Chapter 3

Scanning Transmission Electron Microscopy

Shortly after the invention of the broad-beam illumination transmission electron microscope by Knoll and Ruska (1932), Manfred von Ardenne (1938a), a German physicist, noticed the great potential of using a focused probe rather than a broad beam to study microscopic objects by electrons. Von Ardenne, who was awarded for the invention of the table-top electron microscope by the former Soviet Union, was active as an inventor in various fields of physics, including research in communication and radar technology as well as in medical physics. After the Second World War, he conducted research in nuclear and plasma physics for the Soviet Union, whereupon he returned to Germany in 1953. Without trying to elucidate the historical details of the early years of electron microscopy, it can be summarized that from the invention of the electron microscope in 1932 and von Ardenne’s first electron optical instrument, which made use of a focused electron probe that was scanned across a specimen (von Ardenne, 1938b), it took roughly 30 years to the realization of a dedicated scanning transmission electron microscope which was capable of producing results of similar quality to the broad-beam equivalent. The pioneering work of Albert V. Crewe (1966) marks the beginning of practical scanning transmission electron microscopy (STEM). From then on, the scanning probe mode was developed as a complementary technique to the broad-beam illumination mode. Although the actual realization of STEM has a handicap of roughly 30 years, its fast development has certainly benefited from the electron optical know-how derived from the broad-beam illumination mode. Nowadays, STEM can be regarded as a powerful operation mode which, on many state-of-the-art electron microscopes, i.e. STEM/TEM instruments, provides a wealth of complementary information that elucidates the properties of a material from a slightly different point of view.

3.1 Overview

In STEM, information about the specimen is collected in a serial acquisition mode. The specimen is illuminated with a convergent electron beam which is focused to a small spot at the height of the specimen (see Fig. 3.1). To record an image, the
electron probe is scanned within a rectangular frame on the specimen. On each scan position, the electron probe is propagated through the specimen. As a consequence of electron scattering within the specimen, part of the electrons are scattered away from their initial trajectories. The scattering distribution of the electrons in the far field behind the specimen corresponds to a diffraction pattern. Since the electron probe is convergent, the diffraction pattern is a convergent electron diffraction pattern. If the crystal spacing is large enough, or if the convergence angle of the electron beam is sufficiently large, the diffraction disks in the diffraction pattern partially overlap. Indeed, as will be shown below, the coherent partial overlap of diffraction disks is a requirement for resolving a given crystal spacing in an atomic-resolution scanning transmission electron micrograph.

Each scan position produces a site-specific diffraction pattern. If the electron beam is positioned on an area that contains strong scatterers, like heavy atoms, the intensity at high scattering angles is enhanced, whereas for the case that the electron beam is positioned on an area of weak scatterers or on a thin area, the scattering intensity at high angles is low and the intensity in forward direction and at small scattering angles remains high. This is a very pictorial explanation, and generally the propagation of the focused electron beam through the specimen and the intensity distribution in the diffraction pattern are complex matters which require consideration of dynamical scattering, channelling effects as well as quasi-elastic scattering, such as thermal diffuse scattering. Nonetheless, for our present purpose, where we do not focus on the electron-specimen interaction, this simple explanation shall suffice. For an in-depth discussion of the image formation in STEM and its dependence on the specimen we refer to more specific literature (Rose, 1975; Fertig and Rose, 1981; Nellist and Pennycook, 2000).

A two dimensional STEM micrograph corresponds to a two-dimensional array of data points. Each of these data points reflects the detector signal collected during the dwell time, i.e. during the time the electron probe was stationed on a given scan position. The information contained in a STEM micrograph depends on the position and size of the detector in the diffraction plane. Let us assume we have an infinitely large detector which detects all electrons in the diffraction plane behind the specimen. Neglecting back-scattered electrons, the detector would thus produce a constant signal and we would not learn anything about the specimen. Hence, while scanning the focused electron beam across the specimen, only a certain part of the intensity of the diffraction pattern is recorded as a function of the beam position. The choice of the area of the diffraction pattern that is positioned on the detector determines the image contrast.

Positioning a circular electron detector on the forward scattered beam yields a bright-field (BF) scanning transmission electron micrograph (see Fig. 3.1). For the case where there is no scatterer in the path of the beam, the BF signal reflects the total beam current, whereas for the case that there are scatterers in the path of the beam, the BF signal corresponds to the total beam current minus the integrated
intensity that is scattered to angles beyond the area of the bright-field detector. Alternatively, an annular detector can be used which, instead of detecting the forward scattered beam, records an annular dark-field (ADF) signal. The BF and the ADF signals are in a qualitative way complementary to each other. If there is no scatterer in the path of the beam, the ADF intensity is zero, and for the case that there is scattering, the ADF signal reflects the scattering power — for the selected angular range — of the object that is in the path of the beam. Hence, the ADF signal increases with the scattering factor of the elements in the specimen as well as with the thickness of the specimen. The angular range of the annular detector, which in general is a fixed detector of a given size, can be adjusted by changing the camera length, i.e. the magnification by which the diffraction pattern is projected onto the detector.

A widely applied STEM imaging mode concerns the case for which the ADF detector is setup such that it collects over a large angular area electrons scattered to high angles. The integration of the high-angle scattering over a large area warrants that coherence effects between the diffracted beams are averaged out.
Hence, the corresponding high-angle annular dark-field (HAADF) STEM micrograph essentially reflects an incoherent signal (Rose, 1975; Hartel et al., 1996; Nellist and Pennycook, 1999). Though (incoherent) thermal diffuse scattering contributes to the high-angle scattering, it is not fundamental in explaining the incoherence of the detected signal (Loane et al., 1992; Hartel et al., 1996; Nellist and Pennycook, 1999; Muller et al., 2001). The crucial point that enables a largely incoherent signal is the size of the detector\(^1\). The incoherence of the HAADF STEM signal makes the specimen appear self-luminous. This simplifies image interpretation. Moreover, since the high-angle electron scattering is dominated by Rutherford scattering, the scattered intensity scales with the atomic number \(Z\) of the elements in the sample. For pure Rutherford scattering, one expects a \(Z^2\) dependence of the signal (Schwartz and Cohen, 1987). Experiments and calculations reveal that the actual exponent is around 1.6–1.8 instead of 2 (Hillyard and Silcox, 1995; Rafferty et al., 2001; Erni et al., 2003b). This difference can be explained by the fact that the electron cloud surrounding the nucleus screens the Coulomb potential of the nucleus, which is of relevance for Rutherford scattering (Hartel et al., 1996).

For a specimen of constant thickness, a HAADF STEM micrograph maps the atomic number of the elements in the specimen. Due to its favorable atomic-number dependence, HAADF STEM is usually referred to as \(Z\)-contrast imaging (Nellist and Pennycook, 2000).

Apart from the common BF, ADF and HAADF detector settings, special detector setups have been discussed in the literature which, for instance, are suitable for enhancing the contrast of light atoms (Cowley et al., 1996) or can be used for phase contrast imaging in STEM (Rose, 1974).

However, independent of the detector, the critical part of the scanning transmission mode is the characteristics of the focused electron beam. If the electron beam can be focused to a probe that is of the size of the atomic spacing of a zone-axis oriented crystal, a STEM micrograph reveals modulations which correspond directly to the atomic spacing of the crystal. Hence, it is the electron probe which is decisive for the resolution in STEM; the smaller the electron probe, the better the lateral resolution. Furthermore, similar to HRTEM, it is the characteristics of the objective lens that are of fundamental importance to achieve a small electron probe. However, as can be seen from Fig. 3.1, it is not the post-field that is relevant for the electron probe, but the pre-field of the objective lens.

