NANO 703-Notes

Chapter 28: High-resolution imaging

Linear systems

Linear systems have a very important role in many fields, such as communications. We can represent a linear system S accepting an input function $F(x)$ and returning an output function $G(x)$ as:

$$G(x) = S[F(x)]$$

The definition of a linear system is very particular. It must satisfy two conditions: homogeneity and superposition. If $G_1(x) = S[F_1(x)]$ and $G_2(x) = S[F_2(x)]$, then

$$\alpha G_1(x) + \beta G_2(x) = S[\alpha F_1(x) + \beta F_2(x)]$$

In practice, the operation of a linear system is a convolution:

$$G(x) = F(x) * H(x) = \int_{x'} F(x') \cdot H(x - x') \cdot dx'$$

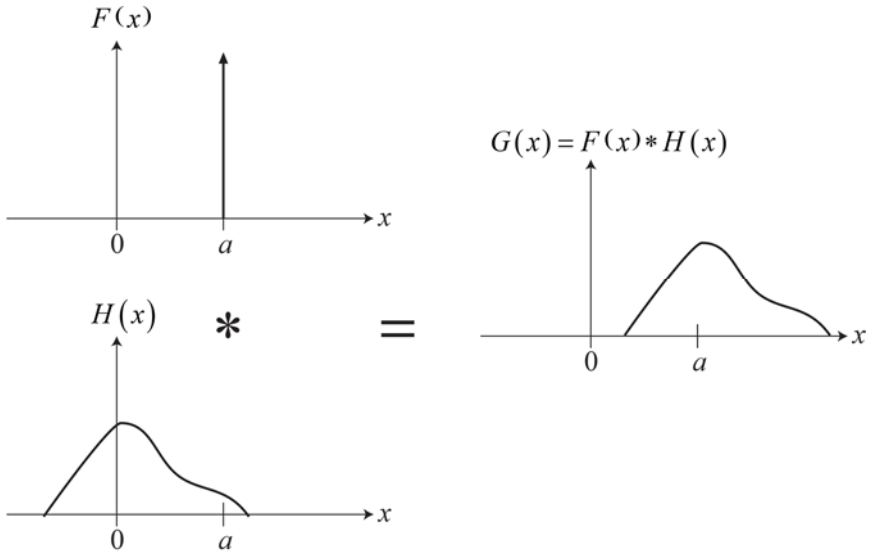
The function $H(x)$ is called the *transfer function*, or either impulse response function, because if the input is a delta function (i.e., $F(x) = \delta(x)$), then the output is $H(x)$:

$$H(x) = \int_{x'} \delta(x') \cdot H(x - x') \cdot dx'$$

In terms of TEM, I will try to show that the the exit wave function below the sample can be treated like the input $F(x)$, which we may call the object function, while the wave function that reaches our viewing screen or camera is equivalent to the output $G(x)$, which we call the image function. So what is the system? It is the microscope itself. Most importantly, it is the objective lens, along with any objective aperture or other components that we may be using, along with the electron source and the detector(s).

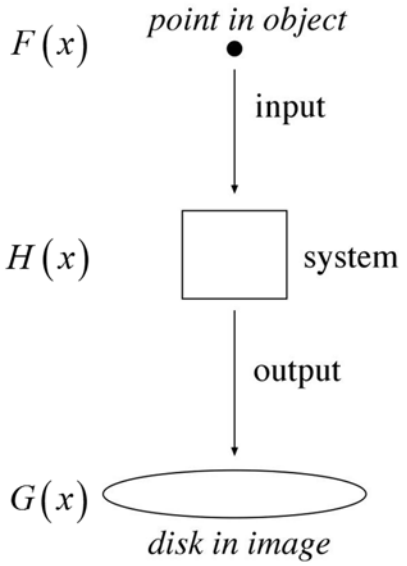
Transfer function

The system takes any input we may give the system and modifies it by the transfer function $H(x)$. For example, if the input is an impulse at $x = a$, then the output is the transfer function centered at $x = a$. There are special requirements that show up in some applications of linear systems theory, such as causality, that are not a concern here.



Microscope as a linear system

Superposition allows us to feed every point of the object function into the system one-by-one and add up their respective outputs to get the resulting image function. A single, infinitesimal point in a 2-D object may result in some sort of disk shaped blob in the image. In other words, the image won't be a perfect replica of the object, but rather a modified representation of it.



If we know that transfer function, we can use some special properties of convolutions to calculate the image function for a given object function. We can find the Fourier transforms

$$F(u) = \Im\{F(x)\} \text{ and } H(u) = \Im\{H(x)\}$$

where u is the reciprocal-space variable. (Note that the Fourier transform is a linear operator, but it turns a function in position space into one in reciprocal space, so it does not describe the type of linear system we are talking about here.) The convolution theorem tells us that

$$G(u) = F(u) \cdot H(u)$$

where

$$G(u) = \Im\{G(x)\}$$

That is, the Fourier transform of the image function is the product of the Fourier transforms of the input and transfer functions. Since we are working in reciprocal space at least half the time anyway, this is very convenient.

Contributions to the transfer function

The goal is to develop a form for $H(u)$ that encapsulates the primary features influences on the image function and, ultimately, the image intensity. Right off the bat, we can break it down into some factors that arise from aspects of the TEM we have already encountered:

$$H(u) = A(u) \cdot E(u) \cdot \mathrm{e}^{-i\chi(u)}$$

$A(u)$: aperture function

$E(u)$: envelope (damping) function

$\chi(u)$: aberration (phase) function

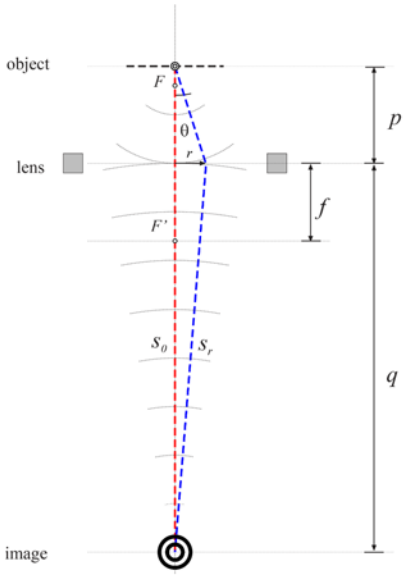
The aperture function $A(u)$ is a fairly simple, binary function that describes the effect of the objective aperture. Recall that the OA is in the BFP of the objective lens. It lets some electrons through and blocks out others, depending on their scattering angle, the diameter of the aperture, and where exactly we position it. (It's usually centered, though.)

The envelope function $A(u)$ can represent a range of contributions that generally limit the high-frequency information we can resolve. For example, chromatic aberration results from energy loss of transmitted electrons. The focal length of an electron lens depends on the electron wavelength, so the image will be smeared out by the energy distribution. It's not that the electrons are prevented from reaching the viewing screen or camera, just that they don't contribute to the image formation, but rather end up as noise in the background. We should also consider the limited spatial coherence, since our incident beam is almost never perfectly parallel. Many other factors, such as vibration in the building, electrical noise, drift, etc., can contribute to the damping effect of the envelope function.

The most important one to consider here, especially in regard to high-resolution phase-contrast lattice imaging, is the phase function $\chi(u)$. I will try to explain what exactly happens when the "Focus" knob focuses, and what focus setting gives us the best image. To do this we will need to see how changing the OL strength shifts the relative phase of scattered electrons.

Path-length correction due to lens

So far, we have only looked at the trajectories of electrons passing through a lens as rays to predict where the image forms for an object near the optical axis and a particular focal length. The trajectories are represented diagrammatically by rays, which indicate the direction of wave propagation. So, let's consider the wave fronts that extend perpendicular to these rays.



Think of the object as a point source on the optic axis that is emitting a spherical electron wave. The wave fronts diverge outward from the object. After passing through the lens plane, the wave fronts must somehow be rearranged so that they are direct back toward the optic axis to reconverge at the image point. The lens must be advancing the phase of the wave front more with increasing radius r from the optic axis. A phase difference corresponds to a difference in distance traveled, which we call a path length difference Δs, with respect to the wavelength, i.e.:

$$\Delta\phi(r) = 2\pi \cdot \left(\frac{\Delta s(r)}{\lambda}\right)$$

Let's see what the difference in path length through the lens would have to be to make all parts of the wave arrive in phase at the image point.

Outside the lens, the path length through the center of the lens is just $s_0 = p + q$. For a ray at a radius r, it is:

$$\begin{aligned} s_r &= \sqrt{p^2 + r^2} + \sqrt{q^2 + r^2} \\ &\approx \left(p + \frac{r^2}{2p}\right) + \left(q + \frac{r^2}{2q}\right) \\ &= p + q + \frac{r^2}{2}\left(\frac{1}{p} + \frac{1}{q}\right) \\ s_r &= p + q + \frac{r^2}{2f_0} \end{aligned}$$

Assuming the radius is small, and using the ideal lens equation. So the path length difference is:

$$\Delta s(r) = s_r - s_0 = -\frac{r^2}{2f_0}$$

The off-axis rays somehow found a short-cut from the front to the back of the lens, allowing them to catch up just to enough to be reconverge at the image point.

