Chapter 3
Linear Image Approximations

Abstract This chapter presents approximations for simple calculation of CTEM and STEM (BF, ADF, and confocal) images using linear image models. A linear image is the convolution of an object function and the point spread function (multiplication in Fourier transform space). This can provide a simple intuitive approach to interpreting the electron micrographs although it might not be particularly accurate.

At some level of approximation human vision is a linear convolution of some function of light intensity with a spatial resolution response function. This simple linear image model is not quantitatively precise but allows for an easy interpretation of everyday observations. Electron microscope images do not in general follow a simple linear-image model to any great precision. However, it is useful to try to find the conditions under which an electron micrograph can be interpreted as linear image of some physical property of the specimen.

In the linear image model the actual recorded image intensity $g(\mathbf{x})$ is related to the ideal image of the object $f(\mathbf{x})$ by a linear convolution of the object function with the point spread function $h(\mathbf{x})$ of the instrument:

$$\begin{aligned} g(x,y) &= f(x,y) \otimes h(x,y) = \int f(x',y')h(x-x',y-y')\mathrm{d}x'\mathrm{d}y' \\ g(\mathbf{x}) &= f(\mathbf{x}) \otimes h(\mathbf{x}) = \int f(\mathbf{x}')h(\mathbf{x}-\mathbf{x}')\mathrm{d}\mathbf{x}', \end{aligned} \tag{3.1}$$

where $\mathbf{x} = (x,y)$ is a two dimensional position vector in the image plane. The point spread function or PSF of the instrument is just the image of an isolated point in the object or specimen. The symbol $\otimes$ represents convolution. Using the Fourier convolution theorem the linear image model can also be written as a product in Fourier or reciprocal space.

$$G(\mathbf{k}) = F(\mathbf{k})H(\mathbf{k}) \tag{3.2}$$

$H(\mathbf{k})$ is the transfer function (or modulation transfer function, MTF) and is the Fourier transform of $h(\mathbf{x})$. $G(\mathbf{k})$ and $F(\mathbf{k})$ are the Fourier transforms of $g(\mathbf{x})$

E.J. Kirkland, *Advanced Computing in Electron Microscopy*,
DOI 10.1007/978-1-4419-6533-2_3,

and $f(\mathbf{x})$, respectively. Capital letters will be used to denote a Fourier transformed quantity and lower case letters will denote a real space quantity. Goodman [126] has given an overview of the linear image model in light optics.

3.1 The Weak Phase Object in Bright Field

Under restricted conditions the BF image in a CTEM may be considered as a linear image model (for example Erikson [95], Hansen [145], Lenz [224], Thon [342], Hoppe [160]). It should be noted that Scherzer [311] stated most of the features of this image model but did not take the final step of writing it as a convolution. This section will discuss image formation in the context of the CTEM although BF-CTEM and BF-STEM are equivalent in this context due to the reciprocity theorem (see Sect. 2.4).

The electrons incident on the specimen have a relatively high energy (approximately 100 keV to 1000 keV) as compared to the electrons in the specimen. If the specimen is very thin, then the incident electrons pass through the specimen with only small deviations in their paths and the effect of the specimen can be modeled as a simple transmission function $t(\mathbf{x})$. The electron wave function after passing through the specimen is:

$$\psi_t(\mathbf{x}) = t(\mathbf{x})\psi_{\text{inc}}(\mathbf{x}) \tag{3.3}$$

where $\psi_{\text{inc}}(\mathbf{x})$ is the incident wave function. In the CTEM the incident electron wave function (see Fig. 2.3 and 3.1) is approximately a plane wave of constant intensity ($\psi_{\text{inc}} \sim 1$). Later chapters will discuss the effects of specimen thickness on the transmission function (i.e., $t(\mathbf{x})$ will no longer be a simple scalar function).

The effect of the aberrations of the objective lens is to shift the phase of each frequency component by a different amount (angle and spatial frequency are proportional). If $\Psi_t(\mathbf{k})$ is the Fourier transform of $\psi_t(\mathbf{x})$ and $\Psi_i(\mathbf{k})$ is the electron wave function in the back focal plane of the objective lens then:

$$\begin{aligned} \Psi_t(\mathbf{k}) &= \text{FT}[\psi_t(\mathbf{x})] \\ \Psi_i(\mathbf{k}) &= \Psi_t(\mathbf{k})\exp[-\text{i}\chi(\mathbf{k})] = \Psi_t(\mathbf{k})H_0(\mathbf{k}) \end{aligned} \tag{3.4}$$

where FT indicates a Fourier transform. The objective lens images this wave function into a virtual image which is equivalent to an inverse Fourier transform of the earlier equation (yielding $\psi_i(\mathbf{x})$ from $\Psi_i(\mathbf{k})$). The projector lenses further magnify this virtual image. Although magnification is the primary function of the microscope this magnification can be ignored (in the math) if the image coordinates are always referred to the dimensions on the specimen.

The actual recorded image is the intensity, or square magnitude of the image wave function. Denoting the intensity in the recorded image as $g(\mathbf{x})$ yields:

$$\begin{aligned} \psi_i(\mathbf{x}) &= \text{FT}^{-1}[\Psi_i(\mathbf{k})] \\ g(\mathbf{x}) &= |\psi_i(\mathbf{x})|^2 = |\psi_t(\mathbf{x}) \otimes h_0(\mathbf{x})|^2, \end{aligned} \tag{3.5}$$

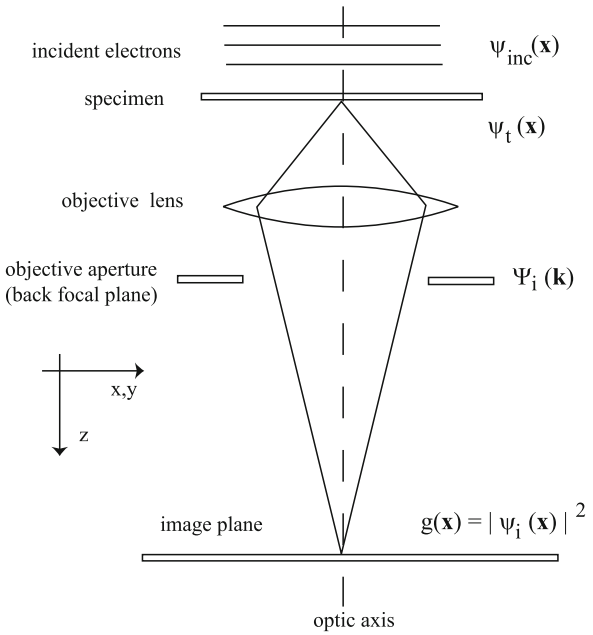

Fig. 3.1 The positions of the imaging wave functions in the CTEM column

where $h_0(\mathbf{x})$ is the complex point spread function of the objective lens (the inverse Fourier transform of $H_0(\mathbf{k})$).

In the weak phase object (WPO) approximation the specimen is assumed to be very thin and composed mainly of light atoms. The primary effect of the specimen is to produce a spatially varying phase shift in the electron wave function as it passes through the specimen. This approximation can also be referred to as the Moliere [252], WKB, or eikonal [312] approximation. If the specimen is also weak then the exponential phase factor can be expanded in a power series where only the low-order terms are important. If the incident wave function is a plane wave ($\psi_{\text{inc}} = 1$) and the specimen is a weak phase object (3.3) becomes:

$$\psi_t(\mathbf{x}) \sim t(\mathbf{x}) \sim \exp[\mathrm{i}\sigma_e v_z(\mathbf{x})] \sim 1 + \mathrm{i}\sigma_e v_z(\mathbf{x}) + \cdots \tag{3.6}$$

In later chapters $v_z(\mathbf{x})$ will be shown to be the projected atomic potential of the specimen and σ_e is an appropriate scaling factor (both are real). It is possible to use the opposite sign convention (reverse the sign of v_z in the exponent, with a corresponding change in $H_0(\mathbf{k})$) because only the square modulus of ψ will be important in the end. The result will be the same if all the signs are consistently changed, however this sign change can lead to some confusion when comparing different published versions of this theory in the literature (different authors use different sign conventions). The sign convention used here is consistent with the forward propagation of electrons (see Self et al. [315] for a discussion of the proper sign). Expanding the

expression for $g(\mathbf{x})$ (3.5) and keeping only the lowest order terms in $v_z(\mathbf{x})$ (i.e., the weak phase object approximation) yields:

$$\begin{aligned} g(\mathbf{x}) &= \left|\left[1+\mathrm{i}\sigma_e v_z(\mathbf{x})+\mathscr{O}(v_z^2)\right]\otimes h_0(\mathbf{x})\right|^2 \\ &= \left|1\otimes h_0(\mathbf{x})+\mathrm{i}\sigma_e v_z(\mathbf{x})\otimes h_0(\mathbf{x})\right|^2+\mathscr{O}(v_z^2) \end{aligned} \tag{3.7}$$

using the Fourier convolution theorem:

$$1\otimes h_0(\mathbf{x}) = \mathrm{FT}^{-1}\left[\delta(\mathbf{k})\exp[-\mathrm{i}\chi(\mathbf{k})]\right] = 1 \tag{3.8}$$

which leaves:

$$\begin{aligned} g(\mathbf{x}) &= 1+\sigma_e v_z(\mathbf{x})\otimes\left[ih_0(\mathbf{x})-ih_0^*(\mathbf{x})\right]+\mathscr{O}(v_z^2) \\ &\approx 1+2\sigma_e v_z(\mathbf{x})\otimes h_{\mathrm{WP}}(\mathbf{x}), \end{aligned} \tag{3.9}$$

where a superscript $*$ denotes complex conjugation and $h_{\mathrm{WP}}(\mathbf{x})$ is the point spread function for BF imaging in the weak phase object approximation. It is easier to state the Fourier transform of the point spread function (i.e., the transfer function) than the PSF itself:

$$\begin{aligned} G(\mathbf{k}) &= \mathrm{FT}[g(\mathbf{x})] = \delta(\mathbf{k})+2\sigma_e V_z(\mathbf{k})H_{\mathrm{WP}}(\mathbf{k}) \\ H_{\mathrm{WP}}(\mathbf{k}) &= \mathrm{FT}\left[h_{\mathrm{WP}}(\mathbf{x})\right] \\ &= \frac{i}{2}\left\{\exp[-\mathrm{i}\chi(\mathbf{k})]-\exp[\mathrm{i}\chi(\mathbf{k})]\right\} \\ &= \sin\chi(\mathbf{k}). \end{aligned} \tag{3.10}$$

An oscillatory transfer function is about the worst thing that can happen. This means that some spacings (i.e., spatial frequencies $\mathbf{k}$) in the specimen will be transmitted as white ($H_{\mathrm{WP}}(\mathbf{k})>0$) at the same time that other spacings are transmitted as black ($H_{\mathrm{WP}}(\mathbf{k})<0$) because the transfer function is both positive and negative.

The minimum of the aberration function $\chi(\mathbf{k})$ remains approximately flat for a significant region near its minimum (see Fig. 2.10). If the defocus is adjusted so than $\sin\chi(\mathbf{k})$ is also near its minimum or maximum (± 1) when $\chi(\mathbf{k})$ is flat then the transfer function will have a significant region of uniformly transferred information (i.e., the transfer function is large and constant in a band of spatial frequencies). Allowing $\sin\chi(\mathbf{k})$ to deviate slightly from unity magnitude in the passband increases the resolution slightly. Therefore, look for conditions in which:

$$\begin{aligned} 0.7 &\le |\sin\chi(k)| \le 1.0 \\ \chi(k) &= -\left[\frac{2n_D-1}{2}\right]\pi \pm \frac{\pi}{4} \\ n_D &= 1,2,3,\ldots \end{aligned} \tag{3.11}$$

The minimum of the aberration function is found by setting its derivative equal to zero. Using the dimensionless form of $\chi(K)$ (2.11) and ignoring the astigmatism (assumed small) yields:

$$\begin{aligned}\frac{\partial\chi(K)}{\partial K} = \pi(2K^3 - 2DK) &= 0\\ K^2 - D &= 0\\ K &= \sqrt{D}.\end{aligned} \tag{3.12}$$

Next substitute this expression for K into the original expression for $\chi(K)$ (2.16) and solve for defocus D to make $\sin[\chi(K)] \sim \pm 1$:

$$\begin{aligned}\chi(K) &= \pi(0.5(\sqrt{D})^4 - D(\sqrt{D})^2)\\ &= \pi(0.5D^2 - D^2) = -0.5\pi D^2\\ &= -\left[\frac{2n_D - 1}{2}\right]\pi - \frac{\pi}{4}.\end{aligned} \tag{3.13}$$

The optimum defocus is therefore:

$$\begin{aligned}D &= \sqrt{2n_D - 0.5}\\ \Delta f &= \sqrt{(2n_D - 0.5)C_s\lambda}\\ n_D &= 1, 2, 3, \ldots\end{aligned} \tag{3.14}$$

The special case of $n_D = 1$ is referred to as Scherzer focus. Each positive integer value of n_D produces a band of uniformly transferred spatial frequencies (Eisenhandler and Siegel [90]). The transfer function for several values of n_D is shown in Fig. 3.2. The broad passband moves to higher spatial frequencies as n_D increases but unfortunately also gets smaller so that the resolution cannot be extended dramatically using large n_D. Eisenhandler and Siegel [90] and Hoppe [160] proposed using zone plates (in the objective aperture) to select appropriate bands in the transfer function to improve resolution. Image reconstruction from a defocus series may use these wide bands to advantage (for example Kirkland et al. [209]).