In the following sections we draw our attention to the central point of STEM imaging which is the formation of the electron probe. Similar to Chapter 2, the

\(^1\)The formation of an incoherent image in HAADF STEM can be explained by employing the principle of reciprocity. Consider the following situation: a source is placed in point A which emits a wave I. The wave is scattered at point P and arrives at point B. The principle of reciprocity states that the amplitude of wave I in point B is equal to the amplitude of a wave II in point A if the source is placed in B (Pogany and Turner, 1968). On the basis of the principle of reciprocity, it can be shown that a large, i.e. spatially incoherent, electron source in TEM is equivalent to a large detector in STEM (Cowley, 1969). Both the large electron source and the large detector provide an incoherent image.
electron-specimen interaction is not discussed in detail, i.e. the propagation of the
electron probe through the sample and the detection of the scattered intensity to
form a scanning transmission electron micrograph are not discussed in this context.

3.2 Geometrical Considerations

In the previous chapter on HRTEM imaging, we saw that the information transfer
in phase contrast imaging is determined by the aperture function, the characteristics
of the objective lens and by the limited degree of coherence of the electron beam,
namely by the partial temporal coherence and the partial spatial coherence. In the
following we will see that these four factors equivalently determine the character-
istics of the electron probe and thus the information transfer in STEM imaging.
We start discussing these effects from a geometrical point of view. The geometrical
treatment of the individual contributions is particularly useful to understand their
impact on certain microscope parameters (Crewe, 1987, 1997). However, in order
to describe the combined effect of these contributions, it is essential to switch to a
wave optical description of the electron probe. This will be done in the subsequent
section.

3.2.1 The diffraction limit

In a twin-type objective lens, the specimen is immersed in the magnetic field formed
by both the pre- and post-field of the objective lens. In analogy to the treatment
of TEM imaging (see Chapter 2), we can simplify this situation by treating the
pre- and post-field of the objective lens separately. With this simplification, the
formation of an electron probe is determined on how the pre-field of the objective
lens focuses the electron beam onto the specimen plane. The focused electron beam
is the electron probe. As illustrated in Fig. 3.1, the electron probe and the specimen
can be considered to be located in the back focal plane of the objective lens’ pre-field.

In the broad-beam TEM mode, the post-field of the objective lens produces
diffraction pattern in the back focal plane of the objective lens’ post-field (see
Fig. 2.1). The back focal plane containing the diffraction pattern is conjugate to
the plane of the electron source and a diffraction spot can be regarded as an image
of the source. For STEM, the electron probe is located in the back focal plane of
the objective lens’ pre-field and, similar to TEM, the electron probe represents a
demagnified image of the electron source. This image is not perfect. One of the
factors explaining why there is no stigmatic image of the source is that there is an
aperture present which limits the angular range of the illuminating electrons that
form the electron probe.

Figure 3.1 shows that the effect of the illumination aperture is to control the il-
lumination (or convergence) angle of the electron probe. As long as the aperture de-
finess the illumination angle of the focused electron beam on the height of the object,
its location along the optical axis is not critical. If the aperture is approximately illuminated by a parallel beam, the electron probe at the object plane is an Airy pattern (see, e.g. Born and Wolf, 2001). In the presence of aberrations, an illumination aperture of finite size is needed to optimize the size of the electron probe. Therefore, an Airy-pattern-type electron probe is practically unavoidable.

The Airy pattern is dominated by a central maximum surrounded by concentric side lobes of distinctly lower intensity (see Fig. 3.2a). In order to relate the characteristics of the Airy pattern to the size of the electron probe, we can choose the first zero of the Airy pattern as the radius $\delta_D$ of the diffraction-limited electron probe. This can be written as

$$\delta_D = \frac{0.61 \lambda}{\alpha},$$

where $\lambda$ is the electron wavelength given in Eq. (2.2) and $\alpha$ is the illumination (or convergence) semi-angle defined by the aperture opening (see Fig. 3.1). The value $\delta_D$ expresses the size of an electron probe, which is solely determined by

![Fig. 3.2 Airy pattern. (a) shows an Airy pattern calculated for 200 keV electrons ($\lambda = 2.5$ pm) and an illumination semi-angle $\alpha$ of 5 mrad. In order to reveal the side lobes of the Airy pattern, it is plotted on a logarithmic scale. (b) shows three line profiles through Airy patterns, calculated for 200 keV electrons and illumination semi-angles of 5, 10 and 20 mrad (dashed, dotted and full lines). The first minimum of the curves defines the diffraction limit according to Eq. (3.1). For 5 mrad $\delta_D$ is 305 pm, for 10 mrad it is 153 pm and for 20 mrad it is 76 pm.](http://www.worldscibooks.com/materialsci/p703.html)
the geometry of the (coherent) illumination. The width of such an electron probe increases with increasing $\lambda$ and decreasing $\alpha$. Only for the case that $\alpha \to \infty$ or $\lambda \to 0$ is the electron probe point-like, i.e. $\delta_D \to 0$. Figure 3.2b plots line profiles across (normalized) Airy patterns for three different illumination semi-angles. It clearly reveals that with increasing illumination semi-angle, the central maximum becomes narrower.

The limitation of the probe size due to the illumination semi-angle $\alpha$ expressed in Eq. (3.1) is called the diffraction limit\(^2\). In fact, Eq. (3.1) is the resolution criterion of an optical system which is solely limited by diffraction; it expresses the Rayleigh limit or Rayleigh criterion. The diffraction limit reveals that in order to increase the resolution in STEM imaging, one should work with a large probe illumination angle and employ electrons of high energy.

The diffraction limit in STEM imaging has an alternative, visual interpretation, which can be regarded as a complementary point of view. Let us assume we do STEM imaging with a crystalline specimen which is in some zone-axis orientation. There shall be the forward scattered beam $0$ and a diffracted beam $g$. The scattering angle of the beam $g$ shall be $\theta$ and the illumination semi-angle of the incident electron probe is $\alpha$. Hence, instead of a sharp diffraction spot, the illumination angle of the illumination causes diffraction disks to appear in the diffraction plane; one for the forward scattered beam and one for the diffracted beam $g$. The radius of both disks in the diffraction plane corresponds with the illumination semi-angle $\alpha$. The diffraction angle $\theta$ between $0$ and $g$, i.e. the angle in respect to the specimen plane connecting the centers of the disks in the diffraction plane, is given by Bragg’s law (see, e.g. Schwartz and Cohen, 1987)

\[
\lambda = 2d_g \sin \left( \frac{\theta}{2} \right),
\]

where $d_g$ corresponds to the crystal spacing which gives rise to the diffraction disk $g$. Now we assume that the diffraction disks are just large enough that they touch each other. Hence, the diffraction angle $\theta$ is equal to twice the illumination semi-angle $\alpha$, i.e. $\theta = 2\alpha$ (see Fig. 3.3a). Neglecting the curvature of the Ewald sphere, we can redraw the triangle ABC indicated in Fig. 3.3a and obtain the triangle shown in Fig. 3.3b. The scattering triangle ABC is an equal-sided triangle; the vector $\overrightarrow{AB}$ corresponds to the incident wave vector, $\overrightarrow{AC}$ is the scattered wave vector and $\overrightarrow{BC}$ is the scattering vector. Hence, the lengths of the two sides $AB$ and $AC$ correspond with the wave vector of the incident and elastically scattered electron, which is $\lambda^{-1}$.

\(^2\)The diffraction limit not only affects STEM imaging but is also essential in TEM. With increasing size of the objective aperture, beams of higher spatial frequency can be transferred to the image plane where they are brought to interference. Since the diffracted beams of high spatial frequencies carry the high resolution information, the objective aperture similarly causes the HRTEM resolution to be limited by diffraction. Selecting, for instance, a very small objective aperture, which transmits only the forward scattered beam, simply implies that there is no lattice information in the micrograph. This mode, which is called bright-field zone-axis imaging, is used to map strain fields at high resolution (Matsumura et al., 1990).
and the third side measures $1/d_g$ (see, e.g. Schwartz and Cohen, 1987). Making the approximation for small scattering angles, Fig. 3.3 leads us to a simplified version of Bragg’s law, which can be written as

$$\lambda \approx 2d_g\alpha. \quad (3.3)$$

Here we used the approximation $\sin(\theta/2) \approx \theta/2 = \alpha$. Rewriting this relation yields

$$d_g = \frac{1}{2\alpha}. \quad (3.4)$$

Comparing this relation with the diffraction limit given in Eq. (3.1) clearly shows that $d_g$ is smaller than the diffraction limit $\delta_D$, i.e. $0.5\lambda/\alpha < 0.61\lambda/\alpha$. Hence, the geometrical setup described in Fig. 3.3 does not allow for resolving the $d_g$-spacing.