Snells' law

Focusing involves changing the direction of rays. whether electrons or light. Refraction is the most common way to bend light rays. Snell's law describes how the angle of propagation of any wave changes when the wave passes at an oblique angle across an interface between two media with different wave speeds. The frequency of the wave does not change at the interface to match this difference; only the wavelength changes. For the wave fronts to be continuous across the interface, the wavelengths parallel to the interface have to match exactly. So

$$\frac{\sin\theta_1}{\lambda_1} = \frac{\sin\theta_2}{\lambda_2}$$

If we multiply both sides by the ratio of c/f

$$\frac{c}{f}\cdot\left(\frac{\sin\theta_1}{\lambda_1} = \frac{\sin\theta_2}{\lambda_2}\right)$$

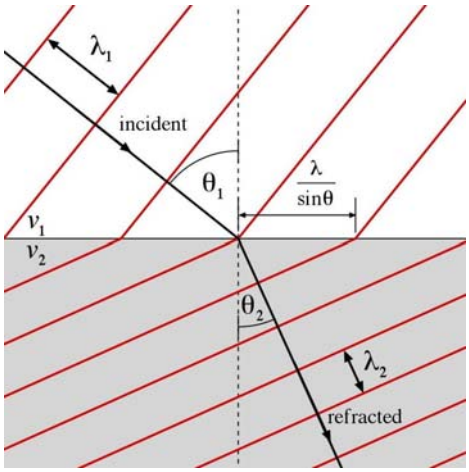
Now we can identify the wave speeds using the refractive indices n_1 and n_2.

$$v_1 = f\cdot\lambda_1 = c/n_1$$
$$v_2 = f\cdot\lambda_2 = c/n_2$$

The result is Snell's law

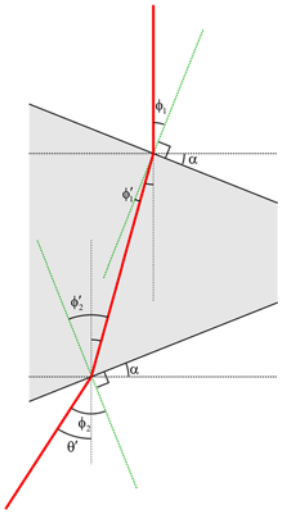
$$n_1\cdot\sin\theta_1 = n_2\cdot\sin\theta_2$$

Note that this law is very general we could be talking about electrons or light, or even surface water waves or sound in air.



Thin optical lens(I)

The function of an electron lens does not seem so magical if we consider how a standard, convex, refracting, optical lens works. The speed of visible light in air is basically the same as it is in a vacuum, so its wavelength is the same, too. But in the lens medium (glass, I presume) the speed is lower, so the wavelength is shorter.



Say a very thin portion of the lens at some radius r, makes an angle α from the lens plane. Let's consider a ray parallel to the optic axis. We have the following relations:

$$\sin\phi_1 = n\cdot\sin\phi_1',\ \sin\phi_2 = n\cdot\sin\phi_2',\ \phi_1 = \alpha,\ \phi_1' + \phi_2' = 2\alpha,\ \theta' = \phi_2 - \alpha$$

where θ' is the angle the ray makes w.r.t. the optic axis on the back of the lens. Let's assume the angles are small:

$$\phi_1 \approx n\phi_1',\ \phi_2 \approx n\phi_2',\ \theta' = n\phi_2' - \alpha = n(2\alpha - \phi_1') - \alpha = 2(n-1)\alpha$$

Since this ray is parallel to the optic axis in front of the lens, it should pass through the focal point in the back of the lens, at a distance f_0 from the lens plane. So

$$\tan\theta' = \frac{r}{f_0} \approx \theta'$$

Now we can relate the lens angle to the focal length:

$$\alpha = \frac{r}{2(n-1)f_0}$$

The lens angle corresponds to a variation in thickness t with radius, which gives

$$\frac{dt}{dr} = -2\tan\alpha \approx -2\alpha = -\frac{r}{(n-1)f_0}$$

Now we can integrate to get the shape of the lens, assuming the center has thickness T:

$$t(r) = T - \int_{r'=0}^{r}\frac{r'\cdot dr'}{(n-1)f_0} = T - \frac{r^2}{2(n-1)f_0} = T - \Delta t(r)$$

So the variation in thickness with radius is

$$\Delta t(r) = \frac{r^2}{2(n-1)f} = \frac{r^2}{2f_p}$$

Like the reflecting lenses we considered earlier, the refracting lens should have a parabolic shape. Since this is a parabolic curve, we went ahead and identified the geometric focal length of the parabola f_p. so we can see that the optical focal length f is related to the geometrical focal length of the lens by

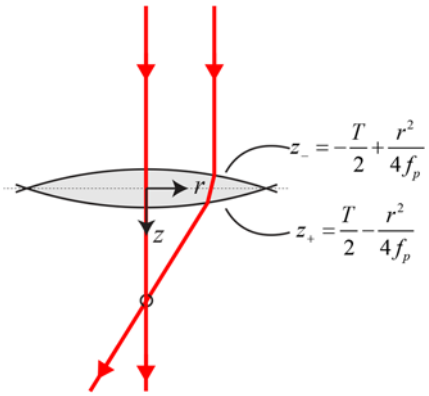
$$f_0 = \frac{f_p}{n-1}$$

Thin Optical Lens(II)

Let's find the relative phase of the wave after passing through the lens at some radius r.

vacuum: $\lambda_0 = c/f$

medium: $\lambda = \frac{c}{nf} = \lambda_0/n$



A phase difference arises from the difference in wavelength between glass and air:

$$\Delta\phi(r) = 2\pi \cdot [T - t(r)] \cdot \left(\frac{1}{\lambda} - \frac{1}{\lambda_0}\right)$$

$$= -2\pi \cdot \frac{(n-1)\cdot \Delta t(r)}{\lambda_0}$$

$$\Delta\phi(r) = 2\pi \cdot \left(\frac{\Delta s(r)}{\lambda_0}\right)$$

where $\Delta s(r)$ is the "optical" path length difference. We already know $\Delta t(r)$, given by

$$\frac{d(\Delta t)}{dr} = \frac{r}{(n-1) f_0}$$

Now we have

$$\Delta s(r) = -(n-1)\cdot \Delta t(r)$$

So

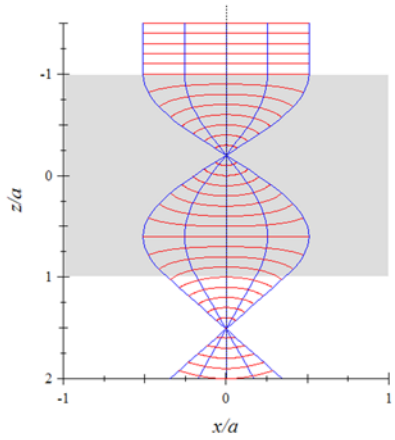
$$\frac{d(\Delta s)}{dr} = -\frac{r}{f_0}$$

To find the optical path length difference, we integrate

$$\Delta s(r) = -\int_{r'=0}^{r} \frac{r' \cdot dr'}{f_0} = -\frac{r^2}{2 f_0}$$

Uniform-field electron lens

The same principle described above must somehow apply to electron lenses. We had earlier developed a model for an electron lens with a uniform field. Focusing for the electrons does not result from refraction in this type of lens. Instead, it involves rather subtle differences in wavelength of the electrons in a magnetic field.



Unfortunately, we also found that this electron lens had significant spherical aberration. So even the figure above is a rather substantial idealization.

Path length correction due to lens

Focusing seems to result from an optical path length difference among various parts of the lens. For an ideal lens, the difference is

$$\Delta s(r) = -\frac{r^2}{2 f_0}$$

For some non-ideal lens

$$f(r) = f_0 - \Delta f(r)$$

To generalize the path length to a non-ideal lens, though, we don't substitute $f(r)$ for f_0 in the expression for $\Delta s(r)$. Instead, we substitute it in the derivative

$$\frac{d(\Delta s)}{dr} = \frac{-r}{f_0} \rightarrow \frac{-r}{f(r)}$$

For a small correction in focal length, we have

$$\frac{d(\Delta s)}{dr} = \frac{-r}{f_0 - \Delta f(r)} \approx \frac{-r}{f_0}\left(1 + \frac{\Delta f(r)}{f_0}\right)$$

If we do the integral

$$\Delta s(r) = \frac{-r^2}{2 f_0} - \int_{r'=0}^{r} \left[\frac{r' \cdot \Delta f(r')}{{f_0}^2} \right] \cdot dr'$$

we get the ideal term back, with a small correction that depends on $\Delta f(r)$.

