Quantification of the Information Limit of Transmission Electron Microscopes

J. Barthel and A. Thust

Institute of Solid State Research, Forschungszentrum Jülich, D-52425 Jülich, Germany
(Received 17 July 2008; published 12 November 2008)

The resolving power of high-resolution transmission electron microscopes is characterized by the information limit, which reflects the size of the smallest object detail observable with a particular instrument. We introduce a highly accurate measurement method for the information limit, which is suitable for modern aberration-corrected electron microscopes. An experimental comparison with the traditionally applied Young's fringe method yields severe discrepancies and confirms theoretical considerations according to which the Young's fringe method does not reveal the information limit.

High-resolution transmission electron microscopes of the latest generation offer the possibility to study the atomic configuration of solid-state objects with a resolution of around 0.08 nm. While a lot of effort is put into improving the instrumental resolution, practical methods for the quantification of this key parameter date from the 1970s and have even come to criticism due to recently performed experiments with 80 kV accelerating voltage yielding an unexpectedly good resolution [1]. The availability of a practicable method for the accurate quantification of the obtained resolution would therefore be highly beneficial in this field.

The resolving power of a high-resolution transmission electron microscope is ultimately limited by the degree of temporal coherence available for the imaging process. A fundamental benchmark parameter, which reflects the effect of the partial temporal coherence, and which is commonly used to characterize the performance of a high-resolution electron microscope, is the information limit. The information limit d_{info} corresponds to the size of the smallest object detail that can be imaged by the electron microscope, and is defined by the spatial frequency $g_{info} = 1/d_{info}$ where contrast damping due to the partial temporal coherence drops to a value of $1/e^2 \approx 13.5\%$ compared with coherent contrast [2]. The frequency dependent damping effect due to the partial temporal coherence is caused by the energy spread of the electron source, by fluctuations of the accelerating voltage, and by objective-lens current fluctuations. All three phenomena have a defocusing effect in the image plane and can therefore be merged together into a net distribution of focal values acting during the imaging process instead of a single sharp focus value. A normalized Gaussian function is commonly assumed as a realistic focal distribution [2–4]. We adopt here the notation as used in Ref. [4], where the defocus-spread parameter Δ denotes the $1/e$-half-width of the Gaussian function. Based on this notation, the Fourier components of the linear image intensity are damped by the temporal coherence envelope $E_t(g)$, with

$$E_t(g) = \exp[-(\pi\Delta\lambda/2)^2 g^4], \qquad (1)$$

where λ is the electron wavelength and g denotes an image frequency. From the condition $E_t(g_{info}) = 1/e^2$ one obtains the following relation for the information limit:

$$d_{info} = (\pi^2\Delta^2\lambda^2/8)^{1/4}. \qquad (2)$$

The Young's fringe method has been used for a long time to determine the information limit in high-resolution transmission electron microscopy [5]. The Young's fringe resolution test is based on the assessment of the power spectrum of an image taken from an amorphous, isotropically scattering object, which is often called a *diffractogram*. Such a diffractogram is used to identify the highest image frequency that can be clearly discerned from detection noise. To distinguish the signal content from the detection noise, two separate images from the same object area are superimposed with a slight real-space displacement, either directly during the experiment by double exposure, or *a posteriori* using digital methods. Because
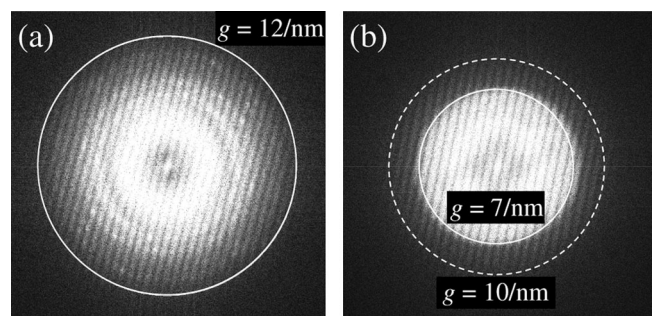

