Ultramicroscopy 49 (1993) 66–80
North-Holland

ultramicroscopy

Ultimate resolution and information in electron microscopy
II. The information limit of transmission electron microscopes

A.F. de Jong [a] and D. Van Dyck [b]

[a] *Philips Research Laboratories WY2, P.O. Box 80000, 5600 JA Eindhoven, The Netherlands*
[b] *University of Antwerp (RUCA), Groenenborgerlaan 171, B-2020 Antwerp, Belgium*

Received 5 August 1992; at Editorial Office 11 September 1992

In this paper the factors affecting the information limit of transmission electron microscopes are described in a general framework. Separate information limits are given for the influence of chromatic aberration and beam convergence, for specimen drift and vibration and for the detection of the image with a limited number of resolved pixels. Their relative importance is discussed and constraints are derived relating all factors to the information limit as given by the chromatic aberration. By taking into account the brightness of the source and the noise explicitly, it has been possible to find optima for the exposure time and the convergence angle of the illumination, yielding the best information limit which can be achieved with a given microscope. The higher brightness of a field-emission gun improves the information limit because a smaller angle of beam convergence can be used, but also because the signal-to-noise ratio is improved.

1. Introduction

In part I of our paper [1] we have described some general ideas about resolution in electron microscopy. In the framework of information theory, the electron microscope can be considered as an information channel which transmits information from the object to the user. The quality of the channel is then given by its capability to transfer information with a high bandwidth. In transmission electron microscopy (TEM), the maximum spatial frequency $G_{\max}$ which is transmitted with sufficient intensity is usually referred to as the inverse of the information limit ρ_i. The information limit is determined by several incoherent effects such as limited spatial and temporal coherence of the electron beam and mechanical instabilities. It signifies the smallest detail for which information can be retrieved by some more or less advanced restoration procedure. The resolution number which is most often quoted for TEMs is the point resolution ρ_s. This resolution represents the smallest detail which can be directly interpreted under special conditions (Scherzer focus and a weak phase object). The point resolution is only related to the spherical aberration coefficient C_s and the wavelength λ of the microscope and is given by $\rho_s = 0.65(C_s\lambda^3)^{0.25}$ (see, e.g., Spence [2]). The microscope transfers information beyond the point resolution, but the interpretation is not straightforward because of the rapid oscillations of the transfer function of the microscope. The information in the resolution range between point resolution and information limit may be retrieved by image reconstruction procedures such as off-axis holography and focal-series restoration [3,4]. These procedures now become feasible because of the availability of high-resolution microscopes equipped with a field-emission gun (FEG) and a slow-scan CCD camera. Thus the relation between information limit and various microscope parameters becomes an important issue.

To obtain the best resolution it is important to balance the various resolution-limiting factors. In this paper we will use one general but simple

framework to describe the most important factors affecting the information limit. To consider the imaging properties of the microscope independently of the specimen, we will adopt the weak phase-object approximation (WPO) for the electron scattering in the object. If dynamical scattering becomes important and non-linear effects have to be included into the imaging process, it becomes much more difficult to define an information limit independently of the object. The WPO approximation is only appropriate for very thin objects, which may seem a severe limitation. Still, in the case of phase retrieval using holography the theory developed here is exact, when applied to the information limit in the reconstructed wave function. In the case of focus-series image reconstruction, the same is true within the limits of a quasi-coherent imaging scheme [2].

An overview of all the factors affecting the resolution of the TEM has been given by Glaeser in 1979 [5]. Here, we will include not only the effects of limited spatial and temporal coherence of the electron beam, but also mechanical instabilities such as specimen drift and vibration, and the limited size of the detector. It is important that all factors are described properly, with one set of consistent equations. In section 2 all factors limiting the bandwidth of the microscope are described and an information limit is derived for each factor separately. In section 3 the relative importance of all factors is discussed. This leads to a set of simple rules that relate the relevant envelope parameters to the desired information limit. In section 4 the influence of noise is taken into account explicitly. By considering the minimum dose needed to obtain a certain resolution (Rose limit [6]), estimates may be given for the actual signal-to-noise ratio and information limit that may be reached. Relations are derived for the optimum exposure time and convergence angle of the illuminating electron beam, for a given set of microscope parameters.

2. Image formation and envelope functions

When the object can be described as a weak phase object, the image intensity can be written in the linear approximation as [2]:

$$I(\boldsymbol{R}) = \mathrm{FT}^{-1}\{I(\boldsymbol{G})\}, \tag{1}$$

$$I(\boldsymbol{G}) = \delta(\boldsymbol{G}) - 2\sigma V(\boldsymbol{G}) E(\boldsymbol{G}) \sin[2\pi\chi(\boldsymbol{G})], \tag{2}$$

where $\boldsymbol{R}$ and $\boldsymbol{G}$ are two-dimensional vectors in real and reciprocal space, respectively. $I(\boldsymbol{G})$ is the complex image spectrum, $V(\boldsymbol{G})$ the Fourier transform of the potential of the object and σ the interaction constant. The transfer function of the microscope consists of a damping function $E(\boldsymbol{G})$ and a phase part $\chi(\boldsymbol{G})$, which are both even functions of G in the absence of astigmatism and misalignment. The phase part of the transfer function has the form

$$\chi(\boldsymbol{G}) = \tfrac{1}{4} C_s \lambda^3 G^4 + \tfrac{1}{2} \lambda \epsilon G^2, \tag{3}$$

with C_s the spherical aberration constant, ϵ the focus (underfocus is negative) and λ the electron wavelength.

We will consider five different effects that attenuate the amplitude of the transfer function $E(\boldsymbol{G})$. They have in common that they model small variances in the image contrast which are added incoherently. In general, they affect the higher spatial frequencies more severely than the lower frequencies, resulting in a cut-off frequency $G_{\max}$ where the power spectrum of the image drops below the noise level. The information resolution ρ_i of the microscope, which is the inverse of the cut-off frequency, is defined as the point where the product of the signal S (more accurately the contrast) and the envelope E drops below the noise N. If we denote the signal-to-noise ratio S/N as s, we have $E(1/\rho_i) = 1/s$. The product of the five separate envelope functions to be described hereafter has to remain above $1/s$:

$$E(\boldsymbol{G}) = E_c(G) E_s(G) E_d(\boldsymbol{G}) E_D(\boldsymbol{G}) E_u(\boldsymbol{G}) \geq 1/s. \tag{4}$$

We will first describe all envelope functions and derive information limits (cut-off frequencies) for all effects separately. In te next section, their interrelationship will be discussed. Here s is treated as a constant, its dependence on other

parameters such as integrated electron dose and specimen contrast is considered in section 4.

2.1. Chromatic aberration envelope function $E_c(G)$

The envelope function for the chromatic aberration reads [2,7]:

$$E_c(G) = \exp\left[-\tfrac{1}{2}(\pi\Delta\lambda)^2 G^4\right], \tag{5}$$

with the defocus spread Δ given by:

$$\Delta = C_c\sqrt{4\left(\frac{(\Delta I)^2}{I^2}\right) + \left(\frac{(\Delta E)^2}{V^2}\right) + \left(\frac{(\Delta V)^2}{V^2}\right)}, \tag{6}$$

where C_c is the chromatic aberration. $\Delta V/V$ and $\Delta I/I$ are the instabilities of the high-voltage supply and the objective-lens current, respectively, while $\Delta E/V$ is the intrinsic energy spread of the electron gun, all defined here as the standard (or root-mean-square) deviations [2].