From this argument we can conclude that in order to resolve a crystal spacing $d_g$ in STEM mode, the corresponding diffraction disk $g$ has to partially overlap with the diffraction disk of the forward scattered beam $0$. The factor 0.61 in Eq. (3.1) essentially describes the amount of overlap that is needed. If the diffraction disks overlap by exactly the amount given by the diffraction limit, the corresponding crystal spacing is theoretically resolved such that it just fulfills the Rayleigh criterion. Provided the contribution of other effects can be ignored, the contrast of the crystal spacing in the micrograph is then 19%.

We can conclude that if diffraction disks do not overlap, the corresponding spatial distance cannot be resolved in a STEM micrograph. This is equivalent to the statement that if the illumination angle of an electron probe is too small, the (diffraction-limited) electron probe is too large to resolve a given spatial distance.
The diffraction limit thus shows that a large illumination angle is needed for a small electron probe, and that in order to resolve a certain crystal spacing, the illumination angle has to be large enough such that the diffraction disk corresponding with the resolvable crystal spacing overlaps with the diffraction disk of the forward scattered beam.

### 3.2.2 Lens effects — spherical aberration

In Chapter 2 we saw that the effect of the rotationally symmetric objective lens in HRTEM can be described by the constant of spherical aberration $C_3$ and by the defocus $C_1$. This is based on the assumption that other effects, like for instance astigmatism, are sufficiently small such that they do not significantly affect the imaging process. While $C_1$ is a variable parameter, $C_3$ is a fixed quantity characteristic for the lens. For TEM imaging, it is the spherical aberration $C_3$ of the post-field of the objective lens which is of importance. For the formation of the electron probe in STEM, it is the spherical aberration of the pre-field of the objective lens. Since $C_1$ can be adjusted, $C_3$ imposes the actual limit.

We can consider the pre-field of the objective lens to produce an image of a point-like electron source. The image of the source is the electron probe. An ideal lens focuses all the rays emerging from the source point in the object space in one single point in the image space. The effect of positive spherical aberration is that rays that pass the lens in a distance from the optical axis are brought to focus closer to the lens than rays that run near the optical axis. One can say that the focal distance of the lens decreases with increasing off-axial distance of the rays entering the lens field. The focal point of the rays that run in an infinitely small distance from the optical axis through the lens defines the Gaussian focal plane. The bundle of electrons in Fig. 3.4 emerges from a point source, passing an aperture, and is brought to focus by a lens suffering from spherical aberration. The aperture defines the illumination semi-angle $\alpha$. In the Gaussian focal plane, where one assumes the specimen to be located on to which the electron beam should be focused, there is a broad disk instead of a point-like image of the point-like electron source. Furthermore, in a certain distance in front of the focal plane there is an area where the envelope of all the rays forms a disk which is clearly smaller than the disk in the Gaussian focal plane. It is this disk which defines the smallest achievable electron probe limited by spherical aberration. This disk of radius

$$\delta_S = \frac{1}{4} C_3 \alpha^3$$

is called the disk of least confusion. It expresses the limitation imposed by the spherical aberration $C_3$ on the achievable probe size for a given probe illumination semi-angle $\alpha$. One easily sees that the smallest disk of least confusion is obtained for vanishing $\alpha$; the smaller the illumination semi-angle, the smaller is $\delta_S$. Apparently, this trend goes in the opposite direction compared with the diffraction limit.
Furthermore, since the defocus $C_1$ describes the deviation of the focus from the Gaussian focal plane, Fig. 3.4 clearly reveals that in order to minimize the effect of spherical aberration, one should work at a finite defocus which essentially moves the disk of least confusion onto the specimen plane.

### 3.2.3 Partial temporal coherence — chromatic aberration

An ideal electron source emits electrons of equal energy. However, real electron sources emit electrons of slightly varying energy and thus exhibit a characteristic energy distribution. The energy distribution of the electron beam can further be influenced by the high-tension ripple. Though not strictly valid, the energy distribution of the electrons can approximately be described by a Gauss function around a nominal electron energy $E_0 = eU$. However, the actual distribution function depends on the type of electron source and its operation condition. For field-emission electron microscopes, the width of the energy distribution, which is often quantified by the full width at half maximum, is better than 1 eV. Employing an electron monochromator, the energy width of the beam can be reduced to values below 100 meV.

The problem of using a non-monochromatic electron beam for STEM (or TEM) imaging lies in the fact that electron lenses are not achromatic. They suffer from the chromatic aberration. This means, as illustrated in Fig. 3.5, that the focal point of the lens depends on the energy of the electrons. The focal point of electrons with
the nominal electron energy $E_0$ lies in the Gaussian image plane. Electrons of energies greater than $E_0$ find a focal point behind the Gaussian focal plane, and electrons of smaller energy in front of this plane.

Hence, if there is a point source in front of a lens, and the lens is used to form a small electron probe, the electron probe is not point-like. In the Gaussian focal plane of the pre-field of the objective lens, i.e. where the specimen for the STEM investigation is located, the point source is imaged into a disk of confusion. The diameter of this disk increases with increasing energy spread. Furthermore, with decreasing opening angle of the aperture and thus increasing illumination semi-angle $\alpha$, the diameter of the disk decreases. In general, the radius of the disk of confusion $\delta_C$ due to the chromatic aberration is given by

$$\delta_C = C_C \frac{\Delta E}{E_0},$$

where $C_C$ is the constant of chromatic aberration of the lens and $\Delta E$ is a measure for the width of the energy distribution.

Scherzer’s theorem (Scherzer, 1936b) states that in rotationally symmetric and stationary electromagnetic lenses, which are free of space charges, the constant of spherical aberration $C_3$ and the constant of chromatic aberration $C_C$ are finite positive and can never be nulled. Hence, in conventional electron microscopes, the disks of confusion due to spherical and chromatic aberrations are unavoidable and thus affect the achievable size of the electron probe used in STEM.
3.2.4 \textit{Partial spatial coherence — the effective source size}

Real electron sources not only show a finite energy spread but are also finite in size and not point-like. Independent of how many lenses are between source and specimen, a STEM probe is always an image of the source. Because the size of the source is finite, the STEM probe can never be point-like. This is independent of the spherical and chromatic aberrations as well as independent of the diffraction limit. One can argue that sufficient, maybe even infinite demagnification can be employed to obtain a point-like electron probe. However, this leads to an electron probe of zero current. If we recall from Chapter 2 the definition of the brightness in Eq. (2.25) given as

\[ B = \frac{I_s}{\Omega A_s} \]

and, by making the approximation that the solid angle \(\Omega \approx \pi \alpha^2\), transform it to

\[ B = \frac{I_{\text{probe}}}{\pi \alpha^2 A_s}, \] \hspace{1cm} (3.7)

we see that the probe current \(I_{\text{probe}}\) depends linearly on the brightness \(B\). We have still assumed that the source is imaged 1:1 to the specimen plane. If we demagnify the source of area \(A_s\) to a circular area of radius \(r_{\text{geo}}\), we see that the current \(I_{\text{probe}}\) of the electron probe is

\[ I_{\text{probe}} = B \frac{\pi \alpha^2 r_{\text{geo}}^2}{M}, \] \hspace{1cm} (3.8)

where we set \(\pi r_{\text{geo}}^2 = MA_s\), with \(M\) expressing the (de-)magnification of the source. If we apply a high demagnification (\(M \ll 1\)), \(r_{\text{geo}} \rightarrow 0\) and we end up with a point-like electron source. However, for a given illumination semi-angle \(\alpha\), Eq. (3.8) shows that if \(r_{\text{geo}} \rightarrow 0\), a finite probe current \(I_{\text{probe}}\) can only be maintained for the unphysical case that \(B \rightarrow \infty\). Electron sources of infinite brightness do not exist. Demagnification comes at the expense of beam current. Yet, this argument explains why electron sources of high brightness are especially important for STEM. The higher the brightness of the source, the larger the probe current that can be maintained for a sufficiently small effective source radius \(r_{\text{geo}}\). Sources of high brightness, like Schottky field-emission sources and cold field-emission sources, which have a greater brightness than Schottky emitters, are thus indispensable for high-resolution STEM instruments. The importance of the source brightness in STEM is underlined by the fact that atomic-resolution STEM imaging has only become feasible with field-emission sources. Indeed, it was essentially the development of the cold field-emission source that enabled the first STEM instrument to produce results comparable to the TEM mode, where brightness is also of high importance but not as crucial as it is for STEM (Crewe, 1966).

