Imaging Theory (Last Revision 3/29/2011)

In an electron microscope the specimen scatters the electron wave and the lenses produce the image. The two are very different processes and have to be dealt with quite separately. The reason why imaging is such an important area is that images can lie. As human beings we are conditioned to interpret what we see in a certain sense: when we look through a window we see what is on the other side and automatically interpret it using what we have learnt about the world since childhood. We already know from simple diffraction theory that this is not true when we look through a electron microscope specimen; we can see light or dark depending upon the orientation and thickness of the specimen. It is therefore necessary to retrain your eye (brain) to interpret electron microscope images appropriately, not for instance light regions as thin areas of the sample.

It is also true that what we see in the electron microscope depends upon the conditions of the microscope. An example that we have already encountered is Fresnel fringes from an edge which depend upon the defocus of the image. The actual minuscule details of how the electromagnetic lenses magnify and rotate the wave is not as important as the sense in which the electron microscope acts as a filter, allowing certain spacings to contribute to the image whilst others do not, and at the same time changing the contrast of the different spacings. This is because the optical system in an electron microscope is comparatively poor (at least compared to an optical microscope or your eyes) with respect to aberrations such as spherical aberration, but at the same time good in terms of the coherence of the electron wave. Therefore we cannot, for instance, consider that an out of focus electron microscope image is simply blurred with respect to a focussed image -- we must use a more sophisticated approach. An analogy of a high resolution electron microscope is a rather poor radio which, for instance, cuts of the treble and distort the bass; different sound frequencies are passed in different ways by the radio.

Before becoming involved with the full wave theory of image formation that we have to use, it is useful to go over a few of the broad details of the HREM imaging system, highlighting the features that we need to consider in more detail later on.

1. Operational Factors

For convenience we will follow the path of the electrons down the column of a conventional HREM.

Illumination system

The electrons start from the electron source or gun. Depending upon the actual details of the source, the electrons are emitted with a spread of energies with a half-width at half height of about 1-2 eV. These are then accelerated to the final operating voltage of the instrument, for instance 300kV. In reality this is 300 $\pm\delta$ kV where δ is due to ripple in the high voltage source, typically about 1 part per million (ppm). These electrons then pass through a number of (condenser) apertures before they reach the specimen. These apertures serve two purposes

1) They limit the angular range of the electrons reaching the specimen. As a rule the electrons from the source are emitted incoherently with a range of different directions. (We will return to the importance of coherence later in this chapter.) Some of these directions are cutoff by the condenser aperture. As a rule we consider the range of angles used to form the image as the convergence of the illumination, characterized by either a Gaussian distribution of directions or as a cone of directions characterized by the half angle of the cone. The former would be when the illumination in the microscope is a little defocussed, the later when it is fully focussed and the half

angle of the cone would depend upon the radius of the condenser aperture.

2) They limit the nett energy spread of the electrons due to both the ripple in the accelerating voltage and the natural energy spread of the electron source. Due to the chromatic aberrations of the condenser lenses, electrons of different energies are focussed at different positions so that the condenser aperture can be used to partially restrict the energy spread.

The two main parameters that we need to carry forward from the illumination system are the energy spread of the electrons and the convergence of the illumination. As we will see, both of these limit the resolution of our images.

Specimen

The full details of what happens to the electron wave as it passes through the specimen are belong the scope of this section. Our main interest is that a wave with one unique wavevector before the specimen becomes split into a wave with a number of different wavevectors by the specimen diffraction. Note that these different wavevectors are all coherent. We should also remember that our specimen is not truly stationary - it will be drifting albeit hopefully rather slowly and perhaps also vibrating in place.

Post specimen imaging

It is conventional to consider all the remaining lenses in the microscope as just one lens, the objective lens of the microscope. (In reality both the objective lens and at least the first intermediary lens are important to the working of the microscope.) With this lens we focus the wave exiting the specimen onto the phosphor screen (or photographic film or TV camera) of the microscope. In fact we do not always use an image which is truly in focus, but instead one which is a little out of focus. In part this is to compensate for the spherical aberration of the microscope which brings waves travelling in different directions into focus at different positions, and in part due to the basic character of the scattering of the electrons as they pass through the specimen. (This will become clearer below.)

Whilst an ideal objective lens would focus the specimen exactly, in reality there is some instabilities in the current of the objective lens. Therefore there is some distribution in the actual focus of the final image, what we refer to as focal spread. The energy spread of the electron source has the same general sense as this focal spread, and we generally refer to the two together as the focal spread of the microscope. (The actual value of this parameter has to be determined experimentally as it will change depending upon the actual illumination conditions.)

Three other effects also have to be considered. Two of these are the astigmatism of the objective lens and the orientation of the incident illumination with respect to the optic axis of the objective lens (post specimen lenses), called the tilt of the illumination. Whilst both of these are hopefully very small and can be corrected by the operator, they are in practice very hard to completely eliminate. The final effect is the objective aperture (if one is used). In many modern high resolution electron microscopes very large apertures are used and there effects can be almost completely ignored.

To summarize from this section, the effects that we have to carry forward for our models of the imaging process in the electron microscope are the convergence, the focal spread, drift, beam tilt, astigmatism, defocus, spherical aberration and objective aperture size.

2. Classical effects of Aberrations

The simplest approach to estimating the role of aberrations in the microscope is to use very simple classical concepts of optics. The resolution given by the Raleigh formula, which is derived by considering the maximum angle of the electron scattering is:

$$R = 0.61\lambda/\alpha \qquad \text{I2.1}$$

where R is the resolution, λ the wavelength and α is the scattering angle. We should combine with this the size of the disk of least confusion due to the spherical aberration, which is given by:

$$D = C_s\alpha^3 \qquad \text{I2.2}$$

where the Spherical Aberration C_s is typically 1mm. A simple argument is that we cannot resolve better than the larger of D and R, and since R decreases with increasing α whilst D behaves in the opposite fashion there is an optimum value of α which will be when R=D, i.e.:

$$0.61\lambda/\alpha = C_s\alpha^3 \qquad \text{I2.3}$$

or $$\alpha = 0.88(\lambda/C_s)^{1/4} \qquad \text{I2.4}$$

with a 'best' resolution of

$$R = 0.69(C_s\lambda^3)^{1/4} \qquad \text{I2.5}$$

This resolution is often quoted, generally overquoted; as we will see later it is not a very good measurement.

To include the effect of instabilities in the lenses and the high voltage, we assume that there are random fluctuations in the voltage ΔV in the total voltage V, and in the current of the lens ΔI with I the total lens current. For random fluctuations we sum the squares of the different terms, with gives us a disc of least confusion of size:

$$C = \alpha\Delta f \text{ where } \Delta f = C_c\sqrt{([\Delta V/V]^2 + [\Delta I/I]^2)} \qquad \text{I2.6}$$

Here Δf is defined as the focal spread and C_c is called the Chromatic aberration of the microscope lens which is typically close to 1 mm.

The final term we can include is the drift or vibration of the instrument. Let us suppose that over the time of the exposure the image drifts a distance L and is vibrating with an amplitude of v. If we assume that these are random fluctuations, we can sum the square of these terms along with the squares of the 'best' resolution and the chromatic disc of confusion to estimate the optimum resolution as:

$$\text{Optimum} = \sqrt{(R^2 + C^2 + L^2 + v^2)} \qquad \text{I2.7}$$

This value can be used for back of the envelope calculations, although it should not be extended beyond this.

3. Wave optics

We now jump a level in our analysis of the imaging process to a more accurate model. Electrons are waves, and any real imaging theory must take this completely into account; a classical imaging approach using solely ray diagrams fails completely to account for image structure. Central to wave models is the Fourier integral. A wave in real space of form $\psi(\underline{r})$ can be decomposed into components $\Psi(\underline{k})$, where $\underline{k}$ is the wavevector of the wave by the Fourier integral:

$$\psi(\underline{r}) = \int \Psi(\underline{k})\exp(2\pi i\underline{k}.\underline{r})d\underline{k} \qquad \text{I3.1}$$

which has the inverse

$$\Psi(\underline{k}) = \int \psi(\underline{r})\exp(-2\pi i\underline{k}.\underline{r})dr \qquad \text{I3.2}$$

As a rule we speak of $\Psi(\underline{k})$ as a spatial frequency of $\underline{k}$. When discussing electron diffraction we generally refer to both $\underline{r}$ the position vector in real space, and $\underline{k}$ the wavevector in three dimensions. For imaging theory we can simplify this to the two dimensions normal to the incident beam direction, and we will use the vector $\underline{u}$ (not $\underline{k}$) to describe the spatial frequency in two dimensions.

For "standard" high resolution electron microscopes all the aberration terms that we described in section 1 fall into one of two classes, namely they are coherent or incoherent effects. (With the newer class of Field Emission sources there are some additional complications, and one cannot simply use these two extremes.) Let us define what these terms mean.