Neglecting the other envelope functions in eq. (4), we obtain an information limit only due to the chromatic aberration:

$$\rho_c = \left(\frac{\pi\lambda\Delta}{\sqrt{2\ln s}}\right)^{1/2}. \tag{7}$$

If we take e^{-2} as the cut-off value, i.e. $\ln s = 2$, we obtain the information limit usually quoted: $\rho_c = \sqrt{\pi\Delta\lambda/2}$.

2.2. Source-dependent envelope function $E_s(G)$

In this paper we will consider a Gaussian-shaped source, which means that the probe on the specimen is slightly defocused. This distribution then leads to an envelope function [2,8]:

$$E_s(G) = \exp\left[-\left(\frac{\pi\alpha}{\lambda}\right)^2\left(\frac{\partial\chi(G)}{\partial G}\right)^2\right]$$

$$= \exp\left[-\left(\frac{\pi\alpha}{\lambda}\right)^2\left(C_s\lambda^3G^3 + \lambda\epsilon G\right)^2\right], \tag{8}$$

where α is the half-angle of the distribution, defined as the value where it reduces to $1/e$ of its value at the origin. Under appropriate conditions, this value may be measured from densitometer traces through diffraction spots of a known specimen. Note, however, that this treatment of the effects of beam convergence or spatial coherence assumes that the illuminating aperture is incoherently filled, which is not always the case for FEG microscopes. In that case, the measured angular width of the source distribution underestimates the degree of spatial coherence.

Clearly, this envelope function is focus-dependent. For negative focus values (underfocus), the C_s term is partly compensated by the focus term, extending the value of G_{max}. The focus can always be adapted in such a way that the expression between brackets in eq. (8) is zero for $G = G_{max}$ (e.g. for testing the information limit). However, when α is too large (roughly above 1 mrad), much of the contrast in the intermediate frequencies is lost, which would limit practical microscopy. In fig. 1 several source-dependent envelopes are given as a function of defocus, illustrating this effect. To extend the information limit to 0.1 nm a half-angle of around 0.1 mrad is probably necessary, which can be obtained using a FEG.

However, we can find a focus value where the envelope function is maximum for the whole *range* of spatial frequencies involved, which means that we have to minimize the argument of the exponential in eq. (8). If we choose the range of frequencies involved as $0 < G < G_{max}$, we obtain the "optimum" focus:

$$\epsilon_{opt} = -\tfrac{3}{4}C_s\lambda^2G_{max}^2 = -\frac{3}{4}\frac{C_s\lambda^2}{\rho_i^2}. \tag{9}$$

This focus value, much further under focus than Scherzer focus, is the same focus which minimizes the effect of delocalization, as reported by Lichte [9], and which will also be used in the next section. As illustrated in fig. 1, at the optimum focus the "dip" in the envelope function at the intermediate frequency $G_{max}/2$ reaches exactly the same value as the value at G_{max}. Close to the

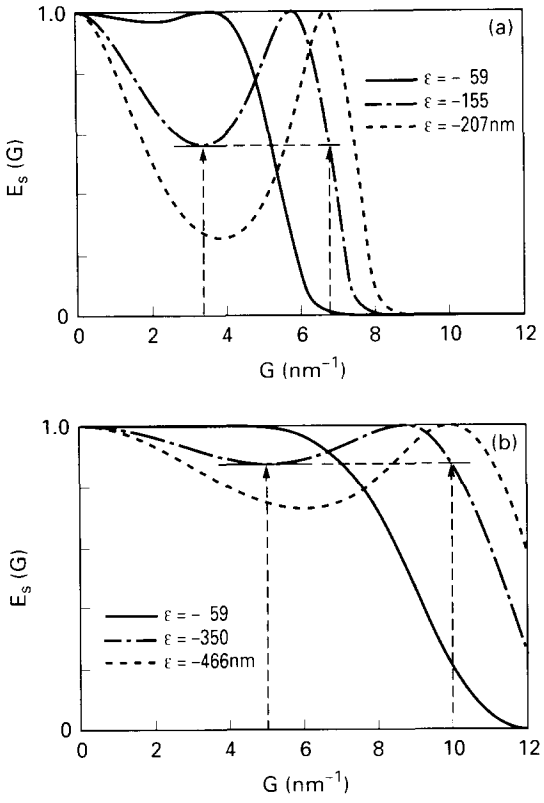


Fig. 1. Envelope function $E_s(G)$ due to a finite angle of beam convergence. The envelope is given for two half-angles of beam convergence: (a) $\alpha = 0.7$ mrad (appropriate for a LaB_6 gun) and (b) $\alpha = 0.1$ mrad (appropriate for a FEG), each with three values of defocus: Scherzer focus, the "optimum" focus for image reconstruction (for $\rho_i = 0.15$ and 0.1 nm, respectively) and the focus which ensures that the source envelope at the information limit is unity. The accelerating voltage chosen was 300 kV, $C_s = 1.2$ mm. Arrows indicate that at the optimum focus the envelope function at $G_{max}/2$ has the same height as at G_{max}.

maximum spatial frequency, we may thus write the source envelope function as

$$E_s(G) = \exp\left[-\left(\frac{\pi\alpha}{\lambda}\right)^2 a^2 \left(C_s\lambda^3 G^3\right)^2\right], \qquad (10)$$

where a varies from 0.25 for the optimum focus given in eq. (9) to 1 for the case of Gaussian focus. The information limit due to the limited coherence of the source only is obtained as:

$$\rho_\alpha = \left(\frac{6\pi\alpha a}{\lambda\sqrt{\ln s}}\rho_s^4\right)^{1/3}. \qquad (11)$$

Via this equation, the information limit is influenced by the point resolution ρ_s of the microscope in the case when the limited spatial coherence of the source is really a resolution-limiting factor. Furthermore, ρ_α is proportional to $a^{1/3}$, which means that in passing from Gaussian focus to the optimum focus, the information limit is improved by about 40%.

2.3. Sample drift envelope function $E_d(G)$

To describe the effect of sample movement on the resolution, we will follow Frank [10]. The image intensity is dependent on time and the recorded image is the time integral of the moving image, which leads to damping envelopes in reciprocal space in a similar way as for focus spread (section 2.1). For the case of a linear sample drift v with a total drift $\boldsymbol{d} = v t_{exp}$, we find for the drift envelope function:

$$E_d(\boldsymbol{G}) = \mathrm{sinc}(\pi \boldsymbol{G}\cdot\boldsymbol{d}) \approx \exp\left[-\tfrac{1}{6}(\pi \boldsymbol{G}\cdot\boldsymbol{d})^2\right]. \qquad (12)$$

$E_d(\boldsymbol{G})$ diminishes appreciably even for small values of G in directions parallel to the drift vector, as illustrated in fig. 2a. For values of the drift which are reasonably small compared with the desired information limit, the sinc function may be approximated by a Gaussian, which is correct up to third order in Gd. From this equation, the information limit due to sample drift may be derived:

$$\rho_d = \frac{\pi d}{\sqrt{6\ln s}}. \qquad (13)$$

2.4. Specimen vibration envelope function $E_u(G)$

For the calculation of the specimen vibration envelope function a vibration of the specimen is assumed with frequencies which are large compared with the inverse exposure time (typically above 1 Hz). When the vibration has an amplitude u (i.e. a mean-square deviation of $u^2/2$), the vibration envelope is [10]:

$$E_u(\boldsymbol{G}) = J_0(2\pi \boldsymbol{G}\cdot\boldsymbol{u}) \approx \exp\left[-(\pi \boldsymbol{G}\cdot\boldsymbol{u})^2\right]. \qquad (14)$$