The demagnification of the electron source is usually done by employing either an (electrostatic) gun lens and/or the first condenser lens, which is the so-called spots-size lens. With increasing excitation of the lens, the demagnification is increased. A
higher demagnification refers to a higher spot size number. This can be seen from the schematics in Fig. 3.6. In reality, of course, the emission area $A_s$ of the source is not a top-hat function which has sharp edges. The source is usually described by a Gaussian source distribution function and $\delta_{\text{geo}}$, the full width at half maximum of the Gaussian function, is a measure for the width of the effective source size, i.e. the size of the image of the source on the specimen.

3.2.5 Stability

An additional factor which can contribute to the effective size of the source imaged onto the specimen is the overall impact of disturbances. We can distinguish between two types of disturbances; electromagnetic disturbances and mechanical vibrations. Electromagnetic fields, which can be present as stray fields in the environment of the microscope or can be caused by an unstable lens or deflector, can cause the electron beam, and thus the electron probe, to jitter. On the other hand, mechanical instabilities become apparent if the specimen is unstable in respect to the electron beam. This, for instance, can be caused by an unstable sample holder or by thermal drift of the specimen.

If disturbances occur in periods shorter than the dwell time of the scan process, the effective electron probe that a particular scan position experiences becomes larger than the actual (instantaneous) geometrical size of the electron probe. In this case, the effective source size is enlarged by the blurring due to the disturbances. If high-frequency disturbances are present which enlarge the effective source size, $\delta_{\text{geo}}$ can be replaced by $\delta_{\text{geo,eff}}$ where $\delta_{\text{geo,eff}}^2 = \delta_{\text{geo}}^2 + \delta_{\text{noise}}^2$. The term $\delta_{\text{noise}}$ describes the blurring due to the disturbances. If disturbances occur in periods longer than the dwell time, they become apparent either as (periodic) noise in the image or
as distortions. The latter effect becomes apparent if disturbances have periods exceeding multiples of the line time\(^3\), or if the specimen and/or the electron beam are prone to a continuous drift. While high-frequency disturbances are difficult to detect in an image, disturbances of lower frequencies can be revealed by taking a fast Fourier transform (FFT) of an atomic-resolution STEM micrograph. Random scan noise is revealed by streaks along the slow scan direction, while periodic scan noise can lead to false crystal reflections in the FFT.

### 3.2.6 Small electron probes

The resolution in STEM imaging is fundamentally limited by the size of the electron probe. Two objects which are at a distance smaller than the size of the electron probe cannot be resolved. A smaller electron probe enables higher resolution. The task of optimizing the resolution of a scanning transmission electron microscope means finding an optical setting for which the overall effect of the probe-limiting factors discussed above is minimal.

For a given microscope high tension and for a given demagnification of the source, the wavelength \(\lambda\) and the effective source size expressed by \(\delta_{\text{geo}}\) are fixed. The remaining parameters which need to be considered in the optimization of an electron probe are the size of the illumination aperture expressed by \(\alpha\), the geometrical lens parameters \(C_3\) and \(C_1\) and the chromatic aberration \(C_C\). Each contribution has a specific dependence on the illumination semi-angle \(\alpha\).

Figure 3.7 shows a log–log plot visualizing the dependencies of \(\delta_{\text{geo}}\), \(\delta_D\), \(\delta_S\) and \(\delta_C\) on the illumination semi-angle \(\alpha\), assuming \(C_3 = 1\) mm, \(C_C = 1\) mm and \(\lambda = 2.5\) pm, i.e. \(E_0 = 200\) keV with \(\Delta E = 1\) eV and \(\delta_{\text{geo}} = 50\) pm. This set of parameters is chosen arbitrarily. Nevertheless, the parameters are quite common for conventional scanning transmission electron microscopes equipped with Schottky field-emission electron sources. Hence, though we cannot deduce general rules from a single set of parameters, we can still see which factors are of relevance for a certain range of \(\alpha\). Furthermore, since the slopes of the curves in Fig. 3.7 depend on the power \(n\) of \(\alpha^n\) in the expressions for \(\delta_D\), \(\delta_S\) and \(\delta_C\), one has to be aware that a change of one of the parameters only leads to a parallel shift of the corresponding line in the log–log plot.

The plot in Fig. 3.7 reveals that for the parameters selected, the diffraction limit imposes the limit at small illumination semi-angles, while for larger \(\alpha\) it is the spherical aberration which becomes the probe-size limiting quantity. The impact of the chromatic aberration is not critical provided, of course, that the constant

\(^3\)We denote the fast scan direction as the direction along the electron probe scans the first line in the frame. The slow scan direction is perpendicular, along the direction which is consecutively filled by scanned lines. The dwell time is the period the beam is stationary on a scan position. The line time is the dwell time multiplied by the amount of pixels along the fast scan direction. Multiplying the line time with the amount of scanned lines gives the approximate frame time. It is the approximate frame time because the actual frame time also depends on the scan synchronization. Often, frames are chosen that are squares of sides which contain \(2^n\) pixels (\(n = 8, 9, 10...\)).
Fig. 3.7 Contributions to the STEM probe. Dependency of the diffraction limit $\delta_D$, the spherical aberration $\delta_S$, the chromatic aberration $\delta_C$ and the effective source diameter $\delta_{geo}$ on the illumination semi-angle $\alpha$, according to Eqs. (3.1), (3.5) and (3.6). The geometrical source size $\delta_{geo}$ is assumed to be 50 pm, $C_3 = C_1 = 1$ mm, $\Delta E = 1$ eV, $E_0 = 200$ keV and thus $\lambda = 2.5$ pm.

of chromatic aberration $C_C$ is of the same order of magnitude as the constant of spherical aberration $C_3$ and provided that the energy spread $\Delta E$ is of the order of 1 eV. This latter condition can be considered to be fulfilled for the case of field-emission electron sources. Hence, the chromatic aberration is not the limiting factor in conventional probe-forming microscopes operated above about 100 kV. This has been investigated in detail by Shao and Crewe (1987). Furthermore, because $\delta_S \propto \alpha^3$ while $\delta_C \propto \alpha$, the impact of the spherical aberration must exceed the impact of the chromatic aberration with increasing illumination semi-angle $\alpha$. The effective source size, or the demagnification, can in principle be chosen such that the finite size of the source is not limiting the probe size. This, of course, comes at the expense of probe current.

From Fig. 3.7 we can conclude that what essentially needs to be considered in the optimization of an electron probe in a conventional scanning transmission electron microscope are the diffraction limit and the spherical aberration. These two contributions need to be balanced. Furthermore, a defocus $C_1$ needs to be chosen, similar to the Scherzer focus in Chapter 2, which translates the disk of least confusion to the specimen plane. Crewe and Saltzman (1982) solved this problem...
and derived that for an optimum defocus $C_{1,\text{opt}}$ of
\[
C_{1,\text{opt}} = -\sqrt{\lambda C_3}
\] (3.9)
and for an optimum illumination semi-angle $\alpha_{\text{opt}}$ of
\[
\alpha_{\text{opt}} = \left(\frac{4\lambda}{C_3}\right)^{\frac{1}{2}}
\] (3.10)
a resolution $\rho_r$ of
\[
\rho_r = 0.43\sqrt{C_3\lambda^3}
\] (3.11)
can be achieved. While experimentally the optimum defocus can be found by optimizing the contrast of an atomic-resolution STEM micrograph, the optimum illumination semi-angle given in Eq. (3.10) needs to be selected carefully in order to achieve the smallest probe size. Only the optimization of the illumination semi-angle in regard to the spherical aberration enables highest resolution in a conventional scanning transmission electron microscope.

