx 1 \mu\text{m}$ and $q_{\text{max}} \approx 1/22 \text{ nm}^{-1}$ ($q_{\text{max}} \approx 1/7.4 \text{ nm}^{-1}$) with the objective lens switched off and the optional 'Lorentz-lens' working as first imaging lens. The Lorentz lens is a small einzel-lens positioned below the specimen such that the specimen is in space free from the magnetic lens field. It was introduced by Zweck and Bormans [52] for the investigation of magnetic specimen undisturbed from outer magnetic fields. The Lorentz-lens has a comparably long focal length of $f = 23 \text{ mm}$, and a spherical aberration given by $C_s \approx 8000 \text{ mm}$. The point resolution is about $q_{\text{scherrz}} = 1/2 \text{ nm}$ and the information limit $q_{\text{lim}} \approx 1/1.3 \text{ nm}$. Image details smaller than 10 nm cannot be interpreted without aberration correction because of delocalisation. The maximum magnification reachable with the Lorentz lens is about 83 000 at the CCD-camera.
- $5 \text{ nm} \leq w \leq 200 \text{ nm}$ with $0.1 \text{ nm}^{-1} \leq q_{\text{max}} \leq 1 \text{ nm}^{-1}$ under combined imaging by means of both objective lens and diffraction lens excited slightly different from the normal imaging mode.

5.2. Detection limits

The specific aspects of the noise figure.

$$\text{Noise figure} = \frac{\sqrt{2\pi}}{|\mu^{\text{sc}}| C_{\text{inst}} C_{\text{MTF}} \sqrt{-\ln(|\mu^{\text{sc}}|) \frac{B}{k^2 e} \varepsilon \tau \cdot \text{DQE}}} \quad (5.1)$$

are considered in the following. At medium resolution, contrast reduction C_{inst} by instabilities is not very critical, since fringe spacing is large and comparably insensitive. This allows the exposure time to be increased to, e.g., $\tau \approx 10$ – 20 s . Mainly, therefore, the phase detection limit is smaller (better) at medium resolution than for high resolution. In the Triebenberg Laboratory, values of Noise figure $\approx 3 \times 10^{-4}$ can be reached.

More consideration than for instabilities is needed for contrast reduction C_{inel} by inelastic interaction to optimize

signal/noise. To obtain a strong phase shift according to

$$\varphi(x, y) = 2\pi \frac{e}{h\nu} V_{\text{proj}}(x, y) - 2\pi \frac{e}{h} \Phi_{\text{mag}}(x, y), \quad (5.2)$$

the object should be as thick as possible; this, however, only makes sense if the object is homogeneous in the projection direction. In any case, with thickness also the probability for inelastic interaction increases and hence damps the fringe contrast. Counterbalancing these two effects gives an optimum object thickness. To be more explicit, assume an object with an electric inner potential IP homogeneously distributed in the z -direction over thickness d . Then the phase shift reads as

$$\varphi(x, y) = 2\pi \frac{e}{h\nu} \text{IP}(x, y) \cdot d := \varphi_0(x, y) d. \quad (5.3)$$

With the inelastic contrast reduction

$$C_{\text{inel}} = \exp(-d/(2\lambda_{\text{inel}})) \quad (5.4)$$

one obtains for the signal/noise ratio

$$\frac{\varphi(x, y)}{\delta\varphi_{\text{lim}}} = \frac{\varphi_0 d C_{\text{inel}}}{\text{Noise figure snr } n_{\text{rec}}}. \quad (5.5)$$

Consequently, one has to maximize the expression

$$d C_{\text{inel}} = d \exp(-d/(2\lambda_{\text{inel}})) \quad (5.6)$$

with respect to d , revealing

$$d_{\text{opt}} = 2\lambda_{\text{inel}} \quad (5.7)$$

as the optimum thickness with $C_{\text{inel}} = e^{-1}$, which would give the best signal/noise properties in the reconstructed phase image. For example, for silicon $d_{\text{opt}} \approx 240 \text{ nm}$ at $U_a = 200 \text{ kV}$.

Optimum phase detection for a given problem. For a given electron microscope, the reachable phase detection limit

$$\delta\varphi_{\text{lim}} = \text{Noise figure} \cdot \text{snr} \frac{n_{\text{rec}}}{C_{\text{inel}}} \quad (5.8)$$

is quantitatively evaluated in the following.

In the Triebenberg Laboratory, which is specially constructed and operated to achieve the minimum possible level of disturbances such as from ac-stray fields, mechanical vibrations, acoustic noise and thermal instability, nearly the best possible ambient conditions for electron holography are realized. Using our Philips 200 keV electron microscope CM200/FEG-ST for medium resolution holography, we reach Noise figure $\approx 3.3 \times 10^{-4}$. For a specimen of optimum thickness, i.e. $C_{\text{inel}} = e^{-1}$, and for an intended signal/noise ratio $\text{snr} = 3$, the phase detection limit is related to the number of reconstructed pixels by $\delta\varphi_{\text{lim}} = 2.7 \times 10^{-3} n_{\text{rec}}$ (figure 29).

Aiming at measuring voltages, for example in semiconductors, the minimum detectable voltage is given by

$$\delta V_{\text{lim}} = \frac{\delta\varphi_{\text{lim}}}{\sigma d_{\text{opt}}} \quad (5.9)$$

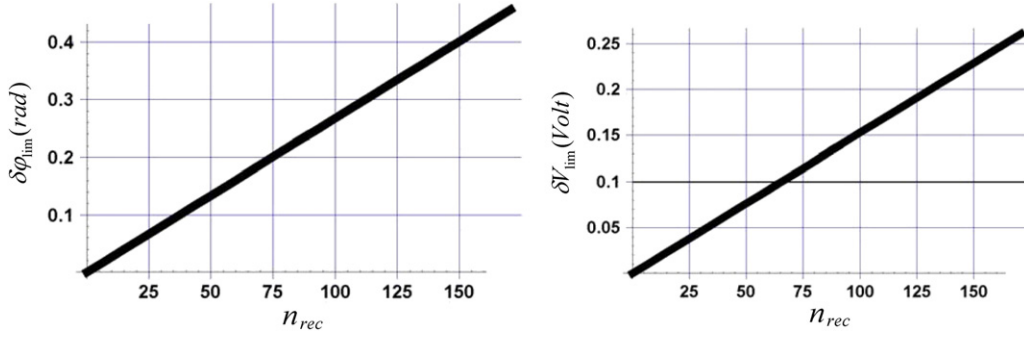


Figure 29. Detection limits for electric potentials at medium resolution. Left: phase detection limit versus number of reconstructed pixels given by $\delta\varphi_{\text{lim}} = 2.7 \times 10^{-3} n_{\text{rec}}$. The more details in space, the less details in signal are detectable. Right: voltage detection limit according to $\delta V_{\text{lim}} = \frac{\delta\varphi_{\text{lim}}}{\sigma d_{\text{opt}}}$ at optimum thickness. To reach $\delta V_{\text{lim}} = 0.1$ V, $n_{\text{rec}} < 65$ is needed.

with σ the interaction coefficient. In silicon with optimum thickness $d_{\text{opt}} = 240$ nm, the relation $\delta V_{\text{lim}} = 1.54 \times 10^{-3} n_{\text{rec}}$ follows. Consequently, to reach a voltage resolution of $\delta V_{\text{lim}} = 0.1$ V, the number of reconstructed pixels must meet $n_{\text{rec}} < 65$. If, for example, the needed lateral resolution is $q_{\text{max}} = 1 \text{ nm}^{-1}$, the width of the hologram must not be larger than $w = n_{\text{rec}}/(2q_{\text{max}}) \approx 32$ nm. This example shows that in order to reach satisfactory results, all relevant parameters have to be taken into account right from the beginning of the holographic procedure.

Aiming at measuring magnetization, one finds the magnetic flux

$$\Phi_{\text{uc}} = \mu_0 M a^2 \quad (5.10)$$

in one unit cell (cubic, lattice constant a) of ferromagnetic material with magnetization M , which is the dipole density; μ_0 is the induction constant. The magnetization can be estimated by means of

$$M = n_{\text{atom}} n_{\text{B}} \frac{\mu_{\text{B}}}{a^3} \quad (5.11)$$

with n_{atom} the number of atoms per unit cell, and n_{B} the number of Bohr magnetons $\mu_{\text{B}} = \frac{eh}{4\pi m_0}$ per atom (m_0 rest mass of electron). Therefore, at orthogonal orientation of M to the electron wave, the phase shift produced by a single unit cell is

$$\varphi_{\text{uc}} = n_{\text{atom}} n_{\text{B}} \mu_0 \frac{e^2}{2m_0 a}. \quad (5.12)$$

For example, inserting $a = 0.4$ nm, $n_{\text{atom}} = 1$ and $n_{\text{B}} = 1$, the magnetic phase shift of a single Bohr magneton follows as $\varphi_{\text{B}} = 4.4 \times 10^{-5}$ rad. Consequently, to obtain a phase shift of $2\pi/50$, the cross-section of a cluster of cobalt ($n_{\text{atom}} = 4$, $n_{\text{B}} = 1.7$) must contain 420 unit cells, i.e. it must have a diameter of about 8 nm. These numbers are in close agreement with experimental results, such as the ones found in figure 46. Smaller magnetic details can hardly be resolved (figure 30).

5.3. Applications for the analysis of electric and magnetic fields

Principal considerations. The object exit wave is modulated both in amplitude and phase. Since for modulation there is a manifold of origins, the interpretation of the findings may be difficult.

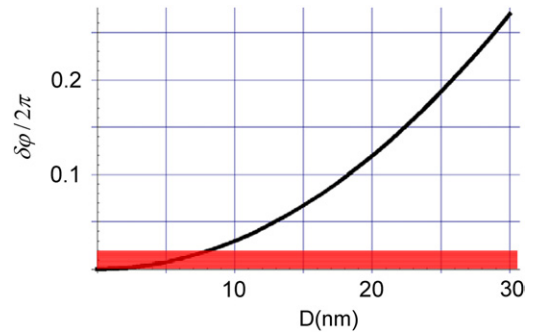


Figure 30. Detection limits for magnetic fields. Phase shift of a cobalt particle of quadratic shape in cross-section versus width D . The hatched area indicates phase shifts smaller than $2\pi/50$, here assumed as detection limit. Evidently, lateral resolution is not determined by the microscope but by the weakness of signals from magnetic specimen. Magnetic details below 8 nm are buried in noise.

Amplitude modulation. Occurs mainly by

- inelastic interaction as $A(x, y) = 1 - T(x, y)$ given by $T(x, y) = \text{Exp}[-t(x, y)/\lambda_{\text{inel}}]$ with λ_{inel} the mean free path for inelastic interaction, and $t(x, y)$ the thickness distribution of the object along the direction of the electron beam.
- scattering into angles larger than the acceptance angle of the optics ('scattering absorption contrast'). This should be minimized by avoiding any tight objective apertures when taking or reconstructing the hologram, because the wave as a whole would be altered both in resolution and signal strength.
- interference effects in a crystal, such as dynamic effects like extinction or bending contours. These can be minimized by tilting away from crystallographic zone axes.
- Fresnel diffraction effects found in the exit face of thick objects.

Phase modulation. The phase modulation is given by the electric and magnetic fields following equation (2.37)

$$\varphi = 2\pi \frac{e}{h\nu} \int V_{\text{obj}}(x, y, z) \, ds - 2\pi \frac{e}{h} \int \vec{A}(x, y, z) \, d\vec{s}.$$

Since the integrals cover the whole optical path from the source to the detector, the phase contains the potential inside and outside the object, such as stray fields. In general, both the electric and the magnetic fields originate in the specimen and, due to the 3D-Laplace equation, may reach far out into the surrounding of the object. Therefore, the integrals measure the whole field distribution provoked by the object, including the stray fields. This has to be taken into account, if one is interested in determining only the intrinsic field inside the specimen. It is particularly a problem at medium resolution, where the comparably large objects produce far reaching stray fields. In any case, the phase image provides only the projected potential $V_{\text{proj}}(x, y) = \int V_{\text{obj}}(x, y, z) dz$ and enclosed magnetic flux $\Phi(x, y) = \oint \vec{A}(x, y, z) dz$, respectively. Therefore, from the phase image, only the field data averaged in the direction of the electron beam can be extracted quantitatively by division with the thickness of the object; the thickness distribution $t(x, y)$ must be determined separately. This means that, for complicated objects, additional thickness measurement is needed, which sometimes may be gained from the inelastic amplitude component [49]. For a complete 3D-determination of the underlying fields, holographic tomography is the solution, which is under development [53, 54].

Stray fields may also leak far sidewise and hence may also influence the reference wave such that it is no more a plane wave [55]. Since the reconstructed phase image always represents the phase difference between object wave and reference wave, this has to be taken into account for a detailed analysis of the object wave. Stray fields do not present a problem with the electric mean inner potential inside solids stemming from averaging over the atomic potentials, because the charges compensate on an atomic level and hence decay at an atomic distance from the object.

Nevertheless, the strong advantage of electron holography prevails in that large area phase contrast—which is completely suppressed under conventional TEM-imaging—allows access to indispensable structure components such as electric and magnetic fields.

Electric phase modulation. The electric phase shift is given by

$$\varphi_{\text{el}}(x, y) = \sigma \int V_{\text{obj}}(x, y, z) dz$$

with the interaction constant σ (see section 2.3).

Possible contributions to electric phase shift are, for example,

- mean inner potential
- contact potentials between different materials
- functional potentials, e.g. from dopants in semiconductors
- ferroelectric polarization
- potentials from charging under the beam
- adsorbed ions.

Additionally, dynamical phase shifts, e.g. at extinction and bending contour lines occur. This may hamper the interpretation in terms of the wanted information and hence should be avoided by tilting out of zone axes.

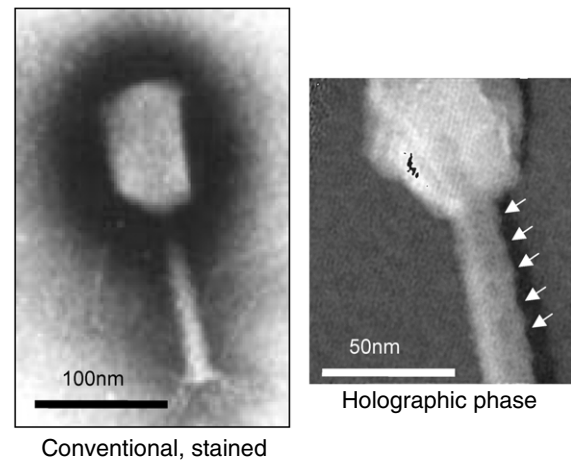


Figure 31. T5-bacteriophage. Viruses are pure phase objects. They are only visible under conventional imaging, if the structure is decorated by staining or defocusing (left). Staining blurs the details and gives rise to misinterpretation by selective adsorption of the stain. Defocusing smears out the fine details. Right: phase image reconstructed from a hologram taken in-focus of an unstained, freeze-dried phase. Details are clearly visible, even windings of the helix in the tail are suggested (arrows). From [56].