Influence of spherical aberration

We previous analyzed the change in focal length due to spherical aberration. In that case

$$\Delta f(r) = C_s \cdot \left(\frac{r}{f_0} \right)^2$$

Let's see what we get for the path-length difference

$$\Delta s(r) = \frac{-r^2}{2 f_0} - \int_{r'=0}^{r} \left(\frac{r'}{{f_0}^2} \right) \cdot \left[C_s \cdot \left(\frac{r'}{f_0} \right)^2 \right] dr' = \frac{-r^2}{2 f_0} - \frac{1}{4} C_s \left(\frac{r}{f_0} \right)^4$$

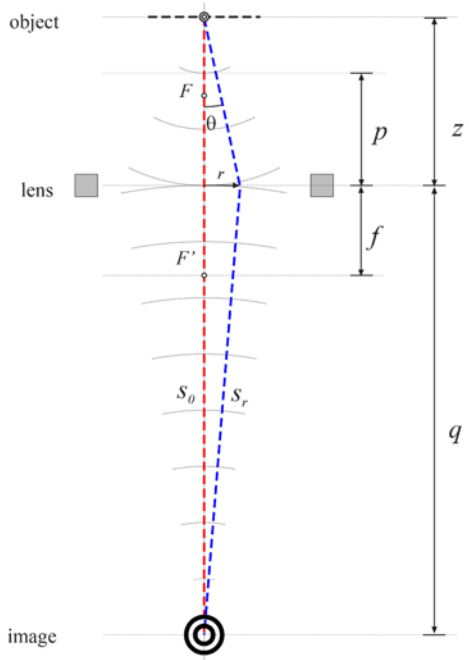
This expression has the effects of defocus and spherical aberration combined.

Path-length difference: general case

Now we know how the path-length differences for a lens at some defocus setting f_0, but which has spherical aberration coefficient C_s. Let's say a focal length f_0 images an on-axis point at a distance p in front of the lens to an on-axis point at a distance q in back of the lens. So

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{f_0}$$

But then we decide to change the focal length a little bit, from f_0 to $f = f_0 - \Delta f$. Also, we get the urge to change our sample height a little bit, from p to $p = z + \Delta z$.



In this case, an on-axis ray travels a distance $s_0 = z + q$. Using the path-length difference we derived for a lens with spherical aberration,

$$\Delta s(r) = \frac{-r^2}{2f} - \frac{1}{4}C_s\left(\frac{r}{f_0}\right)^4$$

an off-axis ray travels a distance

$$\begin{aligned} s_r &= \sqrt{z^2 + r^2} + \sqrt{q^2 + r^2} + \Delta s(r) \\ &\approx z + \frac{r^2}{2z} + q + \frac{r^2}{2q} + \Delta s(r) \\ s_r &= s_0 + \Delta s_{net}(r) \end{aligned}$$

so the net path-length difference in this case is

$$\Delta s_{net}(r) = \frac{r^2}{2}\cdot\left(\frac{1}{z} + \frac{1}{q} - \frac{1}{f}\right) - \frac{1}{4}C_s\left(\frac{r}{f_0}\right)^4$$

Combine defocus and sample height

We can evaluate the preceding expression

$$\begin{aligned} \frac{1}{z} + \frac{1}{q} - \frac{1}{f} &= \frac{1}{p - \Delta z} + \frac{1}{q} - \frac{1}{f_0 - \Delta f} \\ &\approx \cancel{\left(\frac{1}{p} + \frac{1}{q} - \frac{1}{f_0}\right)} + \frac{\Delta z}{p^2} - \frac{\Delta f}{{f_0}^2} \\ &\approx \frac{\Delta z}{p^2} - \frac{\Delta f}{{f_0}^2} \\ \frac{1}{z} + \frac{1}{q} - \frac{1}{f} &\approx \frac{\Delta z - \Delta f}{{f_0}^2} \end{aligned}$$

Recall that, if our nominal magnification $M_0 \gg 1$, then

$$M_0 = \frac{q}{p} = \frac{q}{f_0} - 1 \approx \frac{q}{f_0}$$

So $f_0 \ll q$, which also allows us to write $p \approx f_0$. So only the difference $\Delta z - \Delta f$ is really important to determine how far out-of-focus the object is. In other words, changing the OL strength a little bit and changing the sample height a little bit have approximately the same effect.

Optical path-length difference

Now we can find the net path length difference discussed above, including the effect of changing focus or sample height a little bit:

$$\Delta s \approx \frac{1}{2}\cdot(\Delta z - \Delta f)\cdot\left(\frac{r}{f_0}\right)^2 - \frac{1}{4}C_s\left(\frac{r}{f_0}\right)^4$$

Since our object is very close to the optical axis, at ray at radius r corresponds to a scattering angle θ theta, given by $r = z \cdot \theta$. So then

$$\Delta s = \frac{1}{2} \cdot (\Delta z - \Delta f) \cdot \left(\frac{z}{f_0} \right)^2 \cdot \theta^2 - \frac{1}{4} C_s \left(\frac{z}{f_0} \right)^4 \cdot \theta^4$$

One more trick is needed. We already assumed $p \approx f_0$. Similarly

$$\frac{z + \cancel{\Delta z}}{f_0} = \frac{1}{1 - f_0 / q} \approx 1 + \cancel{f_0 / q} \rightarrow z \approx f_0$$

So the net optical path-length difference as a function of scattering angle is

$$\Delta s \approx \frac{1}{2} \cdot (\Delta z - \Delta f) \cdot \theta^2 - \frac{1}{4} C_s \cdot \theta^4$$

Phase correction

We started off looking for the phase difference. It is very simply evaluated using

$$\Delta \phi = -2\pi \left(\frac{\Delta s}{\lambda} \right) = \frac{\pi}{\lambda} \cdot \left[(\Delta f - \Delta z) \cdot \theta^2 + \frac{1}{2} C_s \theta^4 \right]$$

Actually, we want to know the phase difference as a function of spatial frequency, not angle. This goes back to our analysis of diffraction patterns. For small angles, the scattering angle is proportional to the spatial frequency.

$$\theta = \frac{R}{L} = \frac{\lambda}{d} = \lambda u$$

where R is the radius of scattering and L is the camera length. The distance $d = 1/u$ corresponds to a lattice spacing, or any other diffraction feature. Now $\Delta \phi$ is the phase aberration function we were after

$$\chi(u) = \pi \cdot (\Delta f - \Delta z) \cdot \lambda u^2 + \frac{\pi}{2} C_s \lambda^3 u^4$$

Image function for weak-phase object

What we have so far is applicable to analyzing a range of TEM images without much modification. But for high-resolution imaging, we sometimes want to go even further. We want to not only calculate the image wave function, but to directly relate the image contrast to the projected potential of the specimen. We can really only achieve this if the sample is an extremely thin, weak-phase object, so that

$$F(x) \approx 1 + i\sigma V_t(x)$$

Now

$$G(x) = F(x) * H(x)$$
$$G(x) = [1 + i\sigma V_t(x)] * H(x)$$

We have two terms. Let's take the Fourier transform of the first term:

$$\Im\{1 * H(x)\} = \Delta(u) \cdot H(u)$$

The inverse FT is

$$\Im^{-1}\{\Delta(u) \cdot H(u)\} = H(u)|_{u=0} \approx 1$$

assuming the transfer function is strongly centered at $u = 0$, maybe with a sort of Gaussian shape. Now the image function is

$$G(x) \approx 1 + i\sigma V_t(x) * H(x)$$

Simple situation

We might expect the ideal transfer function to be $H(x) = \delta(x)$. For a weak-phase object, we have

$$G(x) = [1 + i\sigma V_t(x)] * H(x) = 1 + i\sigma V_t(x)$$

In this case, the image intensity is

$$I(x) = |G(x)|^2 = 1 + \cancel{[\sigma V_t(x)]^2} \approx 1$$

In other words, there is no contrast to first order in σ. So, apparently, $H(x) = \delta(x)$ is not the ideal transfer function for a weak-phase object.

Contrast transfer function

The image intensity for a weak-phase object is

$$\begin{aligned} I(x) &= |G(x)|^2 \\ &= [1 - i\sigma V_t(x) * H^*(x)] \cdot [1 + i\sigma V_t(x) * H(x)] \\ &\approx 1 + i\sigma V_t(x) * [H(x) - H^*(x)] \\ &= 1 - 2\sigma V_t(x) * \mathrm{Im}[H(x)] \\ I(x) &= 1 - \sigma V_t(x) * T(x) \end{aligned}$$

We have expanded to lowest order in σ. The new function here is the contrast transfer function (CTF)

$$T(x) \doteq 2\,\mathrm{Im}[H(x)]$$

Notice that we are now linking the image intensity directly to the projected potential of the specimen, keep only first-order terms in σ.