For the set of parameters given above, we obtain for the optimum semi-illumination angle $\alpha_{\text{opt}} = 10$ mrad. This value, which is quite typical for conventional STEM microscopes, roughly coincides with the $\alpha$-value of the point of intersection between the lines $\delta_D$ and $\delta_S$ in Fig. 3.7. Provided the defocus is set to -50 nm according to Eq. (3.9), Eq. (3.11) reveals that the electron probe formed with the optimum illumination semi-angle enables a STEM resolution of 0.15 nm.

Comparing Eqs. (3.9) and (3.11) with the equivalent expressions for HRTEM given in Eqs. (2.21) and (2.32) reveals that both expressions for the optimal focus setting and the expression for the resolution are very similar. Though the relations for an optimum STEM probe given here were derived by Crewe and Salzman (1982), these expressions, as well as the expression for the optimum illumination semi-angle, were essentially derived by Scherzer (1939, 1949). While the optimal focus setting for HRTEM given in Eq. (2.21) is known as the Scherzer focus, the conditions for an optimized STEM probe, which consists of an expression for the focus setting and a requirement for the illumination semi-angle given in Eqs. (3.9) and (3.10) respectively, are sometimes referred to as the Scherzer incoherent conditions (Pennycook and Jesson, 1991).

A final point about the geometrical considerations of STEM probes concerns the quantities $\delta_D$, $\delta_S$, $\delta_C$ and $\delta_{\text{geo}}$, which are given in Eqs. (3.1), (3.5) and (3.6). As pointed out by Crewe (1997), these quantities should be distinguished from actual resolution expressions, like the one for $\rho_r$ given in Eq. (3.11). The values $\delta_D$, $\delta_S$ and $\delta_C$ do not express the achievable STEM resolution for a given illumination semi-angle. We should regard these quantities as measures for the impact of a certain effect, rather than as resolution criteria. This can easily be seen from comparing $\delta_S$ with $\delta_D$. While $\delta_S$ corresponds to the outermost radius of a caustic which in general is sharply peaked and decays fast towards the edge of the caustic, the $\delta_D$-value...
provides a measure for the width of the central peak of an Airy pattern, knowing that there is intensity beyond the $\delta_D$-radius. Hence, $\delta_S$ and $\delta_D$ are not directly comparable; in the first case we measure the probe to the very end of the weak tails, while in the second case we basically ignore the tails of the probe. The important point about $\delta_D$, $\delta_S$, $\delta_C$ and $\delta_{geo}$ is that they express the basic effects of $C_3$, $C_C$ and $\alpha$ on the STEM probe. For this reason, it does not seem appropriate to define the size of the electron probe as well as the STEM resolution simply by taking the geometrical mean of all four quantities $\delta_D$, $\delta_S$, $\delta_C$ and $\delta_{geo}$ as

$$\sqrt{\delta_D^2 + \delta_S^2 + \delta_C^2 + \delta_{geo}^2},$$

which, however, can quite frequently be found in the literature. Indeed, looking at the plot in Fig. 3.7 reveals that there is no point in the plot that would reflect such an electron probe.

Furthermore, the effects of diffraction, the spherical and chromatic aberrations and the finite size of the source can in general not be treated independently. They are highly interrelated, which, in particular, becomes apparent if the electron probe is not solely considered as a two-dimensional focused electron spot but as a three-dimensional entity, which, apart from the lateral extension, also has a longitudinal or vertical component. Hence, the above geometrical considerations about the lateral extension of an electron probe have clear limits. Nonetheless, the simplifications upon which they are based allow us to develop an understanding of the individual components that influence the spatial resolution in STEM imaging. However, in order to understand the electron probe as the result of the collective effect of all four factors mentioned above, we need to describe the electron probe on the basis of wave optics rather than on purely geometrical grounds. The wave optical description of the electron probe, which is the topic of the following section, will enable us to differentiate between the effects of the different probe contributions mentioned above. Still, it is important to note that the above considerations about the individual contributions to the STEM probe and their dependence on the illumination semi-angle qualitatively remain valid. For instance, it is still the spherical aberration and the diffraction limit that define the optimum STEM probe of a conventional scanning transmission electron microscope.

### 3.3 Wave Optical Description of an Electron Probe

In the following we derive a wave optical description of an electron probe, taking into account the four factors mentioned above. The wave optical description allows us to calculate an electron probe with any desired precision and in all three spatial dimensions, provided, of course, all relevant input parameters are known with sufficient precision. It has to be emphasized that even with very detailed knowledge about the lateral and longitudinal extension of the electron probe, the problem of relating the probe characteristics to a resolution criterion is not solved simply. The fundamental problem is that an electron probe shows in principle infinitely long
tails and, as such, it is not a trivial problem to extract a measure on the basis of the calculated electron probe which directly predicts the STEM resolution we recognize in a micrograph. One could, for instance, apply the Rayleigh criterion to estimate the resolution on the basis of a calculated STEM probe. The minimum distance between two calculated STEM probes which enables a central dip between the two superimposed electron probes of 19% is known at the Rayleigh resolution. However, even this very practical but theoretical resolution criterion is not suitable, in case the electron probe has significant side lobes (Fertig and Rose, 1979) or in case the micrographs are affected by noise, which for normal experimental conditions is unavoidable (Van Aert and Van Dyck, 2006). Nevertheless, even though the relation between the geometry of the STEM probe and the actual resolution is intricate, the general rule is clear: the smaller the STEM probe, the better the optical resolution of the instrument. Furthermore, in order to be able to compare and optimize a STEM probe, detailed knowledge about the actual geometry of the STEM probe and its dependence on the set of relevant parameters is crucial.

The way we introduce the wave optical description of an electron probe is equivalent to the geometrical considerations discussed in the previous section. First, we start with the effect of the finite aperture, then we include the lens aberrations, and finally we incorporate the effects of partial spatial and partial temporal coherence. We denote the electron wave incident on the object plane by $\psi_0(r)$, and the electron wave in the aperture plane in front of the object plane by $\psi_0(q)$ (see Fig. 3.1).

Looking at Fig. 3.1, which illustrates the basic experimental setup in STEM mode, it can be seen that the pre-field of the objective lens focuses the electron wave in the aperture plane $\psi_0(q)$ onto the specimen plane. This situation is in principle analogous to the situation in HRTEM (see Fig. 2.1): a plane wave passes through the specimen plane and the post-field of the objective lens produces a focused electron beam in the back focal plane. If a diffracting specimen is located in the specimen plane, each diffracted beam gives rise to a focused spot, i.e. a diffraction pattern is formed. However, similar to an electron probe in STEM, a focused diffraction spot is an image of the source (James and Browning, 1999; Rose, 2009a). For the case of HRTEM as well as for STEM, a lens focuses a parallel beam into a spot. For HRTEM, the focused spot lies in the back focal plane, while for STEM the focused spot lies in the specimen plane, which indeed is the back focal plane of the
pre-field of the objective lens. In both modes, i.e. in HRTEM and in STEM mode, the specimen plane is described by the position coordinate $r$. The plane of the illumination aperture and the plane of the objective aperture are aperture planes which are described by $q$. While in HRTEM the back focal plane containing the diffraction pattern is a plane conjugate to the plane of the electron source, in STEM it is the specimen plane which is conjugate to the plane of the electron source. In order to move the plane which is conjugate to the electron source from the back focal plane to the specimen plane, i.e. in order to switch from TEM to STEM mode, the mini-condenser lens above the objective lens is changed. In STEM mode, the mini-condenser lens is optically off, while in TEM mode, it is optically on (Williams and Carter, 1996; James and Browning, 1999). From this we can conclude that similar to the case of the formation of the diffraction pattern in TEM, in STEM, the electron wave in the aperture plane and the electron wave in the specimen plane are linked to each other by a Fourier transform.

