



ELSEVIER

Ultramicroscopy 55 (1994) 165–170

ultramicroscopy

Partially coherent and holographic contrast transfer theory

L.D. Marks, R. Plass

Science and Technology Center for Superconductivity, Department of Materials Science and Engineering, Northwestern University, Evanston, IL 60208, USA

Received 4 January 1993; in final form 27 April 1994

Abstract

A more complete form of contrast transfer theory is derived using a partial coherency approach for both electron holography and conventional high resolution electron microscopy.

1. Introduction

One of the current directions in high resolution electron microscopy is to attempt to use small electron sources such as a cold field emission tip (e.g. Refs. [1,2]) to extend the envelope. High resolution transmission electron microscopy suffers from information loss and contrast reversal at higher spatial frequencies because of the inability to completely eliminate coherent aberrations such as the objective lens defocus and spherical aberration. These distortions require correction or the use of simulations before HREM images can be fully interpreted, e.g. Ref. [3]. Two methods are currently used to extract higher resolution information from electron microscope images: through-focal series (e.g. Refs. [3–5]) and high resolution electron holography (e.g. Refs. [6–12]). Through the processes of matching experimental images with simulations at different focus values or through focal reconstruction, through-focal series allows for accurate interpretation of the information contained in HREM micrographs. Unlike HREM micrographs which only contain intensity information, a high resolution electron hologram contains information

about the amplitude and phase of the wave which is used to optically or numerically correct the coherent microscope aberrations.

Although the theory of imaging in an electron microscope (the optics rather than the diffraction) is apparently well established, in reality it is only known in detail for two cases:

- (a) For full incoherence in the condenser aperture as first analyzed by Frank [13] which has been developed into what is called a second-order contrast transfer function, e.g. Ref. [3].
- (b) For full coherence in the condenser (for STEM objective) aperture for STEM imaging. Using reciprocity, this can be translated to case (a) with a fully incoherent collector aperture.

In reality, these approximations are highly dubious for any of the newer electron microscopes which employ small sources. The reason is that the approximation of full incoherence in the condenser aperture is not even close to correct for many of these instruments. For instance, with the HF-2000 instrument which we are familiar with (cold FEG source) the illumination used for HREM imaging can be focused down to about 1 nm simply by changing the final condenser.

In this paper we will derive a more complete form of the contrast transfer theory using a partial coherency approach, see for instance Born and Wolf [14], and give some simplified forms for both HREM and electron holography. A related paper [15] looks at one approximation to these equations, the coherent case, and demonstrates quite reasonable agreement between experimental and calculated results; a second related paper [16] will look at the effect of diffraction on partial coherence.

2. Partially coherent contrast (wave) transfer theory

We will sketch here a rather complete contrast transfer analysis using a partial coherency approach. We will focus on the holographic case from which the normal case can be derived without complication. For reference, we will adopt the convention of using lower case for the pre-field optics (the combined condenser and pre-field of the objective lens before the sample) and upper case for the post-field optics (after the sample), with \mathbf{r} or \mathbf{R} in real space and \mathbf{u} or \mathbf{U} in reciprocal space. Such an approach of splitting the microscope optics above and below the specimen can help clarify the mathematics and is conceptually similar to the fashion in which a microscope is operated. However, in the more complete model detailed herein the final image is very much a function of both parts together in a non-simple fashion. The approach that we will use is based upon the idea of a statistical average of the wavefunction $\psi(\mathbf{r})$ at two different points \mathbf{r}_1 and \mathbf{r}_2 called the mutual intensity defined by (e.g. Ref. [14]):

$$\Gamma(\mathbf{r}_1, \mathbf{r}_2) = \langle \psi^*(\mathbf{r}_1)\psi(\mathbf{r}_2) \rangle, \quad (1)$$

where the symbol $\langle \rangle$ represents the appropriate statistical average. In the treatment that will be used here we are assuming quasi-monochromatic irradiation so temporal coherency effects will be omitted and only included at a later stage as an integration over defoci (representing lens and high voltage instabilities and the distribution of energies of the source). There is little additional

complication with this approach (and some mathematical simplifications) relative to a fully coherent analysis using the wavefunction $\psi(\mathbf{r})$ in real space and $\Psi(\mathbf{u})$ in reciprocal space, except that the mutual intensity in reciprocal space is related to that in real space by a double Fourier transformation (rather than a single one), e.g.

$$\begin{aligned} \Gamma(\mathbf{u}_1, \mathbf{u}_2) &= \langle \Psi^*(\mathbf{u}_1)\Psi(\mathbf{u}_2) \rangle \\ &= \iint \Gamma(\mathbf{r}_1, \mathbf{r}_2) \exp[2\pi i(\mathbf{u}_1 \cdot \mathbf{r}_1 \\ &\quad - \mathbf{u}_2 \cdot \mathbf{r}_2)] d\mathbf{r}_1 d\mathbf{r}_2. \end{aligned} \quad (2)$$

Similarly, the effects of the coherent aberration terms such as C_s is included by two multiplicative phase factors, rather than one.

