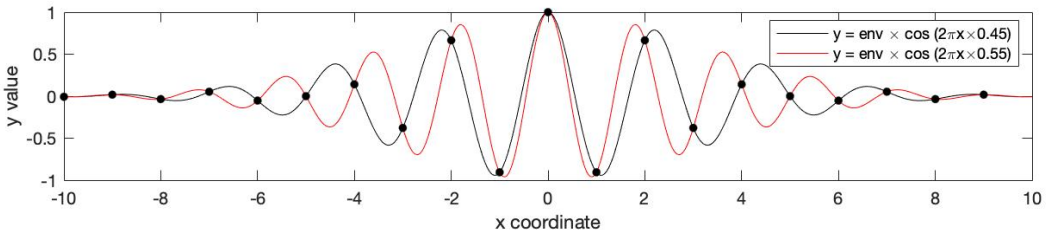
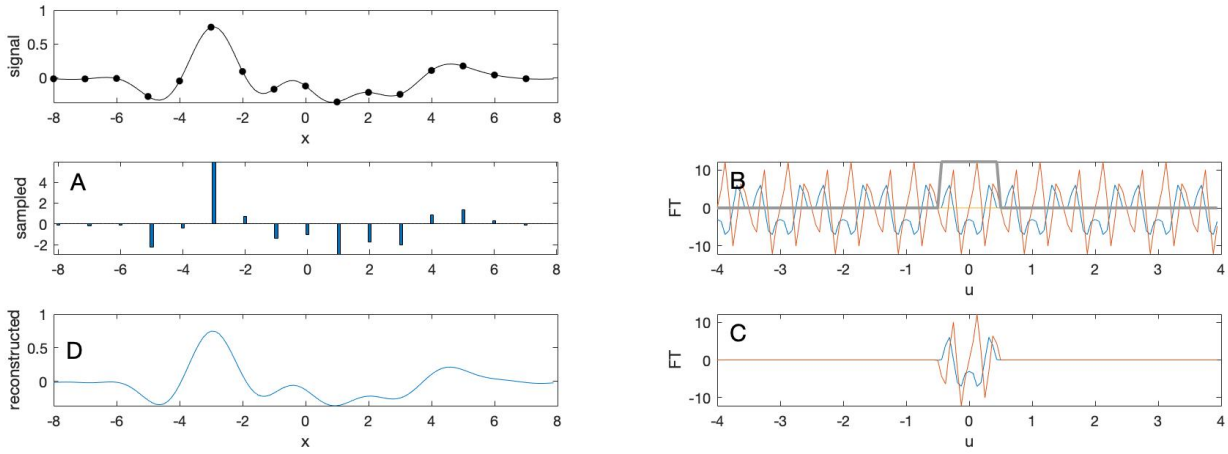
Sampling, 3D Reconstruction and CTF "Correction"

The Sampling Theorem

Up to now we have been talking about Fourier transforms in the platonic ideal of continuous functions on the entire fields of all real (or complex) numbers in 1D, 2D or 3D. But we will be acquiring images with cameras having a limited number of pixels, and will be creating 3D density maps with a limited number of voxels. How can we know that these discrete representations will be accurate ones? The plot below shows an example of a problem. Two cosine waves with different frequencies (u=0.45 u=0.55) yield the same sample points when sampled at unit intervals. This is a phenomenon called "aliasing".



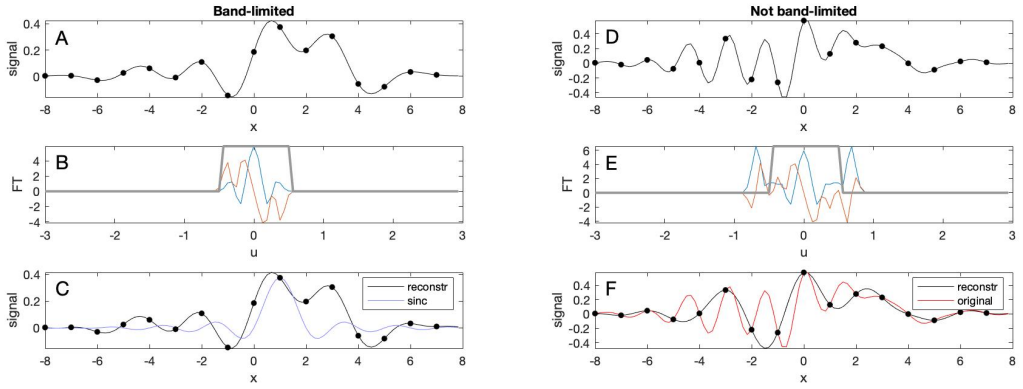
To formally understand sampling, we take a signal as a function of *x* and multiply it by shah(*x*) (Fig. 1A below). From the convolution theorem we know that its FT will be the FT of the original function, convolved with a shah function (Fig. 1B) so it will be periodic in *u.* If we isolate one copy of the periodic FT (using rect(*u*) as shown in gray), then we get the FT in part C. Transforming back to real space we have the reconstructed signal in part D. The trick, it turns