Let us thus start with the electron wave function at the aperture plane in front of the specimen (see Fig. 3.1). We assume that the phase and the amplitude of the electron wave are constant across the aperture opening. For a circular aperture, the electron wave in the aperture plane is a top-hat function described by

$$\psi_0(q) = \begin{cases} 
1 & \text{if } |q| < |q_a| \\
0 & \text{otherwise.}
\end{cases} \quad (3.12)$$

The radius of the aperture opening is denoted by $q_a$, which is related to the illumination semi-angle $\alpha$ by $\alpha = q_a \lambda$. In analogy to the previous chapter, the top-hat function can approximately be described in a closed form by employing the Fermi function. This yields

$$\psi_0(q) = \frac{1}{1 + \exp\left[\frac{|q|^2 - |q_a|^2}{\delta_a^2}\right]}, \quad (3.13)$$

where $\delta_a$ can be chosen as a small fraction of $q_a$. A finite value of $\delta_a$ numerically blurs the edge of the aperture, which reduces the risk of introducing artifacts in calculations (Kirkland et al., 1987; Kirkland, 1998).

Now we could simply derive the Fourier transform of Eq. (3.13) to obtain the electron wave on the specimen plane, i.e. the electron probe. Doing so, we would obtain an electron probe whose geometry is determined solely by the diffraction limit; the larger $q_a$, the smaller the probe (see, Fig. 3.3). Hence, the Fourier transform of Eq. (3.13) for $\delta_a \to 0$ yields an Airy pattern as, for instance, illustrated in Fig. 3.2a.

### 3.3.2 Defocus and spherical aberration

We have already seen that owing to the fact that electron lenses are not perfect, the diffraction limit is not the only contribution to the electron probe. If we only had to consider the diffraction limit, we would simply employ the largest illumination
aperture in order to obtain the smallest probe. However, the next contribution that needs our attention is the effect induced by the phase shifts caused by lens aberrations, namely by the third-order spherical aberration $C_3$, whose effect can be balanced by a proper choice of the defocus $C_1$. The effect of the lens aberrations is taken into account by considering the aberration phase-shifts in the aperture plane. This is equivalent to HRTEM, where we also included the lens aberrations in the aperture plane, which, however, for HRTEM lies behind the object plane, namely in the back focal plane of the objective lens’ post-field. Furthermore, because the effect of defocus and spherical aberration is described in the aperture plane — in STEM mode as well as in HRTEM mode — such aberrations are also called aperture aberrations.

The spherical aberration $C_3$ and the defocus $C_1$ cause a modulation of the phase of the electron wave in the aperture plane. In analogy to HRTEM, the aberration function can be written as

$$\chi(q) = \chi(q) = \frac{1}{2} q^2 \lambda^2 C_1 + \frac{1}{3} q^4 \lambda^4 C_3.$$  \hfill (3.14)

Since we deal with isotropic aberrations, without loss of generality we can switch from the vector notation $\mathbf{q}$ to the scalar notation $q$. The phase shifts $\gamma$ are given by $\gamma(q) = 2\pi/\lambda \cdot \chi(q)$.

As any wave function, the incident electron probe in the aperture plane has the general form $\psi_0 = A \cdot \exp(iB)$, with $B$ the phase of the wave and $A$ the amplitude. So far, we set the amplitude $A = 1$ within the aperture opening, and the phase $B$ we set arbitrarily equal to zero. Incorporating the effect of the geometrical lens aberrations into the probe calculation, we have to include the phase shifts given above. With this addition, the electron wave in the aperture plane given in Eq. (3.13) becomes

$$\psi_0(q) = \exp \left\{ -\frac{2\pi i}{\lambda} \chi(q) \right\} \frac{1}{1 + \exp \left\{ \frac{q^2 - q_{\lambda}^2}{\delta_a^2} \right\}}. \hfill (3.15)$$

While the first term determines the phase of the wave function in the aperture plane, the second term, i.e. the aperture function, determines its amplitude. Equation (3.15) covers all coherent contributions to the wave function. The next step is to take the Fourier transform of Eq. (3.15) to obtain the coherent electron wave on the specimen plane. This can be written as (see, e.g. Nellist and Pennycook, 1998, 2000)

$$\psi_0(r) = \int_{-\infty}^{\infty} \exp \left\{ -\frac{2\pi i}{\lambda} \chi(q) \right\} \exp \left\{ -2\pi i q \cdot r \right\} dq. \hfill (3.16)$$

The integral runs in principle from $-\infty$ to $+\infty$. However, since for $q$ values beyond $q_{\lambda}$ the integrand is zero, the integral can in principle be taken within the aperture.
opening. Furthermore, if the probe needs to be calculated for a specific position \( r_p \) on the specimen plane, \( r \) can be replaced by \( r - r_p \) throughout Eq. (3.16). The intensity \( I_0(r) \) of the electron probe on the specimen plane is then given by taking the modulus of the complex wave function \( \psi_0(r) \). This is written as

\[
I_0(r) = \psi_0(r)\overline{\psi_0}(r) = |\psi_0(r)|^2, \tag{3.17}
\]

where \( \overline{\psi}_0 \) denotes the complex conjugate of \( \psi_0 \).

In the previous section, we discussed the optimization of a STEM probe by balancing the effects of defocus, spherical aberration and the diffraction limit. This is essentially expressed by the Scherzer incoherent conditions given in Eqs. (3.9) and (3.10). For a given wavelength \( \lambda \) and for a given spherical aberration \( C_3 \), those two equations allow us to calculate optimal settings for the illumination semi-angle \( \alpha \) and the defocus \( C_1 \), which are necessary to achieve a small probe size. The wave optical description of the electron probe according to Eq. (3.17) can be used to illustrate the optimization related to Eqs. (3.9) and (3.10). Figure 3.8 depicts three probe intensity profiles calculated according to Eq. (3.17) for 200 keV electrons (\( \lambda = 2.5 \) pm), a constant of spherical aberration \( C_3 \) of 1 mm and an optimized illumination semi-angle \( \alpha_{opt} \) of 10 mrad. For the given \( C_3 \) and \( \lambda \) we can employ Eq. (3.9) to derive the optimal setting for the defocus. This yields \( C_{1 \text{opt}} = -50 \) nm. The electron probes in Fig. 3.8 reveal the effect of a change of defocus on the probe intensity. One probe is calculated for \( C_1 = -75 \) nm, one for the optimal defocus and one for a defocus closer to the Gaussian focus, i.e. \( C_1 = -25 \) nm. It is clear that deviations from the optimal defocus \( C_{1\text{opt}} \) result in probe intensity profiles which have either a wider central maximum or show substantial side lobes that would significantly reduce the achievable image contrast.

![Fig. 3.8](image_url)  

Fig. 3.8 Intensity profiles of electron probes calculated for 200 keV electrons (\( \lambda = 2.5 \) pm) and \( C_3 = 1 \) mm. The probes were calculated according to Eq. (3.17) for a defocus of \(-75 \) nm, for an optimized defocus of \(-50 \) nm (see Eq. (3.9)), and for \(-25 \) nm. The probe illumination semi-angle is in all three cases 10 mrad, corresponding to the optimum angle (see Eq. (3.10)).
3.3.3 Geometrical source size

The coherent contributions to the wave field of the electron probe in the specimen plane are expressed in Eq. (3.16). However, we have already seen that the image formation in HRTEM is not fully coherent such that envelope functions due to partial temporal and partial spatial coherence have to be introduced, which essentially reduce the transfer function at high spatial frequencies. In STEM, the situation is indeed very similar. We also started above with a fully coherent electron probe and now we have to incorporate contributions which take into account the limited coherence of the electron beam.