Ideally, a transfer function would be flat and have the same sign over the range of spatial frequencies that are transmitted to the image. The oscillatory nature of $H_{\mathrm{WP}}(K)$ can cause serious problems because it is not flat and changes sign. Choosing an optimum defocus produces bands of uniform sign in the transfer function but the transfer function is still oscillating. Scherzer [311] realized that it is better to limit the range of spatial frequencies so that the transfer function at least has the same sign over its allowed range. An objective aperture is placed in the back focal plane of the objective lens (see Fig. 3.1) which is conveniently the Fourier transform plane. The radius in the aperture corresponds to spatial frequency in the image. The objective aperture allows all rays within a maximum distance from the optic axis to pass. This means that the objective aperture limits the maximum spatial frequency in the image. If this maximum cutoff is made to coincide with the first zero crossing

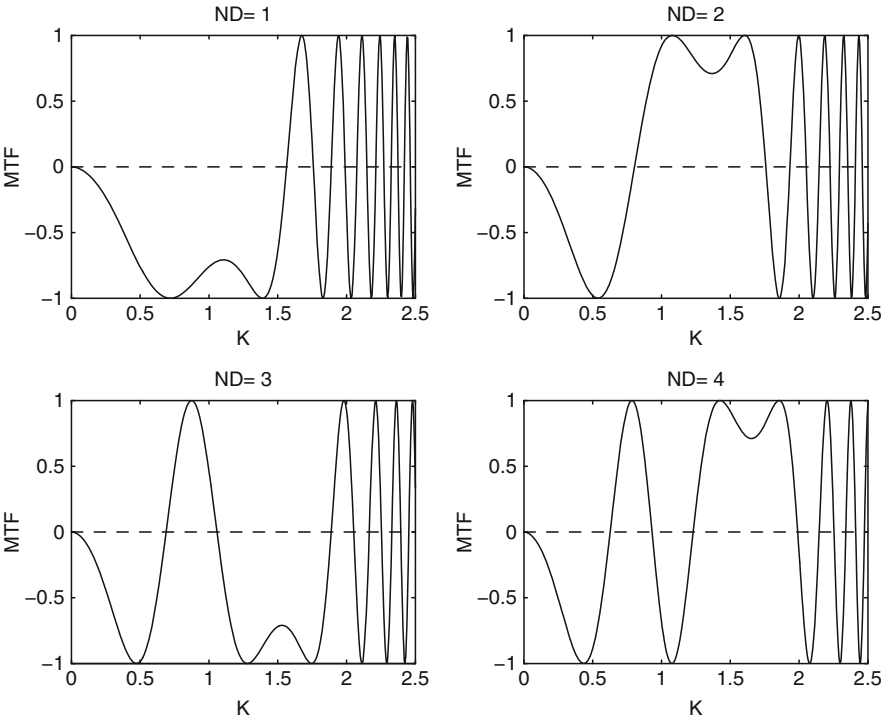

Fig. 3.2 The phase contrast transfer function H_{WP} for a weak phase object in bright field as a function of the normalized spatial frequency K for different values of the defocus index $n_d =$ 1,2,3,4. Astigmatism is assumed to be zero

of the transfer function (for Scherzer focus) then the transfer function will have the same sign over its range. The first zero crossing is found in dimensionless (normalized) form from:

$$\begin{aligned} \chi(K) = \pi(0.5K_{\max}^4 - DK_{\max}^2) &= 0 \\ 0.5K_{\max}^2 - D &= 0 \\ K_{\max} &= \sqrt{2D}. \end{aligned} \tag{3.15}$$

At Scherzer focus $D = \sqrt{1.5}$. Substituting this value of D yields:

$$\begin{aligned} K_{\max} &= \sqrt{2\sqrt{1.5}} = k_{\max}(C_s\lambda^3)^{1/4} \\ k_{\max} &= \left(\frac{6}{C_s\lambda^3}\right)^{1/4} \\ \alpha_{\max} &= \lambda k_{\max} = \left(\frac{6\lambda}{C_s}\right)^{1/4}. \end{aligned} \tag{3.16}$$

This value of $\alpha_{\max}$ is referred to as the Scherzer aperture. Together the Scherzer aperture and Scherzer focus are referred to as the Scherzer conditions. The corresponding resolution with Scherzer conditions is just:

$$d_s > \left(\frac{C_s\lambda^3}{6}\right)^{1/4} = 0.64(C_s\lambda^3)^{1/4} = 1/k_{\max}. \tag{3.17}$$

This resolution is plotted vs. spherical aberration in Fig. 3.3. This is a lower bound because the transfer function $H_{\mathrm{WP}} = 0$ at this spacing and there is no information transferred. There is some ambiguity in the choice of (3.11). Other choices will change the value of various constants presented earlier by small amounts. The differences appearing in the literature mainly reflect the allowed variation of $\sin\chi(K)$ in the pass band (0.7–1.0 chosen here).

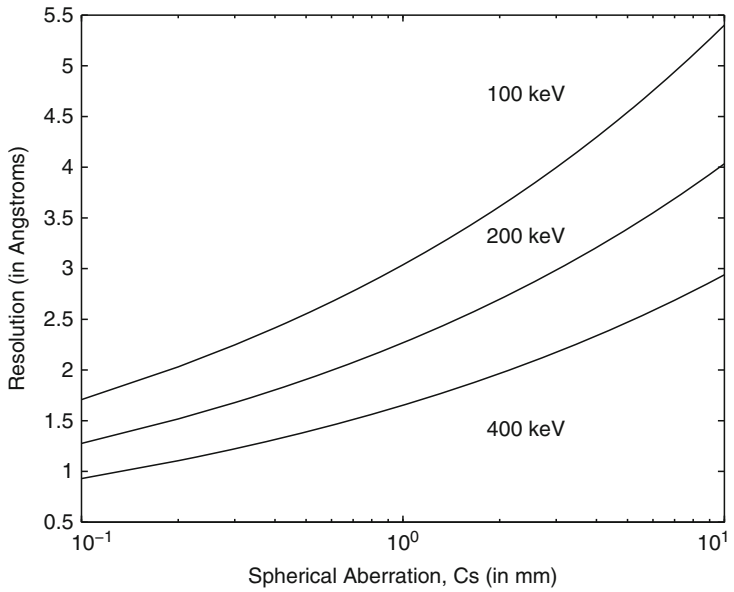

Fig. 3.3 The coherent bright field phase contrast resolution in the weak phase object approximation as a function of spherical aberration C_s for several different electron beam energies

3.2 Partial Coherence in BF-CTEM

In practice the electron microscope always has small deviations from the ideal. The incident illumination is never exactly collimated and parallel to the optic axis of the microscope. The electron energy is not completely monochromatic and the objective and condenser lens currents are never perfectly stable. The degree of collimation of the illumination incident on the specimen is related to the lateral coherence (i.e.,

spatial coherence) of the incident electron wave function and the stability of the beam energy and lens currents is related to the temporal coherence of the imaging process. The imaging described in the previous section assumes that these effects are negligible and that the imaging process is perfectly coherent. When the effects of a small spread in illumination angles and a small spread in beam energy and lens currents are included, the imaging process is said to be partially coherent. An analytical derivation of the effects of partial coherence on the transfer function in the weak phase object approximation has been given by Frank [111], Fejes [99], and Wade and Frank [361].

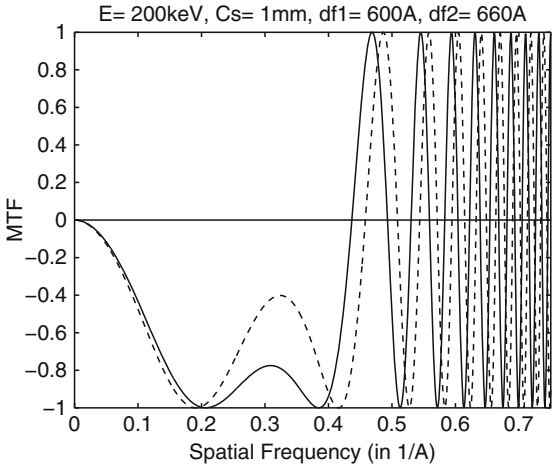

Fig. 3.4 BF-CTEM transfer function with two slightly different defocus values (*solid line* 600Å, and *dashed* 660Å) as might be caused by fluctions in the objective lens current. When these are superposed in the image lower spatial frequencies are little affected but the high spatial frequencies tend to average to zero

The transfer function oscillates both positive and negative. Partial coherence results in the superposition of adjacent portions of the transfer function (as in Fig. 3.4). When the transfer function is oscillating quickly (at high spatial frequencies) then partial coherence tends to reduce the transfer function to zero. Partial coherence limits the maximum information content of the image in the electron microscope by damping the high spatial frequency (large scattering angle) portion of the transfer function.

The electron beam illuminating the specimen (formed by the condenser system) will always have a small distribution of angles (see the right hand side of Fig. 3.5). To calculate the effect on the image first consider an illumination at a single angle as shown in the left hand side of Fig. 3.5. Previously in (3.3) the incident wave function was $\psi_{\text{inc}} \sim 1$ however with the illumination at an angle β:

$$\psi_{\text{inc}}(\mathbf{x}) = \exp(2\pi \mathbf{i} \mathbf{k}_\beta \cdot \mathbf{x}), \tag{3.18}$$

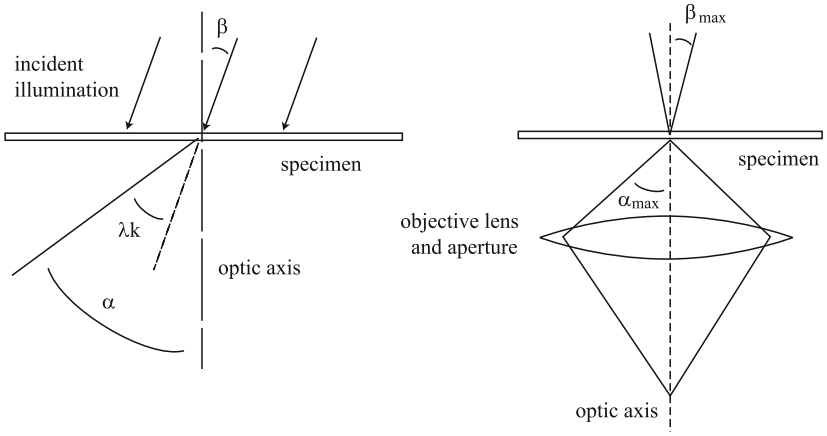

Fig. 3.5 Imaging with nonideal illumination. The incident electrons have an angle β. The specimen scatters at an angle λk and the final angle into the objective lens is α (angles measured with respect to the optic axis). A single electron trajectory is shown on the *left* and the total illumination of a single point on the image is shown on the *right*. $\beta_{\max}$ is typically the condenser aperture and $\alpha_{\max}$ is the objective aperture

where $\mathbf{k}_\beta = \beta/\lambda$ and β is the angle of the incident illumination (with respect to the optic axis). Note that $\mathbf{k}_\beta$ is a two dimensional vector because β can vary in both the polar and azimuthal directions. The transmitted wave function (3.3) now becomes:

$$\psi_t(\mathbf{x}) = t(\mathbf{x}) \exp(2\pi \mathrm{i} \mathbf{k}_\beta \cdot \mathbf{x}). \tag{3.19}$$