3.1 Coherence, Incoherence.

All the aberrations in an electron microscope (except the objective aperture) can be considered in terms of phase shifts of the waves. That is a spatial frequency $\underline{u}$ becomes changed

$$\Psi(\underline{u})\exp(2\pi i\underline{u}.\underline{r}) \dashrightarrow \Psi(\underline{u})\exp(2\pi i\underline{u}.\underline{r} - i\phi) \qquad \text{I3.3}$$

If the phase change ϕ is fixed for each value of $\underline{u}$ we refer to the aberration as coherent, whilst if the value of ϕ is not fixed but has some form of completely random distribution we refer to the aberration as incoherent. As an example, the ripple in the high voltage is incoherent since an electron emitted with one particular energy will pass down the microscope column and reach the image completely independent of a second electron with a different energy that is emitted at some later time. The difference between the two can best be seen by the way they effect the interference between two waves. Let us consider a wave made up of two waves, i.e.

$$\psi(\underline{r}) = A\exp(2\pi i\underline{u}.\underline{r}) + B\exp(2\pi i\underline{u}'.\underline{r} - i\phi) \qquad \text{I3.4}$$

where for convenience we will take A and B as real numbers. The intensity is then

$$I = |\psi(\underline{r})|^2 = A^2 + B^2 + 2AB\cos(2\pi[\underline{u}-\underline{u}'].r + \phi) \qquad \text{I3.5}$$

If ϕ has a fixed value, the term on the right of I3.5 represents fringes in the image with a spatial frequency of $\underline{u}$-$\underline{u}'$ caused by the interference of the two waves. If instead ϕ has a random distribution of values, then when we average over different values the cosine term will average to zero. In this case we have no interference between the two waves. Thus coherent waves interfere whilst coherent aberrations effect the interference between waves, and incoherent waves do not interfere so that

incoherent aberrations can be considered by summing intensities. (A middle ground with partial coherence, i.e. when ϕ is not completely random is important for the most modern electron microscopes and will be dealt with later.) Defocus, astigmatism, tilt and spherical aberration are all coherent aberrations, whereas focal spread, drift and convergence are generaly incoherent aberrations. For instance, the changes in focus due to instabilities in the current in the objective lens windings are completely random. (We are implicitly not considering coherent convergence as in a STEM or the HF2000 for the moment.)

3.2 Coherent Aberrations

We can express rigorously all the coherent aberrations by expanding the phase shift ϕ as a Taylor series, i.e.

$$\phi(\underline{u}) = A + \underline{B}.\underline{u} + Cu^2 + (\underline{D}.\underline{u})^2 + ... \qquad I3.6$$

For the moment let us assume that the electron wave is passing exactly through the center of the lens system along the optic axis. Then, by symmetry, all the odd order terms must vanish and we are left with the series:

$$\phi(u) = A + Cu^2 + (\underline{D}.\underline{u})^2 + Eu^4 + ... \qquad I3.7$$

By comparison with the Fresnel propagator analyzed elsewhere, we can equate

$$C = \pi\lambda\Delta z \qquad I3.8$$

where Δz is the objective lens defocus. The astigmatism of the lens is the term $(\underline{D}.\underline{u})^2$ which is written in the form

$$(\underline{D}.\underline{u})^2 = \pi\varepsilon\{(\underline{\alpha}.\underline{u})^2 \qquad -1/2\alpha^2u^2\} \qquad I3.9$$

where ε is the astigmatism in Angstroms and $\underline{\alpha}$ is a unit vector defining the direction of the astigmatism. (Note that this definition has a zero mean second-order term in u.) The fourth order term E is by definition proportional to the spherical aberration of the microscope defined by the relationship:

$$E = \pi/2\lambda^3Cs \qquad I3.10$$

where Cs is the spherical aberration coefficient. In principle we can extend further, including other Taylor series terms and consider other aberrations. At present there is no evidence that these are particularly important in transmission electron microscopy although correction of, for instance, sixth-order aberrations is important for electron energy loss spectrometers.

Summarizing the above, we can define the phase shift for the electron microscope in the absence of beam tilt and astigmatism as:

$$\chi(\underline{u}) = \pi/\lambda(\Delta z\lambda^2u^2 + 1/2\ Cs\lambda^4u^4) \qquad I3.11$$

(Use of the symbol $\chi(\underline{u})$ is conventional.) In the presence of a small beam tilt we simply shift the origin of the phase shift term from $\underline{u}=\underline{0}$ to a $\underline{u}=\underline{w}$, i.e. consider:

$$\chi(\underline{u}) = \pi/\lambda(\Delta z\lambda^2|\underline{u}-\underline{w}|^2 + 1/2\ Cs\lambda^4|\underline{u}-\underline{w}|^4) \qquad I3.12$$

The phase shift term $\chi(\underline{u})$ is central to understanding the imaging process and will be extensively used later in our analysis.

It is appropriate to talk a little more here about astigmatism and beam tilt. Assuming for the moment that there is no beam tilt, the effective defocus phase shift can be written as:

$$\Delta z u^2 + a u_x^2 + b u_x u_y + c u_y^2 \qquad I3.13$$

Depending upon the sign and magnitude of a, b and c this can be the equation of a circle (no astigmatism), ellipse, hyperbola and so forth. Note that when Δz is large, it will always look close to circular; when Δz is small the astigmatism will be more apparent which is why you go to Gaussian focus ($\Delta z=0$) to correct astigmatism. If you do the same thing for the tilt term, you will find effective defocus and astigmatism terms also appear; tilt can be partially canceled out by astigmatism. In the actual microscope, you have one coil along the x axis which changes the term a, or by going negative is equivalent to the c term, and you have a coil along the x+y direction which, coupled with the first coil, gives you the b term.

3.3 Incoherent Aberrations

The rigorous description of the incoherent aberrations in the electron microscope is in terms of distributions of, for instance, focus for the focal spread defined in equation I2.6. Let us consider that the image for a particular value of the objective lens defocus, beam illumination direction (tilt) is $I(\underline{r}, \Delta z, \underline{w})$ where $\underline{w}$ and Δz have the same meaning as in the previous section. If $F(\Delta f)$ represents the spread of focus and $S(\underline{w})$ the convergence, the final image after these effects are taken into account is:

$$I(\underline{r}) = \iint I(\underline{r}, \Delta z-\Delta f, \underline{w}-\underline{w}')F(\Delta f)S(\underline{w}')df d\underline{w}' \qquad I3.14$$

i.e. an average of different images for a distribution of focus and beam directions. Drift of the specimen can be considered by averaging the image over a variety of positions, similarly specimen vibration.

4. Transfer Theory

We have now established the basic tools that we will need to understand the imaging process in an electron microscope. For a given value of the electron energy, defocus and so forth the wave leaving the specimen is modified by the phase term $\chi(\underline{u})$ which depends upon the vector $\underline{u}$, the spatial frequency of each Fourier component of this exit wave. This modified wave then forms an image after the microscope lenses. We then include all the incoherent effects by averaging images at, for instance, different energies by means of changes in the lens focus to obtain our final image. It is useful to represent this process by means of a flow diagram. If $\psi(\underline{r})$ is the wave exiting the specimen, the imaging process is

	Fourier Transform		Phase Change		Fourier Transform	
$\psi(\underline{r})$	---------->	$\Psi(\underline{u})$	---------->	$\Psi'(\underline{u})$	---------->	$\Psi'(\underline{r})$ ----

Incoherent
Intensity Average
---------> $|\psi'(\underline{r})|^2$ ---------> $I(\underline{r})$, Final Image.

The overall process is somewhat complicated and a relatively large number of approximate models have been generated which provide some insight into what is actually happening in the image. Most of these involve approximations about the form of the wave leaving the specimen $\psi(\underline{r})$ which are relatively restrictive and therefore have only a limited range of applicability. As such they are rather like the Kinematical Theory of diffraction; useful qualitative tools but not to be trusted quantitatively. We will run through some of these approximations leading up to the more general theory.

4.1 Charge Density Approximation

One of the simplest approximations is the charge density approximation, which gives an idea what an image at a relatively large defocus will look like. For a very thin specimen the exit wave can be approximated as:

$$\psi(\underline{r}) = 1 - i\sigma t V(\underline{r}) \qquad \text{I4.1}$$

where t is the crystal thickness, σ a constant which depends upon the electron voltage and is equal to $(2\pi me/h^2k)$ Kinematically, and $V(\underline{r})$ the crystal potential. (Remember that the crystal potential for electrons is negative; the energy drops when they are closer to the nuclei.) Fourier Transforming, we can write

$$\Psi(\underline{u}) = \delta(\underline{u}) - i\sigma t V(\underline{u}) \qquad \text{I4.2}$$

for the decomposition of the wave into spatial frequencies. Multiplying by the phase shift term, the modified wave after the objective lens imaging is

$$\Psi'(\underline{u}) = \Psi(\underline{u})\exp(-i\chi(\underline{u})) \qquad \text{I4.3}$$

Assuming no beam tilt or astigmatism, if we are only interested in small $\underline{u}$ values we can neglect the spherical aberration term and expand the exponential as:

$$\exp(-i\chi(\underline{u})) = 1 - i\chi(\underline{u}) \qquad \text{I4.4}$$

$$= 1 - \pi i\lambda u^2\Delta z \qquad \text{I4.5}$$

so that

$$\Psi'(\underline{u}) = \delta(\underline{u}) - \pi\lambda u^2\Delta z\sigma t V(\underline{u}) - i\sigma t V(\underline{u}) \qquad \text{I4.6}$$

Back Fourier transforming the modified wave in the image plane is

$$\psi'(\underline{r}) = 1 - i\sigma t V(\underline{r}) + \lambda\sigma\Delta z t/4\pi\nabla^2 V(\underline{r}) \qquad \text{I4.7}$$

Taking the modulus squared, the image is then

$$I(\underline{r}) = 1 + \lambda\sigma\Delta z t/2\pi\nabla^2 V(\underline{r}) + ... \qquad \text{I4.8}$$

where we are neglecting all the terms with t^2 contributions as small. We next note that from Poisson's equation the crystal potential and charge density $\rho(\underline{r})$ are related by the equation

$$\nabla^2 V(\underline{r}) = - 4\pi/\varepsilon_o \rho(\underline{r}) \qquad \text{I4.9}$$

so that we can write

$$I(\underline{r}) = 1 - 2\lambda\sigma\Delta z t/\varepsilon_o\rho(\underline{r}) \qquad \text{I4.10}$$

Equation 4.10 tells us that for small spatial frequencies, i.e. a low resolution image (1 nm resolution) and a very thin specimen the image will be proportional to the projected charge density of the specimen. This approximation is useful for Fresnel fringe phenomena, but somewhat restricted in more general use.