CTF in reciprocal space

Now we will take the FT of the CTF

$$
\begin{aligned}
T(u) &= \Im[T(x)] \\
&= \Im\{2\cdot \mathrm{Im}[H(x)]\} \\
&= -i\cdot \Im[H(x) - H^*(x)] \\
&= -i\cdot \int_x dx\cdot \left\{ \int_{u'} du'\cdot [H(u')\cdot e^{2\pi i u' x} - H^*(u')\cdot e^{-2\pi i u' x}] \right\} \cdot e^{-2\pi i u x} \\
&= -i\cdot \int_{u'} du'\cdot \left\{ H(u')\cdot \left[\int_x dx\cdot e^{2\pi i (u'-u)x} \right] - H^*(u')\cdot \left[\int_x dx\cdot e^{-2\pi i (u'+u)x} \right] \right\} \\
&= -i\cdot \int_{u'} du'\cdot \{H(u')\cdot \Delta(u'-u) - H^*(u')\cdot \Delta(u'+u)\} \\
T(u) &= -i\cdot [H(u) - H^*(-u)]
\end{aligned}
$$

Special case: *H*(*u*) is an even function

Let's assume for the moment that $H(u)$ is an even function. In other words $H(-u) = H(u)$. This implies that $H^*(-u) = H^*(u)$. Now find the FT of CTF

$$
\begin{aligned}
T(u) &= -i\cdot [H(u) - H^*(-u)] \\
&= -i\cdot [H(u) - H^*(u)] \\
T(u) &= 2\cdot \mathrm{Im}[H(u)]
\end{aligned}
$$

In this case, we can also say that the FT of the CTF is twice the imaginary part of the FT of the transfer function.

Using the CTF

We proposed that our transfer function has the form

$$H(u) = A(u)\cdot E(u)\cdot \exp[-i\chi(u)]$$

Let's continue to assume $H(u) = H(-u)$. Then

$$T(u) = 2\,\mathrm{Im}[H(u)] = -2A(u)E(u)\sin\chi(u)$$

Now we have

$$
\begin{aligned}
I(u) &= \Im[1 - \sigma V_t(x) * T(x)] \\
&= \Delta(u) - \sigma V_t(u)\cdot T(u) \\
I(u) &= \Delta(u) + 2\sigma V_t(u)\cdot A(u)\cdot E(u)\cdot \sin\chi(u)
\end{aligned}
$$

The delta function at $u = 0$ gives a bright background, upon which the image intensity is combined.

Ideal transfer function for phase object (I)

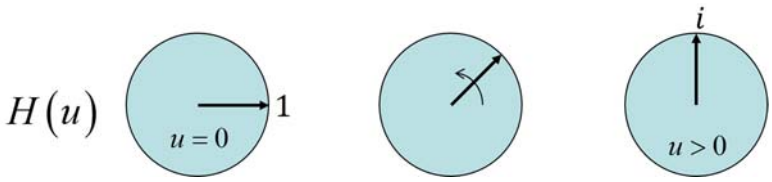
In a fairly general sense, the ideal $H(u)$ will shift the phase of scattered parts of the wave by 90°, without changing the phase of the unscattered wave. In terms of our phase function

$$
\chi(u) = \begin{cases} 0, & |u| < 1/b \\ -\dfrac{\pi}{2}, & |u| \geq 1/b \end{cases}
$$

where b is some length threshold. We want to see phase contrast for lengths smaller than *b*. Now

$$H(u) = \mathrm{e}^{-i\chi(u)} = \begin{cases} 1, & |u| < 1/b \\ i, & |u| \geq 1/b \end{cases}$$

We can imagine this on the phasor-style plots shown below. All spatial frequencies contribute to the image function, but the phases of higher frequencies have been shifted.



The CTF is now

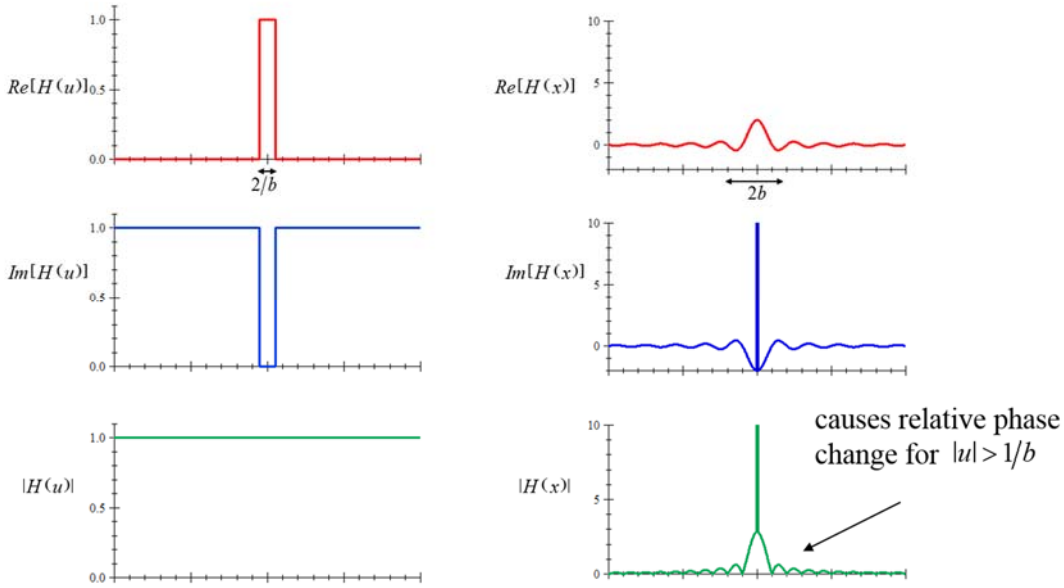
$$T(u) = 2\,\mathrm{Im}[H(u)] = \begin{cases} 0, & |u| < 1/b \\ 2, & |u| \geq 1/b \end{cases}$$

Ideal transfer function for phase object (II)

It is actually easier to picture the effect of $H(u)$ in reciprocal space than it is in direct space. The inverse FT of the preceding, ideal transfer function is

$$H(x) = i\delta(x) + 2(1-i)\cdot \mathrm{sinc}(2\pi x/b)$$

We get an imaginary delta function at $x = 0$, and a complex sinc function everywhere else. The delta function ensures all frequencies contribute to $G(x)$. But the sinc function shifts the relative phase of frequencies $|u| > 1/b$ to give phase contrast.



Ideal CTF

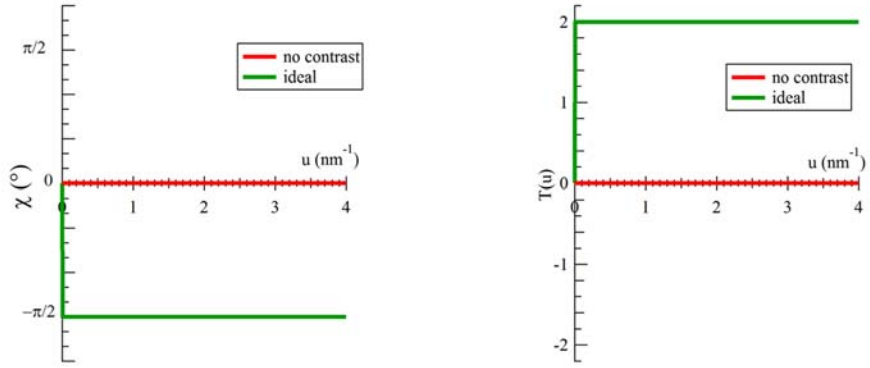
The ideal phase aberration function induces a phase shift of 90° for $|u|>0$ relative to the wave at $u=0$. (Since u is proportional to scattering angle θ, $|u|>0$ corresponds to the scattered wave; $u=0$ is the unscattered wave.)

$$\chi(u)=\begin{cases}0, & u=0\\ -\pi/2, & u>0\end{cases}$$

Let's just assume the aperture and envelope function are both unity for the time being [$A(u)E(u)=1$]. Applying $T(u)=-2\sin\chi(u)$

$$T(u)=\begin{cases}0, & u=0\\ 2, & u>0\end{cases}$$

This should be compared to the case where $\chi(u)=0\ \forall u$, which gives $T(u)=0\ \forall u$, i.e., no contrast.



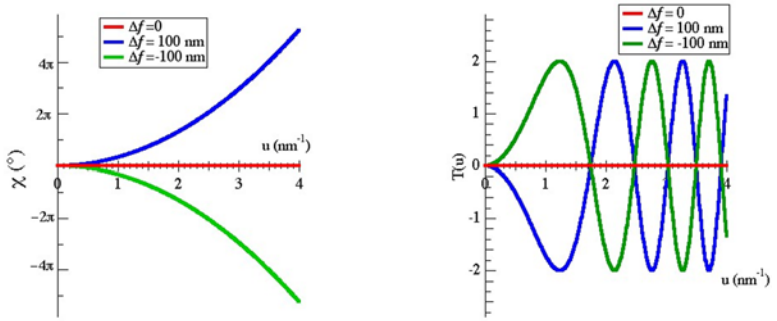
Also note that $T(u)>0$ corresponds to positive phase contrast, which we usually want.