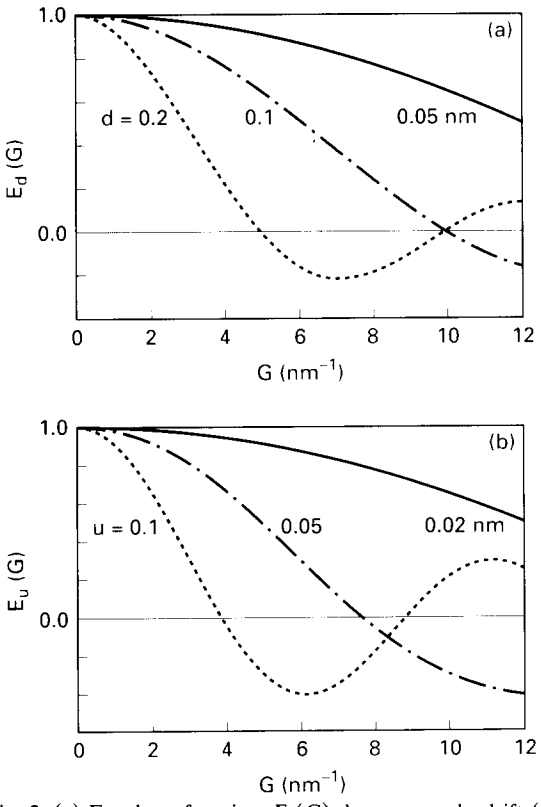


Fig. 2. (a) Envelope function $E_d(G)$ due to sample drift (in the direction of the drift), for three drift values and (b) $E_u(G)$ due to sample vibration (in the direction of the vibration), for three vibration amplitudes.

Here $J_0(x)$ is a Bessel function of zero order, which for small vibration amplitudes may be approximated by a Gaussian [11]. This envelope also already attenuates the contrast for rather small frequencies (fig. 2b). The information limit due to specimen vibration is then given as:

$$\rho_u = \frac{\pi u}{\sqrt{\ln s}}. \tag{15}$$

Compared with the information limit due to specimen drift, the tolerances on the sample vibration are much more severe. This stems basically from the definition of the vibration amplitude: in one cycle, the specimen "drifts" over a distance twice the amplitude.

2.5. Detector envelope function $E_D(\boldsymbol{G})$

The limited number of resolved image points available for detection has up to now not been described as an incoherent damping envelope function affecting the information limit. This becomes necessary when using slow-scan CCD cameras with e.g. 1000^2 pixels, which is needed when further phase retrieval is aimed at. In this section, an envelope function accounting for this effect will be derived. The envelope function is the result of two conflicting effects: delocalization of the information in the imaging process, and a finite pixel size.

The image area which is actually captured may be defined by the circular window function

$$\begin{aligned} w(\boldsymbol{R}) &= 1 \quad \text{for } |\boldsymbol{R}| \le R_w, \\ w(\boldsymbol{R}) &= 0 \quad \text{elsewhere.} \end{aligned} \tag{16}$$

The detected part of the image is the original image multiplied by $w(\boldsymbol{R})$, which yields for the image spectrum $I_s(\boldsymbol{G})$:

$$I_s(\boldsymbol{G}) = \int I(\boldsymbol{G}')W(\boldsymbol{G}-\boldsymbol{G}')\,\mathrm{d}\boldsymbol{G}', \tag{17}$$

with $I(\boldsymbol{G})$ as given by eq. (2), and $W(\boldsymbol{G})$ a sharply peaked first-order Bessel function [11], the width of which is proportional to $(\pi R_w)^{-1}$. The damping envelopes $E(\boldsymbol{G})$ and the specimen potential vary only slowly over $W(\boldsymbol{G})$, so that these functions may be taken as constant. With eq. (2) we then obtain:

$$\begin{aligned} I_s(\boldsymbol{G}) = \delta(\boldsymbol{G}) - 2\sigma V(\boldsymbol{G})E(\boldsymbol{G}) \int \sin(2\pi\chi(\boldsymbol{G}')) \\ \times W(\boldsymbol{G}-\boldsymbol{G}')\,\mathrm{d}\boldsymbol{G}'. \end{aligned} \tag{18}$$

Changing variables $\boldsymbol{G}' = \boldsymbol{G} - \boldsymbol{H}$ and expanding χ to first order around $\boldsymbol{G}$ yields:

$$\begin{aligned} I_s(\boldsymbol{G}) = \delta(\boldsymbol{G}) - 2\sigma V(\boldsymbol{G})E(\boldsymbol{G}) \sin(2\pi\chi(\boldsymbol{G})) \\ \times \int \cos\left(2\pi \frac{\partial\chi}{\partial \boldsymbol{G}} \cdot \boldsymbol{H}\right) W(\boldsymbol{H})\,\mathrm{d}\boldsymbol{H}. \end{aligned} \tag{19}$$

The integrand of this expression is just the Fourier transform of the window function itself, but now in reciprocal space and with $\partial\chi/\partial\boldsymbol{G}$ as the argu-

ment. The passband for transferred spatial frequencies is given by the requirement

$$\left|\frac{\partial\chi(\boldsymbol{G})}{\partial\boldsymbol{G}}\right| = \left|C_s\lambda^3\boldsymbol{G}^3 + \epsilon\lambda\boldsymbol{G}\right| \geq R_w, \tag{20}$$

or, using the same notation as in eq. (10), at the maximum spatial frequency:

$$aC_s\lambda^3G_{max}^3 = R_w. \tag{21}$$

Eqs. (20) and (21) simply mean that the delocalization of the information in the image, given by $\partial\chi/\partial\boldsymbol{G}$, must be smaller than the half-width of the detector. The delocalization depends on the spatial frequency involved and is large when the phase-transfer function oscillates rapidly. Clearly, the same focus that optimizes the source-dependent envelope function also optimizes the image localization, as both are dependent on $\partial\chi/\partial\boldsymbol{G}$ (see also ref. [9]). Eqs. (20) and (21) do not represent an attenuating envelope function, but the window function of the detector transforms to a function with a sharp cut-off frequency in reciprocal space, dictated by the size of the detector.