out, is the multiplication of the FT of the samples by rect(*u*). This of course corresponds to convolving the samples with the FT of rect, which is the sinc function. So to get from A to C in the figure we could also have convolved with sinc(*x*).

The multiplication by rect(*u*) limits the frequencies to *u* values between -½ and ½. If the original signal had contained Fourier components with frequencies outside this range, we would have had a problem, as these components would have overlapped in the periodic function in Fig. 1B.

Here's another example. Panels A-C demonstrate the reconstruction of a "band-limited" signal, that is a signal with no Fourier components beyond $\pm 1/2$ in magnitude, and reconstruction is perfect. (In panel C I've also plotted the sinc(x) interpolation function for one of the points.) If however the signal is not limited to that range, reconstruction is not correct.

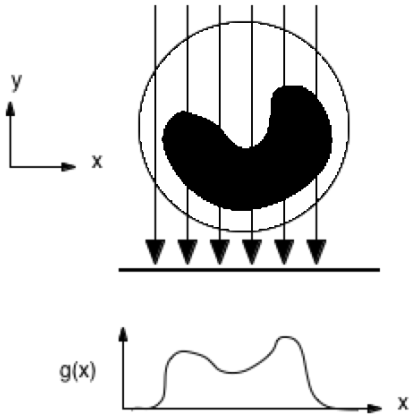


In general, the sampling theorem says that, if a signal is to be sampled discretely at some frequency f_s, the signal is completely represented by the samples if it contains no components at frequencies larger than $f_s/2$. This value $f_s/2$ is called the Nyquist frequency, named after Harry Nyquist (Yale PhD in physics, 1917) who discovered the sampling theorem.

2D Reconstruction

We'll first consider a 2D tomography problem for simplicity. A hospital CAT (computer-assisted tomography) scanner performs 2D reconstruction of slices of your body through the measurement of 1D X-ray projections. From these are computed a 2D tomogram for each slice.

A projection $g(x)$ is taken of a 2D object (say a section of a patient's body):



The projection $g(x)$ is related to the 2D density distribution $f(x,y)$ of the object through the projection integral

$$g(x) = \int f(x,y)dy \qquad (16)$$

where the integral is taken over the full y extent of the object.

Now suppose that we know the Fourier transform of the density distribution, which we will call $F(u,v)$. It can be written as

$$F(u,v) = \iint f(x,y)e^{-i2\pi(ux+vy)}dxdy \qquad (17)$$

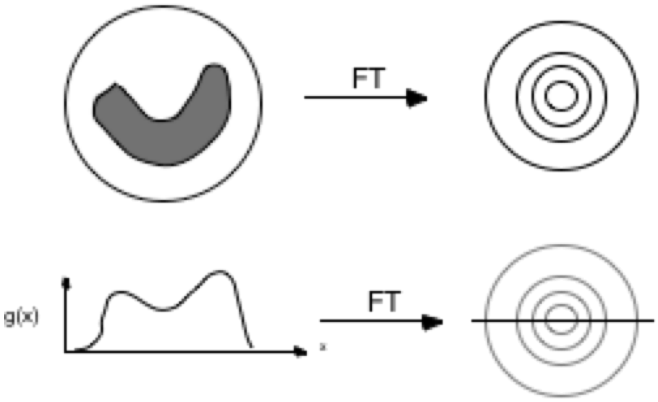
If we evaluate it at v=0, we get

$$\begin{aligned} F(u,0) &= \iint f(x,y)e^{-i2\pi(ux)}dxdy \\ &= \int \left[\int f(x,y)dy\right]e^{-i2\pi ux}dx \end{aligned}$$

which is just the (1D) Fourier transform of the projection $g(x)$,

$$F(u,0) = \int g(x)e^{-i2\pi ux}dx \qquad (18)$$

Thus the projection of an object is a section of its Fourier transform. In pictures:



This, plus the rotation property of Fourier transforms, is all we are going to need. Recall that if we rotate a 2D function, its FT rotates similarly. This means that if we rotate the object and then collect a projection, we will have obtained a different section of the 2D FT. If we collect enough such projections, we can fill in the whole FT. Then by transforming back, we obtain the original density map of the object.