Real CTF: case 1 – *Cs*=0

Let's return to the form of the phase aberration function we predicted would actually represent the TEM objective lens. If $C_s=0$, we have

$$\chi(u)=\pi\cdot\Delta f\cdot\lambda u^2$$

This has odd symmetry in Δf, so the contrast will exactly reverse between underfocus ($\Delta f<0$) and underfocus ($\Delta f>0$).



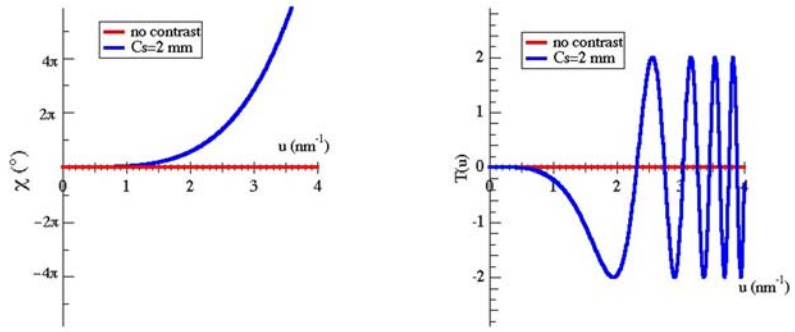
The linear variation in $\chi(u)$ with u causes an oscillation in $T(u)$. So the contrast actually vanishes when $\chi(u) = n\pi$ and reverses every cycle. This does not seem very ideal. But at lease we see that underfocus gives positive phase contrast for low frequencies, indicating that underfocus may be preferable most of the time.

Real CTF: case 2 – Δf=0

Now let's assume we are exactly in focus, but spherical aberration is present ($C_s = 0$), so there is still a phase shift of

$$\chi(u) = \frac{1}{2}\pi C_s \lambda^3 u^4$$

This is always a positive phase shift, giving negative phase contrast for small u.



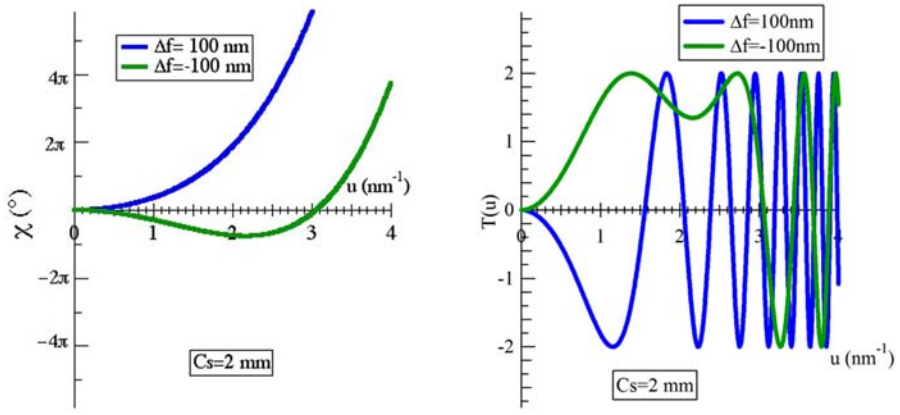
This seems even less suitable for direct interpretation of image contrast. In fact the oscillations increase even faster with frequency, compared to the previous case.

Real CTF: general

When we allow both defocus and spherical aberration to contribute, an interesting effect emerges.

$$\chi(u) = \pi \cdot \Delta f \cdot \lambda u^2 + \frac{1}{2}\pi C_s \lambda^3 u^4$$

This function is quadratic in u^2. If $\Delta f > 0$ (overfocus), the function increases even faster than it would with just one of the two terms. But, if $\Delta f < 0$, the negative u^2 term dominates at small u, but the positive u^4 takes over at higher u. So $\chi(u)$ must hit a minimum at some u^2, which gives a broad range of fairly constant, positive phase contrast.



We want to identify the defocus setting that maximizes this region of positive phase contrast.

Scherzer defocus

Take the first derivative of $\chi(u)$

$$\frac{d\chi(u)}{du} = 2\pi \cdot \Delta f \cdot \lambda u + 2\pi C_s \lambda^3 u^3$$

Now find the defocus, called the Scherzer defocus Δf_{sch}, that sets the frequency of constant phase at a particular value $u_{\min}$.

$$\left.\frac{d\chi(u)}{du}\right|_{u=u_{\min}} = 0 \Rightarrow \Delta f = \Delta f_{sch} = -C_s \lambda^2 {u_{\min}}^2$$

We will need to decide at what frequency the stationary phase should occur. We know that $T_{\max} = 2$. Scherzer decided the best choice is

$$T(u_{\min}) = -2\sin[\chi(u_{\min})] = \frac{\sqrt{3}}{2} \cdot T_{\max} = \sqrt{3}$$

So then

$$\sin[\chi(u_{\min})] = -\frac{\sqrt{3}}{2}$$

In this case we now that

$$\chi(u_{\min}) = -\frac{2\pi}{3} = -\frac{1}{2}\pi C_s \lambda^3 {u_{\min}}^4$$

So we can now find the frequency of constant phase shift

$$u_{\min} = \left(\frac{4}{3C_s\lambda^3}\right)^{1/4}$$

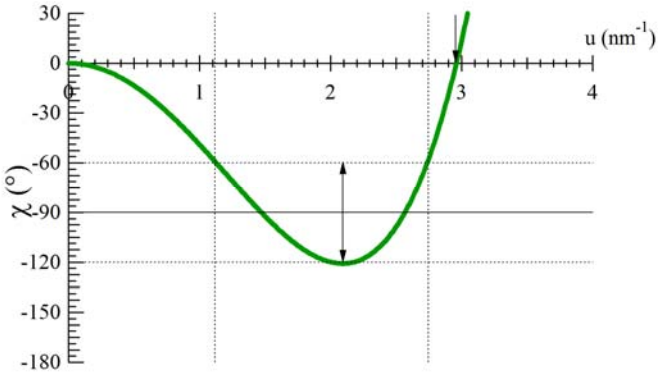
Additionally, we can find the Scherzer defocus

$$\Delta f_{sch} = -\left(\frac{4}{3}C_s\lambda\right)^{1/2} \approx -1.2(C_s\lambda)^{1/2}$$

Notice that it is a negative, corresponding to *underfocus*. Whether or not you are at the precise Scherzer defocus, you usually want the OL slightly underfocused to maintain a broad range of positive phase contrast.

Phase at Scherzer defocus

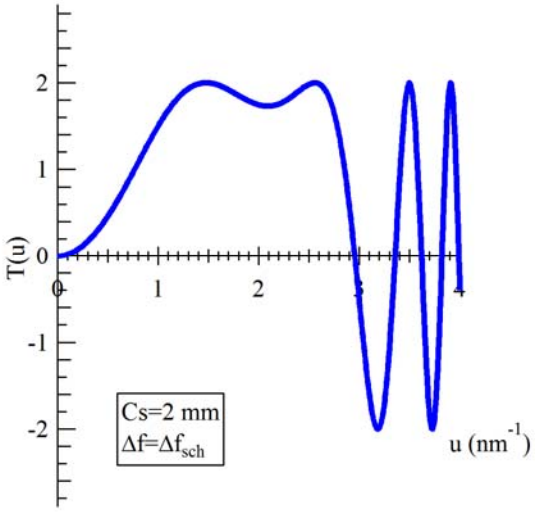
A plot of $\chi(u)$ at Δf_{sch} for the particular microscope parameters $C_s = 2.0$ mm , $E = 125$ KeV is shown below.



It is hard to prove that Δf_{sch} is precisely the optimal defocus setting, but it certainly seems like a good choice.

CTF at Scherzer defocus

For the parameters given above, I get $\Delta f_{sch} = -93.4$ nm . The CTF for these settings is shown below.