An important concept that we can introduce at this stage is phase contrast versus amplitude contrast. In equation I4.1 we are approximating the effect of the specimen as a change in the phase of part of the wave by $\pi/2$. We can visualize this readily by using an Argand (complex plane) diagram. For a completely perfect microscope (no defocus or spherical aberration) the image intensity would be:

$$I(\underline{r}) = 1 + |\sigma t V(\underline{r})|^2 \qquad \text{I4.11}$$

and since the second term is very small, we would see very little contrast. When we include the defocus and spherical aberration this adds an addition phase change to the scattered wave and swings part of it back onto the real axis of the Argand diagram. The contrast that we will see if the phase change from the imaging is $\pi/2$ is the imaginary component of the scattered wave, what we call the phase component. The real component of the scattered wave, which is substantial with more complete diffraction models, is called the amplitude component and will be observed even for 'perfect' imaging conditions. This idea of Phase and Amplitude imaging will reoccur as we further analyze imaging.

4.2 Weak Phase Object Approximation

The next improvement is to again use the essentially Kinematical expression in I4.1 for the exit wave, but now include the lens defocus and spherical aberration more accurately. Using equation I4.3 without approximations the modified wave back in the image plane can be written as:

$$\psi'(\underline{r}) = 1 - i\sigma t \int V(\underline{u}) \exp(-2\pi i \underline{u}.\underline{r} - i\chi(\underline{u}))d\underline{u} \qquad \text{I4.12}$$

Taking the modulus squared to generate the image, and neglecting terms in t^2 as small we have

$$I(\underline{r}) = 1 - i\sigma t \int V(\underline{u}) \exp(-2\pi i \underline{u}.\underline{r} - i\chi(\underline{u}))d\underline{u}$$

$$+ i\sigma t \int V^*(\underline{u}) \exp(2\pi i \underline{u}.\underline{r} + i\chi(\underline{u}))d\underline{u} \qquad \text{I4.13}$$

If the crystal potential is real, which implies that we neglect any absorption (justifiable for a thin

specimen), then $V(\underline{u})$ is conjugate symmetric, i.e.

$$V(-\underline{u}) = V^*(\underline{u}) \qquad \text{I4.14}$$

In addition, if the crystal has a center of symmetry (a reasonable assumption in general) then $V(\underline{u})$ is real so that

$$V(\underline{u}) = V(-\underline{u}) \qquad \text{I4.15}$$

If we finally assume that there is no beam tilt so that $\chi(\underline{u}) = \chi(-\underline{u})$ we can collect the terms with $\underline{u}$ and $-\underline{u}$ which gives us

$$I(\underline{r}) = 1 - i\sigma t/2 \int V(\underline{u})\{ \exp(-2\pi i\underline{u}.\underline{r} - i\chi(\underline{u})) - \exp(2\pi i\underline{u}.\underline{r} + i\chi(\underline{u})) - \exp(-2\pi i\underline{u}.\underline{r} + i\chi(\underline{u})) +$$

$$\exp(2\pi i\underline{u}.\underline{r} - i\chi(\underline{u})) \} d\underline{u} \qquad \text{I4.16}$$

(where we divide by 2 so that we can still use the full range of $\underline{u}$ for the integration)

$$= 1 - \sigma t \int V(\underline{u})\{\sin(2\pi\underline{u}.\underline{r} + \chi(\underline{u})) - \sin(2\pi\underline{u}.\underline{r} - \chi(\underline{u}))\}d\underline{u} \qquad \text{I4.17}$$

$$= 1 + \sigma t \int V(\underline{u})\cos(2\pi\underline{u}.\underline{r}) \{-2\sin(\chi(\underline{u}))\}du \qquad \text{I4.18}$$

We refer to the term $-2\sin\chi(\underline{u})$ as the weak phase object contrast transfer term. Its significance is that with the approximations that we have used (which includes at present the neglect of all the incoherent imaging terms) the sign and amplitude of the contrast for the fringe structure in the image corresponding to the spatial frequency $\underline{u}$ is governed by this term. In this sense we can think of the electron microscope as a filter, changing the sign and contrast of the spacings in the image, filtering out some of them. (Remember that with our definition the potential is negative.)

Which frequencies are passed through will depend upon the exact value of the objective lens defocus Δz that is used, and it is very useful to consider graphs of $-2\sin\chi(\underline{u})$ versus $\underline{u}$ for different values of the lens defocus. (As an exercise it is informative to write a program on a small PC for this purpose.) It is useful for the graphs to use what are called reduced co-ordinates $\underline{u}^*$ and Δz^* which are defined by the relationships

$$u^* = (Cs\lambda^3)^{1/4}u \;;\; \Delta z^* = -(Cs\lambda)^{-1/2}\Delta z \qquad \text{I4.19}$$

which allow the reduction of the phase shift term to

$$\chi(\underline{u}) = \pi (u^{*4} - \Delta z^* u^{*2}) \qquad \text{I4.20}$$

thereby eliminating the machine dependent wavelength and Cs terms at the expense of losing some information about what the true physical parameters are. A large amount of qualitative information can be gleaned from these graphs, and we will only here consider one of the main features of interest, the occurrence of large pass bands where the contrast for a range of spatial frequencies is almost the same. These occur for small negative values of defocus (positive Δz^*), where the transfer function has a value close to 1 for a relatively broad range of values of u^*. Of particular interest is the so-called Shertzer defocus of $\Delta z^* = 1$, often called the 'optimum' defocus. Whilst it is fair to say

that this defocus gives transfer with the same sign and almost the same magnitude over a broad range, and is therefore good for specimens which contain much scattering in this region, it may well be that the spacings of interest physically do not fall within this band. (Generalizations such as 'optimum' defocus are only generalizations!) Other broad pass bands occur for the series $\Delta z^* = \sqrt{n}$ where n is an integer. A slightly larger pass band can be obtained in fact for a slightly larger value of the defocus, approximately $1.2\sqrt{n}$.

4.3 Weak Amplitude Object Approximation

The weak phase approximation assumes that the diffraction is Kinematical, whereas we know in reality that for anything except a very thin specimen the diffraction is really dynamical. Therefore the phase of the scattered component of the exit wave need not be purely imaginary, but could include a real component as well. If we consider that the outgoing wave has the form

$$\psi(\underline{r}) = 1 - \alpha V(\underline{r}) \qquad \text{I4.21}$$

then by using the same arguments as in our derivation of the weak phase object approximation, the final image will be

$$= 1 + \alpha \int V(\underline{u})\cos(2\pi \underline{u}.\underline{r}) \{2\cos(\chi(\underline{u}))\}du \qquad \text{I4.22}$$

We refer to the term $2\cos\chi(\underline{u})$ as the weak amplitude transfer term. It differs from the $-2\sin\chi(\underline{u})$ term in that it is large when $\underline{u}$ is small. This implies that we can see readily large diffraction spacing effects due to dynamical diffraction in our images whereas we cannot see these effects when they are purely due to Kinematical scattering.

4.4 Linear Imaging Theory

Whilst the weak phase and amplitude object forms that we presented above provide us with a lot of useful information, they suffer from one obvious defect - they do not include the effects of any of the incoherent imaging phenomena. Based upon just the weak phase object analysis one might say that the resolution of the microscope is limited by the extent of the broad pass band, and use an objective aperture to remove any of the spatial frequencies beyond this. This type of argument suggests an aperture size of $(C_s\lambda^3)^{-¼}$ corresponding to a microscope resolution of $(C_s\lambda^3)^{¼}$, and is based around the idea of a disc of least confusion. However, if the microscope does pass through any spatial frequency it is inappropriate to eliminate the information which we can in principle use (with a little calculation) just for the sake of simplicity. It is more important to try and push our interpretation of all the information that we can obtain as far as possible. For this it is important to consider the incoherent aberrations which place unavoidable limits on the microscope resolution.

Our next step is therefore to consider the effect of the incoherent averages. There are two ways to approach this problem, namely to use approximate analytical methods or to use a full numerical integration approach. The former is more useful for giving us a physical feel and we will discuss it here; the later is a more accurate procedure that is often used in numerical image simulation programs.