We start by assuming some (spatially incoherent) source described by $s(\mathbf{r})$ which emits electrons in all directions isotropically. It is very tempting to take this as some simple disc, although this may not be accurate enough in all cases and merits further experimental investigation as will be mentioned in the discussion. Prior to the condenser aperture, the mutual intensity in reciprocal space can, with the above approximations, be taken as:

$$\Gamma(\mathbf{u}_1, \mathbf{u}_2) = \int \exp[2\pi i\mathbf{r} \cdot (\mathbf{u}_1 - \mathbf{u}_2)] s(\mathbf{r}) d\mathbf{r}. \quad (3)$$

(There appear to be some differences with different microscopes as to the exact pre-specimen optics, so this equation and some others may need to be modified for a particular instrument.) The effect of coherent aberrations in the condenser lens and objective pre-field is to introduce a spatial frequency dependent phase shift function $C(\mathbf{u})$ where:

$$C(\mathbf{u}) = \pi/\lambda(\Delta z_i \lambda^2 u^2 + \frac{1}{2} O_s \lambda^4 u^4), \quad (4)$$

and Δz_i is the illumination defocus, O_s the pre-specimen spherical aberration, and λ the wavelength. The condenser aperture restricts the range of values of u transmitted, and in the simplest model can be taken as a hard aperture of radius a defined by a function $A(u)$ where:

$$A(u) = \begin{cases} 1, & u < a, \\ 0, & u > a. \end{cases} \quad (5)$$

(Some experimental data presented elsewhere [15] suggests that it is more realistic to incorporate a soft edge rather than this sharp edge.) Using a prime notation to denote the wave after these aberrations are included, the illumination conditions are describable by:

$$\Gamma'(\mathbf{u}_1, \mathbf{u}_2) = A(\mathbf{u}_1)A(\mathbf{u}_2)\Gamma(\mathbf{u}_1, \mathbf{u}_2) \times \exp[i(C(\mathbf{u}_1) - C(\mathbf{u}_2))]. \quad (6)$$

An important caveat should be mentioned here. We are dealing with a very general case, and must therefore also mention the change in angles associated with large defoci of the pre-field. Strictly speaking, if L is some distance in the plane of the condenser aperture, the corresponding value of u in reciprocal space is $L/(f + \Delta z_1)$ where f is the focal length. (Note that in this sign convention for defoci, Schertzer defocus is negative.)

To continue, in general (but see discussion), the diffraction process is linear, and, for an incident plane wave of $\delta(\mathbf{u})$, can be represented after the specimen by a coherent wave $\psi(\mathbf{R})$ of form:

$$\psi(\mathbf{R}) = \sum_{\mathbf{g}} \phi_{\mathbf{g}}(\mathbf{u}) \exp[2\pi i(\mathbf{u} + \mathbf{g}) \cdot \mathbf{R}], \quad (7)$$

where for a crystal \mathbf{g} is a reciprocal lattice vector; for the more general case the sum can be extended to a full integration. Therefore, the post-specimen mutual intensity can be written as:

$$\Gamma(U_1 - \mathbf{g}, U_2 - \mathbf{h}) = \Gamma'(\mathbf{u}_1, \mathbf{u}_2)\phi_{\mathbf{g}}^*(\mathbf{u}_1)\phi_{\mathbf{h}}(\mathbf{u}_2) \quad (8)$$

and the axes for \mathbf{u} and U are the same. (We are conventionally using U below the sample.) We note here that the dependence of the diffraction upon the incident beam direction is commonly omitted. This is valid for very thin samples, but not for samples of approximately 10 nm or thicker (dependent upon the convergence semi-angle) [3,16]. The post-specimen aberrations arise in the form of a phase shift $\chi(U)$ given by:

$$\chi(U) = \pi/\lambda(\Delta z \lambda^2 U^2 + \frac{1}{2}C_s \lambda^4 U^4) \quad (9)$$

and Δz is the post-field defocus, C_s the post-specimen spherical aberration to give (again using a prime notation):

$$\Gamma'(U_1, U_2) = \Gamma(U_1, U_2) \exp[i(\chi(U_1) - \chi(U_2))]. \quad (10)$$