Resolution at Scherzer defocus

We can get a bit more quantitative by defining a resolution of the phase-contrast image. We can say that the range of directly interpretable contrast extends only out to some u_{sch} , where

$$\sin\chi(u_{sch}) = 0 \Rightarrow \chi(u_{sch}) = 0$$

Putting this in our form for $\chi(u)$

$$0 = \Delta f_{sch} \cdot \lambda u_{sch}^{\ 2} + \frac{1}{2} C_s \lambda^3 u_{sch}^{\ 4}$$

and inserting our expression for Δf_{sch}, we have

$$u_{sch} = 2 \cdot \left(3 C_s \lambda^3\right)^{-1/4} \approx 1.51 \left(C_s \lambda^3\right)^{-1/4}$$

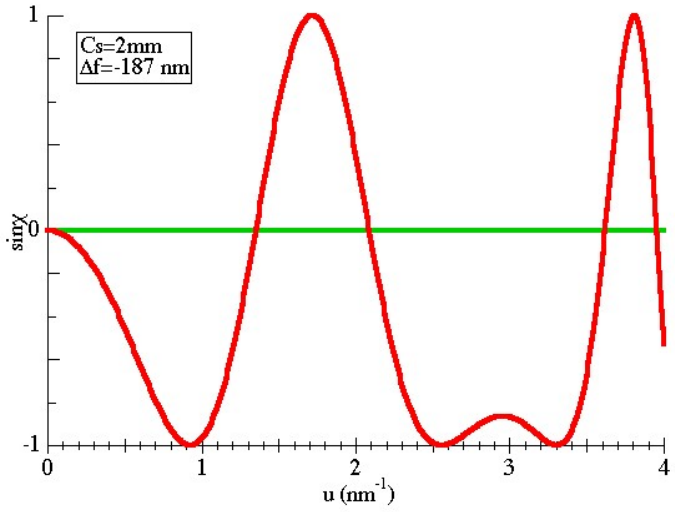
The corresponding length is the inverse of the spatial frequency

$$d_{sch} = \frac{1}{u_{sch}} \approx 0.66 \left(C_s \lambda^3\right)^{1/4}$$

This Scherzer resolution is sometimes called the point resolution. Not that it has the same dependence on C_s and λ as the "practical" resolution we defined earlier, with a different coefficient in front.

Passbands

If we change the decfocus to an even more underfocused setting, we can put the constant phase contrast setting further out in reciprocal space, as in the graph below.



We can certainly do this, but one needs to have a lot of insight into all the parameters affecting the image contrast. It implies an even better approach, which is to use a whole defocus series, sometimes called a "through-focus" series. From the collated set of images, we could conceivably back out the true exit wave function $F(x)$. But the acquisition of the images and the calculation take some time and computer power. So if you just want one image showing the best possible phase contrast, I would suggest a defocus setting near Δf_{sch}.

Damping due to temporal incoherence

The envelope function $E(u)$ contains any factors that diminish high-frequency contributions to the image function. Overall, we expect $E(u)$ to vary from unity at the origin to zero at high frequencies. The most

recognizable contributions arise from incoherence, either in the electron source, or in the transmitted electrons that contribute to the image.

Fundamentally, time and energy are considered conjugate parameters. Temporal incoherence arises from either the actual energy distribution in the incident or transmitted beam, or the equivalent effect of instability in the objective lens. We can certainly imagine that not all of the incident electrons have precisely the beam energy, due to either slight voltage instabilities, thermal effects in the source, inelastic processes, such as Bremsstrahlung, etc. The transmitted electrons will most likely have an even bigger energy spread, due to interactions with the specimen. These different energies will not be focused identically, smearing out the image.

Another contribution may be from instability in the OL current, just due to practical electronic considerations. The focal length may jiggle a little bit, also smearing out the image.

Essentially, these can be contributions combined into a single term

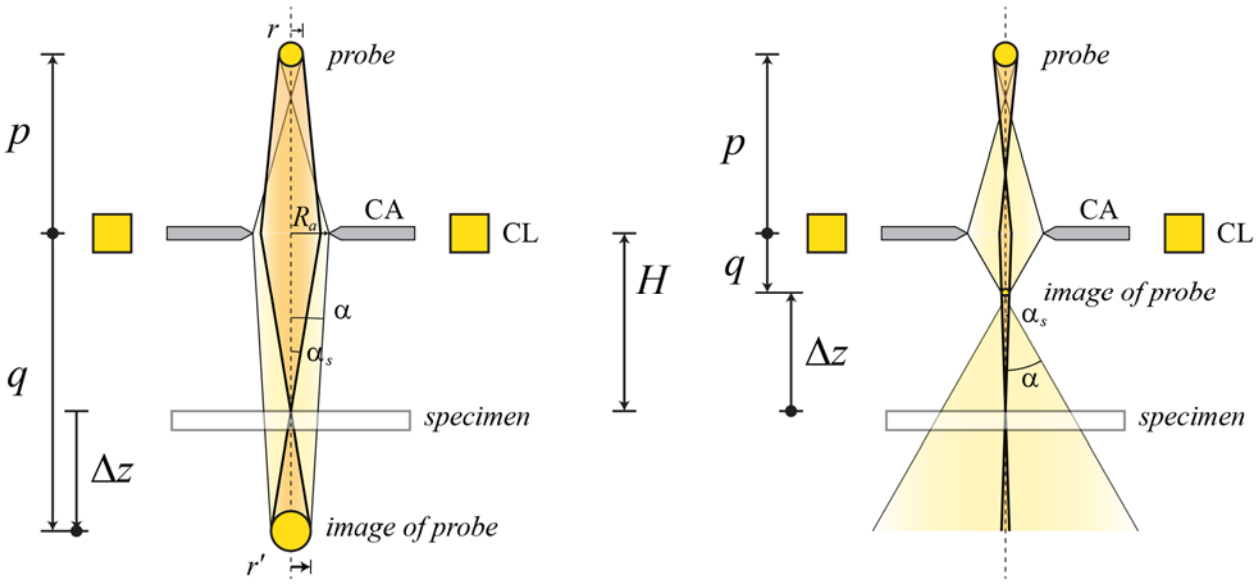
$$E_c(u) = \exp\left(-\pi^2\delta^2\lambda^2u^4/2\right)$$

where

$$\delta = C_c \cdot \sqrt{\left(\frac{\delta E}{E}\right)^2 + \left(\frac{\delta I}{I}\right)^2}$$

Here $\delta E/E$ is the fractional spread in energy and $\delta I/I$ is the fractional spread in lens current. We have assumed these energy and current contributions are small and uncorrelated.

Beam convergence: one-lens condenser (I)
We also assumed parallel illumination in our derivation. In reality, we usually have at least some beam convergence in the TEM. This is generally referred to as spatial incoherence, although the actual coherence in the source is only one characteristic. Consider the two configurations shown below. On the left, the CL is underfocused; on the right it is overfocused.



Notice that the range of angles incident on a point of the specimen is not strictly the same as the convergence semi-angle α of the probe at crossover. Instead, it is given by the semi-angle α_s of the image of probe when viewed from the point on the specimen. Also, it is apparent when going through crossover that the image of the probe is larger when the beam is underfocused than when it is overfocused. This is partly why it is usually recommended to operate with the CL in overfocus for good coherence.

Beam convergence: one-lens condenser (II)

Let's see what exactly determines α_s, starting with a one-lens condenser system. If the specimen is at a distance z below the CL and the crossover (image of the probe), with radius r', is Δz above the specimen, then

$$\alpha_s = \left|\frac{r'}{\Delta z}\right| = \left|\frac{H}{\Delta z} - 1\right| \cdot \alpha_r$$

We define the illumination semi-angle at a point in the lens plane as

$$\alpha_r = \frac{r}{p}$$

where r is the radius of the probe itself (or a prior image) and p is the distance of the probe above the CL.

By inspection, we can see that the maximum value of α_s is

$$\alpha_s^{(\max)} = \frac{R_a}{H}$$

So if the image of the probe is too big, we get $\alpha_s = \alpha_s^{(\max)}$.

$$\alpha_s = \begin{cases} \left|\frac{r''}{\Delta z}\right| \cdot \alpha_r, & \left|\frac{r''}{\Delta z}\right| \cdot \alpha_r < \alpha_s^{(\max)} \\ \alpha_s^{(\max)}, & \text{otherwise} \end{cases}$$

Compare this to the beam convergence semi-angle. We can find the distance q_a from the lens plane for the image of point on the optic axis where the limiting rays cross.

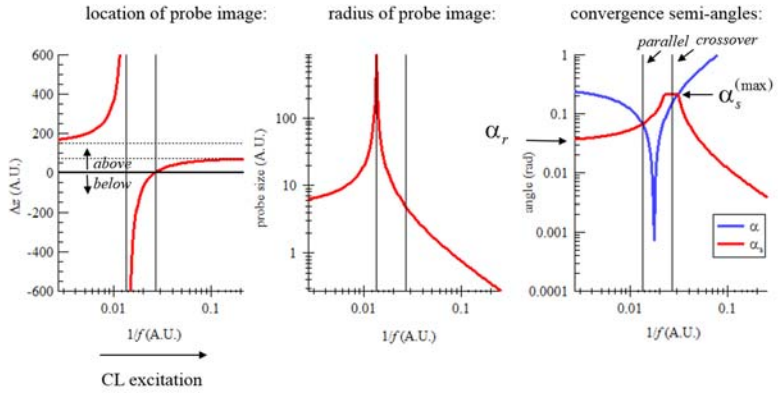
$$q_a = \frac{1}{\frac{1}{q} - \left(\frac{r}{R_a}\right) \cdot \frac{1}{p}}$$

Some algebra gives:

$$\alpha = \left|\frac{R_a}{q_a}\right| = \left|\frac{R_a}{q} - \alpha_r\right|$$

Beam convergence: one-lens condenser (III)