This procedure is how computed tomography works, and is also how 3D molecular structures are obtained. In the latter case, the 3D version of the projection theorem says, a 2D projection is corresponds to a plane (a central section) of the 3D Fourier transform.

To make a 3D reconstruction from 2D projections of an object, you compute the FT of each projection image, which gives you a set of values in a plane. Then you "insert" it into a 3D

Fourier volume; that is, you modify the value at each voxel that intersects the plane, so that it matches the value on the plane. Clearly to do this you need to know at what angle the plane is to be "inserted". That is easy in the case of electron tomography, but harder in the case of single-particle reconstruction.

Electron tomography

Electron tomography (ET) gives you a 3D density map of a micron-sized region of a specimen. The most sophisticated version is cryo-ET, which images a specimen fast-frozen in vitreous ice. The quality of ET data is limited by electron doses. For cryo-ET, people give total electron doses up to about 100 $e^-/Å^2$ which is way too much for high-resolution (better than 1nm) imaging, but works okay for the usual ~5 nm resolution of ET. One accumulates a *tilt series* of micrographs, obtained for example at tilts every 2° from -60° to +60°, with an exposure of 1-1.5 $e^-/Å^2$ at each tilt value. It would be much better to approach 90° but that isn't practical: one would be looking through too thick a specimen, and one starts running into shadows from the grid bars and the specimen holder as well.

The limited range of angles means that there is a *missing wedge* in Fourier space, where the missing planes would have been inserted. The result is that the resolution of the reconstruction is different in different directions. In the figure below, a flagellum (panel a) is imaged by rotating the specimen about a horizontal axis in the plane of this page. The resulting tomograms (b) show good resolution in the horizontal direction, but the top and bottom of each microtubule's cross section is weak due to the missing wedge problem. It is possible to accumulate sub-tomogram averages of objects inside the tomogram. If the objects are present in various orientations within the sample, the missing wedge can be filled in (panel c).

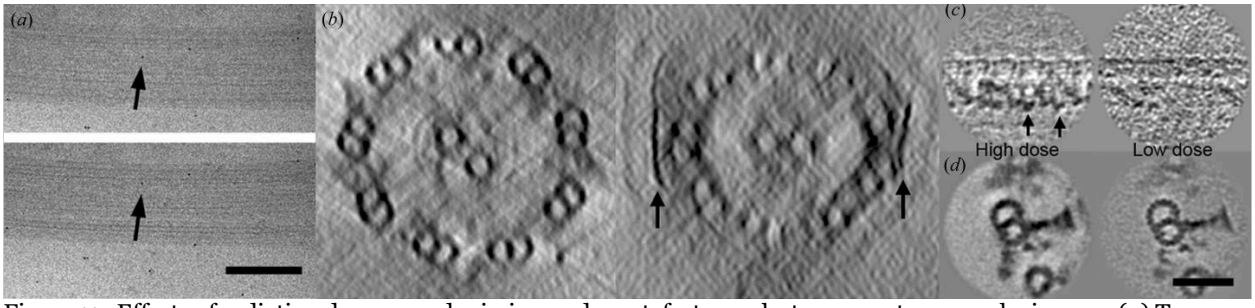
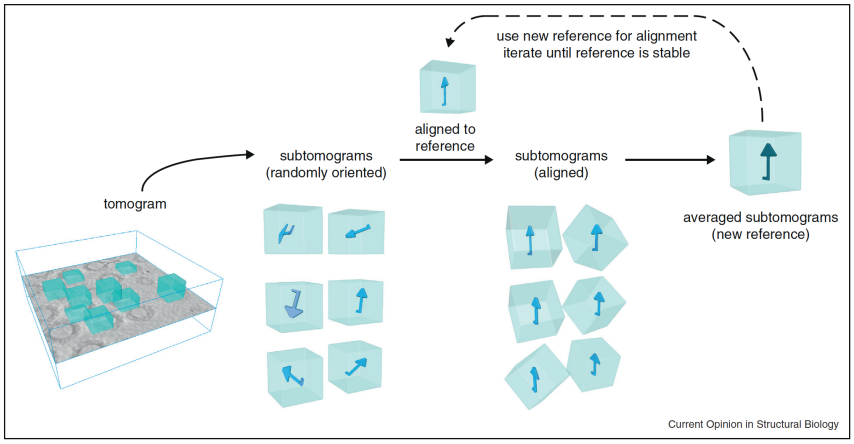