Examples: electric microfields. In the following examples, typical applications of electron holography at medium resolution are given.

1. Phase contrast due to inner potentials in materials. In the life sciences, most objects are phase objects. Therefore, e.g. viruses have to be stained with salt of heavy metals, to translate their structure into an amplitude contrast. The example shows unstained T5-bacteriophages in the phase image reconstructed from a hologram, compared with a stained one at conventional imaging (figure 31).

Likewise, soft matter materials in organic chemistry are imaged in conventional TEM either by the help of staining, or at a strong underfocus to achieve some phase contrast. Alas, both methods bear a strong risk for artifacts in that staining may be misleading by selective adsorption of the stain, and defocusing damages resolution and interpretation of finer details. Therefore, holograms of unstained specimen recorded in focus, give completely new insight into the real structure of these materials, in particular since large area phase contrast is reality now (figure 32).

2. Variation of the mean inner potential in different materials. The mean inner potential of a material varies according to the concentration of constituents. Therefore, the corresponding difference of phase shift reflects the difference of concentration, as shown in the example of $\text{Si}_x\text{Ge}_{1-x}$ in figure 33.

3. Electric fields in biomineralization. At growth processes, the question arises as to *what controls the growth*, for example, the growth of apatite into the gelatine fiber structure resulting, e.g., in a bone. Investigating such intergrowth nuclei at different stages by electron holography, surprisingly in the phase image stray fields of electric fields show up, which strongly suggest that the gelatine fibers carry electric dipoles (figure 34). These findings support the

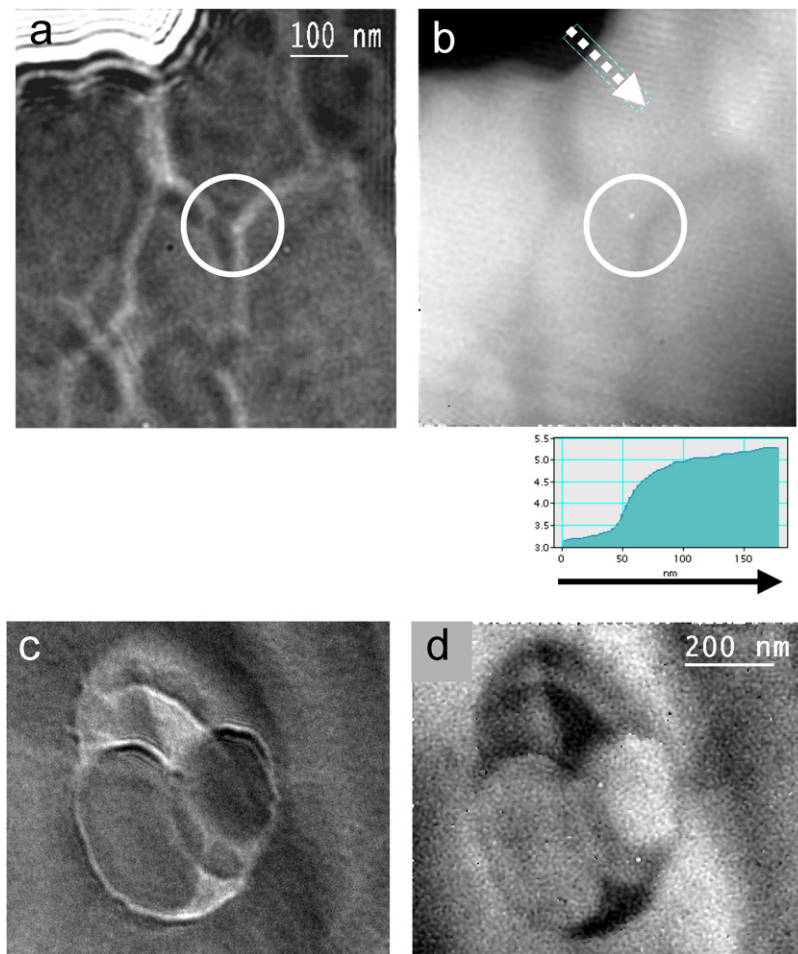


Figure 32. High impact polystyrene (HIPS). Top row: (a) Lorentz micrograph. The image is strongly blurred by Fresnel fringes, e.g. at the edge, by the needed large defocus. The grey areas cannot be interpreted; the brighter lines are Fresnel diffraction phenomena at larger phase gradients of the object. Large area information is missing. (b) Phase image of the same area. The dark region at the top left corner is vacuum. The structure of the bright polystyrene (PS) inclusions separated by narrow darker regions of polybutadiene clearly shows thickness variations. The phase shift due to PS amounts to 1.9 rad (inset at the right bottom) plus 2π , which results in a thickness of 137 nm assuming an inner potential of 8.2 V. Bottom row: (c) Lorentz electron micrograph and (d) electron phase image of high impact polystyrene (HIPS) at the same sample area. Again, in the phase image one discerns thickness variations between and inside the inclusions, which cannot be inferred from (d). From [57].

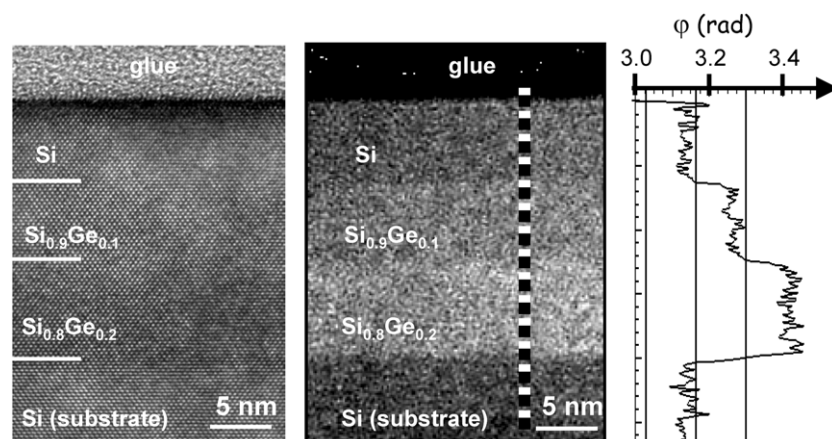


Figure 33. Holographic materials analysis in nanometre dimensions. The layers of $\text{Si}_{1-x}\text{Ge}_x$ with different Ge-content x cannot be distinguished in the conventional micrograph (left). In the phase image reconstructed from the zero beam of a hologram, the mean inner potential changes with x and hence produces a phase contrast, which allows the layers to be identified (middle). From the linescan along the dotted line (right), the mean inner potential hence Ge-concentration can be traced quantitatively. The advantage over other methods such as SIMS is the much better local resolution. From [59].

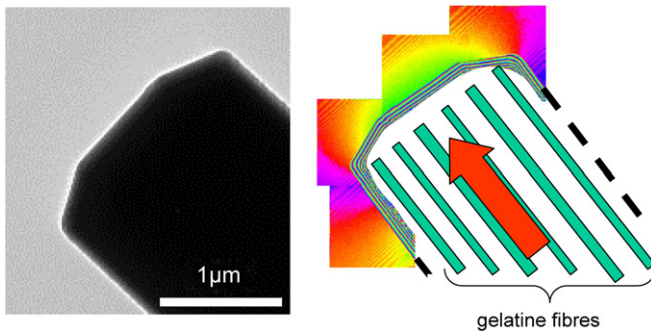


Figure 34. Intergrrowth nucleus of apatite in gelatine fibers. In the conventional image (left), only the shape of the thick tip of the nucleus (about $20\ \mu\text{m}$ long) can be seen. In the phase image (right), the pseudo-coloured equi-phasal lines show the equipotentials of an electric field emerging from the tip. This electric field, never seen before, presumably stems from the gelatine fibers carrying electric dipoles. It suggests the hypothesis that the growth of apatite into the fiber bundles is controlled by the electric fields. Because the field of view of a single hologram is too small to cover the whole tip, the phase image is patch-worked from 4 holograms. From [60].

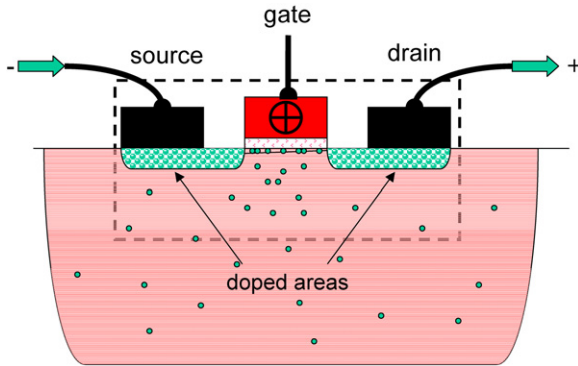


Figure 35. Scheme of a field-effect transistor (FET). Current between source and drain is controlled by gate voltage forming a conducting inversion channel beneath the gate contact. The connection from the metallic or poly-Si contacts (black) to the gate is established by means of doped areas, which have to reach exactly to the gate. For optimal function of the FET, precise shape and doping distribution are needed. These can be measured by electron holography.

hypothesis that intrinsic electric fields control the process of biomineralization [60].

4. Dopant profiling and functional potentials in semiconductors (figures 35–37). In principle, dopants in semiconductors change the scattering properties of the undoped material and hence produce a scattering absorption contrast. However, at the usually extremely low doping levels, this effect is not recognizable. Fortunately, the intrinsic field distribution provoked by the dopants gives rise to a sufficiently strong phase shift. This was shown by Pozzi *et al* [61], and brought to application by Rau *et al* [62]. To make holography a comprehensive and fully quantitative method, problems of sample preparation and interpretation are under investigation [63–68]. Furthermore, *in situ* experiments applying voltages and tomography for a full 3D-analysis are under development [69]. Meanwhile, in face of the achieved degree of integration

in functional components, this method is the only one able to measure the small details of field distributions needed for further improvements of semiconductor technology.

5. Electric polarization in ferroelectrics. Ferroelectric phenomena were observed as early as 1962 [70, 71] in a conventional TEM. However, only domain walls give rise to a contrast, whereas the polarization, which is a phase structure, remains invisible. Therefore, first experiments were conducted to reveal the ferroelectric polarization [72], however the findings and interpretation remain unclear, until [73, 74]. Nowadays, the polarization can be measured by means of holographic phase images (figure 38); even the atomic dipoles in ferroelectrics can be revealed, as shown in section 6.

6. Stray fields around charged nanostructures (figures 39 and 40). Electric nanostructures artificially prepared, e.g., as electrodes or field emitters, can advantageously be investigated by means of holography. With due care, the fields can be measured quantitatively [75, 76].

Magnetic phase modulation. The magnetic phase shift is given by

$$\varphi_{\text{mag}}(x, y) = -2\pi \frac{e}{h} \Phi(x, y)$$

with $\Phi(x, y)$ the total flux enclosed by the electron waves from the source to the detector, i.e. including stray fields along the electron paths. A detailed analysis shows that

$$\nabla \varphi_{\text{mag}}(x, y) \cdot \vec{B}_{\text{proj}}(x, y) = 0$$

holds with

$$\vec{B}_{\text{proj}}(x, y) = \int_{-\infty}^{+\infty} \vec{B}(x, y, z) dz$$

the projected magnetic field. This means that $\nabla \varphi_{\text{mag}}(x, y)$, $\vec{B}_{\text{proj}}$ and the electron trajectories form a righthanded orthogonal system, and the phase contour lines agree with the $\vec{B}_{\text{proj}}$ -lines.

Examples: magnetic microfields. 1. Magnetization in thin films. Magnetic films or thin etched layers are interesting both for basic science, such as correlation of crystallographic and magnetic structures in magnetic shape memory alloys, and technology, e.g. for memory devices. Usually, they can be chosen sufficiently thick to provide a strong phase signal. The phase-wrapping lines often give an excellent impression of magnetization distribution and domain structure (figure 41).

2. Magnetization distribution in small particles. Depending on size and shape, particles are magnetized in single-domain or multi-domains. Single-domain particles always show a strong stray field, whereas at multi-domains, stray fields are weak or close to zero, if the domains form a closed loop (figures 42–45).

3. Magnetic field by assemblies of Co particles (figure 46). Here the field coupling between the particles is of highest interest.

4. Remanence behaviour of arrays of magnetic dots. Macroscopically, measurement of hysteresis gives only an averaged information of a large ensemble. Holography gives

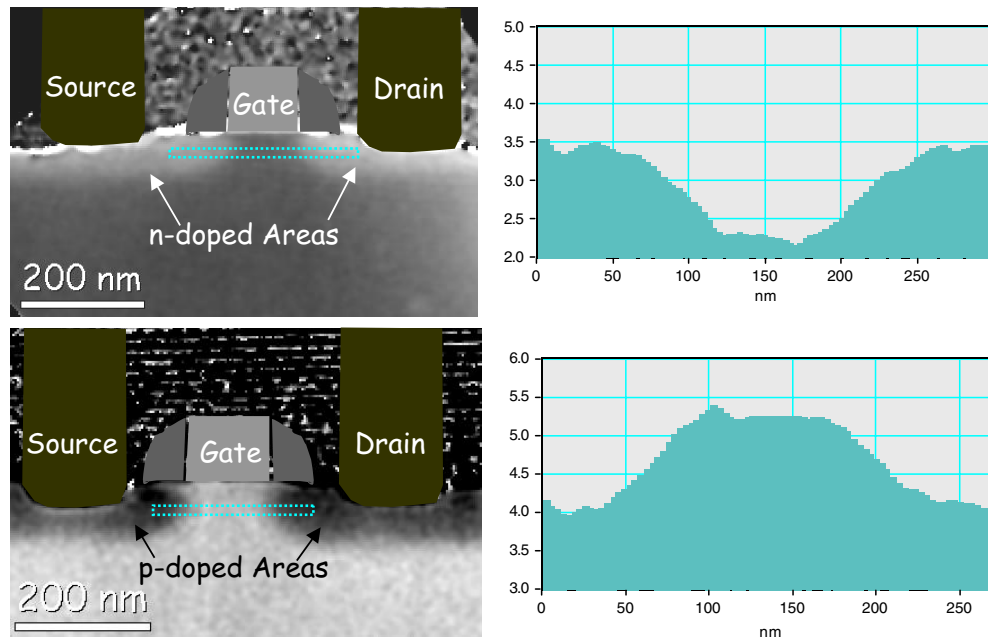


Figure 36. MOSFET. Whereas in the amplitude—as in conventional TEM-intensity—only the metallic contacts can be seen, the phase reveals also the doped areas. The phase values in the linescans give the projected potential of the pn-junction formed between the n-doped area and the p-doped substrate. From [65]. (Cooperation with Dr Hans-Jürgen Engelmann, AMD Dresden and Dr Uwe Mühle, Qimonda Dresden.)