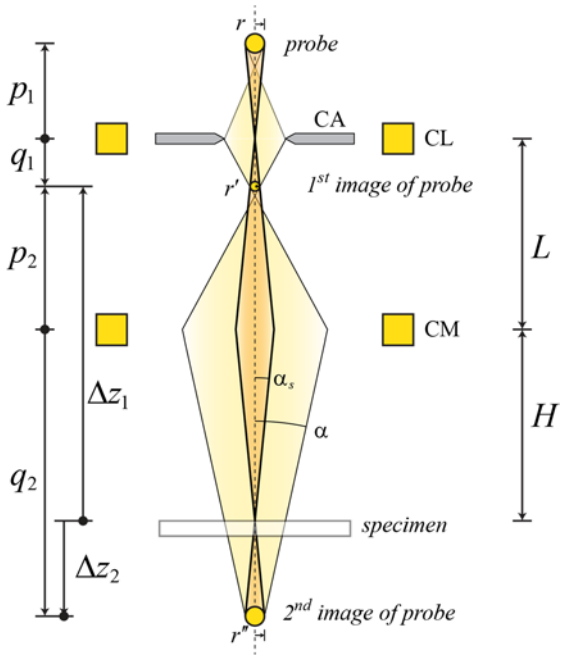
Let's look at how α_s varies with lens strength, indicated by $1/f$. We can see that the CA determines α_s near crossover. By increasing the amount of CL overfocus, we can make α_s arbitrarily small. For underfocus, α_s is never less than α_r.



For a one-lens system, one can make the beam nearly parallel (small α) with a certain amount of underfocus, which puts the probe very close to the focal point. This could be good for a selected-area diffraction pattern of a large area. But for a small small α_s, a very large overfocus, greatly demagnifying the probe, is best.

Beam convergence: two-lens condenser (I)

The actual illumination system has multiple condenser lenses, as well as the OL pre-field, which affect the beam convergence. The main features become apparent with only two lenses. The first is presumably CL3, which is controlled by the Brightness knob. The second may be a composite of the condenser min-lens and the OL pre-field.



We have two separate probe images. The second one is the effective source illuminating the sample. Now the semi-angle of illumination of a point on the specimen is

$$\alpha_s = \left|\frac{r''}{\Delta z_2}\right| = \frac{\alpha_r}{\left|\left(\frac{L}{q_1}-1\right)\cdot\left(\frac{H}{q_2}-1\right)\right|}$$

The maximum value is a bit more complicated now:

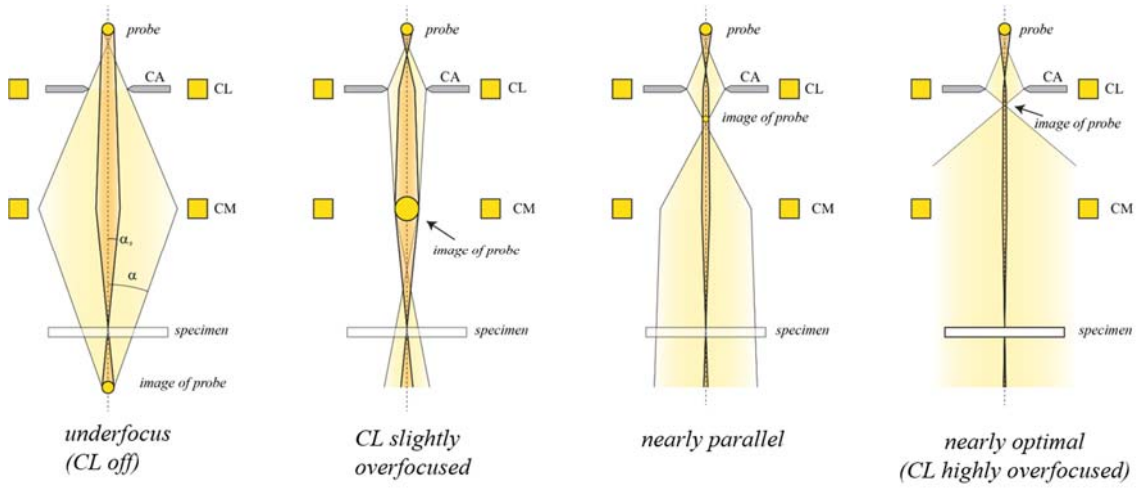
$$\alpha_s^{(\max)} = \frac{R_a}{H\cdot L\cdot\left|\left(\frac{1}{L}-\frac{1}{f_2}-\frac{1}{H}\right)\right|}$$

The form is essentially the same in this case:

$$\alpha_s = \begin{cases} \left|\frac{r''}{\Delta z}\right|, & \left|\frac{r''}{\Delta z}\right|\cdot\alpha_r < \alpha_s^{(\max)} \\ \alpha_s^{(\max)}, & \text{otherwise} \end{cases}$$

Beam convergence: two-lens condenser (II)

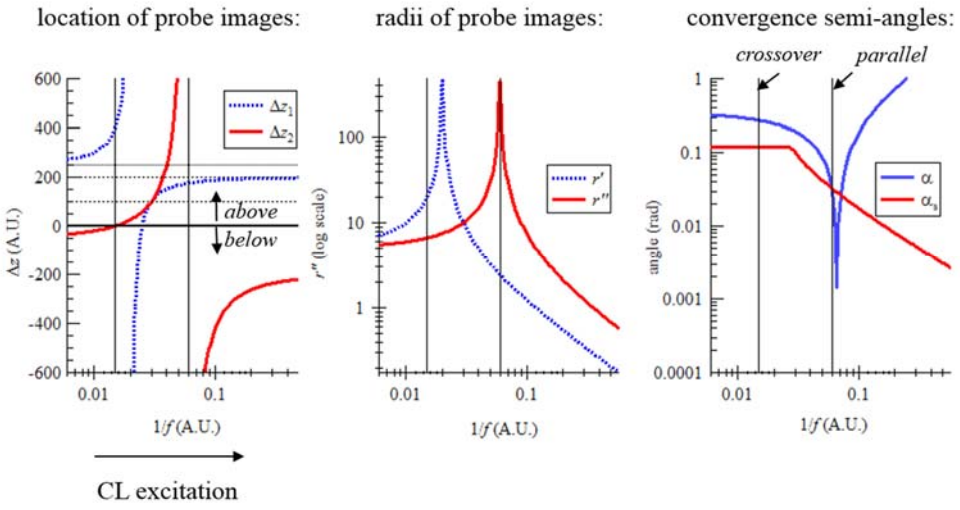
Some ray diagrams showing α_s and α for particular lens settings are shown below.



The same CM strength is assumed in all four diagrams. In the first case, the CL is off. The second shows overfocus, such that the first probe image is precisely in the plane of the CM. In the third case, the first image is at the focal point of the CM, so the beam is nearly parallel on the sample. The last case shows a large overfocus of the CL. This makes α_s very small, although α gets bigger as the CL strength is increased.

Beam convergence: two-lens condenser (III)

Plots for the two-lens case are shown below. Again, α_s decreases indefinitely as the CL strength is increased. The most parallel beam occurs with a moderate overfocus.



To summarize, beyond crossover, as decreases with increasing CL strength ("Brightness"). This can be primarily attributed to the decreasing size of the second image of the probe with overfocus. But a parallel beam is formed with a specific amount of CL overfocus. This is good for selected-area diffraction, but it takes some effort to find the optimal value.

Damping due to beam convergence (I)

If we assume a parallel beam, the image function is

$$G_0(u) = F_0(u) \cdot H_0(u)$$

and the image intensity is

$$I_0(u) = \int_{u'=-\infty}^{\infty} du' \cdot G_0^*(u-u') \cdot G_0(u')$$

Let's now consider some beam convergence. The angular intensity profile of the probe has a roughly Gaussian distribution:

$$S(u) = \frac{1}{\sqrt{\pi} k\alpha_s} \cdot \mathrm{e}^{-\left(\frac{u}{k\alpha_s}\right)^2}$$

Again, the spatial frequency variable corresponds to angle through $u = k \cdot \theta$. It's not entirely clear whether the spread should be applied by convoluting with the image function

$$G(u) = \int_{u'=-\infty}^{\infty} du' \cdot G_0(u-u') \cdot S(u')$$

of with the image intensity

$$I(u) = \int_{u'=-\infty}^{\infty} du' \cdot I_0(u-u') \cdot S(u')$$

If we choose the first case, we have assumed that all points on the source emit coherently. The second case implies there is no coherence among any points on the source. The reality for a thermionic source is somewhere in between. Closely spaced points are more coherent than points with greater separation. For a field-emission source, which is very small, it is often appropriate to assume perfect coherence.