Figure 11. Effects of radiation damage and missing wedge artefacts on electron cryo-tomography images. (*a*) Two images from one series of tomographic acquisition. Top: ice-embedded flagellum tilted by 30°. Bottom: the same flagella without tilt. A gold label is shown by arrows. (*b*) Cross section of *Chlamydomonas* flagella with the membrane removed (left) and intact (right). The membrane is shown by arrows. Missing wedge artefacts generate non-isotropic density distributions. (*c*) Longitudinal (parallel to the microtubule) sections of averaged tomograms (ten particles) along microtubule doublets with high and low doses of electron beam. The averaged image with a high electron dose shows individual dynein molecules (arrows) even with ten particles. (*d*) Vertical sections of averaged tomograms (~1000 particles) with high and low doses. By averaging many particles obtained under a low-dose condition better resolution is obtained. Averaging among nine doublets generates isotropic microtubules (even and round-shaped). (*a*) Scale bar = 250 nm. (*b*)–(*d*) Scale bar = 50 nm. K. H. Bui, G. Pigino and T. Ishikawa, J. Synchrotron Rad. (2011). 18, 2-5.

Electron tomograms generally show strong signals at edges of objects, but do not properly reflect the interiors of areas of high or low density. This is because of the CTF for defocus imaging, which is very small at low frequencies. The situation can be helped somewhat by using a large defocus value (e.g. 10 μm) for low-resolution work. On the other hand, by using smaller defocus and sub-tomogram averaging it has been possible to use ET to obtain surprisingly good 3D structures of macromolecules.



An overview of subtomogram averaging in its simplest form. Subtomograms are extracted from the tomogram. They are rotationally and translationally aligned against a reference. The aligned subtomograms are then averaged to generate a new reference. The new reference is then used for alignment of the subtomograms again. This procedure is repeated until the reference stabilizes.

Current Opinion in Structural Biology 2013, **23**:261–267 www.sciencedirect.com

More about the contrast transfer function

Envelope function

When you use a high defocus value to improve the visibility of your protein, there is a cost in the resolution of the images. It can be difficult to undo all the rapid oscillations in the CTF, for one thing. But then there is a physical limitation that is quite serious. A defocus of 1 μm means that you are focused a very long distance, (some 400,000 wavelengths!) away from the specimen. Now suppose that the effective electron source size is such that some of the incident electrons follow a slightly different path than others. A typical situation in a microscope with a tungsten filament source would be that the incident electrons follow paths that differ in angle by 10^{-3} radians. These different paths can blur out the image of high-resolution features at large distances from the specimen. The variation in electron path is called *spatial incoherence*.

For example, suppose the specimen has a periodicity d = 1nm. At our defocus of 1 μm we are looking for differences in intensity with this same periodicity. But the periodic pattern imaged by electrons traveling at an angle of 10^{-3} radians will be shifted by 10^{-3} x 1μm = 1nm compared to the pattern imaged with zero angle (traveling along the z axis). Thus if the paths of the incident electrons have random angles in this range, the 1 nm pattern will be completely washed out! This is why the field-emission electron gun is so important: it allows the effective electron source size to be so small that angular spreads of 10^{-5} or 10^{-6} radians are attainable, which in turn should allow high resolution at high defocus values.

There is another process that increases angular spread and therefore decreases spatial coherence, called “charging”. When an incident electron is inelastically scattered, it transfers some of its energy to an electron of one of the atoms in the specimen, typically causing it to be ejected from the specimen. The result is that the specimen starts to take on a positive charge. This charge, if it is inhomogeneous or if the sample is tilted, causes a deflection of other incident electrons.

This deflection has the same effect as a large source size: it causes a variation in the electron path angle, and washes out fine details in the image.