Usually, the condenser system will deliver a small cone of illumination angles onto the specimen. Typically, each illumination angle will be incoherent with other illumination angles so the images due to each illumination angle should be summed incoherently by adding intensities $|\psi|^2$ and not amplitudes ψ. It is possible to operate a field emission gun to produce a coherent spread of illumination angles (in which case amplitudes and not intensities would be summed), however this case will not be considered here. If $p(\mathbf{k}_\beta)$ represents the (probability) distribution of (incoherent) illumination angles then (3.5) becomes:

$$\begin{aligned} g(\mathbf{x}) &= \int |\psi_i(\mathbf{x})|^2 p(\mathbf{k}_\beta) \mathrm{d}^2 k_\beta \\ &= \int \left| \left[t(\mathbf{x}) \exp(2\pi \mathrm{i} \mathbf{k}_\beta \cdot \mathbf{x}) \right] \otimes h_0(\mathbf{x}) \right|^2 p(\mathbf{k}_\beta) \mathrm{d}^2 k_\beta \end{aligned} \tag{3.20}$$

A small spread in energy of the incident electron is equivalent to a small (incoherent) spread in defocus values due to the chromatic aberration of the objective lens. Fluctuations in the focusing currents in the objective lens also produce an incoherent

spread in defocus values. When this spread in defocus values is combined and included in the image:

$$g(\mathbf{x}) = \int |\psi_i(\mathbf{x})|^2 \, p(\mathbf{k}_\beta) p(\delta_\mathrm{f}) \mathrm{d}\delta_\mathrm{f} \mathrm{d}^2 k_\beta \tag{3.21}$$

$$= \int \left| \left[t(\mathbf{x}) \exp(2\pi \mathrm{i} \mathbf{k}_\beta \cdot \mathbf{x}) \right] \otimes h_0(\mathbf{x}, \Delta f + \delta_\mathrm{f}) \right|^2 p(\mathbf{k}_\beta) p(\delta_\mathrm{f}) \mathrm{d}\delta_\mathrm{f} \mathrm{d}^2 k_\beta,$$

where δ_f is the fluctuation in defocus and $p(\delta_\mathrm{f})$ is the distribution of this fluctuation. Both $p(\mathbf{k}_\beta)$ and $p(\delta_\mathrm{f})$ are normalized such that their integrated value is unity. This expression for $g(\mathbf{x})$ is formidable and should be treated numerically in general (see later chapters). However, if the deviations from the ideal are assumed very small and the specimen is assumed to be a weak phase object, a modified transfer function can be obtained.

First examine the expression for the wave function in the image plane for one illumination angle and one defocus value:

$$\psi_i(\mathbf{x}) = \left[t(\mathbf{x}) \exp(2\pi \mathrm{i} \mathbf{k}_\beta \cdot \mathbf{x}) \right] \otimes h_0(\mathbf{x}). \tag{3.22}$$

Figure 3.5 shows that the tilt angle can appear in either the objective lens angle α or the incident wave function ψ_inc. To see this mathematically take the Fourier transform of ψ_i and use the Fourier convolution theorem:

$$\mathrm{FT}\left\{ \left[t(\mathbf{x}) \exp(2\pi \mathrm{i} \mathbf{k}_\beta \cdot \mathbf{x}) \right] \otimes h_0(\mathbf{x}) \right\} = T(\mathbf{k} + \mathbf{k}_\beta) H_0(\mathbf{k}). \tag{3.23}$$

Now with a change of variable:

$$T(\mathbf{k} + \mathbf{k}_\beta) H_0(\mathbf{k}) = T(\mathbf{k}') H_0(\mathbf{k}' - \mathbf{k}_\beta). \tag{3.24}$$

The inverse Fourier transform is now an integration over $\mathbf{k}'$ instead of $\mathbf{k}$ (over all space) but gives the same result. This means that the image plane wave function can also be written as:

$$\psi_i(\mathbf{x}) = t(\mathbf{x}) \otimes h_0(\mathbf{x}, \mathbf{k}_\beta) \tag{3.25}$$

Using $h_0(\mathbf{x}, \mathbf{k}_\beta)$ will make the analytical calculation of the transfer function much simpler. This approximation would not be appropriate for thick specimens because the effect of the specimen is no longer a simple multiplicative function.

If the specimen is a weak phase object (3.6) then the expression for the final image intensity will be similar to (3.9) except that the point spread function $h_0(\mathbf{x})$ of the objective lens will be integrated over illumination angles and defocus values as in (3.21). If $\mathbf{k}_\beta$ is small then the leading background constant can also be assumed to be close to unity [as in (3.8)]:

$$1 \otimes h_0(\mathbf{x}, \mathbf{k}_\beta) \sim 1. \tag{3.26}$$

The integral of the transfer function over a spread in illumination angles and defocus values using the dimensionless form of $\chi(\mathbf{K})$ is:

$$H_0(\mathbf{K}) = \int \exp\{-\mathrm{i}\chi(\mathbf{K}+\mathbf{K}_\beta, D+D_p)p(D_p)p(\mathbf{K}_\beta)\mathrm{d}D_p\mathrm{d}^2K_\beta \tag{3.27}$$

where the sign of $\mathbf{K}_\beta$ has been changed to positive for simplicity (i.e., the integrand is symmetric) and:

$$\chi(\mathbf{K}+\mathbf{K}_\beta, D+D_p) = \pi[0.5(\mathbf{K}+\mathbf{K}_\beta)^4 - (D+D_p)(\mathbf{K}+\mathbf{K}_\beta)^2] \tag{3.28}$$

$p(\mathbf{K}_\beta)$ and $p(D_p)$ are the distribution of illuminations angles $\mathbf{K}_\beta$ and defocus deviations D_p. The simplest assumption of a Gaussian distribution for each has the advantage of allowing an analytical solution.

$$p(\mathbf{K}_\beta) = \frac{1}{\pi K_s^2}\exp(-K_\beta^2/K_s^2) \tag{3.29}$$

$$p(D_p) = \frac{1}{\sqrt{\pi}D_s}\exp(-D_p^2/D_s^2), \tag{3.30}$$

where K_s is the $1/e$ width of the spread in illumination angles and D_s is the $1/e$ spread in defocus values. Taylor expanding $\chi(\mathbf{K}, D)$ to lowest order in $\mathbf{K}_\beta$ and D_p:

$$\begin{aligned}\chi(\mathbf{K}+\mathbf{K}_\beta, D+D_p) \sim \\ \chi(\mathbf{K}, D) + \mathbf{K}_\beta \cdot \mathbf{W_1} + D_p\, W_2 + D_p\mathbf{K}_\beta \cdot \mathbf{W}_3 + \cdots\end{aligned} \tag{3.31}$$

where

$$\mathbf{W}_1 = \frac{\partial\chi(\mathbf{K}, D)}{\partial\mathbf{K}} = 2\pi(|\mathbf{K}|^2 - D)\mathbf{K} \tag{3.32}$$

$$W_2 = \frac{\partial\chi(\mathbf{K}, D)}{\partial D} = -\pi K^2 \tag{3.33}$$

$$\mathbf{W}_3 = \frac{\partial^2\chi(\mathbf{K}, D)}{\partial\mathbf{K}\partial D} = -2\pi\mathbf{K}. \tag{3.34}$$

$\mathbf{W}_1$ and $\mathbf{W}_3$ are two dimensional vector quantities. Equation (3.27) now becomes:

$$\begin{aligned}H_0(\mathbf{K}) = \frac{1}{\pi K_s^2\sqrt{\pi}D_s}\exp[-\mathrm{i}\chi(\mathbf{K}, D)]\int \exp[-\mathrm{i}\mathbf{K}_\beta \cdot \mathbf{W_1} \\ - iD_pW_2 - iD_p\mathbf{K}_\beta \cdot \mathbf{W}_3 - \mathbf{K}_\beta^2/K_s^2 - D_p^2/D_s^2]\mathrm{d}^2K_\beta\mathrm{d}D_p\end{aligned} \tag{3.35}$$

First do the integration over $\mathbf{K}_\beta$:

$$\begin{aligned}H_0(\mathbf{K}) \;= \frac{1}{\sqrt{\pi}D_s}\exp[-\mathrm{i}\chi(\mathbf{K}, D)]\int \exp[-0.25K_s^2|\mathbf{W}_1 + D_p\mathbf{W}_3|^2 - \\ iD_pW_2 - D_p^2/D_s^2]\mathrm{d}D_p.\end{aligned} \tag{3.36}$$

Next do the integral over D_p:

$$H_0(\mathbf{K}) = \frac{1}{\sqrt{1+EK^2}} \exp\left[-\mathrm{i}\chi(\mathbf{K},D) + i\frac{D_s^2 K_s^2 \mathbf{W}_1 \cdot \mathbf{W_3} W_2}{4(1+EK^2)}\right]$$
$$\times \exp\left[-0.25K_s^2|\mathbf{W}_1|^2 + \frac{D_s^2[(0.5K_s^2\mathbf{W}_1 \cdot \mathbf{W_3})^2 - W_2^2]}{4(1+EK^2)}\right]. \tag{3.37}$$

Aside: if the cross term $\mathbf{W}_3 = 0$ (implying $E = 0$) is neglected a somewhat more elegant expression results:

$$H_0(\mathbf{K}) = \exp[-\mathrm{i}\chi(\mathbf{K},D) - 0.25K_s^2|\mathbf{W}_1|^2 - 0.25D_s^2 W_2^2]. \tag{3.38}$$

Next substitute for the intemediate variables:

$$H_0(\mathbf{K}) = \exp\left[-\mathrm{i}\frac{\pi K^2}{1+EK^2}\left(0.5K^2(1-EK^2) - D\right)\right]$$
$$\times \frac{1}{\sqrt{1+EK^2}} \exp\left[\frac{-\pi^2 K_s^2 (K^2 - D)^2 K^2 - 0.25\pi^2 D_s^2 K^4}{1+EK^2}\right], \tag{3.39}$$

where $E = \pi^2 K_s^2 D_s^2$

Now take the imaginary part to form the transfer function for the weak phase object image approximation (3.2, 3.10) and substitute the dimensional form of the electron microscope parameters:

$$H_{\mathrm{WP}}(k) = \sin\left[\frac{\pi\lambda k^2}{1+\varepsilon k^2}\left(0.5C_s(1-\varepsilon k^2)\lambda^2 k^2 - \Delta f\right)\right]$$
$$\times \frac{1}{\sqrt{1+\varepsilon k^2}} \exp\left[-\frac{[\pi\lambda k_s k(C_s\lambda^2 k^2 - \Delta f)]^2 + 0.25(\pi\lambda\Delta_0 k^2)^2}{1+\varepsilon k^2}\right], \tag{3.40}$$

where $\varepsilon = (\pi\lambda k_s \Delta_0)^2$, $k_s = K_s(C_s\lambda^3)^{-1/4}$, and $\Delta_0 = D_s\sqrt{C_s\lambda}$. This transfer function (3.40) should be substituted into the image model (3.9 and 3.10). Apart from the extra term with ε (typically small) the oscillatory portion of the transfer functions is the same as the coherent case. The main change is the addition of a damping envelope that attenuates the transfer function at high spatial frequencies. In a practical sense $\beta = \lambda k_s$ is the condenser (illumination) semiangle and Δ_0 is approximately the rms value of all of the appropriate fluctuations multiplied by the chromatic aberration C_c.

$$\Delta_0 \sim C_\mathrm{c}\sqrt{\left(\frac{\Delta E}{E}\right)^2 + \left(2\frac{\Delta I}{I}\right)^2 + \left(\frac{\Delta V}{V}\right)^2}, \tag{3.41}$$

where E, I, and V are the electron energy, lens currents, and acceleration voltage, respectively, and ΔE, ΔI, and ΔV are the $1/e$ width of their fluctuations.

A graph of the WPO transfer function with (3.40) and without (3.10) partial coherence is shown in Fig. 3.6 for typical values of the electron optical parameters.