On the other hand, the size of the detector is limited by the number N and the size D of the pixels: $2R_w = ND$. Assuming that the scintillator causes a Gaussian-shaped spreading of the high-energy electrons [12], the sampling may be described in reciprocal space by an envelope function

$$E_D(\boldsymbol{G}) = \exp\left[-\tfrac{1}{2}\pi^2G^2D^2\right] = \exp\left[-2\pi^2\frac{\left(aC_s\lambda^3G^4\right)^2}{N^2}\right]. \tag{22}$$

In the second part of the equation we have used eq. (21). The quantum detection efficiency is not considered here. The sampling as described here is just according to the Nyquist sampling theorem. In practice, however, an additional oversampling by a factor two seems necessary [9], thus halving the number of independent image pixels. Including the point resolution, we then obtain for the information limit due to the limited detector size (i.e. number of independent image pixels):

$$\rho_D = \left(\frac{12\sqrt{2}\,\pi a}{N\sqrt{\ln s}}\rho_s^4\right)^{1/4}. \tag{23}$$

Thus the detector limits the resolution because of the limited number of resolved image points. This information limit is also dependent on the point resolution of the microscope, in a similar manner as the source-dependent limit (eq. (11)). It should be realized that in practice not all the points of the detector can be used in eq. (23), as this would leave a usable field-of-view of only one pixel. Rather, the value of N should be determined from a realistic border that, although imaged on the detector, cannot be used for valid image interpretation because of image delocalization. For 1000^2 CCDs, this could be a border of 200 points (yielding $N = 400$), leaving a field-of-view of 600^2 points in the reconstructed wave. For off-axis holography an extra magnification of about three is necessary because of the necessity to sample the interference fringes which have a spatial frequency of $3G_{max}$ [9]. This reduces the number of independent pixels in the reconstructed wave by a factor of three. In that case, a realistic border could be 75 independent pixels ($N = 150$), leaving a field-of-view of 180 pixels in the reconstructed wave. Clearly, this resolution limiting factor is much more severe for holography than for normal HREM or focus-variation type image reconstruction.

3. Relative importance of resolution-limiting factors

In the previous section the resolution limits due to several factors have been derived. They are coupled to each other by eq. (4). If the relevant microscope and detection parameters were known, the various resolution limits could be calculated. However, several parameters are generally not known, or can be chosen more or less freely by the microscopist. In this section the relative importance of the factors will be discussed, leading to some simple constraints on parameters such as sample drift and vibration, convergence angle and detector size. All resolution limits given in eqs. (7), (11), (13), (15) and (23) are dependent on the signal-to-noise ratio. In this section we will make the assumption that the contribution of all separate envelope func-

tions to the total attenuation of the wave function is the same. At the resolution limit, we will assume that all envelopes have a value of e^{-1}, which is equivalent to $\ln s = 1$.

Let us first consider the factors which govern the information limit imposed by electrical instabilities combined with the constant for chromatic aberration, as given by eqs. (5)–(7). For LaB_6 guns, the total variance of the energy of the incoming electron beam (ΔV and ΔE together) is in general dominating over the instability of the objective-lens current, even despite the factor of four with which $\Delta I/I$ has to be multiplied. Under conditions used for HREM, the total energy spread is usually between 1.5 and 2.0 eV (FWHH). Assuming that the current has been stabilized to the 1 ppm level (rms value), this energy spread leads to a defocus spread of about 5 nm at 300 kV, for a LaB_6 instrument. For a Schottky Zr/W FEG microscope, a total energy spread has been reported of about 0.8–0.9 eV [13]. This yields a defocus spread of about 3.7 nm (for 300 kV), leading to an improvement of the (chromatic) information limit of about 20%. In this case, the instability of the objective-lens current is the dominating term. The importance of $\Delta I/I$ becomes clear when considering the situation where the total energy spread is only determined by the theoretical minimum of the intrinsic energy spread of the FEG (0.3 eV for a cold FEG, 0.4 eV for a Schottky FEG). In that case the defocus spread is reduced to 3.2 nm, which leads to an improvement of the information limit of only 5%. However, these values cannot be attained in practice because of the Boersch effect and high-tension instabilities. The small theoretical difference between the energy spread of the cold FEG and the Schottky FEG is totally negligible for the information limit.

Table 1
Maximum convergence angle α and minimum number of unusable image points N as a function of the ratio between point resolution and (chromatic) aberration limit, ρ_s/ρ_c, for Gaussian focus (ϵ_0) and optimum focus (ϵ_{opt}) (for the calculation of α from eq. (29), it has been assumed that $\lambda = 0.011\rho_s$)

ρ_s/ρ_c	α (mrad)		N (pix)	
	ϵ_0	ϵ_{opt}	ϵ_0	ϵ_{opt}
1	0.58	2.3	53	13
1.5	0.17	0.69	270	67
2	0.07	0.30	853	213
2.5	0.04	0.15	2082	521
3	0.02	0.09	4320	1080

Because the resolution limit due to chromatic aberration is a given microscope parameter and is usually quoted as "the" information limit, we will use ρ_c as a "yardstick" and impose the condition that all other resolution limiting factors must be better. Thus, from $\rho_\alpha < \rho_c$ and using eq. (11) we find as a constraint for the convergence angle:

$$\alpha \leq \frac{\lambda}{6\pi a \rho_s}\left(\frac{\rho_c}{\rho_s}\right)^3. \tag{24}$$

Likewise, we can find from eq. (23) the constraint for the minimum number of resolved points at the detector:

$$N \geq 12\sqrt{2}\,\pi a\left(\frac{\rho_s}{\rho_c}\right)^4. \tag{25}$$

Both α and N depend on the ratio ρ_s/ρ_c, i.e. on how far beyond the point resolution of the microscope it is desired to retrieve the information. Several values for the maximum convergence angle and minimum number of image points are given in table 1. For the convergence angle we have assumed that in eq. (24) $\lambda = 0.011\rho_s$ (i.e. a point resolution of 0.23 nm at 200 kV and 0.18 nm at 300 kV, which could be obtained with a C_s of about 1 mm). It must be remembered that the number N must be subtracted from the number of points available in the detector to obtain the *usable* number of resolved image points. Clearly, both the maximum angle and the number of border points depend on the focus via the parameter a, which is 0.25 for the optimum focus value defined in eq. (9). Around Scherzer focus the value of this parameters is between 0.8 and 0.9, which means an improvement in the constraints on α and N of 10%–20%. Choosing the optimum focus allows a convergence angle four times as large as the values near Gaussian focus, thus enhancing the current density considerably without losing resolution. Values of α become prohibitively small only at ratios of $\rho_s/\rho_c > 3$ (at

optimum focus). The increase of N with increasing ratio ρ_s/ρ_c is rather dramatic owing to the fourth power in eq. (30). For 1000^2 CCDs, the number of points already becomes prohibitive at values near Gaussian focus when this ratio exceeds 1.75, while at optimum focus 2.5 seems to be the limit. For off-axis holography applications, the detector size is a more serious constraint as three times as many pixels are needed. Taking the hologram at optimum focus is mandatory, but even then a ratio of about 1.75 seems to be the maximum that might be obtained with a 1000^2 CCD camera.

The constraints for specimen drift and vibration can be easily derived from eqs. (13) and (15):

$$d \leq \frac{\sqrt{6}}{\pi}\rho_c \approx 0.8\rho_c, \tag{26a}$$

$$u \leq \frac{1}{\pi}\rho_c \approx 0.3\rho_c. \tag{26b}$$

It should be remembered that the limits given above apply to the case where the attenuation due to all factors is the same. The limits given above can thus be regarded as being rather optimistic. If, on the other hand, it is demanded that the mechanical stability, the spatial coherence and the detector size hardly influence the information limit, the constraints are more severe. For instance, for an allowed increase of the information limit of 10% compared to ρ_c (i.e. $\rho_i = 1.1\rho_c$) one finds $\rho_d < 0.5\rho_c$ and $\rho_u < 0.2\rho_c$.

Two factors may be easily influenced by the microscope operator: spatial coherence via the beam convergence angle, and integrated specimen drift via the exposure time. The other two (number of detector points and specimen vibration) depend very much on the hardware. The convergence angle and the exposure time can be lowered, but only at the cost of a lower signal-to-noise ratio. This problem is explicitly addressed in the next section.