FIG. 1. (a) Young's fringe pattern recorded from thin amorphous tantalum using a FEI Titan 80–300 electron microscope operated at an accelerating voltage of 300 kV. The resulting Young's fringes extend up to a spatial frequency of 12/nm, corresponding to a 0.8 Å spacing in the images. (b) Shortcoming of the Young's fringe method: despite the intentional limitation of the input object spectrum to 7/nm by an objective aperture, the resulting Young's fringes are observed up to 10/nm in the output image spectrum.

of the mutual real-space displacement, a sinusoidal Young's fringe pattern appears in the diffractogram of the superimposed images, which helps to discern between transferred signal and detection noise [Fig. 1(a)].

A pragmatic approach to assess the information limit, which is often used due to the lack of a feasible alternative, is to simply equate g_{info} with g_{max}, where the latter frequency is the highest detectable signal frequency in a Young's fringe pattern. However, this approach is flawed due to a multitude of reasons: (i) A high-resolution image is composed of linear and nonlinear contributions. Whereas the linear image components have the same frequency content as the object spectrum, the nonlinear components are not directly related to the object spectrum and can extend up to twice the highest object frequency [2]. Since nonlinear contrast cannot be distinguished from linear contrast in a Young's fringe pattern, the highest observed image frequency g_{max} cannot be linked to the highest transferred object frequency g_{info}, and consequently it cannot be linked to the information limit. An incontrovertible experimental example is shown in Fig. 1(b), where the highest detected image frequencies extend beyond the highest transferred object frequencies determined by an intentionally inserted objective aperture. (ii) The highest detected image frequency g_{max} is determined by a multitude of equally important effects and not by the partial temporal coherence alone. The observed image spectrum, which in its pure form would reflect the atomic scattering function of the object, is additionally affected by many accumulating effects, such as the object thickness, absorption, modulation transfer of the detector, spatial coherence, mechanical vibration and drift, and electronic instabilities. Competing with the nonlinear frequency-doubling effect mentioned above, the latter effects all have a tendency to narrow the band of observed image frequencies, and thus to lower the highest detectable image frequency g_{max}. (iii) The definition of g_{info} as the frequency where the temporal coherence envelope reaches 13.5% of the unattenuated signal cannot be applied to a Young's fringe pattern since the unattenuated signal is unknown. The Young's fringe method thus lacks any quantitative calibration.

From the above-mentioned points (i)–(iii), it is clear that the traditional Young's fringe method is not suitable for an accurate quantification of the information limit, but provides a qualitative estimate at best. A quantitative measurement procedure for the information limit must be able to isolate the temporal coherence effect from the other attenuation effects mentioned under (ii). This is a challenge, since most of the functions involved are unknown, rotationally symmetric, and overlapping. For our new measurement principle, we take advantage of the symmetry breaking effect evoked by an intentionally applied illumination tilt. Upon tilt the two attenuation functions related to the partial temporal and the partial spatial coherence become direction dependent and therefore separable from all non-coherence-related attenuation functions, since the lat-

ter remain unaffected by the tilt. Additionally, we take advantage of a property of modern aberration-corrected microscopes, where the azimuthal anisotropy of the spatial coherence envelope, as derived by Frank [6], is negligible. For such microscopes, a systematic directional anisotropy under tilted illumination can thus be solely assigned to the partial temporal coherence, which is the phenomenon of interest.