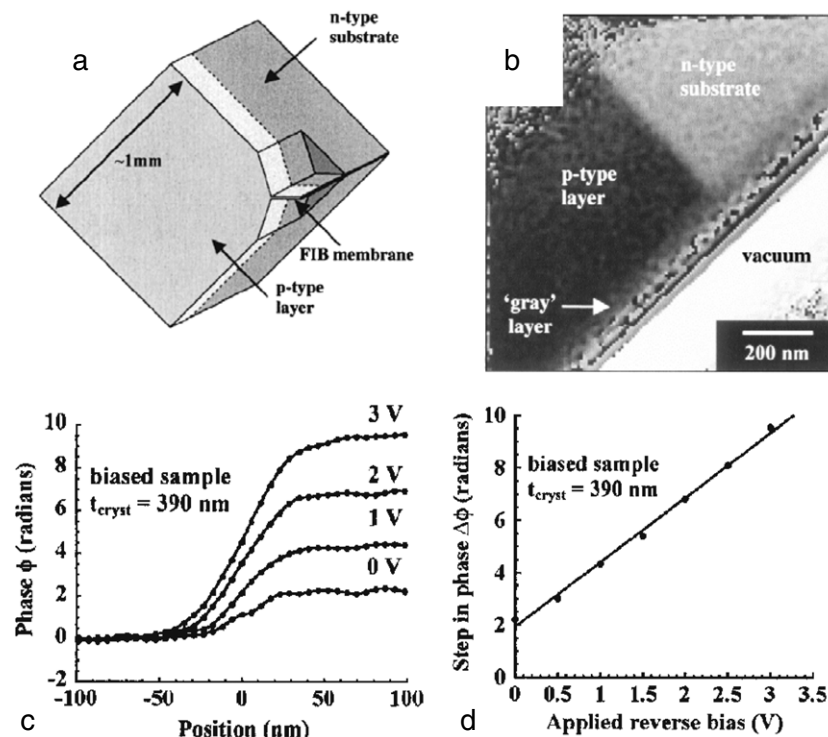


Figure 37. *In situ* investigation of pn-junctions (a) Specimen geometry used for applying external voltages to p–n junctions *in situ*. Focused ion beam milling has been used to machine a membrane of uniform thickness that contains a p–n junction at one corner of a 90° cleaved wedge. (b) Reconstructed phase image acquired from an unbiased Si sample containing a p–n junction. (c) Phase shift measured across a p–n junction as a function of reverse bias for 390 nm crystalline thickness (thickness measured using convergent beam electron diffraction). (d) Height of measured voltage step across junction as a function of reverse bias. From [69].

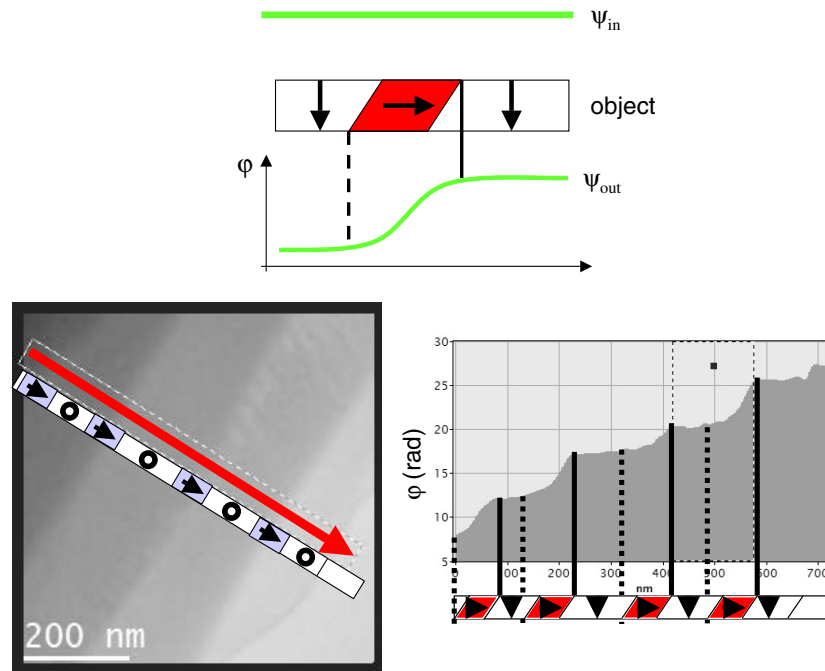


Figure 38. Ferroelectric domains in BaTiO₃. Top: only the in-plane component of the ferroelectric polarization produces a phase shift. Bottom: left: phase image of ferroelectric domains. The bands reveal in-plane polarization oriented as indicated, whereas the flat center of each band carries an out-of-plane polarization ('o'). The typical shoulders found in the linescan (right) suggest the domain structure in the cross-section shown at the bottom of the linescan.

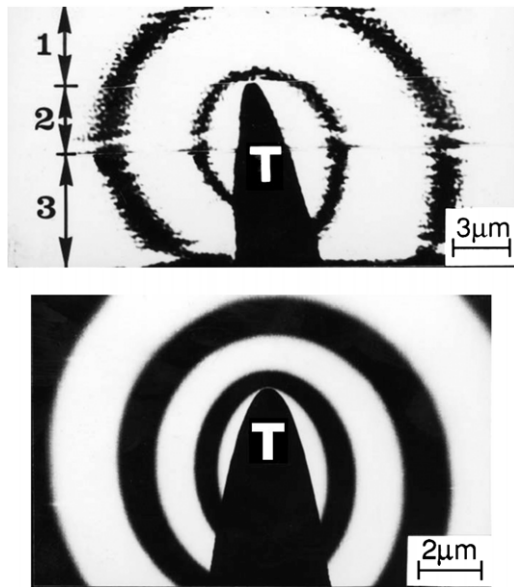


Figure 39. Electric field around biased tungsten tip. The phase image shows contour lines, which do not agree with the electric field, because the reference wave is also modulated, i.e. by the stray field. The experimental result (top) is modeled in good agreement in bottom. Tip voltage 7.5 V. Reprinted from [75], with permission from Elsevier.

deep insight into the details of the magnetic behaviour under and after exposing to external fields, in dependence on particle size, arrangement, etc (figure 47).

5. Embedded Co/Sm-particles. In the case of Co/Sm-particles by ion-implantation it is shown that the arising particles of diameter of 30 nm are magnetic. The findings

suggest that the arising magnetic dipoles are oriented randomly (figure 48).

6. Magnetic flux distribution in superconductors. The distribution of magnetic flux lines under an external magnetic field is most interesting for the understanding of superconductors (figures 49 and 50).

6. Electron holography at atomic resolution

At atomic resolution, the principles of holography elaborated so far also apply. However, they have to be specified according to the special requirements showing up at the intended resolution of 0.1 nm; these arise from the facts that much more complicated details of interaction of the electrons with the object have to be considered, and that the aberrations dominate the imaging process.

6.1. Electron—specimen interaction at atomic dimensions

In transmission electron microscopy, there is an extremely strong interaction of the electron wave with the Coulomb potential of the atoms. On the one hand, it is a blessing, because a strong signal is a prerequisite for measuring very tiny object properties at atomic resolution. On the other hand, it is a curse, because the electron is multiply scattered and hence the scattering event loses its uniqueness. As a result, the object exit-wave represents a complicated two-dimensional summary of all scattering events along the electron beam. This can only be understood by numerical simulation, for example by means of the EMS-package [85].

As long as the specimen has a thickness of only a few monolayers, the approximation of single scattering holds,

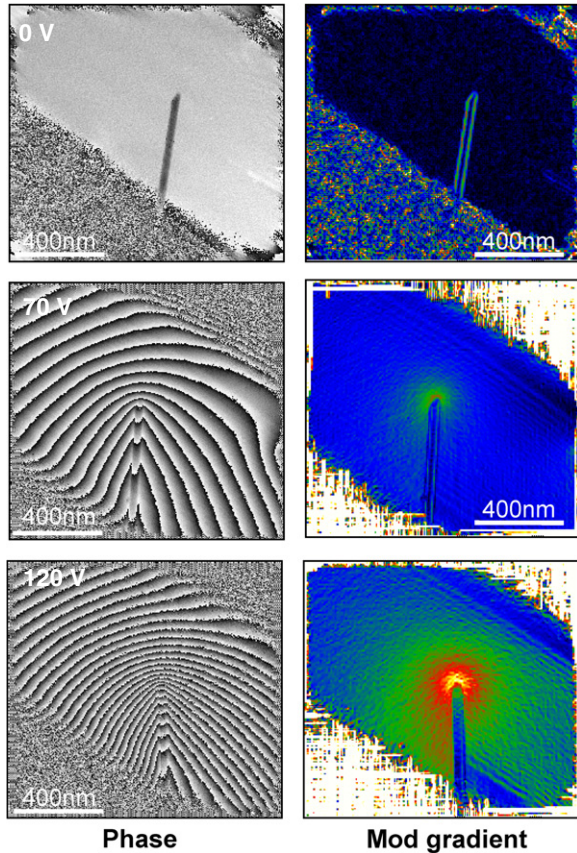


Figure 40. Electric field around biased carbon nanotube. For this study of field emission from carbon nanotubes, the phase lines represent equipotentials, since the reference wave can be assumed comparably flat. The field rapidly decays with distance from tiny tip, and the reference wave is $2 \mu\text{m}$ away. From [76].

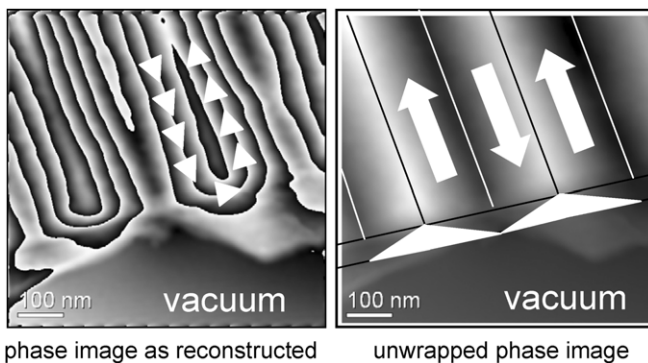


Figure 41. Magnetic phase shift in ferromagnetic domains in cobalt thin film. Left: the phase image as reconstructed shows phase-wrapping lines. Between successive lines the enclosed magnetic flux changes by h/e . Therefore, assuming constant thickness and homogeneous magnetization values in the specimen, and disregarding stray fields above and below the specimen, the phase-wrapping lines can be interpreted as the lines of the intrinsic B -field (arrows). The transition black $\Rightarrow$ white and white $\Rightarrow$ black gives the direction of the phase-gradient, which forms a right-handed system with the B -field and the direction of the electron beam. Right: unwrapped phase image with schematic phase profile suggesting the indicated 180° -domain structure. (Cooperation: W Neumann and R Otto, HU Berlin.)

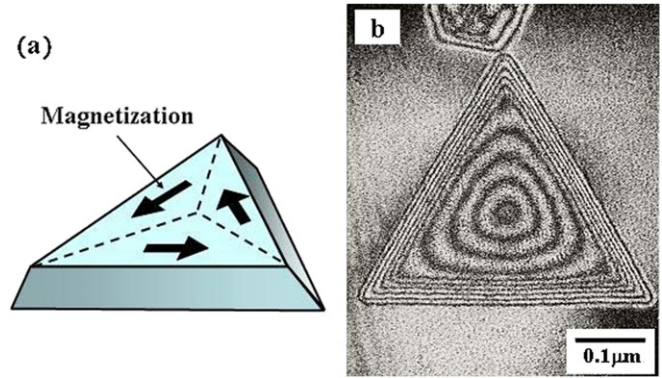


Figure 42. Magnetization in a cobalt smoke particle. (a) Schematic distribution of the magnetization forming a closed loop; (b) Phase image. Two kinds of phase contours appear: narrow fringes parallel to the particle edges indicate thickness contours, and circular fringes in the inner region indicate in-plane magnetic lines of force in $h/2e$ flux units due to the twofold phase amplification. From [77].

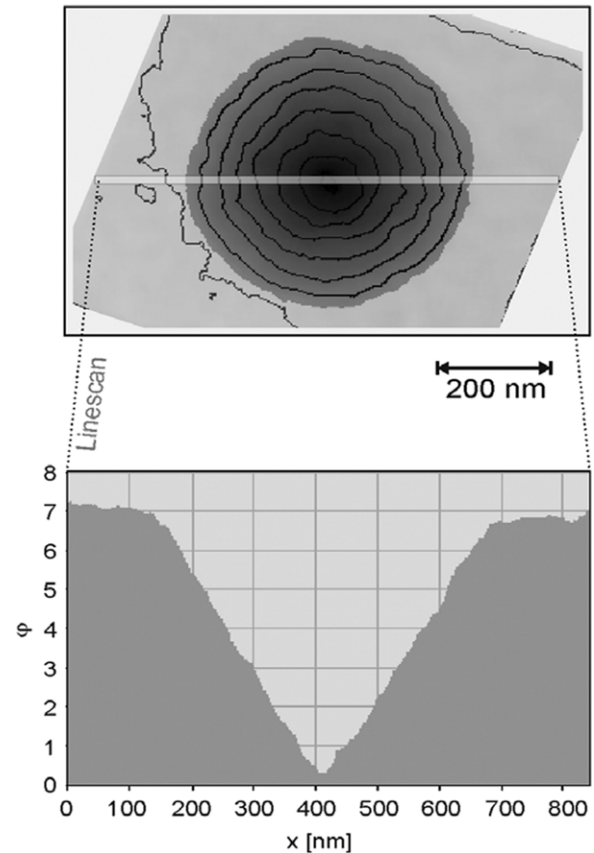


Figure 43. Disc of $\text{Ni}_{80}\text{Fe}_{20}$. Since the circular disc is evenly 20 nm thick, the circular contour lines and the linescan represent the projected B -field lines of the magnetization. The flux is completely closed. From [78].

which is the basis for kinematical theory. Then, amplitude and phase of each reflection in Fourier space depend linearly on the specimen thickness. This can be displayed in so-called beamplots as shown in figure 51. Additionally, it is observed that kinematically forbidden reflections, as, e.g., found in Fourier spectra of centrosymmetric crystals, are not excited. All the waves, emanating from reflections in

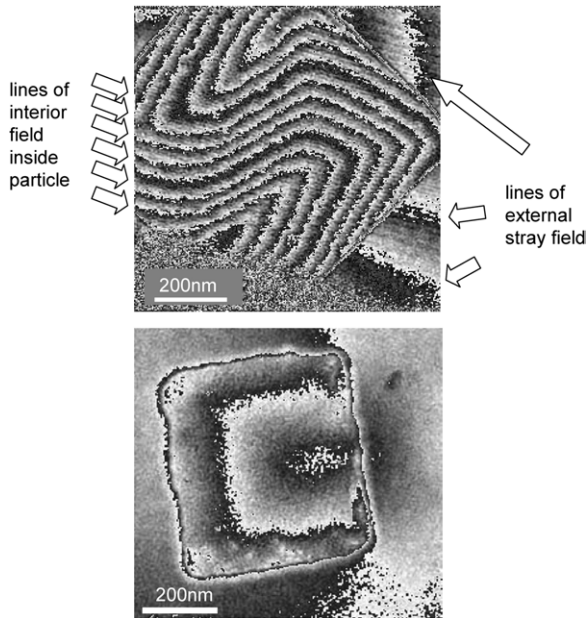


Figure 44. Magnetization in permalloy particles. Magnetization arises in different shapes. Top: S-shape phase amplification, bottom: C-shape. Cooperation of Josef Zweck, Regensburg. For details see [78].