It is interesting to note that the cross term W_3 produced the correction ε in the transfer function which can move the zero crossings in the transfer function but seems to have a negligible effect under Scherzer conditions.

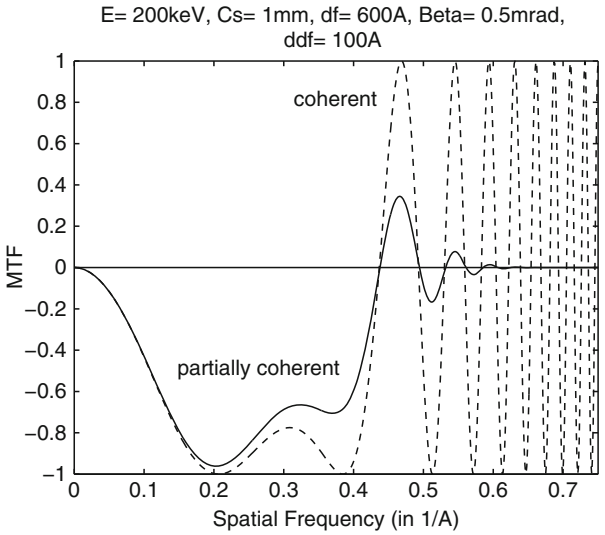

Fig. 3.6 The BF phase contrast transfer function with and without partial coherence

3.2.1 Aberration Correctors and Partial Coherence

It is straight forward but somewhat tedious to add higher order aberrations to the earlier expressions for partial coherence. The transfer function with the fifth-order spherical aberration is:

$$H_{\mathrm{WP}}(k) = \frac{1}{\sqrt{1+\varepsilon k^2}} \times \sin\left[\frac{\pi\lambda k^2}{1+\varepsilon k^2}\left(\frac{1}{3}C_{S5}(1-2\varepsilon k^2)\lambda^4 k^4 + 0.5C_{S3}(1-\varepsilon k^2)\lambda^2 k^2 - \Delta f\right)\right] \times \exp\left[-\frac{[\pi\lambda k_s k(C_{S5}\lambda^4 k^4 + C_{S3}\lambda^2 k^2 - \Delta f)]^2 + 0.25(\pi\lambda\Delta_0 k^2)^2}{1+\varepsilon k^2}\right]. \quad (3.42)$$

However, terms such as the spherical aberration coefficient $C_s = C_{S3}$ are no longer really constant. The spherical aberration of the round objective lens is essentially constant because it is due to the fixed geometry of the magnetic material used to shape the focusing fields and to lowest order insensitive to small fluctuations in the lens current (which instead produces fluctuations in defocus). An aberration

corrector generates a large negative C_{S3} from a combination of strong focusing multipole fields. Fluctuations in the currents in these elements can produce fluctuations in the net C_{S3} coefficients as well as all of the other dozen or so aberration coefficients. There can be of order 100 or more focusing currents (etc.) in the corrector that have there own fluctuations. These potentially cause all of the aberration coefficients to fluctuate (more study is needed here to determine how these all interact, may vary with the specific design of each corrector). Various groupings of the multipoles combine to form each aberrations (some elements may be part of more than one aberration). If the cross terms between aberrations are ignored for simplicity, the aberration phase factor becomes:

$$H_0(\mathbf{K}) = \exp\left[-\mathrm{i}\chi(\mathbf{K},D) - \sum_k 0.25 W_k \left|\frac{\partial\chi}{\partial C_k}\right|^2\right], \tag{3.43}$$

where W_k is the width of the fluctuations for parameter C_k and the summation is over all aberration components (C_k such as defocus, spherical aberrations C_{s3}, C_{s5}, etc.) and the illumination angle. Even though the aberration corrector can effectively negate the existing aberrations in the main (round) objective lens the fluctuations in the corrector causing partial coherence will be centered about the much larger aberrations (such as C_{S3}) of the original objective lens making this effect much larger than might be expected.

It may be easier to numerically integrate over small fluctuation in all parameters. Gauss-Hermite quadrature is very efficient for integration over a Gaussian weighting. With this many parameters it might even be worthwhile performing a Monte-Carlo integration over a range of parameter values (which becomes more efficient when integrating over more than four dimensions). It is also not clear how fluctuations in the various aberration coefficient are coupled (mathematically in the aberration function and practically through common coils and currents).

3.3 Detector Influence (CTEM)

The detector that records the image may itself have a large influence on the image quality. Currently, the detector is usually a CCD camera (plus scintillator). Film (or plates) were the most common detector in the not so distant past. It was common to record the image at not very high magnification and then magnify further in the darkroom enlarger, in which case the film may produce a large effect. Film (or plates) have a transfer function of their own (for example Downing and Grano [80]). CCD detectors also have an associated transfer function which can be significant (Thust [343]). The effects of the detector can be included in the image by convolving the electron image (3.5) with detector transfer function $h_{\mathrm{DET}}(\mathbf{x})$:

$$g(\mathbf{x}) = |\psi_t(\mathbf{x}) \otimes h_0(\mathbf{x})|^2 \otimes h_{\mathrm{DET}}(\mathbf{x}) \tag{3.44}$$

The detector may reduce the contrast in the high resolution component of an image (for example lattice fringes) by a large factor (like two or three).

3.4 Incoherent Imaging of Thin Specimens (CTEM)

The electron microscope image is generated from the electrons scattered by the specimen. For the electron wave scattered from two points of the specimen to interfere the incident electron wave from the condenser system must be coherent over the distance between these points (i.e., the specimen is not self-luminous). The illuminating electrons (using a thermionic source) are produced by a spatially incoherent, quasi-monochromatic source with nonzero size. The source size is typically translated into a maximum condenser angle $\beta_{\max}$ (see Fig. 3.5). The electron rays incident on the specimen are not perfectly parallel but subtend a small cone of angles $\beta_{\max}$ at the specimen plane. The lateral coherence length perpendicular to the optic axis (see Sect. 10.4.2 of Born and Wolf [34]) of the illuminating electron wave function is approximately given by:

$$\Delta x_{\text{coh}} \sim \frac{0.16\lambda}{\beta_{\max}}. \tag{3.45}$$

If the coherence length is much smaller than the resolution element of the image then the imaging will be essentially incoherent. If the coherence length is much bigger than the resolution element then the imaging process will be essentially coherent. The resolution element of the image is approximately $d \sim \lambda/\alpha_{\max}$ where $\alpha_{\max}$ is the maximum objective angle (i.e., the size of the objective aperture). Combining these two expressions yields an approximate criteria for the image coherence:

$$\begin{aligned} \beta_{\max} &\ll 0.16\alpha_{\max} \quad \text{coherent imaging} \\ \beta_{\max} &\gg 0.16\alpha_{\max} \quad \text{incoherent imaging} \end{aligned} \tag{3.46}$$

In between these two extremes the image is partially coherent.

If the image is coherent then the amplitudes ψ of the scattered electrons add and if the image is incoherent then the intensities of the scattered electrons $|\psi|^2$ add. The final image recording process is only sensitive to the intensity of the electrons and not their amplitudes. This means that phase contrast must have a coherent image process to be sensitive to the phase of the electron via some interference process. Reducing the coherence of the imaging process should also reduce the phase contrast transfer function. The ratio $\beta_{\max}/\alpha_{\max}$ can be used to control the coherence of the imaging process.

Because the phase contrast image will likely disappear in an incoherent image the transmission function should include the possibility of amplitude contrast. Assuming that the specimen is a weak-phase, weak-amplitude object yields a transmission function:

$$t(\mathbf{x}) \sim \exp[\mathrm{i}\sigma v_z(\mathbf{x}) - u(\mathbf{x})] \sim 1 + i\sigma_e v_z(\mathbf{x}) - u(\mathbf{x}) + \cdots \tag{3.47}$$

where $u(\mathbf{x})$ is the amplitude component of the specimen transmission function which can arise from scattering outside of the objective aperture (it is preceded by a minus sign because electrons cannot be created with elastic scattering). Alternately $u(\mathbf{x})$ can be considered as the next term in the Taylor expansion of $\exp[i\sigma v_z(\mathbf{x})]$ where $u(\mathbf{x}) \propto v_z^2(\mathbf{x})$. Both $\sigma_e v_z(\mathbf{x})$ and $u(\mathbf{x})$ are small compared to unity. The expression for the image intensity (3.21) now becomes:

$$\begin{aligned} g(\mathbf{x}) &= \int |\psi_i(\mathbf{x})|^2 \, p(\mathbf{k}_\beta) p(\delta_\mathrm{f}) \mathrm{d}\delta_\mathrm{f} \mathrm{d}^2 k_\beta \\ &= \int \left| \left[(1 + \mathrm{i}\sigma_e v_z(\mathbf{x}) - u(\mathbf{x})) \exp(2\pi \mathrm{i} \mathbf{k}_\beta \cdot \mathbf{x}) \right] \otimes h_0(\mathbf{x}, \Delta f + \delta_\mathrm{f}) \right|^2 \\ &\quad p(\mathbf{k}_\beta) p(\delta_\mathrm{f}) \mathrm{d}\delta_\mathrm{f} \mathrm{d}^2 k_\beta . \end{aligned} \tag{3.48}$$

The objective aperture will play a significant role in the following derivation so it should be included with the point spread function $h_0(\mathbf{x})$ or equivalently the transfer function:

$$H_0(\mathbf{k}) = \exp[-\mathrm{i}\chi(\mathbf{k}, \Delta f + \delta_\mathrm{f})] A(\mathbf{k}), \tag{3.49}$$

where $A(\mathbf{k})$ is the aperture function:

$$\begin{aligned} A(\mathbf{k}) &= 1 \quad ; \quad \lambda |\mathbf{k}| = \alpha < \alpha_{\max} \\ &= 0 \quad ; \quad \text{otherwise} \end{aligned} \tag{3.50}$$

and $\alpha_{\max}$ is the maximum semiangle allowed by the objective aperture.

Now expand the integrand keeping only the terms of lowest order in $\sigma_e v_z(\mathbf{x})$ and $u(\mathbf{x})$ and drop the explicit reference to the independent arguments for simplicity:

$$\begin{aligned} g(\mathbf{x}) = \int \{ & |\exp(2\pi \mathrm{i} \mathbf{k}_\beta \cdot \mathbf{x}) \otimes h_0|^2 \\ & + [\exp(2\pi \mathrm{i} \mathbf{k}_\beta \cdot \mathbf{x}) \otimes h_0]^* [i\sigma_e v_z \exp(2\pi \mathrm{i} \mathbf{k}_\beta \cdot \mathbf{x}) \otimes h_0] \\ & - [\exp(2\pi \mathrm{i} \mathbf{k}_\beta \cdot \mathbf{x}) \otimes h_0]^* [u \exp(2\pi \mathrm{i} \mathbf{k}_\beta \cdot \mathbf{x}) \otimes h_0] \\ & + [\text{complex conjugate}] \} p(\mathbf{k}_\beta) p(\delta_\mathrm{f}) \mathrm{d}\delta_\mathrm{f} \mathrm{d}^2 k_\beta . \end{aligned} \tag{3.51}$$

The first term is a constant of order unity that does not vary with position in the image. The rest of the right hand side is rather unpleasant but can be simplified a little. Because $h_0(\mathbf{x})$ is simplest to write in reciprocal space, look at the Fourier transform of the term containing $v_z(\mathbf{x})$, and for simplicity drop explicit reference to the defocus Δf and its fluctuations δ_f:

$$\begin{aligned} & \mathrm{FT}\{ [\exp(2\pi \mathrm{i} \mathbf{k}_\beta \cdot \mathbf{x}) \otimes h_0]^* [i\sigma_e v_z \exp(2\pi \mathrm{i} \mathbf{k}_\beta \cdot \mathbf{x}) \otimes h_0] \} \\ & \quad = [\delta(\mathbf{k} - \mathbf{k}_\beta) H_0(-\mathbf{k})]^* \otimes [\mathrm{i}\sigma_e V(\mathbf{k} + \mathbf{k}_\beta) H_0(\mathbf{k})] \\ & \quad = \mathrm{i}\sigma_e \int \delta(\mathbf{k}' - \mathbf{k}_\beta) H_0^*(-\mathbf{k}') V(\mathbf{k} + \mathbf{k}_\beta - \mathbf{k}') H_0(\mathbf{k} - \mathbf{k}') \mathrm{d}\mathbf{k}' \\ & \quad = \mathrm{i}\sigma_e V(\mathbf{k}) H_0^*(-\mathbf{k}_\beta) H_0(\mathbf{k} - \mathbf{k}_\beta), \end{aligned} \tag{3.52}$$

where $\delta(\mathbf{k})$ is the Dirac delta function.