Partial spatial coherence essentially means that the electron source is not point-like, but has a finite size which we described in the previous section with the source radius \( r_{\text{geo}} \). The value \( r_{\text{geo}} \) was our quantity which measures the radius of the electron source after demagnification projected onto the specimen plane. This implied that the source is homogenous across the emission area. However, real electron sources are not disk-like and they do not emit electrons homogenously over a well defined area (Swanson and Schwind, 1997). Real electron sources have emission characteristics which also include the actual shape of the emitting tip. The intensity distribution of the emission imaged onto the specimen plane is described by a source intensity distribution function (see, e.g. Dwyer et al., 2008). A suitable and simple model for a source intensity distribution function \( S(r) \) is a Gaussian function

\[
S(r) = S(r) = \frac{1}{\sqrt{2\pi}\sigma_s} \exp \left( -\frac{r^2}{2\sigma_s^2} \right). \tag{3.18}
\]

This function describes the distribution of the emitting points of the source which contribute to the electron probe. The magnitude of \( S \) models their relative contribution to the total intensity. Though \( S(r) \) is called the source intensity distribution function, it is indeed the source intensity distribution projected and demagnified onto the object plane. The standard deviation \( \sigma_s \) quantifies the width of the demagnified source on the specimen plane. However, often one is interested in the full width at half maximum (FWHM) of the distribution function, which we identify with \( \delta_{\text{geo}} \). The standard deviation and the FWHM of the Gaussian source intensity distribution function are related to each other by \( \delta_{\text{geo}}^2 = 8 \ln 2 \sigma_s^2 \approx 2.355 \sigma_s^2 \). We call \( \delta_{\text{geo}} \) the geometrical (or effective) source size.

The finite size of the source intensity distribution means that the electron beam does not emerge in a single point, as assumed in Eq. (3.16), but it is emitted from a multitude of points described by \( S(r) \). Each element of the source gives rise to a coherent electron probe as described in Eq. (3.17). The electron probes due to each source element incoherently combine to yield the complete electron probe. Hence, in order to incorporate the source intensity distribution in the probe intensity calculation, we have to convolute the intensity of the coherent probe wave field given in Eq. (3.17) with the source intensity distribution given in Eq. (3.18). We deal solely with the intensity of the electron probe and not with its phase, because
the finite size of the source is an incoherent contribution to the electron probe. Hence, we assume that there is no interference between electrons which are emitted from different points of the source. Or, in other words, the source distribution diminishes the magnitude of the interference between different $q$ vectors within the illumination aperture.

With the addition of the source distribution function, the intensity of the electron probe can now be written as

$$I_0(r) = |\psi_0(r)|^2 \otimes S(r),$$

where $\otimes$ denotes the convolution. Equation (3.19) shows clearly that the effect of the finite source size results in an incoherent blurring of the electron probe.

In analogy to the previous section, the impact of high frequency noise can be incorporated by replacing $\delta_{\text{geo}}$ with an effective geometrical source size $\delta_{\text{geo,e}}$, which includes the blurring of the electron probe by $\delta_{\text{geo,e}}^2 = \delta_{\text{geo}}^2 + \delta_{\text{noise}}^2$.

### 3.3.4 Energy spread of the electron beam

The last step that needs to be done is to include the effect of partial temporal coherence. As we have already seen in Chapter 2, the finite energy spread of the electron beam in combination with the chromatic aberration of the lens essentially leads to a spread of focus (see Eqs. (2.22) and (2.23)). For STEM, this is essentially the same.

Because of the variation $\delta E$ of electron energies around the nominal electron energy $E_0$, the chromatic aberration $C_C$ causes a variation $\delta C_1$ of the defocus $C_1$. An energy offset $\delta E = E - E_0$, where $E_0$ is the nominal electron energy and $E$ the actual energy of a particular electron, results in a change of focus of $\delta C_1$ which is proportional to the constant of chromatic aberration $C_C$: $\delta C_1 = C_C \delta E / E_0$. Hence, it is the spread of focus induced by the finite energy spread of the beam and the chromatic aberration which needs to be incorporated in the probe calculation.

The focus is blurred. This means that for a given nominal defocus $C_1$, the intensity of the electron probe is the incoherent superposition of electron probes integrated over a certain defocus range, which is determined by $C_C$ and $\Delta E$. The focus blur is incorporated via the aberration function $\chi$ given in Eq. (3.14), which is a function of the defocus $C_1$.

Let us assume that the energy distribution $T(E)$ of the electron beam is a Gaussian function. This of course is a simplified model which does not take into account the emission characteristics of a given electron source. For our purpose, however, it is sufficiently accurate. We thus can model the energy distribution by

$$T(E) = \frac{1}{\sqrt{2\pi}\sigma_t} \exp \left\{ -\frac{(E - E_0)^2}{2\sigma_t^2} \right\},$$

with the standard deviation $\sigma_t$ related to the full width at half maximum of the Gaussian energy distribution $\Delta E$ by $\Delta E^2 = 8\ln2 \sigma_t^2 \approx 2.355^2 \sigma_t^2$ (see Fig. 3.9a).
The electron probe can now be calculated by (see, e.g. Haider et al., 2000)

\[ I_0(r) = \int_{-\infty}^{\infty} |\psi_0(r, E)|^2 \otimes S(r) \cdot T(E) dE, \]  

(3.21)

which describes the incoherent superposition of electron probes, weighted by \( T(E) \), spread over a certain focus range which is determined by the energy spread of the beam. This is illustrated in Fig. 3.9.

It is important to emphasize the dependence of \( \psi_0(r, E) \) on the energy \( E \) in Eq. (3.21). For each electron energy \( E \) within the energy distribution \( T(E) \), an electron probe \( \psi_0(r, E) \) needs to be calculated where the effective defocus \( C_1 \) in Eq. (3.14) includes an offset \( \delta C_1 \) with respect to the nominal focus of the electron probe. This focus offset \( \delta C_1 \) is given by \( \delta C_1 = C_C(E - E_0)/E_0 = C_C \Delta E/E_0 \).

### 3.3.5 Concluding remarks

As already mentioned above, an electron probe should be considered as a three-dimensional intensity distribution (see, e.g. Erni et al., 2009). The lateral extension of the electron probe essentially determines the lateral STEM resolution, while the

![Fig. 3.9](image-url)

Fig. 3.9 Effect of the finite energy spread of the electron beam and the chromatic aberration \( C_C \) on the probe intensity. (a) The energy distribution of the electron beam is described by a Gaussian function \( T(E) \) with FWHM \( \Delta E \) centered around the nominal electron energy \( E_0 \). (b) and (c) The energy spread leads to a focus spread of the electron probe which can be considered as an incoherent superposition of electron probes, weighted by \( T(E) \) or \( T(C_1) \) respectively, over the defocus range with FWHM of \( \Delta C_1 \text{ FWHM} = C_C \Delta E/E_0 \).
longitudinal extension of the electron probe, i.e. the extension along the optical axis of the microscope, defines the depth resolution or the depth of field. The chromatic aberration and thus partial temporal coherence affects the electron probe in the lateral and in the longitudinal direction.

Firstly, due to the incoherent superposition of electron probes over a finite defocus range $\Delta C_1$ (see Fig. 3.9), the lateral extension of the electron probe is enlarged. However, for conventional scanning transmission electron microscopes, this effect is not the probe size-limiting factor (Shao and Crewe, 1987). Furthermore, it is not so much the lateral extension of the central maximum of an electron probe that is affected by the chromatic aberration, but the side lobes of the electron probe (see Fig. 3.9). This leads to an increased background intensity and thus to a reduced image contrast (see, e.g. Fertig and Rose, 1979).