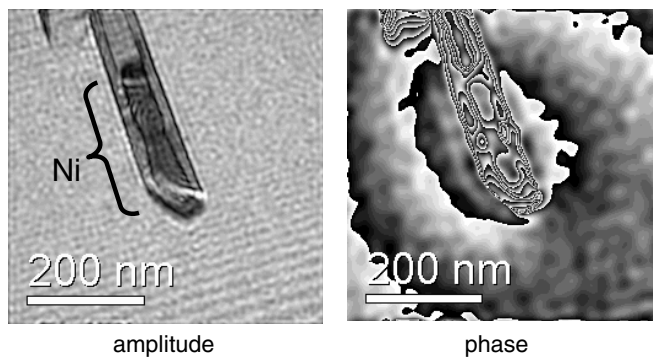


Figure 45. Nickel-filled nanotube. Ni incorporated into a carbon nanotube can easily be recognized in the amplitude image. The phase image shows that, also at these small dimensions, the Ni-rod is magnetic exhibiting a dipole-like magnetic stray field. The magnetic field produces an antisymmetric phase distribution across the rod (see figure 28), whereas an electric field would produce a symmetric one. The undulations in the phase contour lines are an artifact stemming from structure noise of the supporting carbon foil. The fine phase lines inside the nanotube cannot be interpreted easily. They stem from the phase shift mainly from the mean inner potential and magnetization of the nickel, however, the details are undersampled and hence not resolved. From [79].

Fourier space, from the object exit-wave in real space by coherent superposition. In kinematical approximation, the phase shift of atomic columns in real space increases linearly with increasing specimen thickness, which is particularly interesting for material analysis with atomic resolution.

With increasing specimen thickness, however, the probability of multiple scattering rises significantly as described by dynamical theory. In the beamplots (figure 59), strong deviations from linear dependence on specimen thickness are observed in amplitude and phase of each

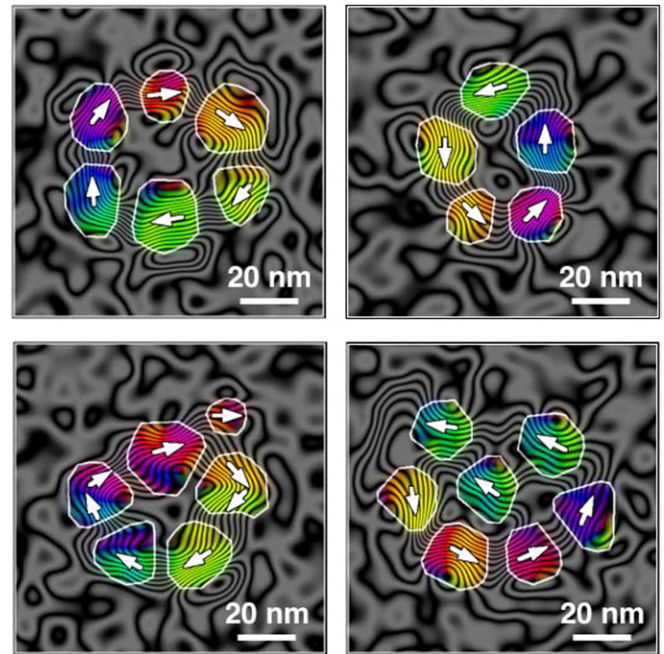


Figure 46. Rings of cobalt-particles. The Co-particles form magnetic rings with flux-closure of the stray field. Phase amplification 128. The phase shift by one particle is about $2\pi/10$. From [80]. Copyright 2004 Wiley-Liss. Reprinted with permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.

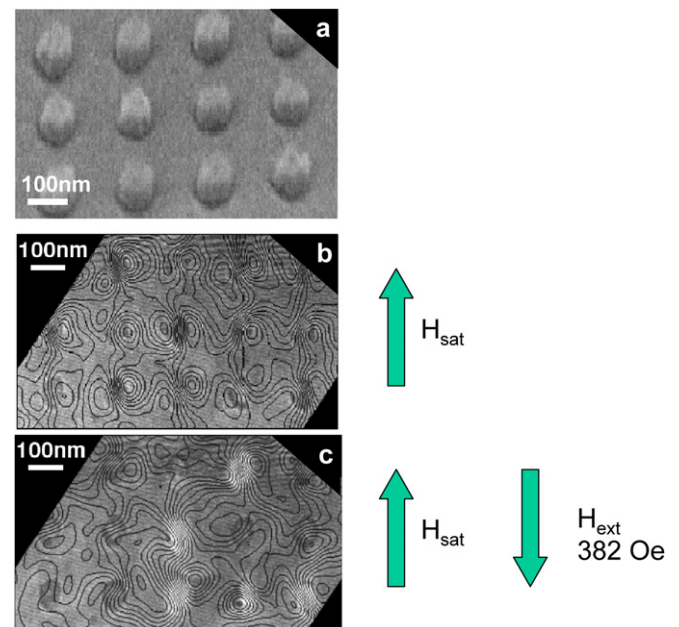


Figure 47. Array of Cobalt-dots (20 nm thick) on Si at remnant states: (a) SEM micrograph. (b) and (c) magnetic phase shift at two remnant magnetic states. Contour spacing 0.033 rad: (b) after saturating dots upwards and then removing the external field. (c) after saturating dots upwards, applying a 382 Oe downward field and then removing the external field. Reprinted with permission from [81].

reflection. Also reflections, which are kinematically forbidden, may have a non-negligible excitation. With increasing specimen thickness, amplitude minima of the zero-beam are observed. They are found as the so-called extinction

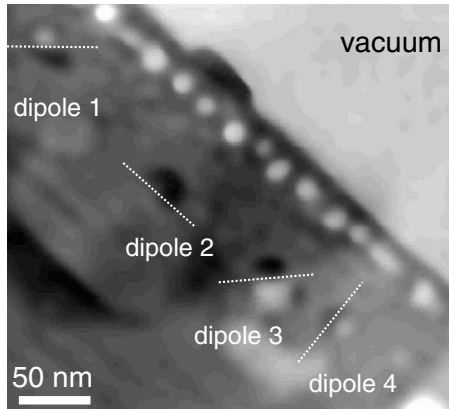


Figure 48. CoSm-particles implanted in Si. Some of the CoSm-particles, ion-implanted in 4H-SiC, show magnetic dipole appearance. The strength of the dipoles differs, because only the in-plane component of magnetization shifts the phase. The orientation of the dipoles seems random. From [82].

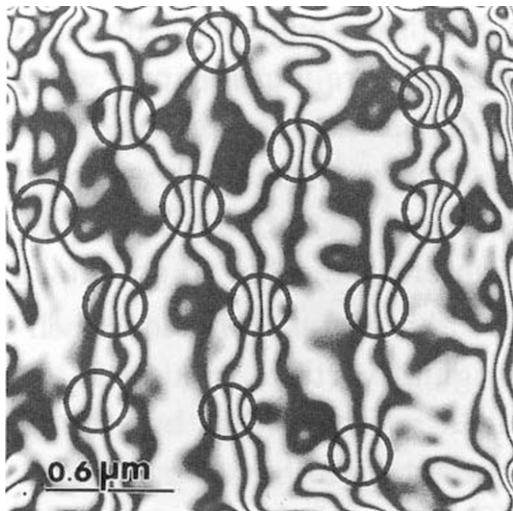


Figure 49. Fluxon lattice in superconducting Nb-film. After cooling to $T = 4.5$ K, under an external field of 15 mT (150 Gauss) applied under 45° to the film plane, the vortices (encircled) arise forming a lattice. Phase amplification 16° . Reprinted with permission from [83]. Copyright 1993 by the American Physical Society.

thicknesses, where most of the beam intensity is found in diffracted reflections. This attenuation of the zero-beam gives rise to a predominant interference of diffracted reflections in real space, which may yield artificial fine details with doubled periodicity in amplitude and intensity images hampering the interpretability. These so-called half-spacings are often referred to as lattice resolution of the electron microscope [86].

Interestingly, since these half-spacings do not occur in phase images, the interpretability in terms of atom species and positions is significantly improved. Additionally, provided an information limit of about 0.1 nm is given, a phase jump of almost 2π is observed at the extinction thickness, which is a very sensitive guideline for specimen thickness determination (figure 52).

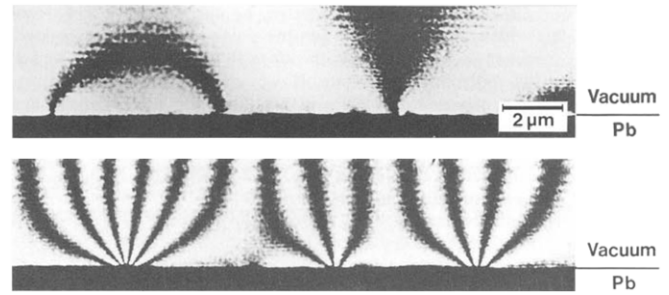


Figure 50. Single fluxons issuing from superconducting Pb-surface. Cross-sections of vortices under an external magnetic field at $T = 4.5$ K. Top: 200 nm thin Pb-layer. Right: single fluxon $h/2e$; left: antiparallel pair of two fluxons. Bottom: the 1000 nm thick Pb-layer accommodates larger vortices consisting of an integral number of bundled fluxons. Phase amplification 2° . Reprinted with permission from [84]. Copyright 1991 by the American Physical Society.

6.2. Benefits of atomic resolution holography

Not only for mesoscopic but also for atomic applications, off-axis electron holography offers advantages for quantitative analysis, when it is compared with conventional HRTEM, namely the ability of zero-loss energy-filtering, linear wave acquisition, and a-posteriori correction of coherent aberrations during hologram reconstruction.

Due to the principle of an object-modulated electron wave interfering with a coherent unscattered reference wave, the recorded image wave can be regarded as exceptionally well zero-loss energy-filtered by 10^{-15} eV. Electrons, which have undergone some inelastic event such as thermal diffuse scattering ('phonon scattering') or plasmon scattering, are 'discarded' from the coherent wave field showing up as 'absorption' like the effect of C_{inel} . Since most of the image simulation programs describe inelastic scattering as absorption, in reality they do not deliver the intended conventional intensity images, instead they provide us with the amplitude image reconstructed from an off-axis hologram. This is because, in conventional images, the inelastic electrons still contribute by a possibly strong background, whereas in holography they are really 'absorbed' by the filtering property. In any case, the available simulations are most favorable for the quantitative interpretation of holographically reconstructed object exit-waves.

Unlike conventional micrographs, the reconstructed image wave is purely linearly transferred due to its indirect encoding as modulation of interference fringes. The Fourier spectrum of the corresponding hologram contains both the linear sideband and the non-linear centerband (figure 53). Except for inelastic contributions forming, e.g. Kikuchi-bands, the intensity of the sideband corresponds to the diffraction pattern in the electron microscope, only limited by the damping envelope functions due to incoherent aberrations. Thus, in combination with the usage of intrinsically highly linear slow-scan CCD-cameras, the evaluation of electron holograms allows a quantitative analysis both in real and reciprocal space at atomic dimensions.

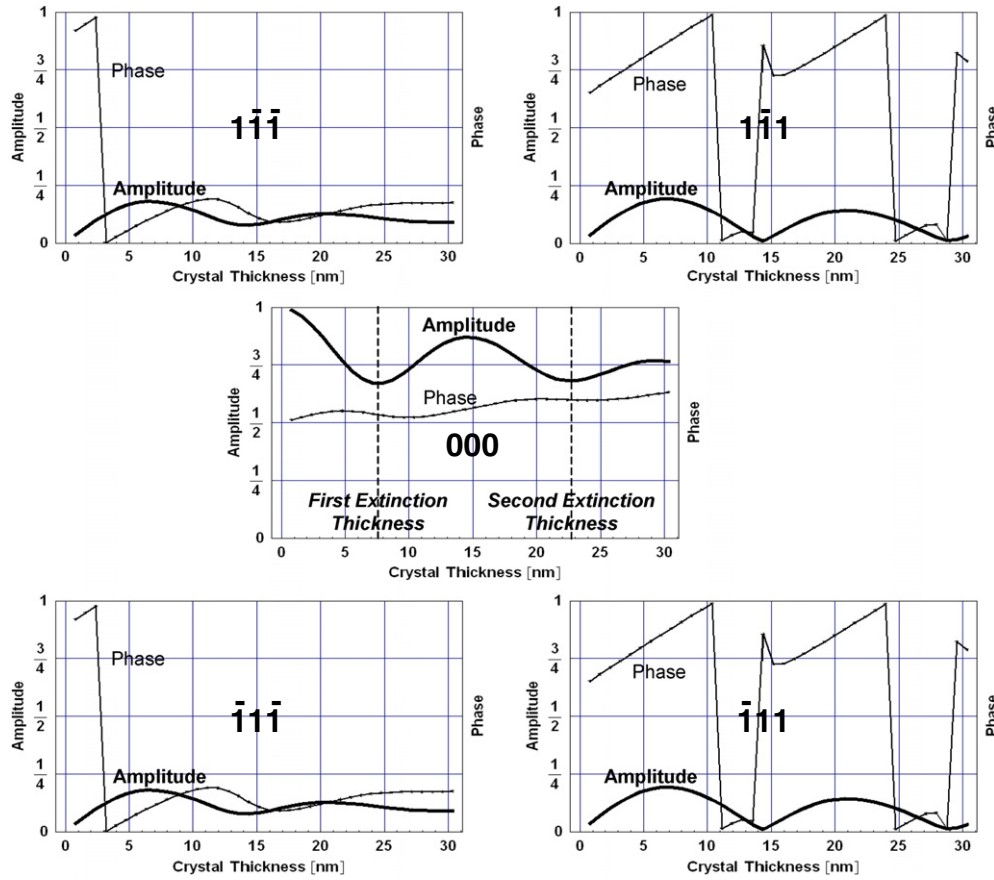


Figure 51. Beamplots of GaAs in the $[1\ 1\ 0]$ -orientation simulated by means of a full dynamical simulation using the EMS program package [85] for a TEM with 300 kV acceleration voltage. Only the zero-beam and the least diffracted $(1\ 1\ 1)$ -reflections are shown.