By restricting our data processing procedure to low spatial frequencies $g \ll g_{\mathrm{info}}$, where the linear contrast contributions are strongly dominant over the nonlinear contributions, the application of a purely linear imaging formalism is still justified. Since the coherence properties of the linear contrast transfer under tilted-beam illumination are treated only rarely in the literature, e.g., in Refs. [7–9], we will now outline the theoretical concept behind our measurement principle by starting from the well-known case of axial illumination. Within the weak-phase-object approximation, which holds for thin amorphous objects, the Fourier component $I(\mathbf{g})$ of the linear image intensity for a spatial frequency vector $\mathbf{g}$ depends linearly on the Fourier component $O(\mathbf{g})$ of the object transmission function:

$$I(\mathbf{g}) = O(\mathbf{g})[T(\mathbf{g}, 0) - T(0, -\mathbf{g})], \qquad (3)$$

where the transmission cross coefficient $T(\mathbf{k}, \mathbf{k}')$ describes the complete influence of the microscope's optical transfer system on the two linear interference products between beams of wave vectors $\mathbf{k} = \mathbf{g}$ and $\mathbf{k}' = 0$, and $\mathbf{k} = 0$ and $\mathbf{k}' = -\mathbf{g}$, respectively [4]. When tilted illumination is considered, the beam tilt $\mathbf{t}$ must be added to the arguments of the transmission cross coefficient, while the arguments of $I(\mathbf{g})$ and $O(\mathbf{g})$ in Eq. (3) remain fixed at the object frequency $\mathbf{g}$. The extension of the axial case described by Eq. (3) to the more general tilted case then yields the following expression for the linear image intensity:

$$I(\mathbf{g}) = O(\mathbf{g})[T(\mathbf{g} + \mathbf{t}, \mathbf{t}) - T(\mathbf{t}, -\mathbf{g} + \mathbf{t})]. \qquad (4)$$

Since the anisotropy of the spatial coherence envelope can be neglected in the present context, we omit this envelope for simplicity and denote the transmission cross coefficient for arbitrary argument vectors $\mathbf{k}$ and $\mathbf{k}'$ as

$$T(\mathbf{k}, \mathbf{k}') = E_t(\mathbf{k}, \mathbf{k}')e^{-i[\chi(\mathbf{k}) - \chi(\mathbf{k}')]}, \qquad (5)$$

where E_t is the temporal coherence envelope and χ is the coherent aberration function. The aberration function χ represents a phase shift imposed by the optics on a beam $\mathbf{k}$ with respect to the undiffracted beam and is given in its simplest form by

$$\chi(\mathbf{k}) = 2\pi(Z\lambda\mathbf{k}^2/2 + C_S\lambda^3\mathbf{k}^4/4) \qquad (6)$$

with the objective-lens defocus Z and the spherical aberration constant C_S. For compactness, further contributions to the coherent aberration function χ, such as twofold astigmatism, axial coma, and further higher-order aberrations have been omitted. This Letter will focus on the long-

range falloff of the observed frequency spectrum produced by the temporal coherence envelope $E_t(\mathbf{k}, \mathbf{k}')$ [4], with

$$E_t(\mathbf{k}, \mathbf{k}') = \exp[-(\pi\Delta\lambda/2)^2(\mathbf{k}^2 - \mathbf{k}'^2)^2], \qquad (7)$$

which in the case of axial illumination reduces again to the formula in Eq. (1).

For tilted illumination, the damping is no longer identical for two opposite beams $\mathbf{g}$ and $-\mathbf{g}$ as in the axial case. Two different damping terms $E_t(\mathbf{g} + \mathbf{t}, \mathbf{t})$, and $E_t(\mathbf{t}, -\mathbf{g} + \mathbf{t})$, respectively, now play a role in an arbitrary image Fourier component $I(\mathbf{g})$. These terms are given by

$$E_t(\mathbf{g} + \mathbf{t}, \mathbf{t}) = e^{S(\mathbf{g},\mathbf{t})}e^{A(\mathbf{g},\mathbf{t})}, \qquad (8)$$

$$E_t(\mathbf{t}, -\mathbf{g} + \mathbf{t}) = e^{S(\mathbf{g},\mathbf{t})}e^{-A(\mathbf{g},\mathbf{t})}. \qquad (9)$$