4. Influence of noise on the resolution

In the preceding sections the signal-to-noise ratio has been considered to be constant. In a more general description the resolution should be defined as the point where the product of the signal S (more accurately: the contrast) and the envelope disappears into the noise N. The S/N ratio (denoted by s) increases with the total dose incident on the specimen, determined by the current density and the exposure time. The current density increases with the square of the convergence angle α, so that $s(\alpha, t) = s_0\alpha\sqrt{t}$. Here s_0 is the normalized signal-to-noise ratio, for a convergence angle of 1 mrad and an exposure time of 1 s. So the resolution may be enhanced by longer exposure times and a larger convergence angle. On the other hand, however, the drift increases linearly with the exposure time, $d(t) = vt$, limiting the resolution. And a larger convergence angle means a strong attenuation because of limited coherence. In this section we will first estimate values for the normalized signal-to-noise ratio, and then find the (approximate) optimum exposure time and convergence angle yielding the best resolution limit.

The normalized signal-to-noise ratio is proportional to the square root of the brightness β of the electron gun (in A cm^{-2} sr^{-1}) and further depends on the detection efficiency and the specimen contrast. Rose [6] has calculated the minimum electron dose which is necessary to obtain a significant signal (i.e. above the noise) within a pixel of a certain dimension (which we will take as the information limit). By analogy, we will derive an expression for the minimum detectable contrast in an image element, which translates into a range of reasonable values for the normalized signal-to-noise ratio s_0. The image intensity in TEM is the combination of a small ripple on a fairly large background, formed by unscattered electrons (and, partly, inelastically scattered electrons). This also follows from eqs. (1) and (2), where the background is represented in reciprocal space by the delta function. In one image element of dimensions ρ_i^2, the background signal is the electron dose multiplied by the dimensions of the element. The dose depends on the brightness of the electron gun β, the exposure time and the angle of convergence used, so that for the background intensity:

$$I_0 = D\rho_i^2 = \beta\pi\alpha^2 t\rho_i^2. \tag{27}$$

We will assume that the noise is white, Poisson-distributed noise, which means that in each element it is of the order of $\sqrt{I_0}$ electrons. The image contrast is equal to DEf, where E is the envelope function at the information limit and f the relative scattering amplitude of the crystal (for that frequency). The minimum contrast that can be detected should be above the noise by some factor k which lies between 1 and 3. This yields the relation

$$DEf\rho_i^2 = k\sqrt{I_0} = k\rho_i\sqrt{D}\,, \tag{28}$$

Combining eqs. (28) and (27) with eq. (4) which relates the envelope function to the signal-to-noise ratio, the normalized initial signal-to-noise ratio can be derived (using the proper units):

$$s_0 = 443\rho_i \frac{f}{k}\sqrt{\beta}\,, \tag{29}$$

with the brightness in A cm^{-2} sr^{-1}. The performance of the microscope is now linked to the scattering inside the specimen by the factor $f = f(G_{max})$. In the WPO approximation, this factor is the relative scattering factor of the material, which for the frequencies concerned (between 7 and 10 nm^{-1}) lies between 5% and 25%, depending on the atomic number of the material. Clearly, a specimen of atoms with small atomic numbers will show less contrast, making it much more difficults to obtain the intrinsic information limit of the microscope. For crystals, the effect of dynamic channelling [1] may enhance the contrast on the atomic cores, but to what extent depends critically on the structure and the thickness of the material. For the value of k we will take 2, which means that the minimal contrast is twice as high as the standard deviation of the noise. Finally, the brightness of the electron guns at high voltages (200–300 kV) is estimated to be around 10^6 A cm^{-2} sr^{-1} for LaB_6 guns and around 10^9 A cm^{-2} sr^{-1} for (Schottky) FEGs; see for example refs. [13,14]. Substituting these values in eq. (29), we then find for LaB_6 instruments at a resolution around 0.15 nm that $\ln(s_0)$ lies between 1.2 and 2.2 (with angles in mrad). For field-emission guns at a resolution around 0.10 nm $\ln(s_0)$ lies between 4.2 and 5.2. It should be noted that when the information limit is assessed from diffractograms of an amorphous specimen, an averaging takes place in one dimension. This enhances the S/N ratio by a factor of 20–30, or about 3 points on a log scale. In the case of image restoration via focus variation, more than one image (typically 32) contribute to the restored image. The S/N of the restored image is then improved by a factor of 6, which means that $\ln(s_0) \approx 6$ for FEG instruments.

In order to make the derivation of the optimum exposure time and convergence angle clearer, we will use the dimensionless parameter $w = (\rho_i/\rho_c)^2$, where $\rho_c = \sqrt{\pi\Delta\lambda/2}$ is the usual criterion for the information limit. Substituting the expressions developed for all envelopes in eq. (4) then yields at the information limit:

$$k_N^2 + \alpha^2 k_s^2 w + 2w^2 + \left(k_d^2 t^2 + k_u^2\right)w^3 = w^4 \ln\left(s_0\alpha\sqrt{t}\right), \tag{30}$$

where $k_d = \pi v/(\rho_c\sqrt{6})$ and $k_u = \pi u/\rho_c$. The constants k_s and k_N are defined as:

$$k_s = \frac{\pi a C_s \lambda^3}{\lambda\rho_c^3} = 6\pi a\left(\frac{\rho_c}{\lambda}\right)\left(\frac{\rho_s}{\rho_c}\right)^4, \tag{31a}$$

$$k_N = \frac{\pi\sqrt{8}\,aC_s\lambda^3}{N\rho_c^4} = \frac{12\pi a\sqrt{2}}{N}\left(\frac{\rho_s}{\rho_c}\right)^4, \tag{31b}$$

which now depend on the point resolution of the microscope. Eq. (30) gives the relation between resolution, exposure time, angle of convergence and several other resolution-limiting parameters. To optimize the exposure time and the angle of beam convergence, we can differentiate eq. (30) with respect to t and α. This yields:

$$\alpha_{opt} = \frac{1}{k_s\sqrt{2}}\left(\frac{\rho_i}{\rho_c}\right)^3 = \frac{1}{6\pi a\sqrt{2}}\frac{\lambda}{\rho_s}\left(\frac{\rho_i}{\rho_s}\right)^3, \tag{32a}$$

$$t_{opt} = \frac{1}{2k_d}\left(\frac{\rho_i}{\rho_c}\right) \cong 0.39\left(\frac{\rho_i}{v}\right). \tag{32b}$$