For reference, the image intensity for conventional HREM is:

$$I(\mathbf{R}) = \iint \Gamma'(U_1, U_2) \exp[-2\pi i \mathbf{R} \cdot (U_1 - U_2)] \times dU_1 dU_2. \quad (11)$$

(It is possible to take the Fourier transform of $I(\mathbf{R})$ in Eq. (11) above which corresponds to the effective source commonly used with conventional LaB₆ sources. However, we should caution that this can drop some of the terms which are important here.)

For holography, the waves now pass through a Möllenstedt type electron biprism [17] mounted in the selected area aperture plane. The biprism is assumed to have a convergent operation (the waves are deflected toward each other), a clean, thin conducting filament placed symmetrically between the plates, and stable electrical and mechanical operation for ideal imaging. The biprism is also assumed to have a sufficient positive voltage applied to the filament that the interference field is much larger than the first Fresnel fringe spacing. Under these conditions, with an incident wave $\psi(\mathbf{R})$ after emerging from the biprism we can write the modified wave $\psi'(\mathbf{R})$ as [18]:

$$\psi'(\mathbf{R}) = \psi(\mathbf{R} - \mathbf{D}) \exp(\pi i \alpha x) H(x) + \psi(\mathbf{R} + \mathbf{D}) \exp(-\pi i \alpha x) H(-x), \quad (12)$$

where $H(x)$ is the Heaviside function ($H(x) = 1$ for $x > 0$, $H(x) = 0$ for $x < 0$) and the biprism is located at $x = 0$ and \mathbf{D} is along the x -axis. Letting $\Gamma'(\mathbf{R}_1, \mathbf{R}_2)$ be the double Fourier transform of $\Gamma'(U_1, U_2)$, after the biprism (with a double prime notation)

$$\begin{aligned} \Gamma''(\mathbf{R}_1, \mathbf{R}_2) &= \Gamma'(\mathbf{R}_1 + \mathbf{D}, \mathbf{R}_2 + \mathbf{D}), \quad x_1, x_2 < 0; \\ &= \Gamma'(\mathbf{R}_1 - \mathbf{D}, \mathbf{R}_2 - \mathbf{D}), \quad x_1, x_2 > 0; \\ &= \Gamma'(\mathbf{R}_1 - \mathbf{D}, \mathbf{R}_2 + \mathbf{D}) \exp[-\pi i \alpha (x_1 + x_2)], \\ &\quad x_1 > 0, x_2 < 0; \\ &= \Gamma'(\mathbf{R}_1 + \mathbf{D}, \mathbf{R}_2 - \mathbf{D}) \exp[\pi i \alpha (x_1 + x_2)], \\ &\quad x_1 < 0, x_2 > 0. \end{aligned} \quad (13)$$

The final intensity is obtained by setting (very carefully) $\mathbf{R} = \mathbf{R}_1 = \mathbf{R}_2$ on the left of this equation (e.g. $\mathbf{R} = \mathbf{R}_1 - \mathbf{D} = \mathbf{R}_2 + \mathbf{D}$ for $x_1 > 0$, $x_2 < 0$). Let us consider that the crystal is located in the half plane $x > 0$. The first two terms in this equation correspond to, respectively, the conventional image of the vacuum wave and a high resolution image of the crystal. The last two will be the two sidebands of interest here. Considering the case when $x_1 > 0$, $x_2 < 0$, we have:

$$I''(\mathbf{R}, \mathbf{R}) = \sum_g \exp[-2\pi i(\mathbf{g} \cdot \mathbf{R} + \alpha x)] P(\mathbf{g}), \quad (14)$$

where $P(\mathbf{g})$ can be considered as the contrast (wave) transfer term and is given by:

$$P(\mathbf{g}) = \iint A(u_1) A(u_2) \left\{ \int s(\mathbf{r}) \times \exp[2\pi i \mathbf{r} \cdot (\mathbf{u}_1 - \mathbf{u}_2)] d\mathbf{r} \right\} \phi_g(\mathbf{u}_2) \times \exp\{i(C(\mathbf{u}_1) - C(\mathbf{u}_2)) + i(\chi(\mathbf{u}_1) - \chi(\mathbf{u}_2 + \mathbf{g})) + 2\pi i[\mathbf{u}_1 \cdot (\mathbf{R} + \mathbf{D}) - \mathbf{u}_2 \cdot (\mathbf{R} - \mathbf{D})]\} d\mathbf{u}_1 d\mathbf{u}_2, \quad (15)$$