This expression is now a simple product so when inverse Fourier transformed back into real space $\sigma_e v_z(\mathbf{x})$ is convolved with a new form of the point-spread function. Repeating this procedure with the amplitude component and combining the complex conjugate terms yield:

$$g(\mathbf{x}) \sim C_0 + 2\sigma_e v_z(\mathbf{x}) \otimes h_{\mathrm{WP}}(\mathbf{x}) - 2u(\mathbf{x}) \otimes h_{\mathrm{WA}}(\mathbf{x}), \tag{3.53}$$

where C_0 is a background constant of order unity:

$$C_0 = \int |\exp(2\pi \mathrm{i}\mathbf{k}_\beta \cdot \mathbf{x}) \otimes h_0|^2 p(\mathbf{k}_\beta) p(\delta_\mathrm{f}) \mathrm{d}\delta_\mathrm{f} \mathrm{d}^2 k_\beta \tag{3.54}$$

and the transfer function for a weak phase object is:

$$\begin{aligned} \mathrm{FT}[h_{\mathrm{WP}}(\mathbf{x})] &= H_{\mathrm{WP}}(\mathbf{k}) \\ &= \mathrm{Imag} \int H_0(\mathbf{k}_\beta - \mathbf{k}) H_0^*(-\mathbf{k}_\beta) p(\mathbf{k}_\beta) p(\delta_\mathrm{f}) \mathrm{d}\delta_\mathrm{f} \mathrm{d}^2 k_\beta \end{aligned} \tag{3.55}$$

and the transfer function for a weak amplitude object is:

$$\begin{aligned} \mathrm{FT}[h_{\mathrm{WA}}(\mathbf{x})] &= H_{\mathrm{WA}}(\mathbf{k}) \\ &= \mathrm{Real} \int H_0(\mathbf{k}_\beta - \mathbf{k}) H_0^*(-\mathbf{k}_\beta) p(\mathbf{k}_\beta) p(\delta_\mathrm{f}) \mathrm{d}\delta_\mathrm{f} \mathrm{d}^2 k_\beta. \end{aligned} \tag{3.56}$$

With the assumption of a small Gaussian spread of defocus [see (3.30)] the integral over defocus can be done analytically yielding an expression for the transfer function for a weak phase object:

$$\begin{aligned} H_{\mathrm{WP}}(\mathbf{k}) = \int & A(\mathbf{k}_\beta) A(\mathbf{k}_\beta - \mathbf{k}) \sin[\chi(\mathbf{k}_\beta - \mathbf{k}) - \chi(\mathbf{k}_\beta)] \\ & \times \exp\{-0.25\pi^2 \lambda^2 \Delta_0^2 [|\mathbf{k}_\beta - \mathbf{k}|^2 - |\mathbf{k}_\beta|^2]^2\} p(\mathbf{k}_\beta) \mathrm{d}^2 k_\beta \end{aligned} \tag{3.57}$$

and an expression for the transfer function for a weak amplitude object:

$$\begin{aligned} H_{\mathrm{WA}}(\mathbf{k}) = \int & A(\mathbf{k}_\beta) A(\mathbf{k}_\beta - \mathbf{k}) \cos[\chi(\mathbf{k}_\beta - \mathbf{k}) - \chi(\mathbf{k}_\beta)] \\ & \times \exp\{-0.25\pi^2 \lambda^2 \Delta_0^2 [|\mathbf{k}_\beta - \mathbf{k}|^2 - |\mathbf{k}_\beta|^2]^2\} p(\mathbf{k}_\beta) \mathrm{d}^2 k_\beta. \end{aligned} \tag{3.58}$$

The only difference between these two transfer functions is the switch between sin and cos. If $\mathbf{k}_\beta$ is large then neither of these integrals can be done analytically and both must be done numerically (in two dimensions).

The phase and amplitude transfer functions are shown in Figs. 3.7 and 3.8, respectively, using Scherzer defocus and the Scherzer aperture. The phase contrast transfer function is initially similar to the coherent transfer (see Fig. 3.2) function but decays to zero as the condenser angle $\beta_{\max}$ is increased as expected (i.e., when the image is incoherent there cannot be any interference to produce phase contrast). The amplitude contrast transfer function however transforms from an oscillatory function into a smoothly falling function similar to the transfer function in normal light

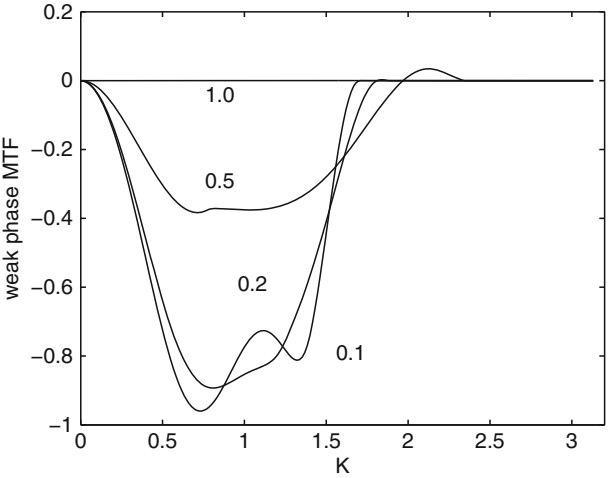

Fig. 3.7 BF-CTEM transfer function for weak phase objects with increasing condenser angle $\beta_{\max}$ as a function of dimensionless spatial frequency K (Scherzer defocus and aperture $\alpha_{\max}$). $\beta_{\max}/\alpha_{\max} = 0.1, 0.2, 0.5, 1.0$. The phase contrast transfer function is zero when the image process is incoherent $\beta_{\max}/\alpha_{\max} \geq 1$

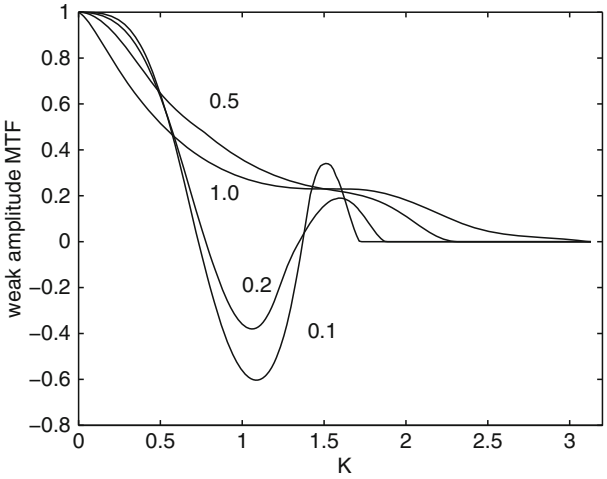

Fig. 3.8 BF-CTEM transfer function for weak amplitude objects with increasing condenser angle $\beta_{\max}$ as a function of dimensionless spatial frequency K (Scherzer defocus and aperture $\alpha_{\max}$). $\beta_{\max}/\alpha_{\max} = 0.1, 0.2, 0.5, 1.0$. The amplitude contrast transfer function is similar to normal light optics when the image process is incoherent $\beta_{\max}/\alpha_{\max} \geq 1$

optics (Black and Linfoot [30]). Note also that the first zero of the transfer function has moved by almost a factor of two in the incoherent ($\beta_{\max}/\alpha_{\max} = 1$) case. The improvement in resolution from incoherent imaging over coherent imaging has been

known for a long time (Rayleigh [292] and Goodman [126]). The essential features of this section were discussed by Hanszen [145], Hanszen and Trepte [146], and Thomson [341].

3.5 Annular Dark Field STEM

The order of the optical components of the STEM (Fig. 2.4) is reversed from that of the CTEM (Fig. 2.3). The objective lens is before the specimen and forms a focused probe on the specimen. The portion of the electrons transmitted through the specimen that fall on the detector form the image brightness at one point in the image. A whole image is built up by scanning the focused probe over the specimen and recording the transmitted intensity at each position of the probe. In bright field the detector integrates (incoherently) over a very small angle centered about zero scattering angle, and in annular dark field (ADF) the detector integrates everything except the center regions. The ADF-STEM image would be equivalent to the BF-CTEM image using incoherent hollow cone illumination by the reciprocity theorem (Engel [92]). In practice a CTEM condenser system may not be able to handle the large angles (typically 100 to 300 mrad at 100 keV) equivalent to the angles in the ADF-detector. Alternately, an image similar to an ADF-STEM image but with reversed contrast may be obtained using BF-CTEM with a large solid cone that is mirror image of hollow cone illumination (Kirkland [204]). The simplified image model discussed in this section can be referred to as the incoherent image model.

A focused probe is calculated by integrating the aberration wave function $\exp[-\mathrm{i}\chi(k)]$ over the objective aperture with translation to a particular point in the image. Figure 3.9 shows the relative placement of the wave functions in the STEM column. The aberrated electron probe wave function in the plane of the specimen when deflected to position $\mathbf{x}_p$ is:

$$\psi_p(\mathbf{x},\mathbf{x}_p) = A_p \int_0^{k_{\max}} \exp[-\mathrm{i}\chi(\mathbf{k}) - 2\pi\mathrm{i}\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}_p)]\mathrm{d}^2\mathbf{k}, \tag{3.59}$$

where $\lambda k_{\max} = \alpha_{\max}$ is the maximum angle in the objective aperture and A_p is a normalization constant chosen to yield;

$$\int |\psi_p(\mathbf{x},\mathbf{x}_p)|^2 \mathrm{d}^2\mathbf{x} = 1. \tag{3.60}$$

With this normalization the total incident intensity in the electron probe is also unity (alternately the probe integral could be scaled to yield the actual value of the beam current in some appropriate choice of units). The probe is really just the demagnified image of the source and this expression assumes that the electron source (the electron gun demagnified by the condenser system) has a negligible size in the plane of the specimen (discussed later in Sect. 3.5.2). The probe size is limited instead by the aberrations of the objective lens. A dedicated STEM would usually use a high

brightness field emission gun with a small virtual source size so that there is enough current left after demagnification to get a large enough signal to detect. Deflecting the beam to different positions on the specimen changes the angles through the objective lens. At high resolution this angle is of order 100 Å divided by the focal length of the objective lens (of order 1 mm) producing a deflection angle of order 0.01 mrad. which is a negligible compared to a typical angle of 10 mrad. in the objective aperture (large deflection angles would change the apparent aberrations but only occur at low magnification).

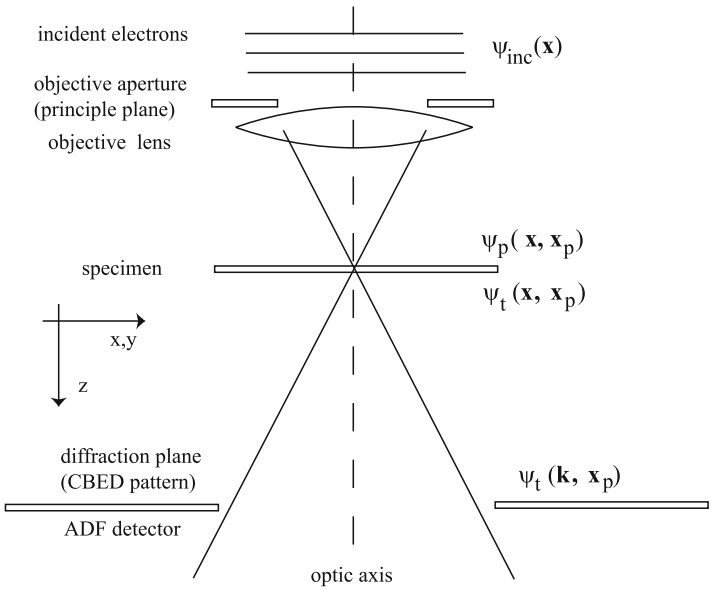

Fig. 3.9 Electron wave functions in the STEM column

The electron probe passes through the specimen and is modulated by the specimen transmission function $t(\mathbf{x})$ (identical to the transmission function for the CTEM). $t(\mathbf{x})$ is in general a complex valued function for a weak phase, weak amplitude object [see (3.47)]. The transmitted wave function is:

$$\psi_t(\mathbf{x}, \mathbf{x}_p) = t(\mathbf{x})\psi_p(\mathbf{x}, \mathbf{x}_p). \tag{3.61}$$

Again a discussion of the effects of specimen thickness on the transmission function will be deferred to later chapters. This wave function is then diffracted onto the detector plane (represented by a Fourier transform).

$$\Psi_t(\mathbf{k}, \mathbf{x}_p) = \mathrm{FT}[\psi_t(\mathbf{x}, \mathbf{x}_p)] = \int \exp(2\pi i \mathbf{k} \cdot \mathbf{x})\psi_t(\mathbf{x}, \mathbf{x}_p)\mathrm{d}^2\mathbf{x}. \tag{3.62}$$

The intensity of this wave function $|\Psi_t(\mathbf{k},\mathbf{x}_p)|^2$ as a function of scattering angle $\lambda\mathbf{k}$ is the Convergent Beam Electron Diffraction CBED pattern.