These relations give the optimum exposure time and convergence angle as a function of the information limit that can actually be reached. Note that t_{opt} only depends on the drift rate v, in such

a way that the *total* drift during the exposure, vt_{opt}, is kept constant at $0.39\rho_i$ (compare eq. (26a)). The optimum convergence angle mainly depends on the ratio between information limit and point resolution, while the focus plays a role via the constant a which varies from 1 (Gaussian focus) to 0.25 (optimum focus). The relation between the information limit and microscope-dependent constants, including the signal-to-noise ratio, is found by substituting eqs. (32) in (30):

$$k_N^2 + 2w^2 + k_u^2 w^3 + \left(\tfrac{3}{4} - K\right)w^4 = \tfrac{7}{4}w^4 \ln w, \tag{33}$$

with

$$K = \ln\left\{\frac{s_0}{2k_s\sqrt{k_d}}\right\}.$$

The values of the microscope constants may vary widely, but in realistic cases we may estimate that k_N varies between 0.4 and 4 and k_u between 0.3 and 1, while the constant K varies between 0 and 6. Of course, w will be of the order of 1. Eq. (33) can be solved numerically using a Newton iteration scheme. The relation between ρ_i/ρ_c and the "general" constant K is given in fig. 3 for several values of k_N and k_u. Fig. 3 may be read as follows: given a set of constants, depending on drift rate, spherical aberration and signal-to-noise ratio, we may calculate K, and find from fig. 3 the best obtainable information limit ρ_i/ρ_c. From this value we may then calculate the optimum exposure time and beam-convergence angle, using eqs. (32). K may change with the normalized signal-to-noise ratio (or gun brightness), with the ratio ρ_s/ρ_c and with the drift rate. An increase of the vibration amplitude (fig. 3a) shifts the curve upwards, decreasing the resolution. Especially above the threshold $u/\rho_c = 0.3$ the resolution deteriorates rapidly. The number of border pixels in the detector only starts to influence the resolution for $N < 250$, i.e. for a 1000^2 CCD camera the field-of-view can easily be about 750 pixels. In the case that we have a magnification of 500 000 (about 0.05 nm per pixel), this leaves a field-of-view of 37.5 nm. However, for holography where the number of necessary pixels is again increased

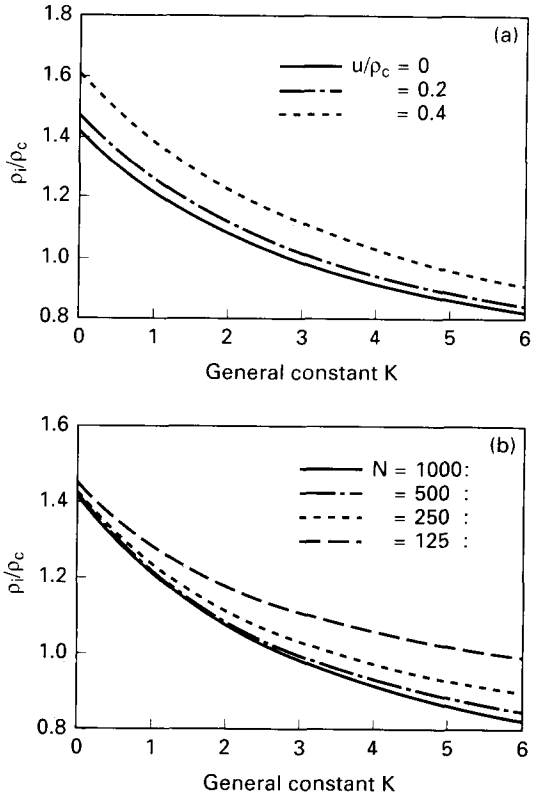
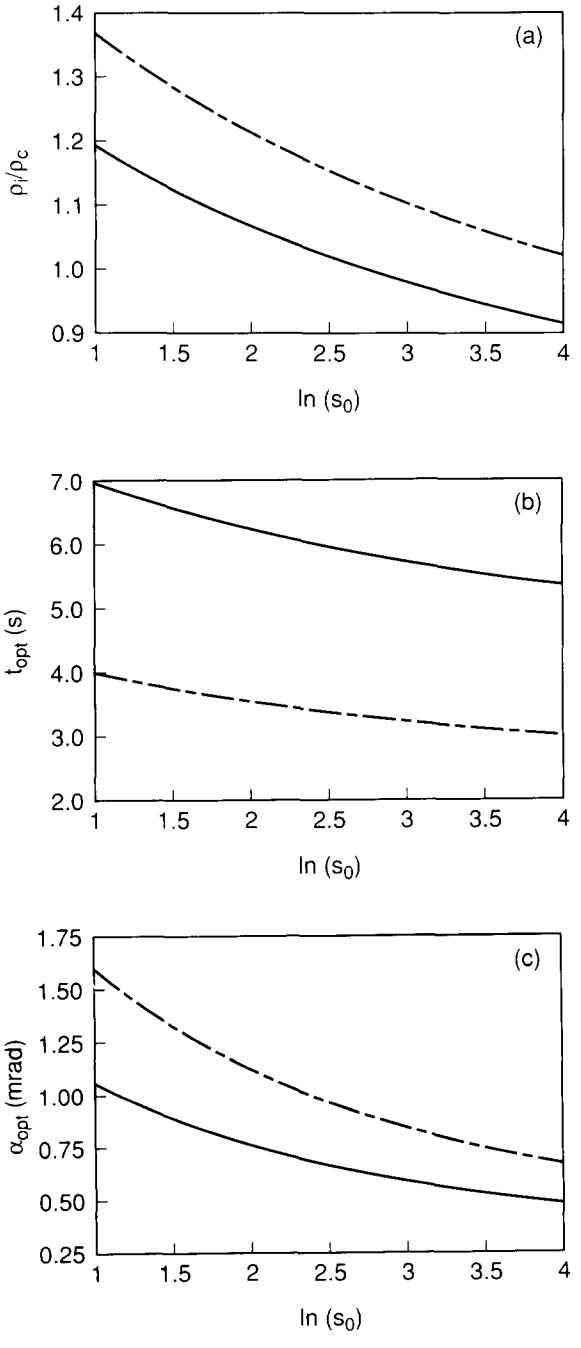


Fig. 3. Relative information limit, ρ_i/ρ_c, which can be obtained as a function of the general microscope constant K, (a) for several values of vibration amplitude u/ρ_c and (b) for several values of the number of unusable pixels N. The constant K has a logarithmic dependence on the (initial) signal-to-noise ratio.

by a factor of three (i.e. 333 pixels in the reconstructed wave function), the field-of-view in the reconstructed image is then only about 80 pixels (about 4 nm), which imposed a serious constraint.

Several typical examples are given in fig. 4 (for an LaB_6 source, with an assumed ρ_c of 0.15 nm) and fig. 5 (for a FEG, with an assumed ρ_c of 0.1 nm), as a function of the normalized signal-to-noise level s_0. All (hypothetical) microscopes operate at 300 kV and have a point resolution of 0.2 nm. For one case (for each source type) the mechanical stability has been chosen twice as good as for the other one, relative to the chromatic aberration limits. We have extended the range of s_0 somewhat further than the limits derived from the Rose limit to cover also cases