where we also need (later) to include an integration over defocus values for the chromatic aberrations. So far we have made no approximations, except with the descriptions of the source and biprism. We will start by neglecting the variation with angle of the diffraction process, which is legitimate for a very thin crystal, and setting this term to unity. Next, we consider two different cases. First, we note that in general the illumination will be well defocussed so that $C(\mathbf{u})$ is rapidly varying, appropriate for a stationary phase approximation. For \mathbf{u}_1 the condition for a stationary phase is:

$$\partial/\partial \mathbf{u}_1 [2\pi(\mathbf{r} + \mathbf{R} + \mathbf{D}) \cdot \mathbf{u}_1 + C(\mathbf{u}_1) + \chi(\mathbf{u}_1)] = \mathbf{0}, \quad (16)$$

with a similar condition for \mathbf{u}_2 . These reduce (for large pre-field defocus) to:

$$\mathbf{u}_1 = \mathbf{u}_2 = (\mathbf{r} + \mathbf{R} + \mathbf{D})/\lambda \Delta z_i. \quad (17)$$

Assuming that the illumination is centered on the region of interest, the condenser aperture drops

out of the equations as does most of the pre-field effects and we have (neglecting terms which to first order simply scale the intensity levels):

$$P(\mathbf{g}) = \iint \delta(\mathbf{u} - (\mathbf{r} + \mathbf{R} + \mathbf{D})/\lambda \Delta z_i) s(\mathbf{r}) \times \exp[i(\chi(\mathbf{u}) - \chi(\mathbf{u} + \mathbf{g})) + 4\pi i \mathbf{u} \cdot \mathbf{D}] d\mathbf{u} d\mathbf{r} \quad (18)$$

$$= \int s(\lambda \Delta z_i \mathbf{u} - \mathbf{R} - \mathbf{D}) \times \exp[i(\chi(\mathbf{u}) - \chi(\mathbf{u} + \mathbf{g})) + 4\pi i \mathbf{u} \cdot \mathbf{D}] d\mathbf{u}, \quad (19)$$

or, alternatively,

$$P(\mathbf{g}) = \lambda \Delta z_i \int s(\lambda \Delta z_i \mathbf{v}) \exp[i(\chi(\mathbf{v} + \mathbf{w}) - \chi(\mathbf{v} + \mathbf{g} + \mathbf{w})) + 4\pi i(\mathbf{v} + \mathbf{w}) \cdot \mathbf{D}] d\mathbf{v}, \quad (20)$$

where

$$\mathbf{w} = (\mathbf{R} + \mathbf{D})/\lambda \Delta z_i. \quad (21)$$

We will break for a moment to discuss, in particular, Eqs. (18)–(21). What we have here is local tilt of $(\mathbf{R} + \mathbf{D})/\lambda \Delta z_i$ across the field of view and the integration over the source size is effectively leading to an integration over tilt in an incoherent fashion. This makes sense from a ray-optical viewpoint, see Fig. 1, and in fact the stationary phase approximation is a ray-optical approach. When the probe is focused ($\Delta z_i = 0$) the integration is effectively over all the tilts so we re-establish the standard incoherently filled aperture result [3].

Using a first-order expansion for the integration over \mathbf{u} , then

$$P(\mathbf{g}) = \exp(-i\chi(\mathbf{g})) E(\mathbf{g}), \quad (22)$$

where the envelope term $E(\mathbf{g})$ is given by:

$$E(\mathbf{g}) = \int s(\lambda \Delta z_i \mathbf{u} - \mathbf{R} - \mathbf{D}) \times \exp[2\pi i \mathbf{u} \cdot (\frac{1}{2}\pi \nabla \chi(\mathbf{g}) + 2\mathbf{D})] d\mathbf{u}, \quad (23)$$