The CBED pattern is incoherently integrated over the detector geometry and the result is the final STEM image signal $g(\mathbf{x_p})$ for one probe position $\mathbf{x_p}$.

$$g(\mathbf{x}_p) = \int |\Psi_t(\mathbf{k},\mathbf{x}_p)|^2 D(\mathbf{k}) \mathrm{d}^2\mathbf{k}, \tag{3.63}$$

where $D(\mathbf{k})$ is the detector function.

$$D(\mathbf{k}) = 1 \text{ for } k_{D\min} \leq k \leq k_{D\max} \tag{3.64}$$

$$= 0 \text{ otherwise}, \tag{3.65}$$

where $\lambda k_{D\min}$ and $\lambda k_{D\max}$ are the inner and outer angles of the ADF detector. This process is repeated for each probe position $\mathbf{x}_p$.

Equation (3.63) is difficult to intuitively relate to any specific structure in the specimen (later chapters will treat this equation more exactly using numerical calculations). However with some approximations a linear image model can be derived that is easier to understand. The ADF detector should go to very large angles (of order 300 mrad or more at 100 keV) so that all electrons that are incident on the specimen either fall on the ADF detector or go through the central hole in the ADF detector. This means that the signal formed by integrating all of the electron in the central hole in the ADF detector is just one minus the ADF signal (where the total incident intensity is assumed to be unity). This effective signal from the central hole in the ADF detector is also equivalent to the BF-CTEM signal (via the reciprocity theorem) with a very large condenser illumination angle $\beta_{\max}/\alpha_{\max} \gg 1$. Therefore, the ADF image process is incoherent if $k_{D\min} \gg k_{\max}$ (see Sect. 3.4). If the outer dimension of the ADF detector is essentially infinite then the large diameter of the central hole in the ADF detector produces an incoherent image. This means that phase contrast should be negligible and the image should be predominately amplitude contrast. Simple imaging approximations for ADF-STEM have been considered by Misell et al. [242], Cowley [60], Spence [332], Jesson and Pennycook [187], Treacy and Gibson [347], and Loane et al. [228].

An approximate linear image model for thin specimens assuming an incoherent image process is:

$$g(\mathbf{x}) = f(\mathbf{x}) \otimes h_{\mathrm{ADF}}(\mathbf{x}), \tag{3.66}$$

where the specimen function $f(\mathbf{x})$ is approximately the probability for scattering to the large angles of the ADF detector.

$$f(\mathbf{x}) \sim \int D(\mathbf{k}) \frac{\partial \sigma(\mathbf{x})}{\partial k_s} \mathrm{d}^2 k_s = \int_{k_{D\min}}^{k_{D\max}} \frac{\partial \sigma(\mathbf{x})}{\partial k_s} \mathrm{d}^2 k_s \tag{3.67}$$

$\partial\sigma/\partial k_s$ is the partial cross section (the square of the scattering factor) for scattering to angle k_s at position $\mathbf{x}$ of the specimen. With the incoherent image assumption an ADF-STEM image of a very thin specimen is essentially a mass thickness map of the specimen. The calculation with this approximation is simple enough to perform

interactively inside a web browser and is useful to quickly build intuition on the imaging process although it may not be quantitatively accurate. This simple incoherent image model also is close to the assumed image models for a variety of sophisticated image restoration algorithms such as the Richardson-Lucy [229, 302] and maximum entropy methods (for example [165]) so these methods can then be easily applied to ADF-STEM images (for example Kirkland [201]).

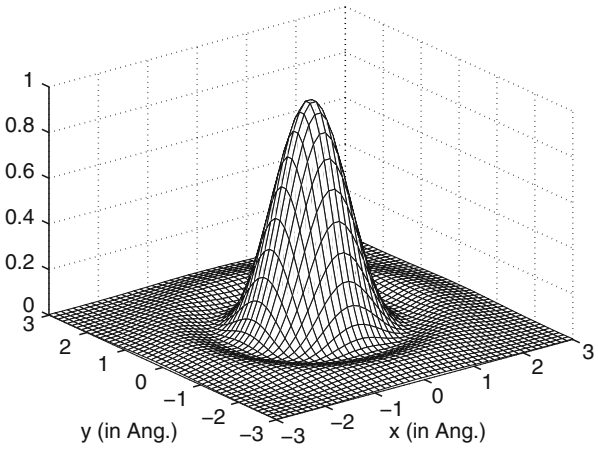

Fig. 3.10 The STEM probe intensity (approximate point spread function) when astigmatism and source size are negligible vs. position (200 keV, $C_s = 1.3$ mm, $\Delta f = 571$Å, $\alpha_{\max} = 9.4$ mrad)

The point spread function (with the assumption of an incoherent image model) is just the intensity distribution in the focused probe. The aberration limited probe is:

$$\begin{aligned} h_{\mathrm{ADF}}(\mathbf{x}) &= |\psi_p(\mathbf{x})|^2 \\ &= A_p \left| \int_0^{k_{\max}} \exp[-\mathrm{i}\chi(\mathbf{k}) - 2\pi\mathrm{i}\mathbf{k}\cdot\mathbf{x}]\mathrm{d}^2\mathbf{k} \right|^2, \end{aligned} \tag{3.68}$$

The constant A_p is chosen to normalize the point spread function to have a total integrated value of unity. Furthermore, if the astigmatism is negligible the azimuthal integral can be done analytically leaving a one dimensional integral for the probe intensity:

$$h_{\mathrm{ADF}}(r) = A_p \left| \int_0^{k_{\max}} \exp[-\mathrm{i}\chi(k)] J_0(2\pi k r) k \mathrm{d}k \right|^2, \tag{3.69}$$

where $J_0(x)$ is the zeroth order Bessel function of the first kind and r is the radial coordinate. $\chi(k)$ is a function of only the magnitude of k when astigmatism is not present. This integral cannot be done analytically and must be done numerically. The point spread function is plotted in Fig. 3.10 vs. position (on the specimen). Figure 3.11 shows a more condensed form of this graph for various values of defocus

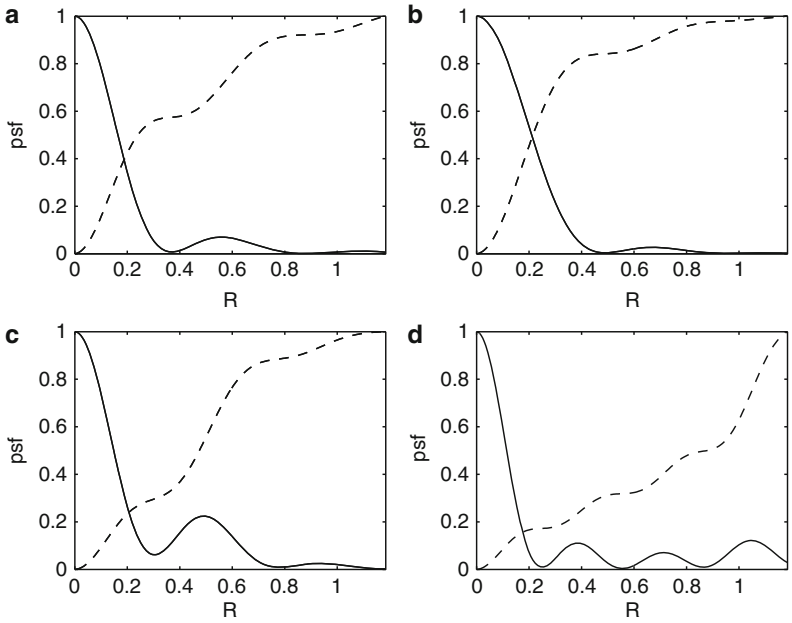

Fig. 3.11 The STEM probe intensity or approximate point spread function (*solid line*) when astigmatism and source size are negligible vs. normalized radius $R = r(C_s\lambda^3)^{-1/4}$ for various values of the normalized defocus $D = \Delta f(C_s\lambda)^{-1/2}$ and objective aperture $K_{\max} = k_{\max}(C_s\lambda^3)^{1/4}$. (a) $D = 1.2$, $K_{\max} = 1.56$ (Scherzer conditions), (b) $D = 0.80$, $K_{\max} = 1.22$, (c) $D = 1.5$, $K_{\max} = 1.5$, (d) $D = 2.5$, $K_{\max} = 2.5$. The total integrated current is shown as a *dashed line*

and objective aperture. The graph has been scaled to make $h_{\mathrm{ADF}}(0) = 1$. Figure 3.11a shows the probe profile for Scherzer defocus and aperture with a full width at half max of approximately $0.43(C_s\lambda^3)^{1/4}$.

The transfer function is just the inverse Fourier transform of the point spread function. With azimuthal symmetry (i.e., no astigmatism), and neglecting the source size the transfer function is:

$$H_{\mathrm{ADF}}(k) = A_p \left| \int_0^\infty h_{\mathrm{ADF}}(r) J_0(2\pi k r) r \mathrm{d}r \right|^2 \tag{3.70}$$

The transfer function is plotted in Fig. 3.12 and BF-CTEM and ADF-STEM are compared in Fig. 3.13.

It is interesting to compare an aberration corrected probe and an uncorrected probe [232]. Figure 3.14 shows a probe at 100 keV with and without an aberration corrector. The curves are normalized to have the same integrated current. It is very surprising that the corrected probe is dramatically higher than the uncorrected probe. The tails of the probe at large radius have a very large contribution to the total current even though they appear small in this type of graph. An aberration corrector can increase the image contrast a lot more than might be expected.