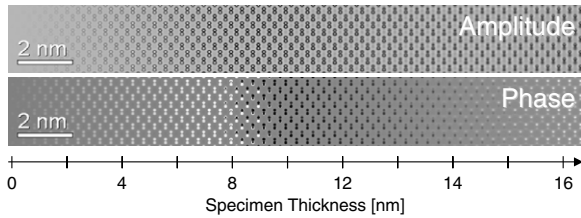


Figure 52. Simulated object exit-wave in amplitude and phase of an $[1\ 1\ 0]$ -oriented GaAs-crystal with a wedge-angle of 45° . Only incoherent aberrations for a Philips CM30FEG UT TEM are taken into account. Because of the wedge-shaped geometry, the distance from the thin crystal edge can be directly interpreted as specimen thickness.

6.3. Special aspects for acquisition of atomic resolution holograms

Fringe spacing. By means of the voltage U_f at the biprism filament, the carrier frequency of the interference fringes has to be chosen as $q_c \geq 2, \dots, 3 q_{\max}$. This guarantees a clear separation of centerband and sideband in reciprocal space (figure 53). Most of the high-resolution holograms are recorded with fringe spacings of about $s_{\text{hol}} = 1/q_c = 0.05$ nm or smaller, which is a challenge concerning disturbances and instabilities.

Width of hologram, number of fringes and pixel number of CCD-camera. A prerequisite for quantitative analysis is the

recording of the image wave as accurately as possible for all spatial frequencies transferred into the image plane well above noise. Therefore, several aspects have to be considered during acquisition of high-resolution electron holograms.

In order to catch all information needed for correction of the coherent wave aberration, the field of view recorded by the CCD-camera must be larger than

$$w_{\min} = 4 \text{ psf} \quad (6.1)$$

with the diameter psf of the point-spread-function $\text{PSF}(\vec{r})$. This condition is equivalent to the well-known Rayleigh-criterion. With $q_c \geq 2, \dots, 3 q_{\max}$, the minimum number of needed hologram fringes results as $n_{\text{fringe}} = 8, \dots, 12 q_{\max} \text{psf}$. Since each fringe has to be sampled by 4 CCD-pixels, the pixel number of the CCD-camera must be larger than $n_{\text{pix}} = 32, \dots, 48 q_{\max} \text{psf}$. This means that finally the CCD-camera determines the resolution limit $q_{\max}$ achievable by high resolution electron holography.

Adaptation of the hologram geometry. To control fringe spacing and width of hologram in terms of the object, the respective formulae (2.12) and (2.13) have to be related from the image plane to the object plane. For this, one has to divide by the magnification $M = (a + b)/f$, with the distances a and b given in figure 1, and f focal length of the objective lens. Related to the object plane one obtains for the spatial frequency

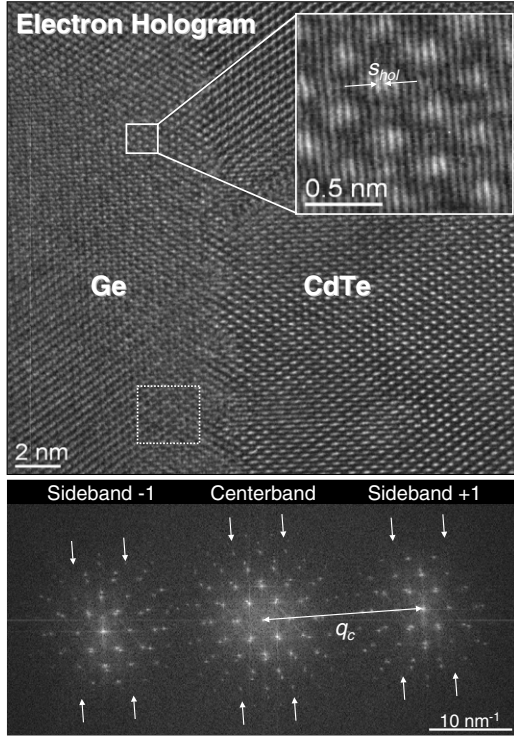


Figure 53. Top: atomic resolution off-axis electron hologram of an [1 1 0]-oriented Ge/CdTe-interface. The fine interference fringes with a spacing of $s = 0.054$ nm can be observed in the magnified insert. The dotted square indicates the area displayed in figure 55. Bottom: the Fourier spectrum consists of centerband and two sidebands. They are sufficiently separated by the carrier frequency $q_c = 1/s = 18.5$ nm⁻¹. The centerband shows pairs of reflections, where the inner reflection belongs to CdTe, the outer one to Ge. Due to the small thickness of the sample, kinematically forbidden reflections of Ge, indicated by arrows, are in fact missing in the sideband, whereas the appearance of these reflections in the centerband proves the non-linearity of image intensity.

of the hologram fringes

$$q_c = \frac{2\gamma_0 U_f a}{\lambda f} \quad (6.2)$$

and for the hologram width

$$w = \frac{2\gamma_0 U_f b f}{a + b} - 2r_f \frac{f}{a}. \quad (6.3)$$

Since $f \approx 1$ mm does not change much under the comparably small defocus values of some 100 nm, the only free parameters are the filament voltage U_f and the distance b , which can be set by means of the excitation of the subsequent intermediate lens. Because q_c does not depend on b , both q_c and w can be controlled independently by means of U_f and b .

Optimum focus of objective lens. Since coherent aberrations have to be corrected anyway a-posteriori, the focus can freely be optimized for taking holograms. It turns out that there is an optimum focus for holography [31] given as

$$D_z^{\text{optimum}} = -0.75 C_s \left(\frac{q_{\text{max}}}{k} \right)^2, \quad (6.4)$$

where q_{max} denotes the highest desirable spatial frequency to be recorded. Generally speaking, this focus keeps the $\text{grad} \chi(q)$ -function as low as possible over the whole range of spatial frequencies $[0, q_{\text{max}}]$. This offers the following advantages:

1. At optimum focus, the spatial coherence envelope is tuned such that the information limit q_{info} of the electron microscope is increased to about $2 q_{\text{Scherz}}$. Then, the information transfer is mainly limited by the temporal coherence damping envelope.
2. At optimum focus, the diameter of the point spread function is minimized as

$$\text{psf} = \frac{1}{2} C_s \left(\frac{q_{\text{max}}}{k} \right)^3 \quad (6.5)$$

(‘disc of least confusion’). Therefore the needed hologram width, which is about a quarter of the one at Scherzer-focus, is the smallest possible for q_{max} . This saves precious coherent electrons and pixel number of the CCD-camera.

3. At optimum focus, the pixel number needed in Fourier space is also minimized, since regions with steep $\text{grad} \chi(q)$ would need many pixels to avoid undersampling. Here, the combination of a low- C_s objective lens with the optimum focus for holography provides a wave aberration with a moderate gradient, which is less demanding in terms of sampling. Consequently, a CCD-camera with $1k \times 1k$ ($2k \times 2k$) pixels allows the reconstruction of sidebands with 256×256 (512×512) pixels and hence a resolution of 0.13 nm (0.1 nm) for a 300 kV electron microscope with a spherical aberration of about $C_s = 0.62$ mm [87].
4. At optimum focus, the phase detection limit is improved:

Phase detection limit. Besides lateral resolution, the phase detection limit

$$\delta\phi = \text{Noise figure} \frac{\text{snr } n_{\text{rec}}}{C_{\text{inel}}} \quad (4.13)$$

is the most important figure of merit for hologram quality and hence for the desired object exit-wave. It should be small enough to allow discerning single atoms at $\text{snr} = 3$.

Of special interest is the contrast factor C_{inel} of inelastic scattering. For reasons of interpretability also in holography, the objects should be thinner than the first extinction thickness of the zero-beam, commonly less than 7–12 nm; this thickness provides predominantly kinematic interaction allowing a direct interpretation of phase distributions in terms of atomic columns. At these thicknesses, there is only a low probability of inelastic scattering (neglecting amorphous surface layers) hence, for the sake of simplicity, $C_{\text{inel}} \approx 1$ is a reasonable assumption.

Last but not least, the number of reconstructed pixels $n_{\text{rec}} = 2q_{\text{max}}w$ is given by intended resolution q_{max} and the hologram width w . Because a minimum hologram width w_{min} (equation (6.1)) is needed, the minimum number of reconstructed pixels is $n_{\text{rec min}} = 2q_{\text{max}}w_{\text{min}}$. Inserting the

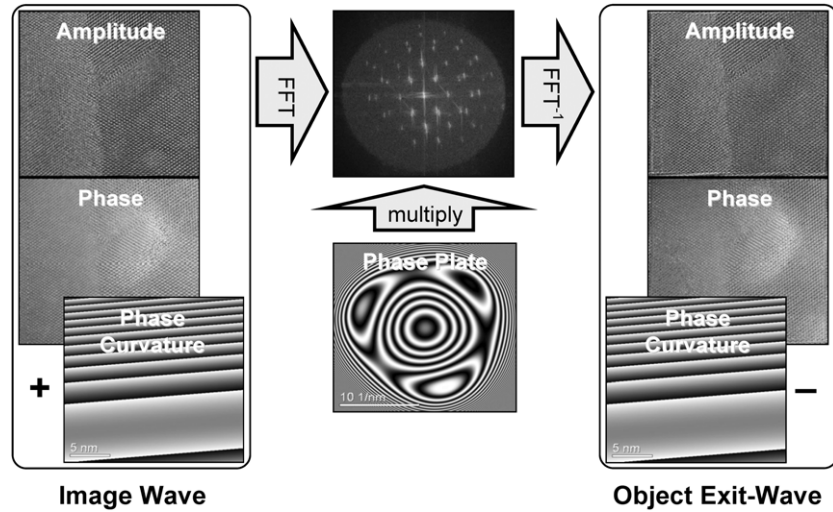


Figure 54. Procedure for the correction of coherent aberrations. The phase curvature added to the image wave simulates the cylindrical wave front of the elliptically shaped holographic illumination. Multiplication of the numerically generated phase plate $\chi_{\text{num}}(\vec{q})$ to the Fourier spectrum and inverse Fourier transformation yields the corrected object exit-wave, after the artificially introduced phase curvature has been subtracted. Aberration parameters: $C_s = 0.623$ mm, $D_z = -154$ nm, $A_2 = 4$ nm, $\alpha_{A2} = 30^\circ$, $A_3 = 1$ μ m, $\alpha_{A3} = -15^\circ$.

relations set up at optimum focus, one finds a simple relation for the minimum needed number of resolved pixels

$$n_{\text{rec min}} = 10.12 \left(\frac{q_{\text{max}}}{q_{\text{schertz}}} \right)^4, \quad (6.6)$$

for a hologram recorded in a TEM with Scherzer-resolution q_{schertz} . From this, one obtains a simple estimate for the phase detection limit

$$\delta\varphi = 10.12 \text{ Noise figure } \frac{\text{snr}}{C_{\text{inel}}} \left(\frac{q_{\text{max}}}{q_{\text{res}}} \right)^4. \quad (6.7)$$

Inserting reasonable values for our Philips CM30FEG UT/Special-Tübingen TEM with Scherzer resolution $q_{\text{schertz}} = 5.9 \text{ nm}^{-1}$, Noise figure $\approx 4 \times 10^{-4}$ at Triebenberg lab, and neglecting inelastic scattering $C_{\text{inel}} \approx 1$ for very thin objects, a phase detection limit of

$$\delta\varphi = \frac{2\pi}{20} \quad (6.8)$$

is reachable for $q_{\text{max}} = 10 \text{ nm}^{-1}$ at a signal-to-noise ratio of $\text{snr} = 3$. This is sufficient to identify in the reconstructed phase image single heavy atoms like e.g. gold with a phase peak of $2\pi/12$; however, single light atoms such as oxygen having a phase peak of $2\pi/50$ are buried in noise.

6.4. Correction of coherent wave aberration

At atomic resolution, the reconstructed image wave $\text{ima}(\vec{r})$ consists of both low and high spatial frequencies. However, as shown in section 3, spatial frequencies larger than $q_{\text{schertz}}/10$ are strongly falsified by the coherent wave aberration $\chi(\vec{q})$ of the objective lens, which shows up in the image wave as a complicated scrambling of amplitude and phase as well as a delocalization of object information, both given in real space by the point-spread-function $\text{PSF}(\vec{r}) = \text{FT}^{-1}\{\exp(-i\chi(\vec{q}))\}$.

In particular, this strongly hampers the interpretation of non-periodic object structures such as grain boundaries, interfaces and small clusters. Therefore, *a posteriori* correction of aberrations is one of the most interesting aspects of holography.

In the following we distinguish between the wave aberration $\chi_{\text{mic}}(\vec{q})$ effective in the TEM during recording the hologram, and $\chi_{\text{num}}(\vec{q})$ modeled for numerical processing of the reconstructed wave.

For correction of the coherent wave aberration $\chi_{\text{mic}}(\vec{q})$, the aberrated image wave $\text{ima}(\vec{r})$ is back-propagated to the object exit plane to obtain the object exit-wave $o(\vec{r})$ according to the scheme in figure 22, [88–91]. This is achieved by means of a computer program establishing a virtual microscope with arbitrarily configurable imaging properties. In particular, arbitrary wave-optical transfer properties of the objective lens can be modeled by a numerically generated phase plate $\chi_{\text{num}}(\vec{q})$ [92]. The actual aberration correction is performed by multiplication of the image wave's Fourier spectrum with $\exp(i\chi_{\text{num}}(\vec{q}))$. By inverse Fourier transformation the desired object exit-wave $o(\vec{r}) = a(\vec{r}) \exp(i\varphi(\vec{r}))$ results, which is free from coherent wave aberration (figure 54). It should be noted that higher spatial frequencies are still damped by incoherent aberrations given by the damping envelope functions for partial spatial and temporal coherence. Due to non-negligible noise contributions, the experimental object exit-wave is normally not deconvoluted from these inelastic aberrations; instead, they are considered in corresponding object exit-wave simulations.