The abbreviation $S(\mathbf{g}, \mathbf{t})$ combines all arguments of the temporal coherence envelope which are symmetric in $\mathbf{g}$, so that $S(\mathbf{g}, \mathbf{t}) = S(-\mathbf{g}, \mathbf{t})$, with

$$S(\mathbf{g}, \mathbf{t}) = -(\pi\Delta\lambda/2)^2[4(\mathbf{g} \cdot \mathbf{t})^2 + \mathbf{g}^4]. \qquad (10)$$

The abbreviation $A(\mathbf{g}, \mathbf{t})$ combines all arguments of the temporal coherence envelope which are antisymmetric in $\mathbf{g}$, so that $A(\mathbf{g}, \mathbf{t}) = -A(-\mathbf{g}, \mathbf{t})$, by

$$A(\mathbf{g}, \mathbf{t}) = -4(\pi\Delta\lambda/2)^2(\mathbf{g} \cdot \mathbf{t})\mathbf{g}^2. \qquad (11)$$

Finally, the diffractogram intensity $D(\mathbf{g}) = |I(\mathbf{g})|^2$ as derived from the linear imaging theory in the case of tilted illumination is given by

$$D(\mathbf{g}) = 4|O(\mathbf{g})|^2 e^{2S(\mathbf{g},\mathbf{t})}\{\sin^2[\chi_s(\mathbf{g} + \mathbf{t}) - \chi_s(\mathbf{t})]$$
$$+ \sinh^2[A(\mathbf{g}, \mathbf{t})]\}. \qquad (12)$$

The diffractogram intensity decreases uniformly towards larger spatial frequencies due to the decreasing power $|O(\mathbf{g})|^2$ of the object transmission function, and due to the part of the exponential damping argument of Eq. (10) that is proportional to $\mathbf{g}^4$. The term $\chi_s(\mathbf{g} + \mathbf{t})$ represents the part of the coherent aberration function $\chi(\mathbf{g} + \mathbf{t})$ which is symmetric with respect to a sign reversal of $\mathbf{g}$ and which leads to the well-known *Thon rings* observed in diffractograms [10]. The terms discussed so far are also common in the axial case.

Because of the tilted illumination, an additional damping argument depending on $(\mathbf{g} \cdot \mathbf{t})^2$ appears in Eq. (10). This term causes an anisotropy of the temporal coherence envelope which is not observed with axial illumination. Whereas the damping in directions perpendicular to the illumination tilt is the same as in the axial case, the diffractogram intensity is damped more strongly in directions parallel to the illumination tilt. Figure 2(a) displays a diffractogram recorded with a large beam tilt of 40 mrad, which exhibits Thon rings due to a strong defocus and which is affected by a strongly anisotropic temporal coherence envelope.

As can be seen from Eq. (12), a further coherence related term appears in the diffractogram intensity due to the
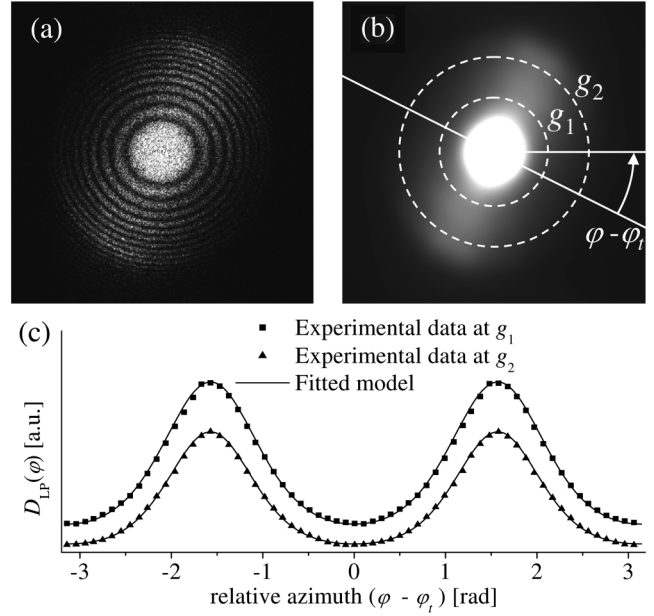