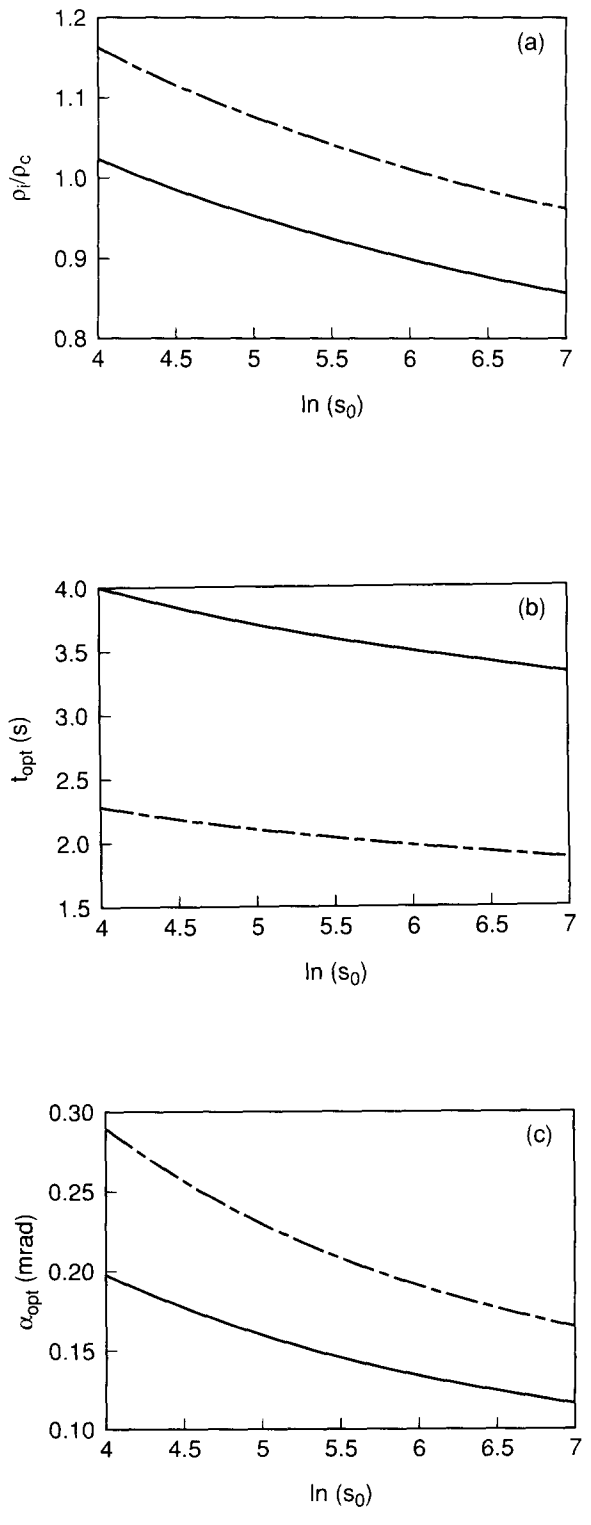
like diffractograms and focal series reconstruction. Note that several values can only be estimated, which causes an uncertainty in the values of ρ_i of approximately 10%. However, trends may be deduced more accurately from figs. 3, 4 and 5.



Concentrating first on the information limit, it may be observed that by improving the mechanical stability by a factor of two, the actual information limit may be increased by about 15%. For LaB_6 instruments the information limit lies in general above ρ_c (up to 30%), while for FEG instruments ρ_i lies around ρ_c (-10% to $+10\%$). Regarded over the total range, the information resolution limit is rather sensitive to the normalized signal-to-noise ratio (which in turn depends on the gun brightness). Within the narrower boundaries dictated by the Rose limit (1.2–2.2 for LaB_6, 4.2–5.2 for FEG), the information limit hardly changes with the signal-to-noise ratio. It then mainly depends on the vibration amplitude (relative to ρ_c). It is clear that the use of a FEG will help to improve the resolution, not only because of a smaller focus spread Δ but also because of the enhanced brightness.

The graphs of the optimum exposure time t_{opt} confirm the previous suggestions, which follow from eq. (26a). t_{opt} is not very sensitive to the normalized signal-to-noise ratio s_0 for a given drift rate. The optimum exposure time is only related to the drift rate v in such a way that the *total* drift $d = vt_{opt}$ is constant ($d/\rho_i \approx 0.4$, so that $E_d(G_{max}) \approx 0.75$). The shorter exposure times for the FEG instruments are caused by the lower values of ρ_c (0.1 instead of 0.15 nm). In general the optimum exposure times are somewhat longer than might be expected from "normal microscopy" experience. This is the case because it appears to be more important to *gain* on the S/N than to *lose* on the drift envelope itself (within certain limits). Because the total (optimum) drift is constant, the constraint on the drift rate, as calculated in the preceding sections, is

Fig. 4. Relative information limit ρ_i/ρ_c, optimum exposure time t (s) and optimum convergence angle α (mrad) as a function of the initial signal-to-noise ratio (logarithmic scale), for microscopes equipped with an LaB_6 gun (300 kV, $\rho_s = 0.2$ nm, $\rho_c = 0.15$ nm). Solid line: $u/\rho_c = 0.2$, $v = 0.01$ nm/s, $N = 600$. Dot-dashed line: $u/\rho_c = 0.4$, $v = 0.02$ nm/s, $N = 300$.



now shifted to a constraint on the signal-to-noise ratio. The drift rate may be high if it is possible to obtain a significant S/N (of the order of e^2) in the very short exposure time necessary to keep the total drift below $0.4\rho_i$. On the other hand, it is not useful to keep the total drift far below this number, as this will unncessarily decrease the S/N and thus the resolution.

The same trend may be observed for the optimum convergence angles: for LaB_6 guns they are somewhat larger than the values commonly used in HREM practice. However, the optimum convergence angles depend more on the signal-to-noise level and on the mechanical (in)stabilities, via the third-power dependence on the information limit (eq. (32a)). For the FEG, they are smaller than those of the LaB_6 guns by only a factor of about 5. This is somewhat smaller than might be expected from the enormous gain in brightness (by a factor of 10^3), which suggests an optimum convergence angle a factor of 30 smaller (order of 0.03 mrad). Thus, the gain of brightness when using a FEG lies only partly in an improved spatial coherence, it is also a gain in current density. It should be remembered, however, that we have chosen the "optimum" focus condition, which is much further under focus than usual for microscopy. Because of this focus value, the source envelope function is not as strongly attenuated around maximum spatial frequencies as would be the case at Scherzer focus (compare fig. 1).

We have also investigated the sensitivity of the actual information limit caused by small deviations from the optimum exposure time and convergence angle. In general, the information limit quickly deteriorates for very short exposure times, while it deteriorates only gradually for longer exposure times. The same general trends holds for the convergence angle: the optimum conver-

Fig. 5. Relative information limit ρ_i/ρ_c, optimum exposure time t(s) and optimum convergence angle α (mrad) as a function of the initial signal-to-noise ratio (logarithmic scale), for microscopes equipped with a FEG (300 kV, $\rho_s = 0.2$ nm, $\rho_c = 0.10$ nm). Solid line: $u/\rho_c = 0.2$, $v = 0.01$ nm/s, $N = 600$. Dot-dashed line: $u/\rho_c = 0.4$, $v = 0.02$ nm/s, $N = 300$.

gence angles are not very critical, but for smaller angles the information limit quickly deteriorates.

The results of this section lead to a strategy that may be followed if the aim is to get the best information limit in a micrograph (e.g. for resolution tests on an amorphous carbon or germanium thin film).

(1) Estimate the information limit. This is slightly above ρ_c for LaB_6 instruments and slightly below that value for FEG instruments.

(2) Using the information limit, the approximate optimum focus may be estimated from eq. (9) and the optimum convergence angle from eq. (32a). The actual convergence angle can be measured in the diffraction pattern of a sample with a known spacing. However, as already mentioned in section 2.2, for a FEG the illuminating source is not completely incoherently filled, which leads to a slight overestimation of the effect of the source envelope.

(3) From the measured drift rate, the optimum exposure time may be calculated with eq. (32b). A somewhat longer exposure time causes less resolution loss than a too short exposure time.