Accuracy requirements. The accuracy of the numerically generated phase plate $\chi_{\text{num}}(\vec{q})$ needed for truthful modeling of the wave aberration $\chi_{\text{mic}}(\vec{q})$ is given by the modified Rayleigh-criterion

$$|\chi_{\text{mic}}(\vec{q}) - \chi_{\text{num}}(\vec{q})| \leq \frac{\pi}{6}, \quad (6.9)$$

which has to be fulfilled for all involved spatial frequencies $\vec{q}$. This criterion makes sure that the residual cross-talk between

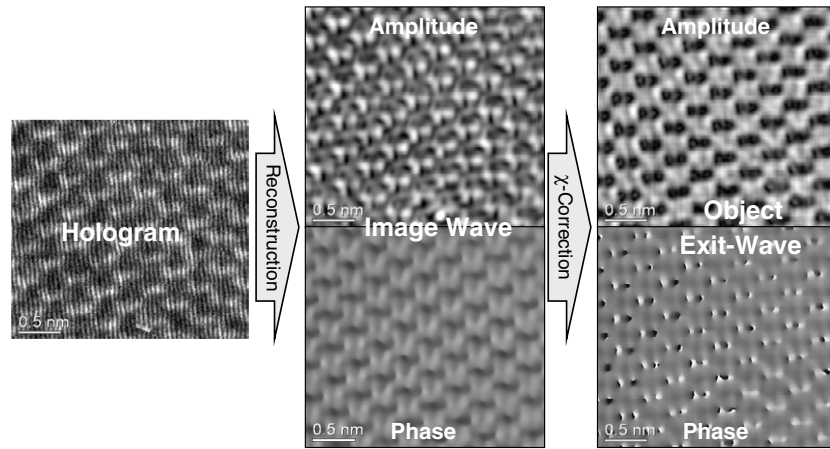


Figure 55. Summary of all reconstruction steps. The correct dumbbell-contrast in amplitude and phase of the object exit-wave, which cannot be observed either in the reconstructed image wave or in the hologram, indicates the successful correction of coherent aberrations. For improved visualization, only a small area (dotted square in figure 53) is displayed in all reconstruction steps. The waves are slightly noise-reduced by Bragg-filtering.

amplitude and phase is sufficiently small (figure 18), i.e.

$$|\sin(\chi_{\text{mic}}(\vec{q}) - \chi_{\text{num}}(\vec{q}))| \leq 0.5. \quad (6.10)$$

This requirement has to be met not only in holography but in all atomic resolution imaging techniques. It poses a challenge in that currently an optimum-search has to be conducted in a 10-parameter space: 8 aberration coefficients (equations (3.34)–(3.38) plus scaling in Fourier space as well as the exact wave number $k_0 = 1/\lambda$ have to be determined with such a high accuracy. For example, to achieve 0.1 nm resolution with a 300 kV TEM, the coefficient for spherical aberration has to be known at about 0.1% accuracy, and the coefficients for defocus and two-fold astigmatism better than 0.5 nm [93].

Evidently, one of the key problems of all atomic resolution methods is the extremely precise determination of the coherent wave aberration $\chi_{\text{mic}}(\vec{q})$, both for *a posteriori* correction and for fine-tuning a hardware corrector. Therefore it would be very helpful to make use of additional criteria derived from general properties of aberration-free imaging, which are not specific to the object structure under investigation. Suitable criteria such as ‘a weak phase object does not produce any amplitude contrast’ or ‘delocalization is minimal’ should be applied.

Using *a priori* knowledge about the object of at least a part of the field of view, special strategies have been developed for aberration assessment: utilizing the criterion that a very thin amorphous edge of a specimen is in good approximation a weak phase object, showing no amplitude contrast at ideal imaging, the amplitude contrast can be minimized; this is a well-established correction guideline, however only for symmetric aberrations such as spherical aberration and defocus [90, 94]. This minimization can be automated by means of a genetic algorithm [91]. Antisymmetric aberration coefficients are normally determined by trial and error using the virtual microscope program [90, 92], which allows an interactive correction of the coherent wave aberration. In comparison with simulated object exit-waves of crystalline samples, all aberration coefficients are tuned such that both amplitude

and phase of the experimental object exit-wave is in good agreement with the simulated ones.

The correction of the coherent wave aberration not only performs an unscrambling of information transfer and a neutralization of delocalization but it also enhances the phase signal by recollecting the phase information from the point spread function back to the point of origin. Since aberration correction opens the imaging aperture, Geiger and Lichte [95] were able to resolve the comparable weak oxygen atom columns in the phase image of a $\text{YBa}_2\text{Cu}_3\text{O}_7$ HTc-superconductor.

Correction of distortions provoked by elliptical illumination. The elliptically shaped electron source used for hologram acquisition imposes an additional need for correction: this special illumination mode gives rise to a variation of beam directions over the field of view, alas, yielding loss of isoplanicity. Therefore, coherent aberrations are successfully corrected only for a small part of the field of view. These imaging artifacts caused by the elliptically shaped illumination are considered in the aberration correction process by means of a phase curvature, which models the shape of the illuminating wave front (figure 54) [96]. Applied in real space during aberration correction, coherent aberrations are compensated for the whole field of view yielding an improved spatial resolution up to the information limit of the electron microscope, which allows a direct interpretation of the resulting object exit-wave regarding the object structure under investigation (figure 55).

6.5. Quantitative analysis

The reconstructed object exit-wave is normally displayed in the form of two images as amplitude and phase. However, the object exit-wave is more than simply an image: it represents a two-dimensional array of quantitative complex data offering the possibility of analyzing the data both in real space and in reciprocal space.

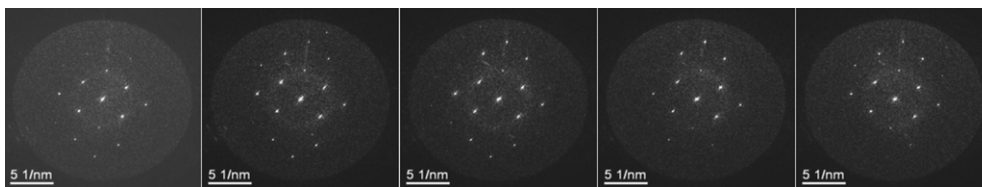


Figure 56. Appearance of crystal tilt in sideband. Series of five sidebands, where the corresponding holograms have been recorded under different crystal tilt, in steps of 0.1° about the horizontal tilt axis. The sideband in the middle shows a good orientation of the GaAs-crystal on the $[1\ 1\ 0]$ -zone-axis. Evidently, crystal tilt can be examined under reconstruction from symmetric excitation of reflections.

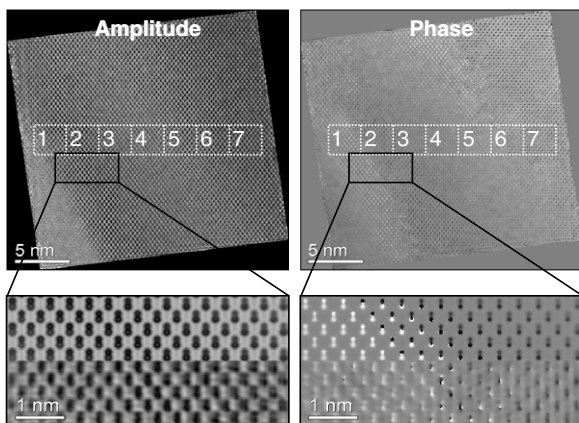


Figure 57. Holographically reconstructed object exit-wave (noise-reduced) in amplitude (display range: 0–1.4) and phase (display range: 0– 2π) of a wedge-shaped GaAs-crystal in the $[1\ 1\ 0]$ -orientation. The thinnest edge is located in the lower left half of the field of view, and the specimen thickness increases to the right according to its 45° wedge-angle. In the magnified areas, the upper parts have been replaced by a simulated image showing good agreement in both structure and quantity. The numbered boxes indicate the areas, from which the nanodiffraction patterns, shown in figure 58, have been determined.

The accurate zone-axis orientation, important for quantitative analysis such as materials characterization at atomic dimensions, can already be judged from the Fourier transform of the electron hologram. A small residual misorientation shows up as an asymmetric excitation of reflections in the sideband (figure 56), which is similar to effects observable in diffraction patterns taken at the electron microscope. The sideband, however, offers quantitative measures, whereas in the electron microscope the exact zone-axis orientation can only be found by subjective estimations of reflection intensities.

Holographic nanodiffraction. Local variations of crystal tilt and thickness can easily be visualized by holographic nanodiffraction. By means of a numerical mask, arbitrary areas as small as a few unit cells are selected from the object exit-wave in real space, Fourier transformed and squared yielding diffraction patterns from these small areas (figures 57 and 58). They reproduce all the effects of asymmetry and different excitation of reflections caused by non-centrosymmetric crystals, local tilt and thickness variations.

Imaging in the light of single reflections. The reconstruction in the light of single reflections allows the analysis of their

contribution to the object exit-wave in real space. Here, a single reflection is masked out and centered in Fourier space, inverse Fourier transformed and displayed as amplitude and phase. This is of special interest for wedge-shaped samples, where the excitation of reflections can be measured depending on the specimen thickness and compared with corresponding object exit-wave simulations (figure 59). These experimentally obtained beamplots show all theoretically known effects of dynamical interaction.

Stobbs factor problem. In general, quantitative comparisons of lattice fringe contrasts in image intensities with correspondingly simulated ones have the result that the lattice fringe structure is well-reproduced by the simulation but the contrast is about a factor of 3–5 higher in the simulation. This contrast mismatch is also known as Stobbs factor problem. It severely hampers the possibility of materials characterization with atomic resolution. Many possible reasons for the contrast mismatch are in discussion, and apparently there are many contributions. Fortunately, all holographic investigations indicate that the reconstructed object exit-wave is hardly affected by the Stobbs factor. These findings suggest a major contribution to the contrast mismatch of inelastically scattered electrons due to e.g. thermal diffuse scattering since the holographically reconstructed object exit-wave is extremely well zero-loss energy filtered. Consequently, a sophisticated materials analysis with atomic resolution is within reach.

6.6. Applications

Materials analysis on the level of atoms. In order to solve the problem ‘which atoms are where’, two prerequisites have to be fulfilled: a sufficient spatial resolution in combination with a reasonable signal-to-noise ratio. In holography, both quantities are characterized by the maximum spatial frequency $q_{\max}$ available after reconstruction, and the phase detection limit $\delta\varphi_{\lim}$ giving an estimation for discernibility of neighboring atom columns. For holographic materials analysis, a challenge is the characterization of zinc-blende structures in $[1\ 1\ 0]$ -projection, where typical pairs of atom columns show up, the so-called dumbbells, with a distance less than the point resolution of most TEMs. Within a dumbbell, each column is exclusively built up with one specific atom species of the crystal under investigation. The task of holographic materials analysis is the identification of atom species within the dumbbell structure.

Figure 60 shows the holographically reconstructed object exit-wave of a cross-sectional ZnSe/BeTe-multilayer. The

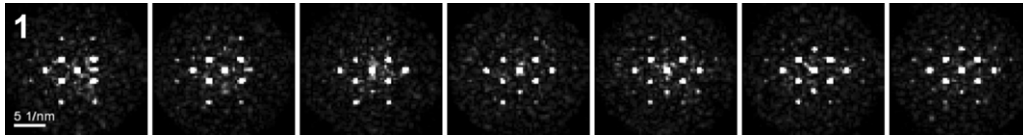


Figure 58. Nanodiffraction patterns determined from the areas shown in figure 57. Due to the increasing specimen thickness from left to right, the excitation of reflections varies according to dynamical diffraction theory.

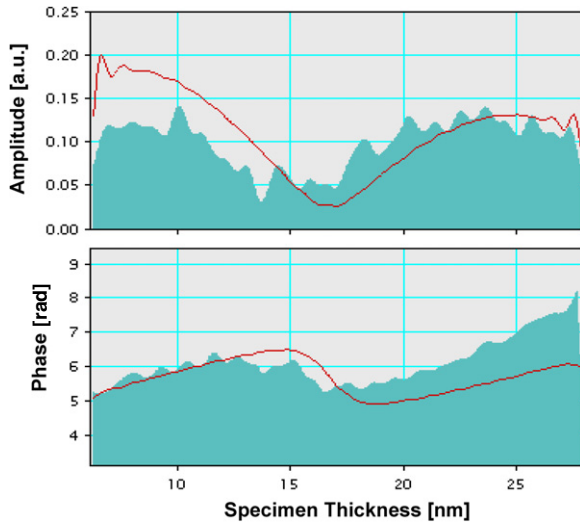


Figure 59. Beamplot of $1\bar{1}\bar{1}$ -reflection in amplitude and phase as determined from the object exit-wave of GaAs shown in figure 57. The filled areas represent the experimental data, whereas the line is the corresponding simulation. The modulations represent the variations of reflection excitation due to increasing specimen thickness known from dynamical diffraction theory.

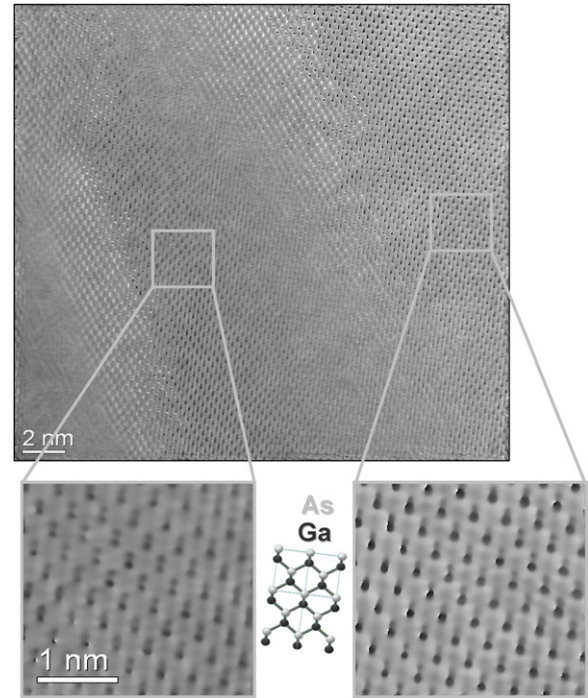


Figure 61. Phase image of the reconstructed object exit-wave of the $[1\bar{1}0]$ -oriented GaAs shown in figure 57. The different atom species within the dumbbell structure can be clearly distinguished due to their characteristic phase shift.