FIG. 2.   (a) Diffractogram of amorphous tantalum recorded with a FEI Titan 80–300 electron microscope operated at 300 kV. An anisotropic intensity damping envelope caused by an intentionally applied beam tilt of 40 mrad is observed. (b) Diffractogram intensity after digital removal of the Thon ring pattern. (c) Experimental data extracted along the circular paths marked in (b) together with the model-function fitted according to Eq. (13).

illumination tilt. The anisotropic term $\sinh^2[A(\mathbf{g}, \mathbf{t})]$ occurs because the two linear interference terms involving the $\mathbf{g}$ and the $-\mathbf{g}$ beams experience a different degree of damping. As a consequence, the Thon ring pattern can only be formed with an amplitude related to the weaker (stronger damped) interference. The remaining part of the stronger interference is imaged separately without producing Thon rings, comparable to holographic single sideband imaging [11].

The strength of the anisotropic effects caused by the illumination tilt depends on the wavelength λ, on the tilt modulus $t = |\mathbf{t}|$, and on the defocus-spread parameter Δ. This dependency is exploited here to measure the defocus-spread parameter. The separation of anisotropic tilt-dependent contributions to the diffractogram intensity from rotationally symmetric contributions is achieved by evaluating the diffractogram data along circular paths with the center at the spatial frequency $\mathbf{g} = (0, 0)$. On such paths, any rotationally symmetric function like atomic scattering functions, the $\mathbf{g}^4$-dependent part of the temporal coherence envelope, and the modulation transfer function of the detector are constant. A second measure for signal separation concerns anisotropic signal contributions which are constant in time and are not caused by the applied beam tilt. Such unwanted directional modifications of the diffractogram intensity can be caused by an object drift during exposure time or by a strong directional confinement of the amorphous object region. By considering a series of dif-

fractograms recorded under different azimuthal beam-tilt directions instead of one single diffractogram, and by transforming all recorded diffractograms into a coordinate system which is at rest with respect to the beam tilt, such an initially steady signal varies sinusoidally over the tilt series and cancels out to a sufficient degree.

Our data processing procedure starts by removing the coherent Thon ring pattern from the diffractogram data with a low-pass frequency filter. In this way, the term $\sin^2[\chi_s(\mathbf{g} + \mathbf{t}) - \chi_s(\mathbf{t})]$, which causes a high-frequency modulation of the diffractogram intensity, only contributes to the low-pass result by its mean value $\frac{1}{2}$ as shown in Fig. 2(b). On a circular path of radius $g = |\mathbf{g}|$, the low-pass filtered diffractogram intensity $D_{LP}(\varphi)$ follows a general functional dependency according to

$$D_{LP}(\varphi) = c_0 + c_1 \exp[-\kappa \cos^2(\varphi - \varphi_t)]$$
$$\times \cosh[\kappa' \cos(\varphi - \varphi_t)], \quad (13)$$

where a polar notation for the vectors $\mathbf{g} = (g, \varphi)$ and $\mathbf{t} = (t, \varphi_t)$, and the explicit expressions of $S(\mathbf{g}, \mathbf{t})$ and $A(\mathbf{g}, \mathbf{t})$ have been used in Eq. (12). The constants c_0 and c_1 represent rotationally symmetric effects which are constant on any circular path, as discussed above. The general function of Eq. (13) is fitted with a least-squares procedure to the experimental low-pass diffractogram data extracted from several circular paths [Fig. 2(c)], yielding the parameter $\kappa = 2(\pi \Delta \lambda g t)^2 = \frac{t}{g} \kappa'$, and thereby the desired defocus-spread parameter Δ for each path. The final measurement result is obtained by averaging over all circular paths per single diffractogram and additionally over all diffractograms, which yields a massive redundancy and therefore a high statistical confidence.