For the moment at least we will stay with our earlier use of a weakly scattering object for our analysis. This gives us a good idea of the effects that will be introduced, although it is a severely limited model. In the weak phase object approximation, the image (without any incoherent aberrations) is:

$$I(\underline{r}) = 1 + \sigma t \int V(\underline{u})\cos(2\pi \underline{u}.\underline{r}) \{-2\sin(\chi(\underline{u}))\}du \quad \text{I4.23}$$

We must integrate this expression for a range of values of defocus (focal spread) and beam tilt values. (We ignore any changes in the diffraction when the incident beam direction changes, a problem that we will return to later.) Provided that the spread in defocus values and tilts are small we can expand $\chi(\underline{u})$ as a Taylor series in the form (neglecting astigmatism)

$$\chi(\underline{u},\Delta f,\underline{w}) = \pi/\lambda([\Delta z+\Delta f]\lambda^2|\underline{u}-\underline{w}|^2 + 1/2\ Cs\lambda^4|\underline{u}-\underline{w}|^4) \quad \text{I4.24}$$

$$= \chi(\underline{u},0,\underline{0}) + \underline{w}.\underline{\nabla}\chi(\underline{u},0,\underline{0}) + \pi\lambda\Delta f u^2 + ... \quad \text{I4.25}$$

Using $\chi(\underline{u})$ rather than $\chi(\underline{u},0,\underline{0})$ for simplicity, we can write the expression for our final image as

$$I(\underline{r}) = 1 + \sigma t \int V(\underline{u})\cos(2\pi \underline{u}.\underline{r})$$

$$\{ (-2) \iint \sin(\chi(\underline{u}) + \underline{w}.\underline{\nabla}\chi(\underline{u}) + \pi\lambda\Delta f u^2)F(\Delta f)S(\underline{w})d\Delta f d\underline{w} \}du \quad \text{I4.26}$$

(We are assuming that both the focal spread and convergence distributions are normalized to unity, and using the same notation as in section I3.3.) As a rule the focal spread and convergence distributions will be symmetrical functions, so if we expand the sin term and only retain the symmetrical contribution we have:

$$I(\underline{r}) = 1 + \sigma t \int V(\underline{u})\cos(2\pi \underline{u}.\underline{r})\{-2\sin(\chi(\underline{u}))\}E(\underline{u})d\underline{u} \quad \text{I4.27}$$

where $E(\underline{u})$, called the linear envelope term is given by the equation

$$E(\underline{u}) = \int \cos(\underline{w}.\underline{\nabla}\chi(\underline{u}))S(\underline{w})d\underline{w} \int \cos(\pi\lambda\Delta f u^2)F(\Delta f)d\Delta f \quad \text{I4.28}$$

$$= S(\underline{\nabla}\chi(\underline{u})/2\pi)F(\lambda u^2/2) \quad \text{I4.29}$$

where S and F are the cosine Fourier transforms of the convergence and focal spread distributions respectively. Our main reason for using the above analysis is that we end up with a relatively simple analytical form for the effects of the convergence and focal spread which we can look up using Fourier transform tables.

We will complete our analysis by using Gaussian distributions for the convergence and focal spread. Other distributions can be used, but there is no clearcut evidence at present indicating that any particular form is best. We then write:

$$S(\underline{w}) = (\alpha/\pi)\exp(-\alpha w^2)\ ;\ F(\Delta f) = \sqrt{(ß/\pi)}\exp(-ß\Delta f^2) \quad \text{I4.30}$$

in which case

$$E(\underline{u}) = \exp(-|\underline{\nabla}\chi(\underline{u})|^2/4\alpha)\ \exp(-\pi^2\lambda^2u^4/4ß) \quad \text{I4.31}$$

The envelope function $E(\underline{u})$ is a damping function which limits in a very fundamental way what spatial frequencies we will obtain in our image. Whilst if $\sin(\chi(\underline{u}))$ happened to be zero for a

particular spacing and defocus value so that this spacing does not appear in our image, we could always use a different defocus to increase the contrast. This is not the case for the envelope term (except to some extent for the convergence contribution). If the envelope is small, there is no way that we will see this information in our image and this therefore represents the more fundamental limit to our attainable resolution.

It is informative to consider, physically what these envelope terms are due to. First let us consider the convergence contribution. When we change the beam direction we are in effect changing the effective value of the spatial frequency to use in our phase shift term $\chi(\underline{u})$. How big a difference this makes to our image will depend upon how fast $\chi(\underline{u})$ is varying - if it is fast then positive and negative terms can cancel out and our resultant image contrast will be very small. There is a special condition when the convergence vanishes which is when

$$\underline{\nabla}\chi(\underline{u}) = \underline{0} = 2\pi\lambda\underline{u}(\Delta z + Cs\lambda^2u^2) \qquad \text{I4.32}$$

i.e. $$\Delta z = -Cs\lambda^2u^2 \qquad \text{I4.33}$$

As we will see later when we look at what contrast transfer means in the image plane, this is a rather special condition corresponding to the case when information from the object returns to the correct place in the image. This particular defocus value is called the overlap defocus for the spacing $\underline{u}$. Note that as we go to more negative defoci we can therefore reduce the convergence contribution for larger values of $\underline{u}$, thereby effectively increasing our resolution.

For the focal spread, there is no defocus contribution and we cannot therefore gain anything by changing our defocus. We can therefore consider the envelope term as a soft aperture which fundamentally limits our resolution, and similarly consider the convergence as an aperture. As an estimate, if the combined effect of these two terms is less than 0.1 we probably will not be able to resolve the fringe structure in the image so the value of $\underline{u}$ where this occurs is the true information limit of the microscope.

The linear theory that we have just described is particularly useful on two counts; it contains the leading terms that effect images in a sensible fashion and can also be used to explain the image contrast from amorphous films. The later are often used for this reason to test the resolution of an electron microscope and are also very useful for correction of astigmatism. This is discussed in Appendix A. However it should not be forgotten that linear imaging theory is analogous to Kinematical diffraction theory; good as a guide but not to be trusted quantitatively.

It is straightforward to extend from the above to consider a weak amplitude object; all that is required is the replacement of $-2\sin\chi(\underline{u})$ by $\cos\chi(\underline{u})$ and this can be generalized for a weak object by using instead $\exp(-i\chi(\underline{u}))$.

5. Non-Linear Imaging Theory

The linear theory that we have just described is qualitatively very useful but is limited to the case when the scattering is very weak. In practice it is necessary to avoid this approximation and consider what happens when the scattering is stronger. We start from the general form of our phase shifted wave

$$\Psi'(\underline{u}) = \Psi(\underline{u})\exp(-i\chi(\underline{u})) \qquad \text{I5.1}$$

for a specific value of the objective lens focus and beam tilt. The corresponding intensity is

$$I(\underline{r}) = |\int \Psi(\underline{u})\exp(2\pi i\underline{u}.\underline{r} - i\,\chi(\underline{u}))d\underline{u}|^2 \qquad I5.2$$

$$= \iint \Psi^*(\underline{u}')\,\Psi(\underline{u})\exp(2\pi i[\underline{u}-\underline{u}'].\underline{r} - i\,\chi(\underline{u}) + i\,\chi(\underline{u}'))d\underline{u}'d\underline{u} \qquad I5.3$$

We now make the substitution (to simplify the mathematics)

$$\underline{v} = \underline{u} - \underline{u}' \qquad I5.4$$

to give

$$I(\underline{r}) = \int \exp(2\pi i\underline{v}.\underline{r})\,d\underline{v}\int \Psi^*(\underline{u}-\underline{v})\,\Psi(\underline{u})\exp(-i\,\chi(\underline{u}) + i\chi(\underline{u}-\underline{v}))d\underline{u} \qquad I5.5$$

The first integral is a Fourier transform, so if we write the Fourier transform of the image as $P(\underline{v})$ where $P(\underline{v})$ denotes the amplitude of the spatial frequency $\underline{v}$ in the image,

$$P(\underline{v}) = \int \Psi^*(\underline{u}-\underline{v})\,\Psi(\underline{u})\exp(-i\,\chi(\underline{u}) + i\,\chi(\underline{u}-\underline{v}))d\underline{u} \qquad I5.6$$

We will now work solely with $P(\underline{v})$. In order to include all the incoherent effects we have to integrate over our focal spread and beam tilt. To generate an analytical expression, we expand the phase shift term as in the linear theory, i.e. use a phase shift

$$\chi(\underline{u}) + \underline{w}.\underline{\nabla}\chi(\underline{u}) + \pi\lambda\Delta f u^2 + ... \qquad I5.7$$

so that

$$P(\underline{v}) = \int\Psi^*(\underline{u}-\underline{v})\,\Psi(\underline{u})\exp(-i\,\chi(\underline{u}) + i\,\chi(\underline{u}-\underline{v}))d\underline{u}$$

$$\iint\exp(-i\underline{w}.(\underline{\nabla}\,\chi(\underline{u})-\underline{\nabla}\,\chi(\underline{u}-\underline{v})) - \pi i\lambda\Delta f(u^2-|\underline{u}-\underline{v}|^2))F(\Delta f)S(\underline{w})d\Delta f d\underline{w} \qquad I5.8$$

$$= \int \Psi^*(\underline{u}-\underline{v})\,\Psi(\underline{u})\exp(-i\,\chi(\underline{u}) + i\,\chi(\underline{u}-\underline{v}))E(\underline{u},\underline{u}-\underline{v})d\underline{u} \qquad I5.9$$

with

$$E(\underline{u},\underline{u}-\underline{v}) = S([\underline{\nabla}\,\chi(\underline{u})-\underline{\nabla}\,\chi(\underline{u}-\underline{v})]/2\pi)F([u^2-|\underline{u}-\underline{v}|^2]/2) \qquad I5.10$$

The result is qualitatively the same as that for the linear theory, the difference being that we have changed the form of the term in S and F, the Fourier transforms of the convergence and focal spread distributions. To be more concrete, it is useful to introduce specific forms for these distributions. Using the same forms as for the linear theory, i.e.

$$S(\underline{w}) = (\alpha/\pi)\exp(-\alpha w^2)\ ;\ F(\Delta f) = \sqrt{(ß/\pi)}\exp(-ß\Delta f^2) \qquad I5.11$$

then

$$E(\underline{u},\underline{u}-\underline{v}) = \exp(-|\underline{\nabla}\chi(\underline{u})-\underline{\nabla}\chi(\underline{u}-\underline{v})|^2/4\alpha)\exp(-\pi^2\lambda^2(u^2-|\underline{u}-\underline{v}|^2)^2/4ß) \qquad I5.12$$

This analytical expression can be used to generate a lot of useful information. Physically what we have to do in our integral of equation I5.9 is consider all the different interference terms

which can give us a particular spatial frequency. For instance, let us consider a simple [110] pole of a face centered material and assume that we are using the (111) and (200) spots to produce our image. In linear theory we only obtain (200) spatial frequencies in our image when the (200) diffracted wave interferes with the transmitted beam. The non-linear theory gives us a more complete picture; we can also obtain (200) spatial frequencies when the (111) and (111) beams interfere. Indeed, we can even obtain spatial frequencies such as (222) and (311) which are not allowed through the objective aperture. We refer to these additional spatial frequencies as second order effects, as they are second order in the magnitude of the diffraction.