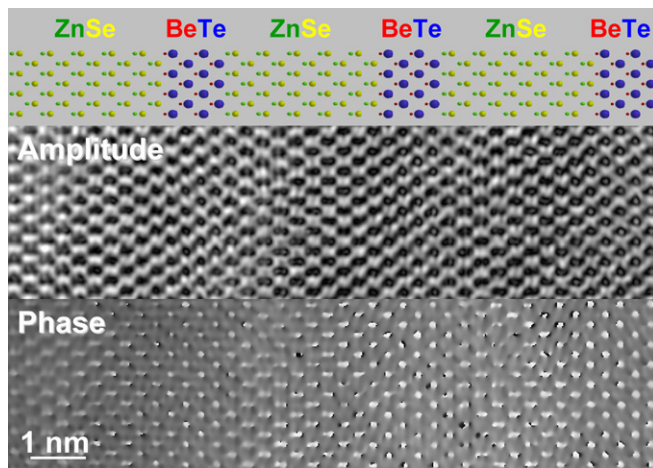


Figure 60. Holographically reconstructed object exit-wave of a cross-sectioned ZnSe/BeTe-multilayer in the $[1\bar{1}0]$ -orientation. The separate layers with a thickness of only a few monolayers can easily be identified both in amplitude and phase. (Cooperation with Dagmar Gerthsen, Karlsruhe, and Andreas Waeg, Braunschweig.)

different layers can easily be identified, because the Be-columns are hidden due to their very small weight compared with the Te-columns. A slight atomic intermixing at the interface, expected from layer deposition, causes a structural transition zone of about two monolayers.

In $[1\bar{1}0]$ -oriented GaAs, both types of atomic columns within the dumbbell structure can hardly be distinguished, because there is only a small difference in atomic numbers for Ga and As with $Z_{\text{Ga}} = 31$ and $Z_{\text{As}} = 33$. From object exit-wave simulations, it is known that only the phase of the electron wave is sensitive enough to identify the difference of both atomic columns due to the stronger phase shift of the As-column [97]. The phase difference between both atomic columns becomes larger with increasing specimen thickness. According to our experiences, Ga and As columns can satisfactorily be discriminated at a specimen thickness of over 10 nm, which is slightly beyond the first extinction thickness of the zero-beam (figure 61). At this thickness, a phase difference of $2\pi/12$ is calculated by means of a full dynamical simulation, which is larger than the phase detection limit of our CM30FEG UT/Special Tübingen TEM and hence detectable. This is caused by amorphous surface layers, which additionally modulate amplitude and phase of the object exit-wave. Nevertheless, the phase of the object exit-wave is sensitive enough to allow materials characterization on an atomic level currently unreachable by analytical methods.

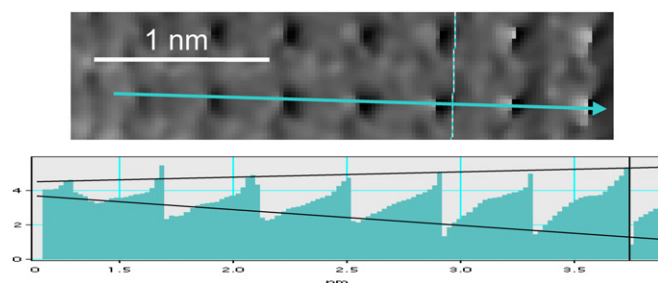


Figure 62. Phase of an object exit-wave reconstructed from a hologram of BaTiO₃ in the [1 0 0]-orientation. At this special specimen thickness, the phase signal stemming from dynamical diffraction is very small, whereas a strong phase shift due to atomic dipoles can be observed. The phase signal becomes larger to the right, because the specimen thickness gradually increases.

Atomic dipoles in ferroelectrics. With improving spatial resolution and phase detection limit, one can think about challenges beyond the question ‘which atoms are where’. An even more demanding question is the determination of local charge density distributions and bonds, important for a basic understanding of solid state physics on an atomic level.

A first step in this direction has already been taken with holographic analysis of ferroelectric crystals such as BaTiO₃ on an atomic level. Here, the asymmetric local charge distribution plays an important role: caused by a tiny tetragonal distortion, the centers of positive and negative atomic charges are shifted against each other within a unit cell such that as a net effect an atomic dipole is built up between two virtual charges with opposite signs. The in-plane component of such an atomic dipole causes a characteristic phase difference at these two charges, when a plane electron wave penetrates a thin foil of the ferroelectric specimen. For BaTiO₃, a phase difference of $2\pi/30$ per unit cell thickness is expected, however, too small for investigations of a single dipole with present TEMs. Fortunately, real-world samples have a foil thickness of several monolayers, where in projection the phase shift is significantly increased by adding up several atomic dipoles. Therefore, atomic dipoles have already been identified in phase images of BaTiO₃ (figure 62) using high-resolution off-axis electron holography in our CM30FEG UT/Special Tübingen TEM. Moreover, ferroelectric nanodomains with the size of only a few nanometres have been found in thin specimen areas, only explainable as a metastable state, which is not expected in bulk crystals.

6.7. Future challenges

Improvement of Signal/Noise by a Cs-corrector. Off-axis electron holography offers unique possibilities of recovering completely the aberration corrected object wave also with uncorrected microscopes and hence would not need a Cs-corrected microscope for improved lateral resolution. However, in particular at atomic resolution, the performance of holography is affected by the aberrations of the recording TEM in that the Signal/Noise-properties (‘phase detection limit’) of the reconstructed wave are damaged. To improve not only lateral resolution but also signal resolution, we realized off-axis electron holography with a C_s-corrected TEM [98].

It turns out that the phase detection limit improves by a factor of four at atomic resolution. A further advantage is the possibility of fine-tuning the residual aberrations by a-posteriori correction. Therefore, combination of both the methods, i.e. C_s-correction of a TEM and off-axis electron holography, opens new perspectives for a complete TEM-analysis on an atomic scale.

Interpretation: need for better modeling. The usual simulation programs use free-atom models, mostly the Weickenmeier–Kohl models [99]. While this is appropriate to interpret the position of the atom positions even at sub-Angstrom resolution, insufficiencies are expected for the correct modeling of intrinsic fields on an atomic scale, such as those occurring in ferroelectrics. A solution is offered by *ab initio* calculations which also consider the binding properties between the atoms [100].

Likewise essential is the correct consideration of thermal atomic motions for the calculation of waves and intensities. It turned out that the consequences are in principle quite different in holography, conventional imaging and diffraction. In particular at heavy atoms, the description by means of the usual Debye–Waller damping factor is insufficient, from the viewpoint of both the underlying physics and the numerical results [101].

Interpretation: solution of the inverse problem. Like the conventional image intensity, the reconstructed object exit-wave in amplitude and phase represents only the projection of the object, i.e. a two-dimensional measurement of object properties with atomic resolution. Despite this limitation, the object exit-wave has advantages compared with the conventionally gained image intensity. Besides the possibility of correcting coherent aberrations, it might allow solving the inverse problem of electron scattering, i.e. the unique three-dimensional determination of atom species and positions from a two-dimensional object exit-wave. First attempts for the determination of local tilt and thickness of the object have already been made [102]. This could be the foundation for a true three-dimensional reconstruction of the object structure using a combined holographic/tomographic acquisition technique also at atomic resolution.

7. Conclusions and outlook

Electron waves are unique in their high sensitivity against electric and magnetic structures on a length scale reaching from several micrometres to atomic dimensions. Due to the highly developed electron optical components such as sources, lenses and detectors, object information can be analyzed quantitatively in real space from highly resolved images, as well as in Fourier space from accurate diffraction patterns; furthermore, for materials analysis, energy losses can be analyzed by means of energy spectrometers and filters at an unrivaled high resolution both in energy and position. Therefore electron microscopic methods are very successfully exploited for structure determination and are hence indispensable for future nanoscience.

Unfortunately, transmission electron microscopy suffers from the drawback that the usually recorded intensity images solely display the squared amplitude (modulus) of the image wave, but not the phase. Therefore, most interesting items of the object structure, which are purely encoded in the phase, are virtually invisible. Consequently, the object information found in the intensity image is incomplete, which may result in a misleading interpretation of the findings.

This restriction is overcome by means of electron holography, which allows recording both amplitude and phase of the image wave. After about 40 years of development, electron off-axis holography using the Möllenstedt electron biprism as beam splitter has reached a very powerful performance for the analysis of the complete image wave.

Holography is very advantageous, since the reconstructed wave offers the following advantages over the intensity:

- linearity
- zero loss
- complete analysis in real space and in Fourier space
- *a posteriori* aberration correction
- intrinsic quantitative evaluation of all data.

The field of applications exploiting the whole wave information now covers nearly the same range as conventional TEM. Typical examples of holographic analysis of electric and magnetic nanofields show the benefits for solid state physics and materials science. At medium resolution magnetic and electric fields become accessible, which represent a new dimension in structure analysis. At atomic resolution the quantitative interpretation of atomic phase shift allows determination of atomic species and interatomic fields, e.g. from ferroelectric dipoles in single unit cells.

Correspondingly, for further dedication, electron holography is being improved in several aspects:

- It will be more flexible by adapting the field of view and lateral resolution to the special needs of the different tasks; this will be achieved by providing different positions for the electron biprism along the path of rays.
- Increase of the obtainable amount of object information, given by the number of pixels and the signal/noise ratio, will be achievable by recording the holograms in most advanced TEMs, e.g. with electron sources of higher brightness, and aberration correction. The better the TEM, the better will be the performance also of holography.
- In-situ experiments are under development for a more unique measurement of the different object data by variation of corresponding parameters in and around the object, such as temperature, external fields and mechanical strain.
- Quantitative interpretation in terms of physics needs more reliable modeling of the object exit wave starting from the atomic properties. Therefore, ab-initio methods are increasingly used in areas of electron microscopy and also in electron holography.

Most interestingly, thanks to the quality of today's electron microscopes, electron holography is applicable by everyone, who carefully operates such a microscope according to the

state of the art. The needed special hard- and software is available on the market. Therefore and because of the high potential for structure determination, the number of holographic investigations is rapidly growing worldwide. The future expectations comprehend more detailed and more precise investigations of e.g. electric, magnetic and mechanic fields, which are indispensable for understanding the structure-properties relations of modern materials.

Acknowledgments

Developing electron holography in the past decades, we have experienced much stimulation and many discussions with colleagues. At the very beginning in Tübingen, Gottfried Möllenstedt, Friedrich Lenz, Karl-Heinz Herrmann and Herbert Wahl offered a lot of encouragement and insight into the basics. Later on, steady discussions within the scientific community with the groups around Tom Mulvey, Karl-Josef Hanszen, Akira Tonomura, Giulio Pozzi, John Cowley, Dirk Van Dyck and Peter Schattschneider helped to develop and follow common visions. Our present work is coined by detailed and lively arguments between the members of the Triebenberg-Laboratory, Bernd Eienkel, Petr Formánek, Dorin Geiger, Andreas Lenk, Martin Linck, Christopher Matzeck, Marianne Reibold, Falk Röder, Axel Rother, Karin Vogel and Daniel Wolf. Special thanks go to Harald Rose for his critical remarks considerably improving the manuscript, as well as to the referees for their very stimulating comments.

We are well aware that, beyond the papers we have cited, a lot of very fruitful contributions have been published in electron holography. We ask these colleagues for understanding that we cannot provide here a complete bibliography.

Our progress in the field was possible only by substantial financial support obtained over many years, such as that obtained from Deutsche Forschungsgemeinschaft, Volkswagen-Stiftung, Körber-Stiftung Hamburg, Francqui Foundation Bruxelles, the German Ministry for Research (BMBF) and the European Union (Framework 6 Integrated Infrastructure, Reference 026019 ESTEEM).

References

- [1] Reimer L 1989 *Transmission Electron Microscopy* 2nd edn (Berlin: Springer)
- [2] Hawkes P W and Kasper E *Principles of Electron Optics, vol. 3: Wave Optics* (London: Academic) ISBN 0-12-333354-7
- [3] Möllenstedt G and Düker H 1956 Beobachtungen und Messungen an Biprisma-Interferenzen mit Elektronenwellen *Z. Phys.* **145** 377–97
- [4] Wu Y Z, Schmid A K, Altman M S, Jin X F and Qiu Z Q 2005 Spin-dependent Fabry-Perot interference from a Cu thin film grown on fcc Co(001) *Phys. Rev. Lett.* **94** 027201
- [5] Born M and Wolf E 1980 *Principles of Optics* 6th edn (Oxford: Pergamon)
- [6] Schmid H 1985 Ein Elektronen-Interferometer mit 300 μm weit getrennten kohärenten Teilbündeln zur Erzeugung hoher Gangunterschiede und Messung der Phasenschiebung durch das magnetische Vektorpotential