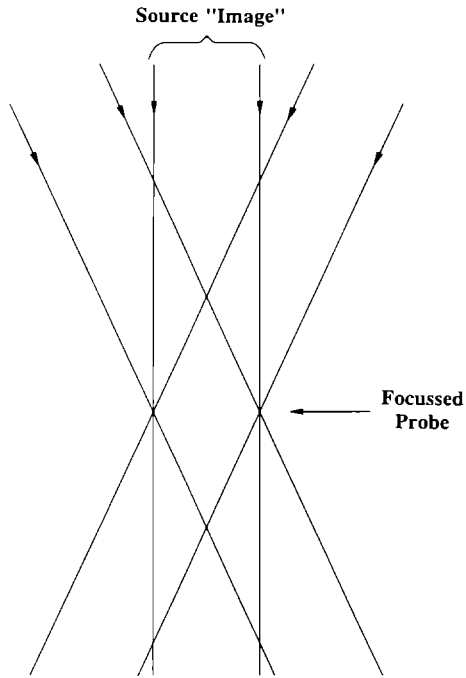


Fig. 1. Ray-optical interpretation of the convergence with somewhat coherent illumination. We can think of the illumination as leading to a set of "images" of source, each at a different angle as sketched in the figure.

and taking a simple approximation for the source such as a (normalized) disc of radius β :

$$E(\mathbf{g}) = J_1(\pi\beta|\frac{1}{2}\pi\nabla\chi(\mathbf{g}) + 2\mathbf{D}|) / \pi\beta|\frac{1}{2}\pi\nabla\chi(\mathbf{g}) + 2\mathbf{D}| \times \exp[2\pi i(\mathbf{R} - \mathbf{D}) \cdot (\frac{1}{2}\pi\nabla\chi(\mathbf{g}) + 2\mathbf{D})], \quad (24)$$

For reference, in this limit the conventional image contrast transfer can be written as:

$$I(\mathbf{r}) = \sum_{\mathbf{g}} \exp(-2\pi i\mathbf{g} \cdot \mathbf{R}) P(\mathbf{g}), \quad (25)$$

where

$$P(\mathbf{g}) = \sum \phi_{\mathbf{g}-\mathbf{h}}^* \phi_{\mathbf{h}} \exp[i(\chi(\mathbf{g}-\mathbf{h}) - \chi(\mathbf{h}))] \times E(\mathbf{g}-\mathbf{h}, \mathbf{h}) \exp[i\mathbf{R} \cdot (\nabla\chi(\mathbf{g}-\mathbf{h}) - \nabla\chi(\mathbf{h}))] \quad (26)$$

with

$$E(\mathbf{g}-\mathbf{h}, \mathbf{h}) = 2J_1[(\beta/2)|\nabla\chi(\mathbf{g}-\mathbf{h}) - \nabla\chi(\mathbf{h})|] / \beta|\nabla\chi(\mathbf{g}-\mathbf{h}) - \nabla\chi(\mathbf{h})|. \quad (27)$$

(Note that all these envelope terms need to be multiplied by the standard linear (for holography) or second-order chromatic envelopes.)

The second limiting case we can consider is when the convergence aperture is small, the illumination is very incoherent and the pre-field is focused. (With coherent illumination, i.e. a small source size, and focused illumination the illuminated region is too small to be relevant here.) In this case we can reduce:

$$P(\mathbf{g}) = \int A(\mathbf{u}) \exp[i(\chi(\mathbf{u}) - \chi(\mathbf{u}-\mathbf{g})) + 4\pi i\mathbf{u} \cdot \mathbf{D}] d\mathbf{u}, \quad (28)$$

which gives the same result as above excepting the omission of the positional dependent term and β needs to be changed to the radius of the condenser aperture. (This result was previously given by Coene [19].) This gives the same result as in Eq. (26) above for a second-order analysis with the omission of the \mathbf{R} dependent term on the right, matching the original work with an incoherently filled condenser aperture [3].

3. Discussion

The form that we have derived herein is quite general and relevant to small source electron microscopes where there is substantial non-isoplanarity of the imaging. The form also seems to handle moderately well the effects of defocusing the electron beam above the sample, which can only be roughly included with the conventional, fully incoherent condenser aperture approach. For the two limits that we have looked at, the holographic case has essentially the same resolution limits as conventional high resolution since it has the same type of envelope function; however, holography has an intrinsic lower signal-to-noise ratio since only part of the total recorded wave contains holographic information. We will leave to the future which approach will turn out to be more experimentally useful.