The intensities of the different spatial frequencies can also be radically different in the non-linear theory compared to the linear theory. Let us consider first the focal spread term. In the simple linear case the larger the spatial frequency, the larger the damping from the focal spread. Compare this to the (200) fringes produced by interference of the (111) and (111) beams. In this case u and $|\underline{u}-\underline{v}|$ are the same and there is no damping at all. It is in fact quite conceivable that the second order contribution is far larger than the first order frequency! Note that by choosing the overlap defocus for (111) fringes we can also make the convergence envelope term go to one.

What therefore comes out of the full non-linear theory is that we cannot simply neglect the interference between two diffracted beams as small using as our argument the fact that it is second order in the diffraction amplitude, since the attenuation due to the envelope terms is generally larger for the linear contributions (diffracted wave with the straight through wave), compensating for the diffraction amplitude to some extent. The linear theory can be workable for a very thin specimen, but not for a thicker specimen when the diffracted beams can approach or exceed the intensity of the straight through beam.

6 Imaging in real space

Whilst our treatment of imaging theory so far is mathematically correct, it has one major limitation - all it tells us is the magnitude of the different spatial frequencies passed to the image. The action of the microscope in scrambling the phases of the different beams has another important consequence; it also changes the position where information appears in the images. It is important to consider what the effects are in the image plane to develop a feel for how to interpret images in detail. We will explore here three different approaches to the problem of the image form in real space.

6.1 Point response function.

In the linear theory, we end up with an expression for the image intensity of

$$I(\underline{r}) = 1 + \sigma t \int V(\underline{u})\cos(2\pi\underline{u}.\underline{r})\{-2\sin(\chi(\underline{u}))\}E(\underline{u})d\underline{u} \qquad \text{I6.1}$$

Writing

$$I(\underline{r}) = 1 + \sigma t \int T(\underline{u})\cos(2\pi\underline{u}.\underline{r})V(\underline{u})d\underline{u} \qquad \text{I6.2}$$

with

$$T(\underline{u}) = -2\sin(\chi(\underline{u}))E(\underline{u}) \qquad \text{I6.3}$$

where we call $T(\underline{r})$ the point response function of the microscope. We can now rewrite the image

intensity as

$$I(\underline{r}) = 1 + \sigma t \int T(\underline{r}-\underline{r}')T(\underline{r}')d\underline{r}' \qquad \text{I6.4}$$

Physically we can interpret equation I6.4 by saying that the information in our scattered wave is averaged in position over the function $T(\underline{r})$. For an ideal microscope $T(\underline{r})$ would be simply a delta function. As a gauge of our resolution we could use the half-width at half height of $T(\underline{r})$, relating $T(\underline{r})$ to the size and form of the disc of confusion.

6.2 Dispersive expansion

Another way of considering the imaging problem which yields far more hard information in real space is to consider each term in the Taylor series expansion of the phase shift and examine its effect upon the image. Let us start with an expansion of the electron wave leaving our specimen as a set of diffracted waves whose amplitudes vary with position, i.e.

$$\psi(\underline{r}) = \sum_{\underline{g}} \phi_{\underline{g}}(\underline{r})\exp(-2\pi i \underline{g}.\underline{r}) \qquad \text{I6.5}$$

This is a realistic model of the wave after a crystal (as against that of simple linear theories which are aimed at amorphous materials). In reciprocal space the wave is therefore

$$\Psi(\underline{u}) = \sum_{\underline{g}} \Psi_{\underline{g}}(\underline{g}+\underline{u}) \qquad \text{I6.6}$$

which corresponds to a wave with diffuse intensity around the diffraction spot. Including our phase shift we have

$$\Psi(\underline{u}) = \sum_{\underline{g}} \Psi_{\underline{g}}(\underline{g}+\underline{u})\exp(-i\chi(\underline{g}+\underline{u})) \qquad \text{I6.7}$$

Going back into the image plane the modified wave is

$$\Psi(\underline{u}) = \sum_{\underline{g}} \{\int \Psi_{\underline{g}}(\underline{r}-\underline{r}')T(\underline{r}')d\underline{r}'\}\exp(-2\pi i \underline{g}.\underline{r}) \qquad \text{I6.8}$$

where $T(\underline{r}')$ is the transform of $\exp(-i\chi(\underline{g}+\underline{u}))$. We again have an averaging of the information in the wave leaving our specimen over position. We next consider the effect of each of the different terms in our phase shift in terms of averaging this information. First we expand (using reduced coordinates)

$$\chi(\underline{g}+\underline{u}) = \chi(\underline{g}) + 2\pi \underline{g}.\underline{u}D + \pi u^2 D + 2\pi(\underline{g}.\underline{u})^2 + 2\pi(\underline{g}.\underline{u})u^2 + 1/2\pi u^4 \qquad \text{I6.9}$$

where $D = \Delta z^* - g^2$. Now taking each term in turn:

a) The first, $\exp(-2\pi i \underline{g}.\underline{u}D)$, corresponds to a shift of the information by $D\underline{g}$. This shift is equivalent to the shift of an off axis beam with defocus in a simple ray diagram analysis of an electron microscope, and is readily apparent in images by the motion (as a function of defocus) of the

diffracted beams across the specimen (called ghost images).

b) The second, $\exp(-\pi i u^2 D)$ the Fresnel term with a defocus of D, not simply Δz^*. This indicates that the off axis diffracted beam has a different defocus to that of the transmitted beam. Physically we can think of this term as a Gaussian spreading the information equally in all directions.

c) The third, $\exp(-2\pi i(\underline{g}.\underline{u})^2)$ is an astigmatism term. This indicates that the information is really spread anisotropically in the two directions normal and perpendicular to $\underline{g}$. Note that this term can be canceled by some genuine astigmatism in the microscope operating conditions.

d) The fourth, $\exp(-2\pi i(\underline{g}.\underline{u})u^2)$ is another directional term which resembles Siedel Coma and spreads the information anisotropically.

e) The last, $\exp(-\pi i u^4)$ is the spherical aberration term which leads to an isotropic spreading of information.

We see that the microscope therefore scrambles to some extent the position of information, not simply allows some waves through but not others. It is of interest to note that in many respects the 'best' defocus will be when

$$D = 0, \text{ i.e. } \Delta z^* = g^2 \text{ or } \Delta z = -Cs\lambda^2 g^2 \qquad \text{I6.10}$$

This is the overlap defocus mentioned earlier when discussing the effect of convergence on image contrast.

The above arguments carry over at least quantitatively into the more complete non-linear theory. To a good approximation one can write the image intensity as (L.D.Marks, Ultramicroscopy 12, 237, 1984)

$$I(\underline{r}) = \sum_{g,q} \exp(-2\pi i[\underline{g}-\underline{q}]) \int\phi_g(\underline{r}-\underline{r}')A_{g,q}(\underline{r}')d\underline{r}'\int\phi_q(\underline{r}-\underline{r}')A_{q,g}(\underline{r}')d\underline{r}' \qquad \text{I6.11}$$

where $A_{g,q}(\underline{r})$ is the Fourier transform of a modified envelope term.

6.3 Image localization

To conclude our discussion of where information is in the images, we will use an approach which specifically looks at the positions of the information in images. We substitute for the modulation of the wave as a function of position in the image using the equation

$$\phi_g(\underline{r}) = N^{-½} \sum_{r_n} \phi_g(\underline{r}_n)w(\underline{r}-\underline{r}_n) \qquad \text{I6.12}$$

where

$$w(\underline{r}-\underline{r}_n) = N^{-½} \int\exp(-2\pi i\underline{u}.[\underline{r}-\underline{r}_n])d\underline{u} \qquad \text{I6.13}$$

with N the number of unit cells in the image (introduced to normalize the integrals) and the positions $\underline{r}_n$ are the atomic sites in a perfect crystal. The expansion in equations I6.12 and I6.13 is called a Wannier expansion and is a standard trick in solid state physics for switching from reciprocal to real space. Carrying out a weak phase object approximation the image intensity can be written as

(L.D.Marks, Ultramicroscopy 18,33,1985)

$$I(\underline{r}) = 1\text{-}2N^{-½} \sum_{g} \cos(2\pi \underline{g}.\underline{r}) \sum_{r_n} \phi_g(\underline{r}-\underline{r}_n)L(\underline{r}_n) \qquad \text{I6.14}$$

$$\text{with } L(\underline{r}_n) = N^{-½} \int (T(\underline{g}+\underline{u}) + T^*(-\underline{g}-\underline{u}))\exp(-2\pi i \underline{u}.\underline{r}_n)d\underline{r}_n \qquad \text{I6.15}$$

We have generated a form in equations I6.14 and I6.15 where we are specifically considering an average of the information over the different lattice points for each of the diffracted beams that we are interested in. We refer to $L(\underline{r}_n)$ as the localization function and it tells us the information that we need. For instance, if $L(\underline{r}_n)$ is a delta function then all the information in our object for the particular lattice spacing is correctly positioned in the image, whereas if it is large then the information at any given 'black dot' in the image is really the average of the information at many different positions. Note that one can still obtain strong fringe contrast when the localization is large (i.e. a delocalized image), but point information for instance point defect information will be spread out over a large area.