- bei metallisch abgeschirmtem Magnetfluss *Dissertation* University of Tübingen
- [7] Byatt W J 1956 Analytical representation of Hartree potentials and electron scattering *Phys. Rev.* **104** 1298
- [8] Kirkland J 1998 *Advanced Computing in Electron Microscopy* (New York: Plenum)
- [9] Linck M, Lichte H and Lehmann M 2006 Electron holography: materials analysis at atomic resolution *Int. J. Mater. Res* (formerly *Z. Metallk.*) **97** 890
- [10] Jönsson C, Hoffmann H and Möllenstedt G 1965 Messung des mittleren inneren Potentials von Be im Elektronen-Interferometer *Phys. Kondens. Mater.* **3** 193
- [11] O'Keefe M and Spence J H C 1994 On the average Coulomb potential (Φ_0) and constraints on the electron density in crystals *Acta Cryst. A* **50** 33–45
- [12] Ehrenberg W and Siday R E 1949 The refractive index in electron optics and the principles of dynamics *Proc. Phys. Soc. Lond B* **62** 8–21
- [13] Aharonov Y and Bohm D 1959 Significance of electromagnetic potentials in quantum theory *Phys. Rev.* **115** 485
- [14] Chambers R G 1960 Shift of an electron interference pattern by enclosed magnetic flux *Phys. Rev. Lett.* **5** 3
- [15] Möllenstedt G and Bayh W 1962 Kontinuierliche Phasenschiebung von Elektronenwellen im kraftfeldfreien Raum durch das magnetische Vektorpotential eines Solenoids *Physikalische Bl.* **18** 299
- [16] Tonomura A, Osakabe N, Matsuda M, Kawasaki T, Endo J, Yano S and Yamada H 1986 Evidence for the Aharonov-Bohm effect with magnetic field completely shielded from electron wave *Phys. Rev. Lett.* **56** 792
- [17] Wohlleben D 1967 Diffraction effects in Lorentz microscopy *J. Appl. Phys.* **38** 3341
- [18] Cohen M S 1967 Wave-optical aspects of Lorentz microscopy *J. Appl. Phys.* **38** 4966
- [19] Tonomura A, Matsuda T, Endo J, Arie T and Mihama K 1986 Holographic interference electron microscopy for determining specimen magnetic structure and thickness distribution *Phys. Rev. B* **34** 3397
- [20] Harscher A, Lichte H and Mayer J 1997 Interference experiments with energy filtered electrons *Ultramicroscopy* **69** 201–9
- [21] Lichte H and Freitag B 2000 Inelastic electron holography *Ultramicroscopy* **81** 177–86
- [22] Verbeeck J, van Dyck D, Lichte H, Potapov P and Schattschneider P 2005 Plasmon holographic experiments: theoretical framework *Ultramicroscopy* **102** 239–55
- [23] Schattschneider P and Lichte H 2005 Correlation and the density matrix approach to inelastic electron holography in solid state plasma *Phys. Rev. B* **71** 045130
- [24] Potapov P L, Lichte H, Verbeeck J and van Dyck D 2006 Experiments on inelastic electron holography *Ultramicroscopy* **106** 1012–18
- [25] Howie A 1979 Image contrast and localised signal selection techniques *J. Microsc.* **117** 11
- [26] Rose H and Kohl H 1985 Theory of image formation by inelastically scattered electrons in the electron microscope *Adv. Electron. Electron Opt.* **65** 173
- [27] Verbeeck J, Bertoni G and Schattschneider P The Fresnel effect of a defocused biprism on the fringes in inelastic holography *Ultramicroscopy* at press
- [28] Zernike F 1935 Das Phasenkontrastverfahren bei der mikroskopischen Beobachtung *Z. Tech. Phys.* **16** 454
- [29] Scherzer O 1936 Über einige Fehler von Elektronenlinsen *Z. Phys.* **101** 593–603
- [30] Scherzer O 1949 The theoretical resolution limit of the electron microscope *J. Appl. Phys.* **20** 20–9
- [31] Lichte H 1991 Optimum focus for taking electron holograms *Ultramicroscopy* **38** 13–22
- [32] Rose H 1990 Outline of a spherically corrected semiplanatic medium- voltage transmission electron microscope *Optik* **85** 19–24
- [33] Haider M, Uhlemann S, Schwan E, Rose H, Kabius B and Urban K 1998 Electron microscopy image enhanced *Nature* **392** 768–9
- [34] Haider M, Rose H, Uhlemann S, Schwan E, Kabius B and Urban K 1998 A spherical-aberration-corrected 200 kV transmission electron microscope *Ultramicroscopy* **75** 53–60
- [35] Jia C L, Lentzen M and Urban K 2003 Atomic-resolution imaging of oxygen in perovskite ceramics *Science* **299** 870–3
- [36] Gabor D 1948 A new microscopic principle *Nature* **161** 563–4
- [37] Gabor D 1949 Microscopy by reconstructed wave-fronts *Proc. R. Soc. A* **197** 454–87
- [38] Leith E H and Upatnieks J 1962 Reconstructed wavefronts and communication theory *J. Opt. Soc. Am.* **52** 1123–30
- [39] Haine M E and Mulvey T 1952 The formation of diffraction image with electrons in the Gabor diffraction microscope *J. Opt. Soc. Am.* **42** 763
- [40] Möllenstedt G and Wahl H 1968 Elektronenholographie und Rekonstruktion mit Laserlicht *Naturwissenschaften* **55** 340–1
- [41] Wahl H 1975 Bildebenen holographie mit Elektronen *Thesis* University of Tübingen
- [42] Cowley J M 1992 Twenty forms of electron holography *Ultramicroscopy* **41** 335–48
- [43] Lichte H 1996 Electron holography: optimum position of the biprism in the electron microscope *Ultramicroscopy* **64** 79–86
- [44] Meyer R R and Kirkland A 1998 The effects of electron and photon scattering on signal and noise properties of scintillators in CCD cameras used for electron detection *Ultramicroscopy* **75** 23–33
- [45] Lenz F 1988 Statistics of phase and contrast determination in electron holograms *Optik* **79** 13–14
- [46] Lichte H, Herrmann K-H and Lenz F 1987 Electron noise in off-axis image plane holography *Optik* **77** 135–40
- [47] Harscher A and Lichte H 1996 Experimental study of amplitude and phase detection limits in electron holography *Ultramicroscopy* **64** 57–66
- [48] Lichte H Performance limits of electron holography *Ultramicroscopy* at press
- [49] McCartney M R and Gajdardziska-Josifovska M 1994 Absolute measurement of normalized thickness, t/λ , from off-axis electron holography *Ultramicroscopy* **53** 283–9
- [50] Lenk A, Lichte H and Muehle U 2005 2D-mapping of dopant distribution in deep sub micron CMOS devices by electron holography using adapted FIB-preparation *J. Electron Microsc.* **54** 351–9
- [51] Matteucci G, Missiroli G F, Nichelatti E, Migliori A, Vanzi M and Pozzi G 1991 Electron holography of long-range electric und magnetic fields *J. Appl. Phys.* **69** 1835
- [52] Zweck J and Bormans B 1992 The CM30 Lorentz lens *Philips Electron Opt. Bull.* **132** 1–8
- [53] Twitchett A C, Yates T J V, Dunin-Borkowski R E, Newcomb S B and Midgley P A 2006 Three-dimensional electrostatic potential of a Si p-n junction revealed using tomographic electron holography *J. Phys. Conf. Ser.* **26** 29
- [54] Wolf D 2007 Towards quantitative electron-holographic tomography *Microsc. Microanal. (Suppl.)* **13** 112–13
- [55] Matteucci G, Missiroli G F, Nichelatti E, Migliori A, Vanzi M and Pozzi G 1991 Electron holography of long-range electric and magnetic fields *J. Appl. Phys.* **69** 1835

- [56] Harscher A 1999 Electron holography of biological objects: basics and examples of applications *PhD Thesis* University Tübingen, Germany
- [57] Simon P, Adhikari R, Lichte H, Micheler G H and Langela M 2005 Electron holography and AFM studies on styrenic block copolymers and a high impact polystyrene *J. Appl. Polym. Sci.* **96** 1573–83
- [58] Simon P, Lichte H, Formanek P, Lehmann M, Huhle R, Carrillo-Cabrera W, Harscher A and Ehrlich H Electron holography of biological samples *Micron* at press
- [59] Formánek P 2005 TEM-holography on device structures in microelectronics *PhD Thesis* TU Cottbus
- [60] Simon P, Zahn D, Lichte H and Knip R 2006 Intrinsic electric dipole fields and the induction of hierarchical form developments in fluorapatite-gelatin nanocomposites: a general principle for morphogenesis of biominerals? *Angew. Chem. Int. Ed.* **45** 1911–15
- [61] Frabboni S, Matteucci G and Pozzi G 1985 Electron holographic observations of the electrostatic field associated with thin reverse-biased p-n junctions *Phys. Rev. Lett.* **55** 2196
- [62] Rau W-D, Schwander P, Baumann F H, Höppner W and Ourmazd A 1999 Two-dimensional mapping of the electrostatic potential in transistors by electron holography *Phys. Rev. Lett.* **82** 2614
- [63] McCartney M R, Gribelyuk M A, Jing Li, Ronsheim P, McMurray J S and Smith D J 2002 Quantitative analysis of one-dimensional dopant profile by electron holography *Appl. Phys. Lett.* **80** 3213–15
- [64] Twitchett A C, Dunin-Borkowski R E and Midgley P A 2002 Quantitative electron holography of biased semiconductor devices *Phys. Rev. Lett.* **88** 238302
- [65] Lenk A, Lichte H and Muehle U 2005 2D-mapping of dopant distribution in deep sub micron CMOS devices by electron holography using adapted FIB-preparation *J. Electron Microsc.* **54** 351–9
- [66] Muehle U, Lenk A, Weiland R and Lichte H 2005 Characterisation of dopants distribution using electron holography and FIB-based lift-off preparation *Microelectron. Reliab.* **45** 1558–61
- [67] Wang Z, Sasaki K, Kato N, Urata K, Hirayama T and Saka H 2001 Examination of electrostatic potential distribution across an implanted p-n junction by electron holography *J. Electron Microsc.* **50** 479–84
- [68] Formanek P and Kittler M 2005 Potential and limitations of electron holography in silicon research *Solid State Phenom.* 108–09 603–8
- [69] Twitchett A C, Dunin-Borkowski R E and Midgley P A 2002 Quantitative electron holography of biased semiconductor devices *Phys. Rev. Lett.* **88** 238302
- [70] Pfisterer H, Fuchs E and Liesk W 1962 Elektronenmikroskopische Abbildung ferroelektrischer Domänen in dünnen BaTiO₃-Schichten *Naturwissenschaften* **49** 178
- [71] Tanaka M, Kitamura N and Honjo G 1962 Electron optical studies of barium titanate single crystal films *J. Phys. Soc. Japan* **17** 1197
- [72] Zhang X, Hashimoto T and Joy D C 1992 Electron holographic study of ferroelectric domain walls *Appl. Phys. Lett.* **60** 784
- [73] Lichte H 2000 Are ferroelectric crystals blaze-gratings for electrons? *Cryst. Res. Technol.* **35** 887–98
- [74] Lichte H, Reibold M, Brand K and Lehmann M 2002 Ferroelectric electron holography *Ultramicroscopy* **93** 199–212
- [75] Matteucci G, Missiroli G F, Muccini M and Pozzi G 1992 Electron holography in the study of electrostatic fields: the case of charged microtips *Ultramicroscopy* **45** 77–83
- [76] Cumings J, Zettl A, McCartney M R and Spence J H C 2002 Electron holography of field-emitting carbon nanotubes *Phys. Rev. Lett.* **88** 056804
- [77] Tonomura A, Matsuda T, Endo J, Arai T and Mihama K 1980 Direct observation of fine structure of magnetic domain walls by electron holography *Phys. Rev. Lett.* **44** 1430–3
- [78] Heumann M, Uhlig T and Zweck J 2005 True single domain and configuration-assisted switching of submicron permalloy dots observed by electron holography *Phys. Rev. Lett.* **94** 077202
- [79] Simon P, Lichte H, Drechsel J, Formanek P, Graff A, Wahl R, Mertig M, Adhikari R, and Michler H G 2003 Electron holography of organic and biological materials *Adv. Mater.* **15** 1475–81
- [80] Dunin-Borkowski R E, Kasama T, Wei A, Tripp S L, Hÿtch M J, Snoeck E, Harrison R J and Putnis A 2004 Off-axis electron holography of magnetic nanowires and chains, rings, and planar arrays of magnetic nanoparticles *Microsc. Res. Technol.* **64** 390
- [81] Dunin-Borkowski R E, Newcomb S B, McCartney M R, Ross C A and Farhoud M 2001 *Inst. Phys. Conf. Ser.* **168** 485
- [82] Biskupek J, Kaiser U, Lichte H, Lenk A, Gemming Th, Pasold G and Witthuhn W 2005 TEM-characterization of magnetic samarium- and cobalt-rich-nanocrystals formed in hexagonal SiC *J. Magn. Magn. Mater.* **293** 924
- [83] Bonevich J E, Harada K, Matsuda T, Kasai H, Yoshida T, Pozzi G and Tonomura A 1993 Electron holography observation of vortex lattices in a superconductor *Phys. Rev. Lett.* **70** 2952
- [84] Matsuda T, Hasegawa H, Igarashi M, Kobayashi T, Naito M, Kajiyama H, Endo J, Osakabe N, Tonomura A and Aoki R 1989 Magnetic field observation of a single flux quantum by electron holographic interferometry *Phys. Rev. Lett.* **62** 2519
- [85] Stadelmann P A 1987 EMS—a software package for electron diffraction analysis and HREM image simulation in materials science *Ultramicroscopy* **21** 131
- [86] O’Keefe M A 1992 Resolution in high-resolution electron microscopy *Ultramicroscopy* **47** 282–97
- [87] Lichte H 1993 Parameters for high-resolution electron holography *Ultramicroscopy* **51** 15–20
- [88] Franke F J, Herrmann K-H and Lichte H 1988 Numerical reconstruction of the electron object wave from an electron hologram including the correction of aberrations *Scanning Microsc.* **2** (Suppl.) 59–67
- [89] Lichte H 1991 Electron image plane off-axis holography of atomic structures *Adv. Opt. Electron. Microsc.* **12** 25–91
- [90] Fu Q, Lichte H and Völkl E 1991 Correction of aberration of an electron microscope by means of electron holography *Phys. Rev. Lett.* **67** 2319–22
- [91] Lehmann M 2000 Determination and correction of the coherent wave aberration from a single off-axis electron hologram by means of a genetic algorithm *Ultramicroscopy* **85** 165–82
- [92] Lehmann M and Lichte H 1994 Interactive computerbased holographic correction of aberrations *Proc. 13th Int. Congr. on Electron Microscopy, ICEM 13 (Paris, 1994)* pp 293–4
- [93] Lichte H 1992 Electron holography: I. Can electron holography reach 0.1 nm resolution? *Ultramicroscopy* **47** 223
- [94] Ishizuka K, Tanji T, Tonomura A, Ohno T and Murayama Y 1994 Aberration correction using off-axis holography: I. Aberration assessment *Ultramicroscopy* **53** 361–70
- [95] Geiger D and Lichte H 1998 Can electron holography reveal oxygen atoms in YBa₂Cu₃O₇? *Proc. 14th Int. Congr. on Electron Microscopy (Cancun, Mexico)* vol 1 p 535
- [96] Lehmann M 2004 Influence of the elliptical illumination on acquisition and correction of coherent aberrations in

- high-resolution electron holography *Ultramicroscopy* **100** 9–23
- [97] Lehmann M and Lichte H 2005 Electron holographic material analysis at atomic dimensions *Cryst. Res. Technol.* **40** 149–60
- [98] Geiger D, Lichte H, Linck M and Lehmann M Electron holography with Cs-corrected transmission electron microscope *Microsc. Microanal.* at press
- [99] Weickenmeier A and Kohl H 1991 Computation of absorptive form factors for high-energy electron diffraction *Acta Cryst. A* **47** 590
- [100] Rother A, Reibold M, Lichte H, Leisegang T, Levin A A, Paufler P, Meyer D C, Gemming S, Chaplygin I and Seifert G 2006 X-ray investigation, high-resolution electron holography and density functional calculations of single-crystalline BaTiO₃ *Phys. Rev. B* **74** 134116
- [101] Rother A, Gemming Th and Lichte H The statistics of the thermal motion of the atoms during imaging process in transmission electron microscopy and related techniques *Ultramicroscopy* submitted
- [102] Scheerschmidt K 2003 Parameter retrieval in electron microscopy by solving an inverse scattering problem *Proc. 6th Int. Conf. Mathematical and Numerical Aspects of Wave Propagation (WAVES 2003) (Germany, 2003)* ed G C Cohen (Berlin: Springer) p 607