(4) The total electron dose is now determined by the optimum exposure time and the optimum angle of beam convergence, not by the exposure-meter and emulsion setting of the microscope! However, these may be limited to some extent by over-exposure of the plate or the CCD detector. If possible, these should be adjusted to the electron dose obtained, and not the other way around.

If we calculate the *actual* signal-to-noise ratio by including the convergence angle and the exposure time we find for the LaB_6 instruments near the Rose limit that $\ln(s) \approx 2$–3, which seems reasonable for HREM images. This means that the cut-off usually taken for the envelope function (e^{-2}) is reasonable, but not only the chromatic aberration envelopes but also the other envelopes should be taken into account. When the information limit is examined using diffractograms, an effective averaging takes place over many pixels (effectively in one dimension), so that the effective s_0 is enhanced by a factor of 20–30. This then yields $\ln(s_0) \approx 4$, which from the upper limit of fg. 4 leads to an information limit somewhat below ρ_c. For the FEG instruments, the actual signal-to-noise ratio becomes $\ln(s) \approx 3$–4, which means that the envelope function extends somewhat further than the e^{-2} limit usually taken. Compared to the situation with LaB_6 sources, the actual signal-to-noise ratio is improved by a factor of about 3. Again, if we examine the information limit using diffractograms, $\ln(s_0)$ is enhanced to about 7 in which case we will find information limits about 10% below ρ_c.

5. Conclusions

We have investigated factors limiting the information limit ρ_i of a transmission electron microscope, which is defined as the inverse of the spatial frequency G_{max} where the envelope of the contrast transfer function disappears below the noise. This limit is dependent on the signal-to-noise ratio, which in turn depends on the electron dose incident on the specimen, on the specimen contrast and on the detector efficiency. The electron dose is determined by the brightness of the gun used, the convergence angle and the exposure time. For normalized conditions of 1 s exposure time and 1 mrad half-angle of beam convergence, we use the Rose limit for the detectable contrast and estimate that the signal-to-noise ratio is between e^1 and e^2 for LaB_6 instruments and between e^3 and e^4 for FEG instruments.

Apart from the resolution-limiting factors usually taken into account (limited lateral coherence caused by a finite amount of beam convergence and focus spread caused by chromatic aberration), we have also considered specimen drift and vibration. Moreover, we have derived an envelope function related to the limited number of resolved image points on the detector (e.g. 1000^2 for a CCD). This number imposes a resolution limit because for an accurate description of the highest spatial frequencies both a small sampling interval and a large field-of-view are needed. For all incoherent effects we have derived separate information limits. Subsequently, we have found tolerances for all the parameters involved in the incoherent damping effects, relative to the chromatic information limit ρ_c. Both the maximum allowable angle of beam convergence and the

minimum number of image points needed depend markedly on the ratio between point resolution and information limit. Working under conditions of optimum focus as defined by Lichte [9], the beam convergence angle does not impose a limit on the information. However, for CCD detectors with 1000^2 pixels, the detector envelope becomes limiting when $\rho_s/\rho_i > 2.5$. For off-axis holography applications the detector is a much more important limiting factor already at $\rho_s/\rho_i > 1.75$. The maximum allowable vibration amplitude u and the specimen drift d during the exposure are linearly related to the desired information limit: $u_{max} \approx 0.2\rho_i$, $d_{max} \approx 0.5\rho_i$. If these thresholds are exceeded, specimen drift and vibration become important resolution-limiting factors.

From a consideration of the signal-to-noise ratio, which varies with the angle of beam convergence and with the exposure time, it has been possible to derive the *optimum* values for exposure time and convergence angle that yield the best information limit. The optimum exposure time is only related to the drift rate in such a way that the *total* drift is $d = 0.4\rho_i$. The optimum convergence angle is related to the point resolution of the microscope and is somewhat more dependent on the S/N, but typical values are 0.7–1.3 mrad for LaB_6 instruments and 0.15–0.25 mrad for FEG microscopes (both at 300 kV with 0.2 nm point resolution). For both the exposure time and the convergence angle the calculated values are somewhat higher than those commonly used in practical HREM. In general, it seems more important to gain on the signal-to-noise ratio than to lose on the drift or the beam convergence envelope.

Using these optimum values, the best obtainable information limit may be calculated. With realistic values for specimen vibration and drift rate, we then find an actual information limit which for LaB_6 instruments lies above ρ_c by about 15% but for FEG instruments is very close to ρ_c. When the information limit is experimentally assessed using diffractograms of an amorphous object, the image is usually averaged (by computer or by the eye) in the angular direction. This effectively enhances the signal-to-noise ratio, yielding an apparent information limit about 15% better than that obtained from the Rose limit. In the special case of image restoration via focus variation, more than one image (typically 32) contribute to the restored image. The signal-to-noise ratio of the restored image is then improved by a factor 6 which may improve the information limit in the restored wave function by about 10%.

Comparing LaB_6 with FEG microscopes in this optimization procedure, we find that the increased brightness of the FEG is only *partly* used for a better spatial coherence (smaller convergence angle). An important improvement on the resolution is obtained just from a better signal-to-noise ratio. If we assume that the brightness of the FEG is better than that of the LaB_6 by a factor of 10^3, it seems favourable to use a factor 25 to obtain a better spatial coherence (α smaller by a factor of 5). Part of the improved brightness is needed to maintain enough counting statistics at a smaller information limit: lowering ρ_i from 0.15 to 0.1 nm means an enhancement in the needed dose of 2.25. The remaining part of the brightness (a factor of about 20) is then used to decrease the (relative) noise level, so that $\ln(S/N)$ is enhanced by about 1.5. The cut-off value of e^{-2} usually taken to indicate the information limit of the transfer function must be extended to $e^{-3.5}$ for FEG instruments.

The above analysis applies to single micrographs and in the framework of a linear imaging approximation. For most crystalline specimens, where non-linear imaging effects become more important, it may be that some crystal spacings below the information limit are nonetheless visible because of secondary interference. Still, the above analysis may be applied to the reconstructed wave function obtained by holography or, within the limits of a quasi-coherent imaging scheme, with focus-series restoration.

References

[1] D. Van Dyck and A.F. de Jong, Ultramicroscopy 47 (1992) 266.

[2] J.C.H. Spence, Experimental High-Resolution Electron Microscopy, 2nd ed. (Oxford University Press, New York, 1988).
[3] H. Lichte, Adv. Opt. Electron Microsc. 12 (1991) 25.
[4] W. Coene, A. Janssen, M. Op de Beeck and D. Van Dyck, Phys. Rev. Lett. 69 (1992) 3743.
[5] R.M. Glaeser, J. Microscopy 17 (1977) 77.
[6] A. Rose, Adv. Electron. Electron Phys. 1 (1948) 133.
[7] P.L. Fejes, Acta Cryst. A 33 (1977) 109.
[8] J. Frank, Optik 38 (1973) 519.
[9] H. Lichte, Ultramicroscopy 38 (1991) 13.
[10] J. Frank, Optik 30 (1969) 171.
[11] M. Abramowitz and A. Stegun, Handbook of Mathematical Functions (Dover, New York, 1970) ch. 9.
[12] I. Daberkow, K.H. Herrmann, L. Liu and W.D. Rau, Ultramicroscopy 38 (1991) 215.
[13] P.M. Mul, B.J.H. Bormans and M.T. Otten, Philips Electron Opt. Bull. 130 (1991) 53.
[14] J.T. Whitney, Trans. Roy. Microsc. Soc. 1 (1991) 43.