>> Some expansion to be added here

6.4 CTF with large convergence (coming soon)

6.5 CTF in STEM BF and Z-contrast (coming soon)

7 Tilted beam contrast transfer.

It is informative to consider the weak phase object approximation in the case of beam tilt. This shows one of the experimental problems in current high resolution electron microscopy - precise correction of beam tilt. We start from our expression for the intensity:

$$I(\underline{r}) = 1 - i\sigma t \int V(\underline{u}) \exp(-2\pi i \underline{u}.\underline{r} - i\chi(\underline{u}))d\underline{u}$$

$$+ i\sigma t \int V^*(\underline{u}) \exp(2\pi i \underline{u}.\underline{r} + i\chi(\underline{u}))d\underline{u} \qquad \text{I7.1}$$

with the tilted beam form of the phase shift

$$\chi(\underline{u}) = \pi/\lambda(\Delta z\lambda^2|\underline{u}-\underline{w}|^2 + 1/2Cs\lambda^4|\underline{u}-\underline{w}|^4) \qquad \text{I7.2}$$

As before we take terms with $\underline{u}$ and $-\underline{u}$ together, but we must now remember that $\chi(\underline{u})$ is not equal to $\chi(-\underline{u})$. We therefore have

$$I(\underline{r}) = 1 - i\sigma t/2 \int V(\underline{u})\{ \exp(-2\pi i \underline{u}.\underline{r} - i\chi(\underline{u})) -$$

$$\exp(2\pi i \underline{u}.\underline{r} + i\chi(\underline{u})) + \exp(-2\pi i \underline{u}.\underline{r} + i\chi(-\underline{u})) -$$

$$\exp(2\pi i \underline{u}.\underline{r} - i\chi(-\underline{u})) \} d\underline{u} \qquad \text{I7.3}$$

$$= 1 + \sigma t \int V(\underline{u})\{ \sin(2\pi \underline{u}.\underline{r}+\chi(\underline{u})) - \sin(2\pi \underline{u}.\underline{r} - (-\underline{u})\}d\underline{u} \qquad \text{I7.4}$$

$$= 1 - \sigma t\int V(\underline{u})\{ \cos(2\pi\underline{u}.\underline{r})[\sin\chi(\underline{u})+\sin\chi(-\underline{u})]$$

$$- \sin(2\pi\underline{u}.\underline{r})[\cos\chi(\underline{u}) - \cos\chi(-\underline{u})] \}d\underline{u} \qquad \text{I7.5}$$

The form in equation I7.5 is different from that for the case without any tilt. As a further step, it is useful to consider the case when the beam tilt is small. Then we can expand

$$\chi(\underline{u}) = \chi(\underline{u}) + \underline{w}.\underline{\nabla}\chi(\underline{u}) + ... \qquad \text{I7.6}$$

and in our analysis of the envelope terms above. (On the right in equation I7.6 we are using $\chi(\underline{u})$ to denote the phase shift without tilt. Remembering that $\underline{\nabla}\chi(\underline{u}) = -\underline{\nabla}\chi(-\underline{u})$, I7.5 simplifies to

$$I(\underline{r}) = 1 + \sigma t \int V(\underline{u})(-2\sin\chi(\underline{u}))[\cos(2\pi\underline{u}.\underline{r})\cos(\underline{w}.\underline{\nabla}\chi(\underline{u}))$$

$$-\sin(2\pi\underline{u}.\underline{r})\sin(\underline{w}.\underline{\nabla}\chi\underline{u}))]d\underline{u} \qquad \text{I7.7}$$

If $\underline{\nabla}\chi(\underline{u})$ is large the difference relative to the non-tilted case appears at first sight to be large, but this is also the same condition for the Envelope term due to convergence to be small. Thus the difference at least for an amorphous specimen will tend to be hidden by the attenuation from the envelope term. (It turns out that the same tilt for an crystalline specimen has a large effect on the appearance of the images, destroying the symmetry.)

The fact that the transfer function changes when the beam is tilted is currently exploited as a method for correcting beam tilt within the electron microscope. The method involves superimposing an additional tilt $\pm\underline{t}$ and comparing these two images. Writing the phase shift term as

$$\chi(\underline{u}) = \pi/\lambda (\Delta z\lambda^2|\underline{u}+\underline{w}\pm\underline{t}|^2 + 1/2Cs\lambda^4|\underline{u}+\underline{w}\pm\underline{t}|^4) \qquad \text{I7.8}$$

$$= \pi/\lambda \{ \underline{u}.[2\lambda^2 \Delta z(\underline{w}\pm\underline{t}) + 2Cs\lambda^4(\underline{w}\pm\underline{t})^3]$$

$$+ \lambda^2u^2 [\Delta z + 2Cs\lambda^2|\underline{w}\pm\underline{t}|^2]$$

$$+ 4Cs\lambda^4(\underline{u}.[\underline{w}\pm\underline{t}])^2$$

$$+ 2Cs\lambda^4\underline{u}.(\underline{w}\pm\underline{t})^{u2}$$

$$+ \Delta z\lambda^2 |\underline{w}\pm\underline{t}|^2 + 1/2Cs\lambda^4(u^4+(\underline{w}\pm\underline{t})^4)\} \qquad \text{I7.9}$$

where we have deliberately ordered the terms in order of the u coefficients, we can generate substantial information by considering the order of the terms. Starting from the first line of equation I7.9, the first term is linear and corresponds to a shift in the image plane. This shift depends upon the magnitude of both the net tilt $\underline{w}\pm\underline{t}$ and the defocus. Minimizing this shift when the focus is varied is the principle of current centering. The next term is the same in character as the defocus term in the untilted phase shift, but contains an apparent additional focus of $2Cs\lambda^2|\underline{w}\pm\underline{t}|^2$. When $\underline{w}$ is non zero, the images at $\pm t$ will look as if they are at different defoci. Visual inspection of the images to make the granularity of the amorphous film similar can therefore be used to reduce $\underline{w}$. The next term has the same structure as astigmatism. Therefore when $\underline{w}=\underline{0}$ the images should appear to have

the same directionality. (Note that this is why astigmatism can correct for beam tilt to some extent.) The fourth term is cubic in u and has the structure of what is called Coma. It again leads to directionality in the image. Finally there is the untilted spherical aberration term and a phase shift which will be the same for all u values (which can therefore be neglected).

We can therefore correct for beam tilt by superimposing $\pm\underline{t}$ and correcting visually images of an amorphous film until the two images appear to have the same degree of directionality. It is also often possible to use the effective focus change, i.e. Fresnel fringes if there is no amorphous film. In practice this is far more accurate than simple current centering.

>> Some expansion here on coma-free methods

8 Partial Coherence

So far we have looked only at the case where the aberrations are either completely coherent, or completely incoherent. These are two extreme cases and the approximation of full incoherence for the convergence works well provided that the illumination is fully focussed, what is often used in practice. However, with modern instruments that use a field emission source or when the illumination is not focussed we cannot use these approximations. Under these circumstances one has to include the effects of partial coherence. (This was in fact suspected very early on by, but was not found to be relevant when the original imaging theory was developed for tungsten filament sources.) The current instrumental trend is for microscopes to handle both HREM and small probe (STEM) imaging modes, and to understand even approximately what is going on with a small probe (of the order of 1nm) one again needs to go away from these approximations.

When we looked earlier at the question of coherency, we used the idea that there is some phase difference ϕ between two waves with different wavevectors, then stated that this was either fixed or (statistically) completely random. We do the same thing for partial coherency, but make no assumptions about this phase change. We consider the relationship between the wave at some point $\underline{\mathbf{r}}_1$ and that at another point $\underline{\mathbf{r}}_2$. Everything of interest can be found using the product $\psi(\underline{\mathbf{r}}_1,t_1)^*\psi(\underline{\mathbf{r}}_2,t_2)$, where t_1 and t_2 are two different times. In general, we need to understand the average value of this product over, for instance, the exposure time of an electron micrograph. All the processes whereby we measure electrons are based around electronic excitations which take place fast, on the order of 10^{-13}-10^{-15} seconds, and the distances we are interested in for the electron column are in meters. Therefore two electrons (or parts of one electron wave) which leave the source at the same time with different energies arrive at out detectors sufficiently well seperated that they do not interfere. Thus the product will only be substantialy non-zero if $t_1=t_2$. What is being measured by us is then the statistical average over time of the product, which we write as:

$$\Gamma(\underline{\mathbf{r}}_1,\underline{\mathbf{r}}_2) = \langle\psi(\underline{\mathbf{r}}_1)^*\psi(\underline{\mathbf{r}}_2)\rangle \qquad \text{I8.1}$$

The function $\Gamma(\underline{\mathbf{r}}_1,\underline{\mathbf{r}}_2)$, called the mutual intensity of the electron beam contains all the information that will be of use. The form used here is for the image plane; a completely analogous form exists in reciprocal space, i.e.

$$\Gamma(\underline{\mathbf{u}}_1,\underline{\mathbf{u}}_2) = \langle\Psi(\underline{\mathbf{u}}_1)^*\Psi(\underline{\mathbf{u}}_2)\rangle \qquad \text{I8.2}$$

Just as $\Psi(\underline{\mathbf{u}})$ and $\psi(\underline{\mathbf{r}})$ are related by a Fourier transform, these two forms are related by double Fourier transforms, i.e.

$$\Gamma(\underline{\mathbf{r}}_1,\underline{\mathbf{r}}_2) = \iint\Gamma(\underline{\mathbf{u}}_1,\underline{\mathbf{u}}_2)\exp(-2\pi i[\underline{\mathbf{u}}_1.\underline{\mathbf{r}}_1-\underline{\mathbf{u}}_2.\underline{\mathbf{r}}_2])d\underline{\mathbf{u}}_1 d\underline{\mathbf{u}}_2 \qquad \text{I8.3}$$

$$\Gamma(\underline{\mathbf{u}}_1,\underline{\mathbf{u}}_2) = \iint\Gamma(\underline{\mathbf{r}}_1,\underline{\mathbf{r}}_2)\exp(+2\pi i[\underline{\mathbf{u}}_1.\underline{\mathbf{r}}_1-\underline{\mathbf{u}}_2.\underline{\mathbf{r}}_2])d\underline{\mathbf{r}}_1 d\underline{\mathbf{r}}_2 \qquad \text{I8.4}$$

We would "measure" in an image an intensity at any given point **r** by setting $\underline{\mathbf{r}}_1=\underline{\mathbf{r}}_2=\underline{\mathbf{r}}$, i.e.

$$I(\mathbf{r}) = \Gamma(\underline{\mathbf{r}},\underline{\mathbf{r}}) \qquad \text{I8.5}$$

and use $\underline{\mathbf{u}}$ similarly for the intensity of a diffraction spot. If after the sample we use an electron biprism to produce a translation of part of the image by $\underline{\mathbf{D}}$, the holographic interference terms would be of the form $\Gamma(\underline{\mathbf{r}},\underline{\mathbf{r}}-\underline{\mathbf{D}})$.