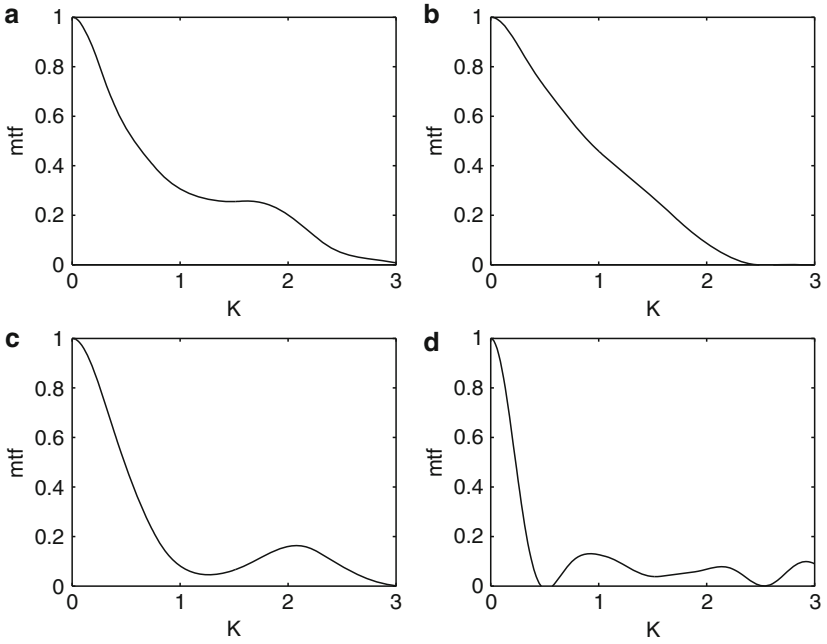

Fig. 3.12 The approximate STEM transfer function corresponding to the defocus values and objective apertures (a) $D = 1.2$, $K_{\max} = 1.56$ (Scherzer conditions), (b) $D = 0.80$, $K_{\max} = 1.22$, (c) $D = 1.5$, $K_{\max} = 1.5$, (d) $D = 2.5$, $K_{\max} = 2.5$ vs. the normalized spatial frequency $K = k(C_s\lambda^3)^{1/4}$. Source size is assumed negligible

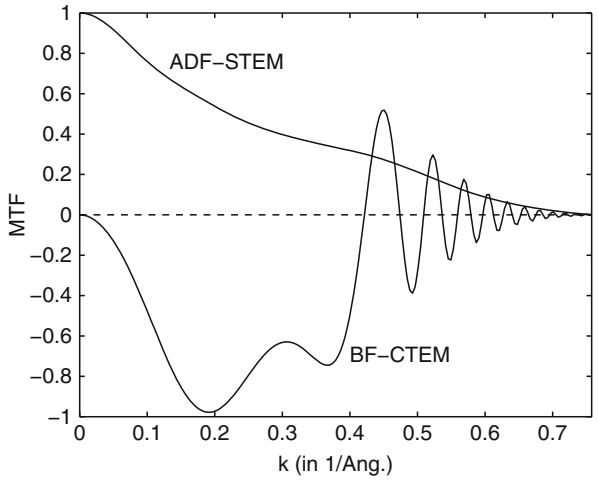

Fig. 3.13 Comparison of ADF-STEM and BF-CTEM transfer functions for the same spherical aberration, $C_s = 1.2$ mm at 200 keV. Δf= 550 Å and $\alpha_{\max} = 9.5$ mrad for STEM and $\Delta f = 670$ Å for BF-CTEM with a condenser half angle of 0.1 mrad and a defocus spread of 100Å

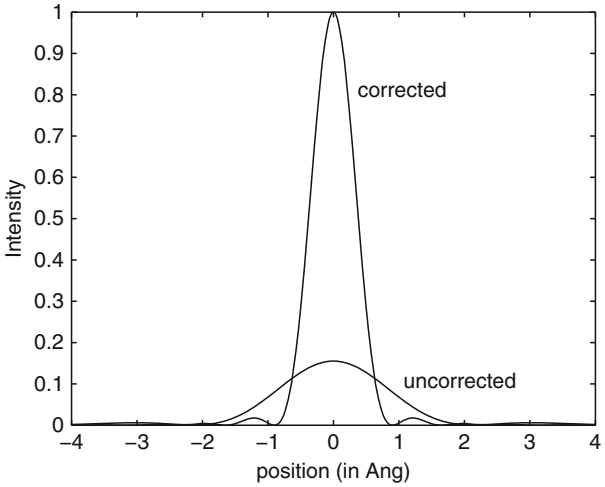

Fig. 3.14 Comparison of ADF-STEM probe with and without an aberration corrector at 100 keV. The uncorrected probe has $C_s = 1.3$ mm, $\Delta f = 694$ Å and $\alpha_{\max} = 10.3$ mrad. The corrected probe has $C_s = \Delta f = 0$ and $\alpha_{\max} = 25$ mrad, Both curves are normalized to have the same total (integrated) current

3.5.1 Minimum Probe Conditions

It is difficult to define an optimum defocus and aperture to produce the best probe for the highest resolution. In Fig. 3.11d the full width at half maximum (FWHM) is clearly much smaller than the FWHM for Scherzer conditions (Fig. 3.11a), however the tails (small wiggles at large radius) are dramatically increased in size. A small value at large radius can produce a large signal when integrated over a larger circumference. The total current inside a radius r in the probe is:

$$I(r) = 2\pi \int_0^r |\psi_p(\mathbf{x}')|^2 r' \mathrm{d}r' \tag{3.71}$$

This current is shown in Fig. 3.11 as a dashed line. The extra factor of r inside the integral (3.71) can shift the signal to a larger radius in a significant manner. The radius containing half of the current is a reasonable definition of the probe size (many other definitions are possible). This FWHM of the integrated intensity is a good measure of where the signal comes from.

A plot of this FWHM radius vs. both defocus and objective aperture size is shown in Fig. 3.15. The minimum probe rms radius appears to be at defocus $D = 0.87$ and objective aperture $K_{\max} = 1.34$ (Fig. 3.11b). This produces the smallest tails but increases the FWHM (minimum rms radius approx. $0.43(C_s\lambda^3)^{1/4}$) to about twice that of the probe with Scherzer conditions. Scherzer conditions seem to be a compromise between a small full width half maximum and large tails. Mory et al. [254] have

also considered the optimum probe defocus for STEM imaging and microanalysis. Intaraprasonk et al. [171] have discussed optimizing the probe with spherical aberration through fifth order. There is also a long history of balancing aberrations in incoherent light optics similar to the treatment here, starting with Maréchal [234] (see also Sect. 4.4 of O'Neil [275] for balancing 3rd and 5th order spherical aberrations) and Black and Linfoot [30] and including annular objective apertures (for example Barakt and Houston [18]).

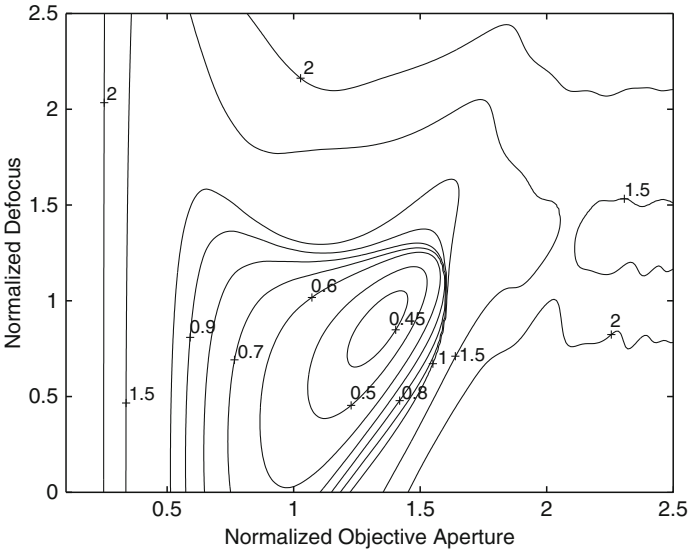

Fig. 3.15 The normalized rms radius $r_{\text{rms}}(C_s\lambda^3)^{-1/4}$ of the STEM probe as a function of the normalized objective aperture $k_{\max}(C_s\lambda^3)^{1/4}$ and the normalized defocus $\Delta f(C_s\lambda)^{-1/2}$ with a large $C_S = C_{S3}$ (no aberration corrector)

3.5.2 Source Size

The probe is just the image of the electron source, which can also contribute to the probe size. The brightness of a source is defined as:

$$\beta = \frac{j}{\pi\alpha^2} \tag{3.72}$$

where j is the current density in the probe and α is the convergence half angle ($\pi\alpha^2$ is the solid angle). Brightness is conserved in a magnetic lens but may vary with beam energy. Various condenser lenses and the objective lens are used to demagnify the image of the source onto the specimen. More source demagnification produces less current (and a smaller source contribution to the probe size) in a predictable manner. If the probe is approximated as a disk of diameter d_S (the source size) and

uniform intensity then the current in the probe $I_P = j(\pi d_S^2/4)$ and the probe size are related as:

$$I_P = \frac{1}{4}\pi^2\alpha^2\beta d_S^2. \tag{3.73}$$

The probe current is shown in Fig. 3.16 for two different types of electron sources. A probe current of about 10–100 pAmp is needed for practical imaging. Only a field emission source can produce enough current to produce a source size of about 1 Angstrom (or smaller) and is preferred. This simple approximation is a convenient means of estimating the source contribution from just a measurement of the probe current and aperture size although it may not be that accurate.

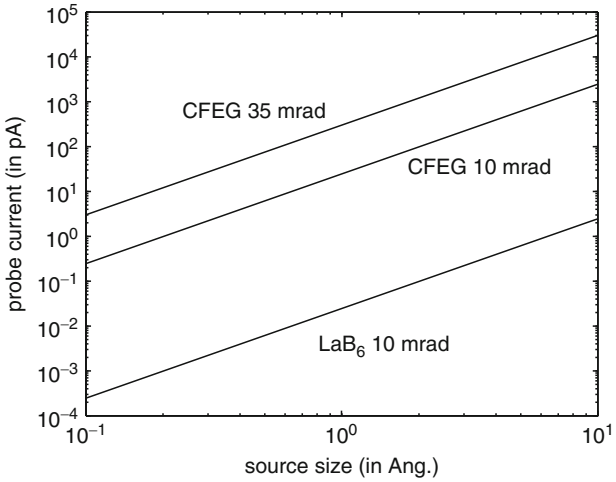

Fig. 3.16 Approximate probe current vs. source size from brightness for a cold field emitter (CFEG) with brightness $\beta = 10^9$ amp/(cm^{2}sr) for two different aperture sizes and a typical LaB_6 source with $\beta = 10^6$ amp/(cm^{2}sr)

Each part of the electron source can emit electrons that travel through the microscope and form an image of their own in some way. Each of these images is offset in position equivalent to the shift in position on the source demagnified by the lenses in the microscope. This can be summarized as a further convolution with an effective source size in the specimen plane (including appropriate demagnification).

$$g(\mathbf{x}) = f(\mathbf{x}) \otimes h_{\mathrm{ADF}}(\mathbf{x}) \otimes h_{\mathrm{source}}(\mathbf{x}) \tag{3.74}$$

The source contribution is typically taken as a Gaussian, and may become the dominate factor at high resolution. If there are other nondirectional instabilities in the microscope then some of these may be treated in a similar manner. For example if the stage has some small random vibrations convolving the final image with a small Gaussian may be an appropriate way to model this effect (stage vibrations likely have a preferred direction so the convolution kernel should match this asymmetry).

3.5.3 Defocus Spread

As in the CTEM, small fluctuations in the accelerating voltage, lens current and the thermal energy spread in the source itself produce a small spread in defocus values. The ADF-STEM transfer function does not vary as dramatically with defocus as the BF-CTEM transfer function so the effect of defocus spread is small on the ADF-STEM transfer function and can frequently be ignored. In a simple approximation the ADF-STEM transfer function (3.70) can be integrated over a small range of defocus values to approx. the effects of these small fluctuations, leaving an effective transfer function:

$$H_{\mathrm{ADF}}^{\mathrm{eff}}(k) = \int_{-\infty}^{\infty} H_{\mathrm{ADF}}(k,\Delta f)p(\Delta f)d(\Delta f), \tag{3.75}$$

where $p(\Delta f)$ is the probability distribution of defocus values. This integration may also be done in real space in this linear imaging approximation. Typically $p(\Delta f)$ is a Gaussian distribution about its mean value, which is easily performed numerically using a Gauss-Hermite quadrature formula (5–9 points are probably sufficient in most cases), which conveniently includes a Gaussian weighting of the integrand (for example Sect. 4.6 of Press et al. [289], or Chap. 25 of Abramowitz and Stegun [1]). Sheppard and Wilson [319] have considered partial coherence in scanning microscopy in a more general manner.

3.6 Confocal Mode for Weak Phase Objects

The light optical scanning transmission microscope (in confocal mode) has exhibited improved performance relative to a conventional microscope, but what happens in the electron microscope? A simple imaging theory is similar to that of the light optical microscope (Wilson and Sheppard [383]). Nellist et al. [262] have also described the theory of imaging in a double corrected instrument.

Confocal mode is a combination of STEM and CTEM (Fig. 2.8). A focused electron probe is raster scanned across the specimen (much like a STEM). The electrons transmitted through the specimen are imaged by a collector lens (much like CTEM) onto a detector (usually a small point like detectors).