Damping due to beam convergence (II)

Let's proceed by assuming no OA is in, and that the only effect of the OL is a phase shift:

$$G_0(u) = F(u) \cdot \mathrm{e}^{-i\chi(u)}$$

and that the source is perfectly coherent

$$G(u) = \frac{1}{\sqrt{\pi} k\alpha_s} \cdot \int_{u'=-\infty}^{\infty} du' \cdot F(u-u') \cdot \mathrm{e}^{-i\chi(u-u')} \cdot \mathrm{e}^{-\left(\frac{u'}{k\alpha_s}\right)^2}$$

To find the transfer function, we assume our input is a delta function, in which case

$$H(x) = S[\delta(x)]$$

We know the FT of delta function

$$F(x) = \delta(x) \rightarrow F(u) = 1$$

So

$$H(u) = \frac{1}{\sqrt{\pi} k\alpha_s} \cdot \int_{u'=-\infty}^{\infty} du' \cdot \mathrm{e}^{-i\chi(u-u')} \cdot \mathrm{e}^{-\left(\frac{u'}{k\alpha_s}\right)^2}$$

Damping due to beam convergence (III)

It is reasonable to assume $k\alpha_s$ is small, so the gaussian function is sharply peaked. So we expand

$$\chi(u-u') \approx \chi(u) - u' \cdot \left.\frac{\partial \chi(u'')}{\partial u''}\right|_{u''=u} = \chi(u) - u' \cdot C(u)$$

Now we can do the integral using some standard identities

$$H(u) \approx \mathrm{e}^{-i\chi(u)} \cdot \frac{1}{\sqrt{\pi} k\alpha_s} \cdot \int_{u'=-\infty}^{\infty} du' \cdot \mathrm{e}^{-iC(u)\cdot u'} \cdot \mathrm{e}^{-\left(\frac{u'}{k\alpha_s}\right)^2}$$

$$= \mathrm{e}^{-i\chi(u)} \cdot \frac{1}{\sqrt{\pi} k\alpha_s} \cdot \int_{u'=-\infty}^{\infty} du' \cdot \cos[C(u) \cdot u'] \cdot \mathrm{e}^{-\left(\frac{u'}{k\alpha_s}\right)^2}$$

$$H(u) = \mathrm{e}^{-i\chi(u)} \cdot \exp\left[-\left(\frac{k\alpha_s}{2} \cdot C(u)\right)^2\right] = \mathrm{e}^{-i\chi(u)} \cdot E_s(u)$$

We see that the beam convergence is described by a damping, envelope function

$$E_s(u) = \exp\left[-\left(\frac{k\alpha_s}{2} \cdot \frac{\partial \chi}{\partial u}\right)^2\right]$$

It is evident why this is called spatial incoherence. The result is to smear out the finer features in the image.

Combine incoherence effects

The temporal and spatial incoherence contributions can be combined, along with any other sources of attenuation we can identify (e.g., specimen drift, electromagnetic interferences, poor control of air flow in the laboratory, etc.)

$$E(u) = E_c(u) \cdot E_s(u) \cdots$$

Image of periodic specimen (I)

Let's return to the transfer function, without applying the WPOA, and assuming the object function is periodic.

$$F(x) = \sum_g F_g \cdot \mathrm{e}^{2\pi i g x}$$

Its FT is

$$\begin{aligned} F(u) &= \Im\{F(x)\} \\ &= F_g \cdot \Im\left\{\sum_g \mathrm{e}^{2\pi i g x}\right\} \\ F(u) &= \sum_g F_g \cdot \Delta(u-g) \end{aligned}$$

Let's forget damping for the moment and assume the transfer function is

$$H(u) = A(u) \cdot \mathrm{e}^{-i\chi(u)}$$

Now the FT of the image function will be

$$G(u) = \left[\sum_g F_g \cdot \Delta(u-g)\right] \cdot A(u) \cdot \mathrm{e}^{-i\chi(u)}$$

Taking the inverse FT

$$\begin{aligned} G(x) &= \lim_{K\to\infty}\left[\int_{u=-K}^{K} G(u) \cdot \mathrm{e}^{2\pi i u x} \cdot du\right] \\ &= \lim_{K\to\infty}\left\{\int_{u=-K}^{K} \sum_g F_g \cdot \Delta(u+g) \cdot A(u) \cdot \mathrm{e}^{-i\chi(u)} \cdot \mathrm{e}^{2\pi i u x} \cdot du\right\} \\ G(x) &= \sum_g F_g \cdot A(g) \cdot \mathrm{e}^{-i\chi(g)} \cdot \mathrm{e}^{2\pi i g x} \end{aligned}$$

Image of periodic specimen (II)

We can probably spot the Fourier coefficients of G already, but let's work through it:

$$
\begin{aligned}
G(u) &= \lim_{L\to\infty}\left[\int_{x=-L}^{L} G(x)\cdot e^{-2\pi iux}\cdot dx\right] \\
&= \lim_{L\to\infty}\int_{x=-L}^{L}\left\{\sum_g\left[F_g\cdot A(g)\cdot e^{-i\chi(g)}\cdot e^{2\pi igx}\right]\cdot e^{-2\pi iux}\cdot dx\right\} \\
&= \sum_g\left\{\lim_{L\to\infty}\int_{x=-L}^{L}\left[F_g\cdot A(g)\cdot e^{-i\chi(g)}\cdot e^{2\pi igx}\right]\cdot e^{-2\pi iux}\cdot dx\right\} \\
&= \sum_g F_g\cdot A(g)\cdot e^{-i\chi(g)}\left\{\lim_{L\to\infty}\int_{x=-L}^{L} e^{-2\pi i(g-u)x}\cdot dx\right\} \\
&= \sum_g F_g\cdot A(g)\cdot e^{-i\chi(g)}\Delta(u-g) \\
G(u) &= \sum_g G_g\cdot\Delta(u-g)
\end{aligned}
$$

So

$$G_g = F_g\cdot A(g)\cdot e^{-i\chi(g)} \text{ and } G(x)=\sum_g G_g\cdot e^{2\pi igx}$$

Image of periodic specimen (III)

Let's assume we are only concerned with the terms 0, g and $-g$. Then

$$F(u)=F_{-g}\cdot\Delta(u+g)+F_0\cdot\Delta(u)+F_g\cdot\Delta(u-g)$$

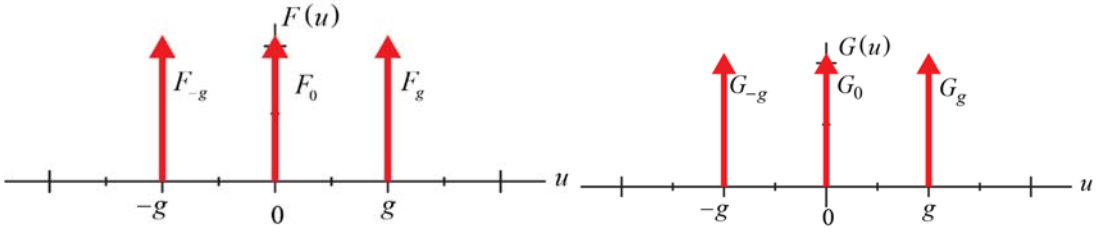
So

$$G(u)=G_{-g}\cdot\Delta(u+g)+G_0\cdot\Delta(u)+G_g\cdot\Delta(u-g)$$

We could make a table of these

	$-g$	0	g
$F(u)$	F_{-g}	F_0	F_g
$A(u)$	$A(-g)$	$A(0)$	$A(g)$
$\chi(u)$	$\chi(-g)$	$\chi(0)$	$\chi(g)$
$e^{-i\chi(u)}$	$e^{-i\chi(-g)}$	$e^{-i\chi(0)}$	$e^{-i\chi(g)}$
$G(u)$	$F_{-g}\cdot A(-g)\cdot e^{-i\chi(-g)}$	$F_0\cdot A(0)\cdot e^{-i\chi(0)}$	$F_g\cdot A(g)\cdot e^{-i\chi(g)}$

A generic plot is also useful



Example

Here is an example. Take

$$F\left(x\right)=1+iB\cdot\sin\left(2\pi gx\right)$$

This can be written

$$F\left(x\right)=-\frac{B}{2}\mathrm{e}^{-2\pi igx}+1+\frac{B}{2}\mathrm{e}^{2\pi igx}=F_{-g}\mathrm{e}^{-2\pi igx}+F_0+F_g\mathrm{e}^{2\pi igx}$$

Let's say $A\left(u\right)=1$ (no OA) and

$$\chi\left(u\right)=\begin{cases}0, & \left|u\right|\le g/2\\ -\pi/2, & g/2<\left|u\right|\end{cases}$$

Our completed table is:

	$-g$	0	g
$F(u)$	$-B/2$	1	$B/2$
$A(u)$	1	1	1
$\chi(u)$	$-\pi/2$	0	$-\pi/2$
$\mathrm{e}^{-i\chi(u)}$	i	1	i
$G(u)$	$-iB/2$	1	$iB/2$

Then

$$G\left(x\right)=-\frac{iB}{2}\mathrm{e}^{-2\pi igx}+1+\frac{iB}{2}\mathrm{e}^{2\pi igx}=1-B\sin\left(2\pi gx\right)$$