A possibly confusing point is that we are suddenly switching from using "classical" quantum mechanical ideas of an electron wavefunction to something rather more complicated. Why, you may ask, is there any need? What we are really doing is something much better, and in fact the classical approach is really not right. We cannot say that the electrons have a specific phase; to do that would be violating the basic premises of quantum mechanics. We can only say what we have via some sort of measurement, which is a statistical sampling. For instance, on a TV screen (or through the binoculars) we don't see a constant intensity but instead a randomly fluctuating one. At any given time there is only (roughly) one electron in the column at a time, so we are seeing the statistics of electrons reaching any particular point. Handling everything in such a statisticaly fashion is what is done in Quantum Field Theory. While our single wavefunction methods are generally O.K., they can in fact go wrong!

Up to now we have assumed that if $\underline{\mathbf{u}}_1$ is different from $\underline{\mathbf{u}}_2$ then the waves at these two points are completely incoherent, i.e.

$$\Gamma(\underline{\mathbf{u}}_1,\underline{\mathbf{u}}_2) = \delta(\underline{\mathbf{u}}_1-\underline{\mathbf{u}}_2)S(\underline{\mathbf{u}}_1) \qquad \text{I8.6}$$

where $S(\underline{\mathbf{u}}_1)$ is the source distribution that was used previously. To see how this approximation breaks down we will next derive a better form which shows properly how the size of the source effects the convergence.

8.1 Effect of the Source Size

Derivation of the mutual intensity function for a given source size is one of the classic cases which can be found in many text books. Suppose that our source emits electrons over a particular region, and at any point $\underline{\mathbf{r}}$ of the source the amplitude is $s(\underline{\mathbf{r}})$; the size of this region can be included by letting $s(\underline{\mathbf{r}})=0$ outside a certain range of values for $\underline{\mathbf{r}}$. I will further assume that all different points on the source emit incoherenty with respect to all others. This is a reasonable approximation with a LaB_6 source where the emitting region is something like a micron in size, and there is no coherence between the electrons in the solid of such a size regime; with a small field emitter tip it may not be such a good approximation!

We can immediately write down the mutual intensity function for the source, which is simply:

$$\Gamma(\underline{\mathbf{r}}_1,\underline{\mathbf{r}}_2) = \delta(\underline{\mathbf{r}}_1-\underline{\mathbf{r}}_2)|s(\underline{\mathbf{r}}_1)|^2 \qquad \text{I8.7}$$

Transforming to reciprocal space we have

$$\Gamma(\underline{\mathbf{u}}_1,\underline{\mathbf{u}}_2) = \iint\delta(\underline{\mathbf{r}}_1-\underline{\mathbf{r}}_2)|s(\underline{\mathbf{r}}_1)|^2\exp(+2\pi i[\underline{\mathbf{u}}_1.\underline{\mathbf{r}}_1-\underline{\mathbf{u}}_2.\underline{\mathbf{r}}_2])d\underline{\mathbf{r}}_1 d\underline{\mathbf{r}}_2 \quad \text{I8.8}$$

$$= \int|s(\underline{\mathbf{r}})|^2\exp(+2\pi i[\underline{\mathbf{u}}_1-\underline{\mathbf{u}}_2].\underline{\mathbf{r}})d\underline{\mathbf{r}} \quad \text{I8.9}$$

$$= S(\underline{\mathbf{u}}_1-\underline{\mathbf{u}}_2) \quad \text{I8.10}$$

where $S(\underline{\mathbf{u}})$ is the Fourier Transform of $|s(\underline{\mathbf{r}})|^2$. If the source is relatively large, the mutual intensity approaches a delta function so we can recover the approximation used in the previous sections, although there are some additional conditions that we will come to later. For an intrinsicaly small source such as a field-emission tip there is no way that the result is the same.

It is informative to look at the case where we have complete incoherence in the diffraction plane, with a condensor aperture. The result is the same as what we have above, with u and r switched, i.e.:

$$\Gamma(\underline{\mathbf{r}}_1,\underline{\mathbf{r}}_2) = \int|A(\underline{\mathbf{u}})|^2\exp(+2\pi i[\underline{\mathbf{r}}_1-\underline{\mathbf{r}}_2].\underline{\mathbf{u}})d\underline{\mathbf{u}} \quad \text{I8.11}$$

$$= a(\underline{\mathbf{r}}_1-\underline{\mathbf{r}}_2) \quad \text{I8.12}$$

with $A(\underline{\mathbf{u}})$ the condensor aperture ($|A(\underline{\mathbf{u}})|^2=A(\underline{\mathbf{u}})$ if the aperture has a hard edge) and $a(\underline{\mathbf{r}})$ its Fourier transform. This type of imaging condition where the wave coherence only depends upon the relative seperation between any two points is call isoplanaric illumination. If the aperture has a width of u_o is reciprocal space, the width of its transform is approximately $1/u_o$. This the transverse coherence length, and points in the image plane seperated by more than this do not coherently interfere with each other. For instance, in a holographic experiment you will only see interference effects if the distance between the points translated such that they overlap (in the final image) is smaller than the transverse coherence lenght. Most of you have already come across this in some early physics experiments. You start with a very incoherent light source (e.g. a light bulb) and put in front of it a small aperture. By doing this you get a somewhat coherent source that you can then do simple optics experiments with.

However, there is a trap here that we have to be cautious about. In the form given above all possible points seperated by less than the transverse coherence width interfere coherently. In reality, only a certain region of the sample is illuminated and we have somehow lost this information. The source of this is that we have taken a limit of fully incoherent illumination too early in the derivation which we are really not allowed to do.

8.2 Specimen Illumination in HREM mode

The next step that we will take is to generate a more proper description of the illumination of the sample. This is the same in principle as what we have done for the post-specimen imaging, now using the lens aberrations above the sample and the mutual intensity rather than the wavefunction. The aberrations of the lenses above the sample can be included as a phase shift term $\gamma(\underline{\mathbf{u}})$ where:

$$\gamma(\underline{\mathbf{u}}) = \pi/\lambda(\Delta f\lambda^2u^2 + 1/2C_p\lambda^4u^4) \quad \text{I8.13}$$

with Δf the defocus of the condensor-objective prefield region and C_p the spherical aberration of this lens. (Tilt and astigmatism have been taken as zero here.) The wave at the sample can then be described as:

$$\Gamma(\underline{\mathbf{r}}_1,\underline{\mathbf{r}}_2) = \iint\Gamma(\underline{\mathbf{u}}_1,\underline{\mathbf{u}}_2)\exp(-2\pi i[\underline{\mathbf{u}}_1.\underline{\mathbf{r}}_1-\underline{\mathbf{u}}_2.\underline{\mathbf{r}}_2]+i\gamma(\underline{\mathbf{u}}_1)-i\gamma(\underline{\mathbf{u}}_2))d\underline{\mathbf{u}}_1d\underline{\mathbf{u}}_2 \qquad \text{I8.14}$$

Taking our fully incoherent source and a condensor aperture (inside of which $A(\underline{\mathbf{u}})=1$) this can be rewritten as:

$$\Gamma(\underline{\mathbf{r}}_1,\underline{\mathbf{r}}_2) = \iint S(\underline{\mathbf{u}}_1-\underline{\mathbf{u}}_2)A(\underline{\mathbf{u}}_1)A(\underline{\mathbf{u}}_2)\exp(-2\pi i[\underline{\mathbf{u}}_1.\underline{\mathbf{r}}_1-\underline{\mathbf{u}}_2.\underline{\mathbf{r}}_2]+i\gamma(\underline{\mathbf{u}}_1)-i\gamma(\underline{\mathbf{u}}_2))d\underline{\mathbf{u}}_1d\underline{\mathbf{u}}_2 \qquad \text{I8.15}$$

For the moment, let us take $\gamma(\underline{\mathbf{u}})=0$. This "simplifies" to:

$$\Gamma(\underline{\mathbf{r}}_1,\underline{\mathbf{r}}_2) = \{\int s^*(\underline{\mathbf{r}}_1-\underline{\mathbf{r}}')a^*(\underline{\mathbf{r}}')d\underline{\mathbf{r}}'\}\{\int s(\underline{\mathbf{r}}_2-\underline{\mathbf{r}}'')a(\underline{\mathbf{r}}'')d\underline{\mathbf{r}}''\} \qquad \text{I8.16}$$