Since our criticism of the Young's fringe method also concerns the neglect of nonlinear intensity contributions, we checked this issue for the present method on the basis of simulated diffractograms. We found no significant differences between the information limit determined from linearly calculated diffractograms and those calculated with the full nonlinear imaging formalism.

We applied the present method to measure the information limit of two aberration-corrected high-resolution transmission electron microscopes installed at the Jülich Ernst Ruska Centre. These microscopes are a Philips CM200 operated at 200 kV accelerating voltage, and a FEI Titan 80–300 operated at 80, 200, and 300 kV. The respective measurement results are listed in Table I in comparison to the outcome of respective Young's fringe tests. In general, the information limit determined quantitatively by our new method differs significantly from the results of the traditional Young's fringe method. The most prominent discrepancy is observed at an accelerating voltage of 80 kV. Whereas the traditional (and incorrect) interpretation of the Young's fringe method leads to a seemingly good information limit near 1.1 Å, the value

TABLE I. Information limit measured by the present method in comparison to results of the Young's fringe method for two microscopes installed at the Ernst Ruska Centre in Jülich.

Instrument	Information limit	Young's fringes
FEI Titan 80–300 (80 kV)	1.89($\pm$0.09) Å	$\approx$ 1.1 Å
FEI Titan 80–300 (200 kV)	1.11($\pm$0.01) Å	$\approx$ 1.0 Å
FEI Titan 80–300 (300 kV)	0.83($\pm$0.02) Å	$\approx$ 0.8 Å
Philips CM200 (200 kV)	1.10($\pm$0.02) Å	$\approx$ 1.3 Å

determined by the present method is 1.89 Å. We explain this drastic discrepancy by the fact that the nonlinear frequency-doubling effect is almost completely observable because the damping effects competing with the nonlinear frequency-doubling effect play a less significant role at the resolution achievable with 80 kV. The situation is different for 200 and 300 kV accelerating voltage. Here, only moderate to no differences occur between the two methods, which indicates that the above-mentioned competing effects are balanced by chance, hiding the principal weakness of the Young's fringe method. With future improvements of the information limit towards 0.5 Å again a massive discrepancy can be expected between the two methods. In contrast to the method proposed here, the Young's fringe method is then likely to yield a worse information limit than actually available because the attenuation effects not related to the coherence of the electron beam may hide a further improvement of the coherence properties.

In conclusion, complementary information on the resolution limitation of transmission electron microscopes can be obtained by the traditional Young's fringe method and the new approach presented here. The Young's fringe method reveals a qualitative kind of net resolution limit as a result of a mixing of several accumulating effects related to the object, to the microscope's optical transfer properties, to the camera, and to environmental influences. In contrast, our new approach allows us to isolate the resolution limiting effect caused by the partial temporal coherence and thereby to quantify precisely the information limit according to its theoretical definition.

[1] S. Kujawa *et al.*, Microsc. Microanal. **12**, 1470 (2006).
[2] M. A. O'Keefe, Ultramicroscopy **47**, 282 (1992).
[3] K.-J. Hanßen and L. Trepte, Optik **32**, 519 (1971).
[4] K. Ishizuka, Ultramicroscopy **5**, 55 (1980).
[5] J. Frank, Optik **44**, 379 (1976).
[6] J. Frank, Optik **38**, 519 (1973).
[7] R. H. Wade, Optik **45**, 87 (1976).
[8] R. H. Wade and W. K. Jenkins, Optik **50**, 1 (1978).
[9] P. W. Hawkes, Optik **55**, 207 (1980).
[10] F. Thon, Z. Naturforsch. **21a**, 476 (1966).
[11] M. Hohenstein, Appl. Phys. A **54**, 485 (1992).