The longitudinal extension of the electron probe is essentially determined by the geometrical restriction imposed by the finite probe illumination angle. For electron probes of conventional field-emission scanning transmission electron microscopes, the depth of field is typically larger than the thickness of the specimen. However, as already pointed out by Rose (1975), with increasing probe illumination angle and thus improved resolution, the depth of field becomes smaller. This opens the way to access not just projected information about the specimen, but three-dimensional information. Three-dimensional information can be obtained from the specimen by recording focal series on specimens whose thickness exceeds the depth of field of the electron probe. The depth of field $\Delta C_1$ can be expressed as (Born and Wolf, 2001)

$$\Delta C_1 \approx \frac{\lambda}{\alpha^2}. \quad (3.22)$$

For our virtual electron microscope with $C_3 = 1$ mm operated at 200 keV ($\lambda = 2.5$ pm), we obtain for an optimized illumination semi-angle $\alpha = 10$ mrad a depth of field $\Delta C_1$ of 25 nm. This value is in the order of magnitude of the thickness of a typical high-resolution STEM specimen. Hence, we do not expect to see a distinct depth dependence on a conventional STEM instrument. Moreover, the chromatic aberration further contributes to the blurring of the defocus. However, if we consider a spherical aberration-corrected instrument, which can be used to form an electron probe with an illumination semi-angle of 25 mrad, the depth of field estimated by Eq. (3.22) is 4 nm for 200 keV electrons. This can be significantly smaller than the thickness of a specimen. Indeed with the advent of aberration-corrected scanning transmission electron microscopes, depth sectioning by means of recording through focal series in STEM mode has become feasible and has been brought to application (van Benthem et al., 2005, 2006).

This topic will be discussed in more detail in the last part of this book, where we deal with applications of small electron probes formed with aberration-corrected scanning transmission electron microscopes. For now, we shall be concerned solely with the effect of partial temporal coherence on the spatial extension of the electron probe.
probe. The spread of focus, which is caused by the energy spread of the beam and the finite value of \( C_C \), increases the depth of field and thus decreases the achievable depth resolution in STEM imaging. However, the focus spread due to partial temporal coherence is clearly smaller than the effect given by the geometry of the electron probe (see, e.g. Eq. (3.22)). Hence, we do not expect that on conventional scanning transmission electron microscopes the focus blur due to partial temporal coherence will be significant.

We conclude that for the case of conventional probe-forming instruments operated between about 100 kV and 300 kV employing field-emission sources, the impact of the chromatic aberration is not critical. This is in agreement with the purely geometrical considerations summarized in Fig. 3.7. The electron probe in conventional field-emission scanning transmission electron microscopes is limited by the spherical aberration. The defocus and the aperture opening are set according to the Scherzer incoherent conditions given in Eqs. (3.9) and (3.10), in order to minimize the impact of the spherical aberration \( C_3 \) and thus to obtain a small electron probe.

Figure 3.10 summarizes the individual contributions to the electron probe discussed above. We employ the microscope parameters from above, i.e. the acceleration voltage is 200 kV and the constant of spherical aberration of third order is \( C_3 = 1 \) mm. Equation (3.10) allows us to calculate the optimum illumination

![Fig. 3.10 - Electron probe intensity profiles; (a) linear scale, (b) logarithmic scale. The full line considers only the finite size of the aperture, the dashed line includes aperture aberrations \( C_1 \) and \( C_3 \) according to Eq. (3.17), the dotted line considers in addition a finite source size \( \delta_{\text{geo}} \) of 0.08 nm, and the dashed-dotted line is calculated according to Eq. (3.21) considering \( C_C \) of 2 mm and \( \Delta E = 5 \) eV in order to amplify the effect of partial temporal coherence. The other parameters are \( C_3 = 1 \) mm, \( E_0 = 200 \) keV (\( \lambda = 2.5 \) pm) and \( \alpha = \alpha_{\text{opt}} = 10 \) mrad.](image-url)
semi-angle, which is \( \alpha_{\text{opt}} = 10 \) mrad. The full line in Fig. 3.10 (a and b) shows a probe-intensity profile calculated according to Eq. (3.13), where the only contribution to the probe is the effect of the finite aperture size. This is essentially a line profile through an Airy pattern, as already illustrated in Fig. 3.2. The dashed line in Fig. 3.10 corresponds to a probe intensity profile incorporating, in addition to the finite aperture angle, the aperture aberrations \( C_1 \) and \( C_3 \) according to Eq. (3.16). The defocus \( C_1 \) was set to \(-50 \) nm in agreement with the optimum defocus which can be derived from Eq. (3.9). The finite source size of 0.08 nm is taken into account as an additional factor in the dotted curve in Fig. 3.10. Since the effect of partial temporal coherence is small, in order to illustrate its impact on the electron probe, we considered an energy spread of \( \Delta E \) of 5 eV and a constant of chromatic aberration \( C_C \) of 2 mm. This yields the dashed-dotted curves in Fig. 3.10.

Figure 3.7a shows clearly that incorporating stepwise the individual contributions to the electron probe, the intensity of the central maximum of the electron probe decreases. However, none of the effects substantially increases the width of the central maximum. Even the unrealistically large energy spread of 5 eV for a field-emission source combined with a \( C_C \) of 2 mm does not substantially degrade the width of the central maximum. The diffraction limit \( \delta_D \) due to \( \alpha = 10 \) mrad would in principle allow for a resolution of 0.15 nm (see Eq. (3.1)). The FWHM of the full curve measures 0.13 nm. Though the FWHM is smaller than the \( \delta_D \), we use the FWHM to compare the probe profiles in Fig. 3.10. Incorporation of \( C_3 \) including \( C_{1,\text{opt}} \) leads to a FWHM of 0.133 nm; including \( \delta_{\text{geo}} \) of 0.08 nm leads to a FWHM of 0.14 nm, and including partial temporal coherence, i.e. the chromatic aberration, leads to a FWHM of 0.14 nm. Hence, provided the defocus \( C_1 \) and the probe illumination semi-angle \( \alpha \) are set according to the optimum settings given in Eqs. (3.9) and (3.10), neither the spherical aberration, the chromatic aberration nor the finite source size of 0.08 nm substantially reduces the probe size and thus the achievable STEM resolution. The overall effect due to \( C_3 \), \( \delta_{\text{geo}} \) and \( C_C \) is less than 10 pm. This, of course, depends strongly on the assumption of the effective source size, but 0.08 nm is not an unrealistically small value (see, e.g. LeBeau et al., 2009; Dwyer et al., 2008). Hence, the Scherzer incoherent conditions make it possible that for the optimum illumination semi-angle the effect of the spherical aberration is minimized such that the resolution is nearly determined by the diffraction limit imposed by \( \alpha_{\text{opt}} \).

Although the intensity of the central maximum decreases when incorporating the different contributions to the electron probe (see Fig. 3.10a), the total probe intensity needs to be a constant. Hence, the intensity which is lost in the central maximum is transferred to side lobes. This can be seen in the logarithmic plot in Fig. 3.10b. The main impact of the spherical aberration under optimized probe conditions and the influence of partial coherence is to increase the intensity of the tails of the electron probe. The tails of the probe do not primarily impact the achievable STEM resolution, but the image contrast. Because the central maximum
is hardly affected by the individual probe contributions, the probe remains sensitive to nearly the same object spacings. However, with growing tails of the probe, the background intensity of a micrograph increases and, as a result, the contrast decreases. Fertig and Rose (1979) show that for an electron probe which provides zero contrast due to its side lobes, the Rayleigh criterion can still be fulfilled. This clearly shows that relating the probe characteristics to an achievable resolution is not a trivial issue. Fertig and Rose (1979) suggest relating the diameter $d_{59}$ which contains 59% of the total probe intensity with the STEM resolution. This is essentially the diameter corresponding to the first minimum of the electron probe and is thus equivalent to the diffraction limit, which also relates the first minimum of the Airy pattern with the achievable resolution.

3.4 Summary

This chapter focused on one particular aspect of STEM imaging; the formation of the electron probe. Contributions which affect the characteristics of the electron probe were discussed from a purely geometrical as well as from a wave optical point of view. Keywords addressed in this chapter are: STEM imaging modes; electron probe; diffraction limit; isotropic aberration function; chromatic aberration; effective source size; resolution and depth of field.