At any pair of points $\underline{\mathbf{r}}_1,\underline{\mathbf{r}}_2$ we have a convolution (weighted average) of the source and the transform of the condensor aperture. Since we have previously assumed that no two different points in the source are coherent with respect to each other, this function is zero unless:

$$\underline{\mathbf{r}}_1-\underline{\mathbf{r}}'=\underline{\mathbf{r}}_2-\underline{\mathbf{r}}'' \qquad \text{I8.17}$$

Therefore

$$\Gamma(\underline{\mathbf{r}}_1,\underline{\mathbf{r}}_2) = \int|s(\underline{\mathbf{r}}_1-\underline{\mathbf{r}}')|^2a^*(\underline{\mathbf{r}}')a(\underline{\mathbf{r}}_2-\underline{\mathbf{r}}_1+\underline{\mathbf{r}}')d\underline{\mathbf{r}}' \qquad \text{I8.18}$$

We are retaining a term rather like $a(\underline{\mathbf{r}}_1-\underline{\mathbf{r}}_2)$ for a transverse coherence. In addition we only have illuminated a region equivalent to the size of the source. If we assume that $a(\underline{\mathbf{r}})$ is negligably small outside a region that is small compared to the source, and further that we are working towards the center of the illuminated region, this reduces to:

$$\Gamma(\underline{\mathbf{r}}_1,\underline{\mathbf{r}}_2) = \int a^*(\underline{\mathbf{r}}')a(\underline{\mathbf{r}}_2-\underline{\mathbf{r}}_1+\underline{\mathbf{r}}')d\underline{\mathbf{r}}' \qquad \text{I8.19}$$

We can further back transform this to give $|A(\underline{\mathbf{u}})|^2$ in reciprocal space and in then come back to the image plane giving:

$$\Gamma(\underline{\mathbf{r}}_1,\underline{\mathbf{r}}_2) = a(\underline{\mathbf{r}}_1-\underline{\mathbf{r}}_2) \qquad \text{I8.20}$$

In other words, if the illumination is in focus ($\Delta f=0$) and we can neglect the spherical aberration of the pre-field, justifiable for small illumination angles, and further assuming that we are towards the center of the illuminated region then we recover the very simple isoplanaric condition given earlier.

8.3 Specimen Illumination in STEM mode

Suppose that instead of doing HREM, we want to do STEM or focus down the illumination to obtain chemical information from a small area. We can not now ignore the spherical aberration in the pre-field, and we may want to use something different from exact defocus. If our source size is very small (which we can often make it by using a large demagnification via condensor/gun lenses) the answer is simple. In this case the mutual intensity is generated by setting $S(\underline{\mathbf{u}}_1-\underline{\mathbf{u}}_2)=1$ in equation I8.15, i.e.

$$\Gamma(\underline{\mathbf{r}}_1,\underline{\mathbf{r}}_2) = \iint A(\underline{\mathbf{u}}_1)A(\underline{\mathbf{u}}_2)\exp(-2\pi i[\underline{\mathbf{u}}_1.\underline{\mathbf{r}}_1-\underline{\mathbf{u}}_2.\underline{\mathbf{r}}_2]+i\gamma(\underline{\mathbf{u}}_1)-i\gamma(\underline{\mathbf{u}}_2))d\underline{\mathbf{u}}_1d\underline{\mathbf{u}}_2 \qquad \text{I8.21}$$

There are no cross terms, so we can write

$$\Gamma(\underline{\mathbf{r}}_1,\underline{\mathbf{r}}_2) = T^*(\underline{\mathbf{r}}_1)T(\underline{\mathbf{r}}_2) \qquad \text{I8.22}$$

where

$$T(\mathbf{r}) = \int A(\underline{\mathbf{u}}) \exp(-2\pi i \underline{\mathbf{u}}.\underline{\mathbf{r}} - i\gamma(\underline{\mathbf{u}}))d\underline{\mathbf{u}} \qquad \text{I8.23}$$

This is nothing more than the point-response function discussed in section 6.1 in linear imaging theory; we can include instabilities in the high-voltage (and energy spread in the source) similar to what we did before to add an envelope term in. This approximation is often used for dedicated STEM instruments. How good it is, and to what extent one should include finite size effects for the source has not (yet) been explored in much detail.

8.4 General Solution

It is useful (I think) to go one step further and include the postspecimen lens aberrations. With the approximation that the diffraction conditions do not change as the incident angle changes (good for very thin crystals), we can include the postspecimen lens aberrations by writing:

$$\Gamma'(\underline{\mathbf{r}}_1,\underline{\mathbf{r}}_2) = \iint \Psi(\underline{\mathbf{u}}_1)^* \Psi(\underline{\mathbf{u}}_2) G(\underline{\mathbf{u}}_1,\underline{\mathbf{u}}_2) \exp(i\chi(\underline{\mathbf{u}}_1) - i\chi(\underline{\mathbf{u}}_2) + 2\pi i(\underline{\mathbf{u}}_1.\underline{\mathbf{r}}_1 - \underline{\mathbf{u}}_2.\underline{\mathbf{r}}_2))d\underline{\mathbf{u}}_1 d\underline{\mathbf{u}}_2 \qquad \text{I8.24}$$

expanding $\chi(\underline{\mathbf{u}}+\underline{\mathbf{w}}) = \chi(\underline{\mathbf{u}}) + \underline{\mathbf{w}}.\underline{\nabla}c(\underline{\mathbf{u}})$ for the integral over different values of u_1 and u_2, valid for a small angular range, and considering a set of diffraction spots this can be rewritten as:

$$\Gamma'(\underline{\mathbf{r}},\underline{\mathbf{r}}) = \Sigma \exp(2\pi i \underline{\mathbf{g}}.\underline{\mathbf{r}}) \Sigma \Gamma(\underline{\mathbf{r}} + \underline{\nabla}c(\underline{\mathbf{q}})/2\pi, \underline{\mathbf{r}} + \underline{\nabla}c(\underline{\mathbf{g}-\mathbf{q}})/2\pi) \Psi(\underline{\mathbf{q}-\mathbf{g}})^* \Psi(\underline{\mathbf{q}}) \exp(i\chi(\underline{\mathbf{q}}) - i\chi(\underline{\mathbf{g}-\mathbf{q}})) \qquad \text{I8.25}$$

where $\Gamma(\underline{\mathbf{r}}_1,\underline{\mathbf{r}}_2)$ is the mutual intensity above the sample. In other words, the Envelope function that was defined earlier in linear and non-linear imaging theory <u>is</u> the mutual intensity function, i.e.

$$E(\underline{\mathbf{q}},\underline{\mathbf{g}-\mathbf{q}}) = \Gamma(\underline{\mathbf{r}} + \underline{\nabla}\chi(\underline{\mathbf{q}})/2\pi, \underline{\mathbf{r}} + \underline{\nabla}\chi(\underline{\mathbf{g}-\mathbf{q}})/2\pi) \qquad \text{I8.26}$$

This relationship is completely general, and is stronger than the particular approximations used previously. To put this into physical terms, what we are really doing in our final image is interfering the wave at different positions and the results that we get out are determined by both a simple phase shift and the coherence between the two positions (in the image plane). As such this unites the ideas of "contrast changes" in reciprocal space and worrying about where the information comes from in real space discussed earlier.

Unfortunately, at this stage we run into problems with going much further. The thing is that to do our imaging theory right we need to be working with 4-dimensional Fourier Transforms. While mathematically there is no difficulty with these, numerically on a computer they are a headache. Perhaps in a few years computers will have grown large and fast enough that.....

8.5 An Approximate Model

Having an aperture with a hard edge and spherical aberration both above and below the sample creates a lot of (mathematical) problems. We can avoid some of these by approximating the condensor aperture and the source as Gaussians -- how valid this is remains to be tested. It is useful

to use this to see what we will expect in an image of the probe. Let:

$$s(\underline{\mathbf{r}}) = \exp(-\underline{\mathbf{r}}^2/4\pi a) \qquad \text{I8.27}$$

$$A(\underline{\mathbf{u}}) = \exp(-\underline{\mathbf{u}}^2 b) \qquad \text{I8.28}$$

Then, in reciprocal space (ignoring spherical aberrations) we have

$$\Gamma(\underline{\mathbf{u}}_1,\underline{\mathbf{u}}_2) = \exp(-a|\underline{\mathbf{u}}_1-\underline{\mathbf{u}}_2|^2 -b(\underline{\mathbf{u}}_1^2+\underline{\mathbf{u}}_2^2) + \pi\lambda i[\Delta f+\Delta z](\underline{\mathbf{u}}_1^2-\underline{\mathbf{u}}_2^2)) \qquad \text{I8.29}$$

where we have (in order) the source-size effect, the condensor aperture and the pre- and post-field defoci. While integrating this looks a little formidable, it is actually simple since for a matrix **M** and some general vector **v** (in any number of dimensions)

$$\int \exp(- \mathbf{v}.\mathbf{M}.\mathbf{v} +2\pi i\ \mathbf{v}.\mathbf{r}) = 1/\mathrm{Det}(\mathbf{M})\exp(- \mathbf{r}.\ \mathbf{M}^{-1}.\mathbf{r}/4\pi) \qquad \text{I8.30}$$

where Det(**M**) is the determinant and $\mathbf{M}^{-1}$ is the inverse. Without working it out in detail, note that the pre- and post-field defoci add. Hence we will obtain identical images for any sum of the two. In order to "see" the true size of the probe, we therefore need to have $\Delta z=0$ otherwise endless confusion can arise.