There are several phenomena which are appropriate to discuss since, at least to these authors, they raise some questions. First, there is

the issue of the exact description of the source. With a small, physical size of the source, it becomes unclear whether every point in the source should be considered as a separate, incoherent emitter; the mean free path of the conduction electrons at room temperature is about 1 nm. More detailed analyses of the source optics (excluding coherency in the emission) have been published, e.g. Refs. [20–24], and the simple source form used here is quite questionable. For instance, a recent analysis of the transverse coherence of a nanotip has suggested that it is perfectly coherent [23]. (Remember that the source form comes in as an effective incoherent convergence for relatively defocused illumination.)

The other point that merits attention is the implicit approximation that the diffraction process is linear in character. The fact that this is an approximation may be a little surprising, but is rigorously correct [16]. The point is that incoherent scattering (e.g. thermal diffuse and inelastic) does not appear to separate in a reasonable fashion in certain cases. However, this appears only to be the case for small, partially coherent probes and provided that we are dealing with collimated illumination such as for HREM this does not seem to be an issue.

A third point to note is that the illumination optics for various microscopes are somewhat different. For instance, on the HF-2000 the condenser aperture is before the final condenser lens; in our conventional HREM (H-9000) it is after the condenser. From an experimental viewpoint it is not at all clear to us whether either illumination system is really optimized (or has enough lenses) for HREM imaging with partially coherent illumination.

Finally, it should be noted that the approximations that we have given are only approximations, and the full forms such as Eq. (11) in the text should be used to describe the imaging process. At least numerically this is not much harder than the conventional approaches used; multi-dimensional fast Fourier transform programs are avail-

able. To what extent one has to go to these extremes is an issue for future research.

Acknowledgements

The authors would like to thank Dr. William M.J. Coene for his helpful correspondence. This work was supported by the Science and Technology Center for Superconductivity under grant No. DMR 91-20000.

References

- [1] A.V. Crewe, D.N. Eggenberger, J. Wall and L.M. Welter, *Rev. Sci. Instrum.* 39 (1968) 576.
- [2] A. Tonomura, T. Matsuda, J. Endo, H. Todokoro and T. Komoda, *J. Electron Microsc.* 28 (1979) 1.
- [3] P. Self and M.A. O'Keefe, in: *High-Resolution Transmission Electron Microscopy*, Eds. P. Busek, J. Cowley and L. Eyring (Oxford University Press, Oxford, 1988).
- [4] E.J. Kirkland, *Ultramicroscopy* 15 (1984) 151.
- [5] W. Coene, G. Janssen, M. Op de Beeck and D. Van Dyck, *Phys. Rev. Lett.* 69 (1992) 3743.
- [6] D. Gabor, *Proc. Roy. Soc. (London) Ser. A* 197 (1949) 454.
- [7] A. Tonomura, T. Matsuda and J. Endo, *Jpn. J. Appl. Phys.* 18 (1979) 9.
- [8] H. Lichte, *Ultramicroscopy* 20 (1986) 293.
- [9] E. Volkl and H. Lichte, *Ultramicroscopy* 32 (1990) 177.
- [10] T. Kawasaki, Q.X. Ru, T. Matasuda, Y. Bando and A. Tonomura, *Jpn. J. Appl. Phys. Lett.* 30 (1991) 1830.
- [11] J.M. Cowley, *Ultramicroscopy* 41 (1992) 335.
- [12] H. Lichte, *Ultramicroscopy* 47 (1992) 223.
- [13] J. Frank, *Optik* 38 (1973) 519.
- [14] M. Born and E. Wolf, *Principles of Optics* (Pergamon, New York, 1959) ch. X.
- [15] B. Storey and L.D. Marks, *Ultramicroscopy*, submitted.
- [16] L.D. Marks, *Ultramicroscopy*, submitted.
- [17] G. Möllenstedt and H. Düker, *Naturwissenschaften* 42 (1954) 41.
- [18] G.F. Missiroli, G. Pozzi and U. Valdrè, *J. Phys. E* 14 (1981) 649.
- [19] W.M.J. Coene, private communication (September 1992) (this unpublished analysis is dated 9 July 1991).
- [20] T.E. Everhart, *J. Appl. Phys.* 38 (1967) 4944.
- [21] J.C. Wiesner and T.E. Everhart, *J. Appl. Phys.* 44 (1973) 2140.
- [22] J.R.A. Cleaver, *Int. J. Electron.* 38 (1975) 513.
- [23] N.D. Lang, *Phys. Rev. Lett.* 63 (1989) 1499.
- [24] M.R. Scheinfein, W. Qian and J.C.H. Spence, *J. Appl. Phys.* 73 (1993) 2057.