The incident probe focused onto the specimen has a complex wavefunction as given by (3.59). The wave transmitted through the specimen is given by (3.61). The probe is scanned across the specimen, and in confocal mode there is also a collection lens that images the transmitted probe onto a detector. The collection lens adds a second complex points spread function. If the probe is scanned (moved) across the specimen the transmitted beam must be inversely scanned so that there is no net motion on the detector. Alternately the electron beam may be fixed and the specimen moved in a raster fashion. The theory is slightly less complicated in this alternate mode (used here). The transmitted wave function is:

$$\psi_T(\mathbf{x},\mathbf{x}_{\mathrm{P}}) = h_{\mathrm{P}}(\mathbf{x})t(\mathbf{x}+\mathbf{x}_{\mathrm{P}}), \tag{3.76}$$

where $\mathbf{x}$ is position in the specimen plane and $\mathbf{x}_\mathrm{P}$ is the position of the specimen (or probe). $t(\mathbf{x})$ is the complex transmission function of the specimen, and $h_\mathrm{P}(\mathbf{x}) = \psi_P(\mathbf{x})$ is the point spread function of the probe forming lens (3.59). The wave function incident on the detector is:

$$\begin{aligned}\psi_D(\mathbf{x},\mathbf{x}_\mathrm{P}) &= h_\mathrm{C}(\mathbf{x}) \otimes [h_\mathrm{P}(\mathbf{x}) t(\mathbf{x}+\mathbf{x}_\mathrm{P})] \\ &= \int h_\mathrm{C}(\mathbf{x}') h_\mathrm{P}(\mathbf{x}'-\mathbf{x}) t(\mathbf{x}'-\mathbf{x}-\mathbf{x}_\mathrm{P}) \mathrm{d}^2\mathbf{x}' \end{aligned} \tag{3.77}$$

where $h_\mathrm{C}(\mathbf{x})$ is the point spread function of the collector lens. The subscript P or C refer to the probe or collector lens, respectively. If the detector is a point at position $\mathbf{x} = 0$ then:

$$\begin{aligned}\psi_D(\mathbf{x}=0,\mathbf{x}_\mathrm{P}) &= \int h_\mathrm{C}(\mathbf{x}') h_\mathrm{P}(\mathbf{x}') t(\mathbf{x}'-\mathbf{x}_\mathrm{P}) \mathrm{d}^2\mathbf{x}' \\ &= [h_\mathrm{C}(\mathbf{x}) h_\mathrm{P}(\mathbf{x})] \otimes t(\mathbf{x}). \end{aligned} \tag{3.78}$$

In practice the detector must have a nonzero size so this is an unphysical approximation to simplify the mathematics. The transmitted intensity for a given probe position with this approximation is therefore:

$$g(\mathbf{x}) = \left|[h_p(\mathbf{x}) h_c(\mathbf{x})] \otimes t(\mathbf{x})\right|^2, \tag{3.79}$$

where the distinction between $\mathbf{x}$ and $\mathbf{x}_\mathrm{P}$ has been dropped for simplicity.

A transfer function is only defined for a linear system but this is still rather nonlinear. Approximate the transmission function of the specimen as a weak-phase object as in (3.6) to obtain a simplified linear theory:

$$t(\mathbf{x}) = \exp[\mathrm{i}\sigma v_z(\mathbf{x})] \sim 1 + \mathrm{i}\sigma v_z(\mathbf{x}) \tag{3.80}$$

or weak-amplitude object (3.47):

$$t(\mathbf{x}) = \exp[u(\mathbf{x})] \sim 1 + u(\mathbf{x}) \tag{3.81}$$

Using the weak phase object approximation, and keeping only lowest terms in v_z, the image model becomes:

$$g(\mathbf{x}) \sim \left|[1 + \mathrm{i}\sigma v_z(\mathbf{x})] \otimes (h_p(\mathbf{x}) h_c(\mathbf{x}))\right|^2 \tag{3.82}$$

$$\sim 1 \otimes (h_p h_c) + 2Re[i\sigma v_z \otimes (h_p h_c)] \tag{3.83}$$

$$\sim 1 + 2\sigma v_z \otimes h_\mathrm{WPA} \tag{3.84}$$

$$h_\mathrm{WPA}(\mathbf{x}) = Re[i h_p(\mathbf{x}) h_c(\mathbf{x})]. \tag{3.85}$$

The transfer function is just the 2D Fourier transform of h_{WPA}. Using azimuthal symmetry the transfer function for weak phase objects in confocal mode is:

$$H_{\mathrm{WPA}}(k) = \int_0^{r_{\max}} \mathrm{Re}[ih_p(r)h_c(r)]J_0(2\pi kr)r\mathrm{d}r \tag{3.86}$$

Remember that both $h_p(\mathbf{x})$ and $h_c(\mathbf{x})$ are complex valued functions. To repeat this derivation in the weak amplitude approximation just remove the factor of i, to obtain the transfer function for weak amplitude objects:

$$H_{\mathrm{WAA}}(k) = \int_0^{r_{\max}} \mathrm{Re}[h_p(r)h_c(r)]J_0(2\pi kr)r\mathrm{d}r \tag{3.87}$$

Confocal mode has twice as many optical parameters and can produce a larger variety of features in the transfer function. Some examples are shown in Fig. 3.17 for the same set of parameters as ADF-STEM in Fig. 3.12. The confocal transfer function is compared to the BF-CTEM transfer function (both in the weak phase approximation) in Fig. 3.18. This particular choice of parameters produces a transfer function similar to ADF-STEM except that it is reversed in contrast (negative instead of positive). There is most likely a difference in the signal strength as well which is not apparent in this representation.

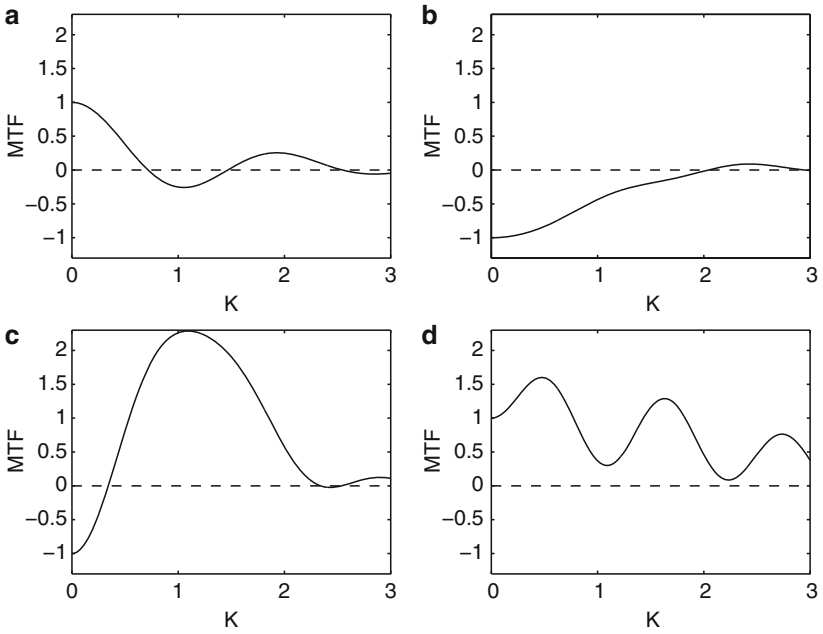

Fig. 3.17 Confocal transfer functions in the weak phase approximation (3.86) for the same aberrations as ADF-STEM (a) $D = 1.2$, $K_{\max} = 1.56$ (Scherzer conditions), (b) $D = 0.80$, $K_{\max} = 1.22$, (c) $D = 1.5$, $K_{\max} = 1.5$, (d) $D = 2.5$, $K_{\max} = 2.5$. Both the probe and collectors lens aberration are the same in each graph

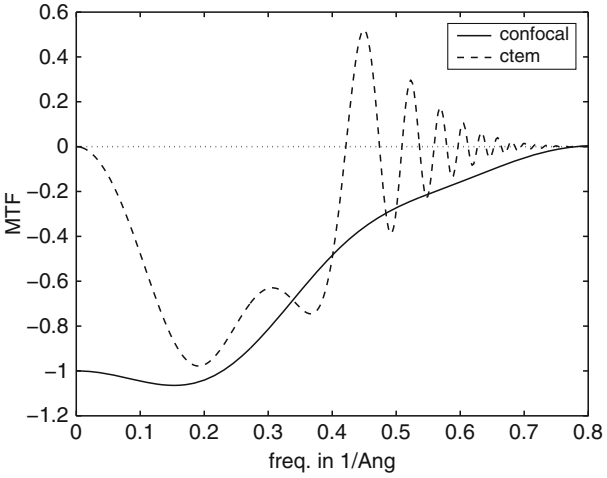

Fig. 3.18 Comparison of confocal and CTEM transfer functions in the weak phase approximation for the same spherical aberration, $C_s = 1.2$ mm, at 200 keV. $\Delta f = 550$ Å and $\alpha_{\max} = 9.5$ mrad for confocal. Both the probe and collectors lens aberration in confocal were the same. $\Delta f = 670$ Å for BF-CTEM with a condenser half angle of 0.1 mrad and a defocus spread of 100Å (Scherzer conditions)

3.7 Phase and Amplitude Contrast Revisited

The theory presented earlier is a traditional view and seems plausible, but it is worth checking a little. Figure 3.19 shows how a pure amplitude and pure phase object (combined into a single image) are imaged in BF-CTEM and ADF-STEM, using methods that will be discussed later. The letters "AMP." are a amplitude object and the letters "PHASE" are a phase object. The image transmission function is given by:

$$t(\mathbf{x}) = \exp(\mathrm{i}p - a), \tag{3.88}$$

where a and p are the amplitude and phase components (respect.) with a value of 0.05 on the appropriate letters and zero otherwise. The BF-CTEM and the ADF-STEM images contain both amplitude and phase contrast features. It is more the specimen that determines which form of contrast will be in the image. The ADF-STEM image only renders the high frequency components from the edges (scattered at high angle onto the ADF detectors). However, every specimen is composed of many small atoms, each of which scatters to high angle so this edge effect is not in a real specimen composed of many individual atoms (sharp edges, or points are everywhere). The simple theory is very helpful in developing an intuitive understanding of electron microscope image but there is still a need for more detailed simulations to more completely understand what the images mean. The property of coherence and incoherence also plays a large role in image formation.

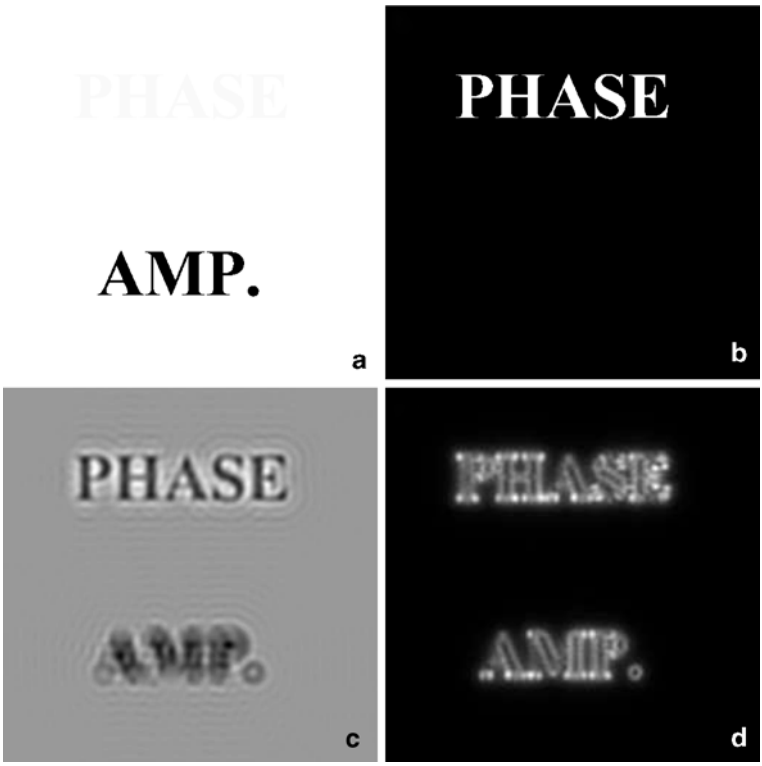

Fig. 3.19 Phase and amplitude contrast in BF-CTEM and ADF-STEM. (**a**) Pure amplitude object, (**b**) pure phase object. The object was formed from the superposition of (**a**) and (**b**) and imaged as a (**c**) BF-CTEM image and (**d**) ADF-STEM image