These mechanisms both have the effect of blurring the image. They are typically modeled as a Gaussian decay of the CTF at high spatial frequencies. When we include this term, the CTF looks like

$$\mathrm{CTF} = \sin(\chi - \alpha) e^{-Bs^2/4} \qquad (2.1)$$

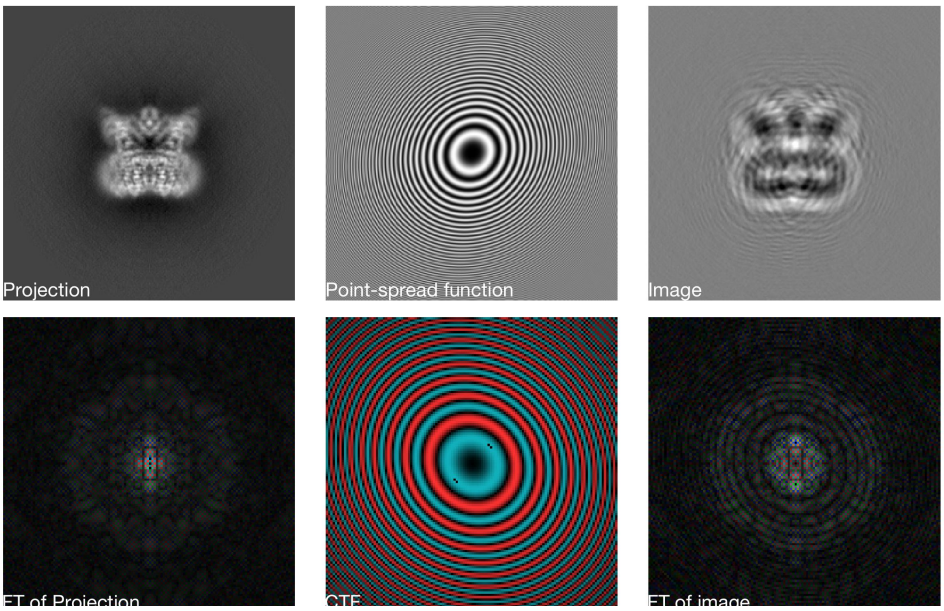
where

$$\chi = -\pi\lambda\delta f^2 + \frac{\pi}{2} C_s \lambda^3 f^4$$

and B has units of nm^2 or Å^2 and is called the "B-factor" or "envelope factor". The best cryo-EM images have B values of 20-80 Å^2, but even these values are not so good. At 60 Å^2 spatial frequencies of 4 Å are attenuated to $1/e$ of their original amplitude, and the power in the signal (the square of the amplitude) is reduced to about 1/10 of the original value. Higher spatial frequencies are attenuated even more.

Effect of CTF at high resolution

The figure below shows the effect of the CTF on a high-resolution image. The image in this case is a projection of the TRPV1 ion channel map. The CTF-modified image, as obtained with 2µm of defocus, looks inverted in contrast (due to the CTF being negative at low frequencies) and considerably distorted. Especially interesting is the dispersion of high-frequency information away from the center of the particle image. This results from the diffraction mechanism that we considered in the first lecture, and shows that, especially when a small particle is being imaged, ample space must be left around the particle to include the high-frequency fringes. In this simulation I also included some astigmatism, which can be seen from the elliptical rings in the CTF. This phenomenon arises from, in effect, a different defocus value holds for different orientations.



Effect of CTF on a high-resolution image. Top row are real space, lower row are corresponding Fourier transforms. Top left is a projection of the 3D map for the TRPV1 ion channel (Liao…Cheng, Nature 2013); the image is 310Å square. The simulated image, filtered according to the CTF for a defocus of 2 µm, is shown top right. The CTF also reflects 0.2 µm of astigmatism.

CTF "Correction"

How can we un-distort images that have been severely distorted by the effects of high defocus? It turns out that the worst part is the alternating polarity of contrast transfer. We can fix this by taking the Fourier transform of the image, multiplying that by the sign of the CTF, and transforming back. Here is what the CTF and its sign look like in two dimensions, for $\delta = 1\mu m$:

This operation, called "phase flipping", goes a long way toward making a high-defocus image interpretable, but it clearly isn't perfect. It reduces the worst part of the dispersion, but leaves an undershoot (white border) around the particle and does not eliminate the fringes entirely.

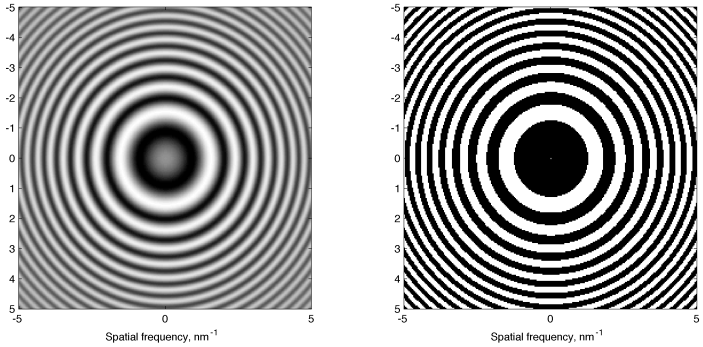


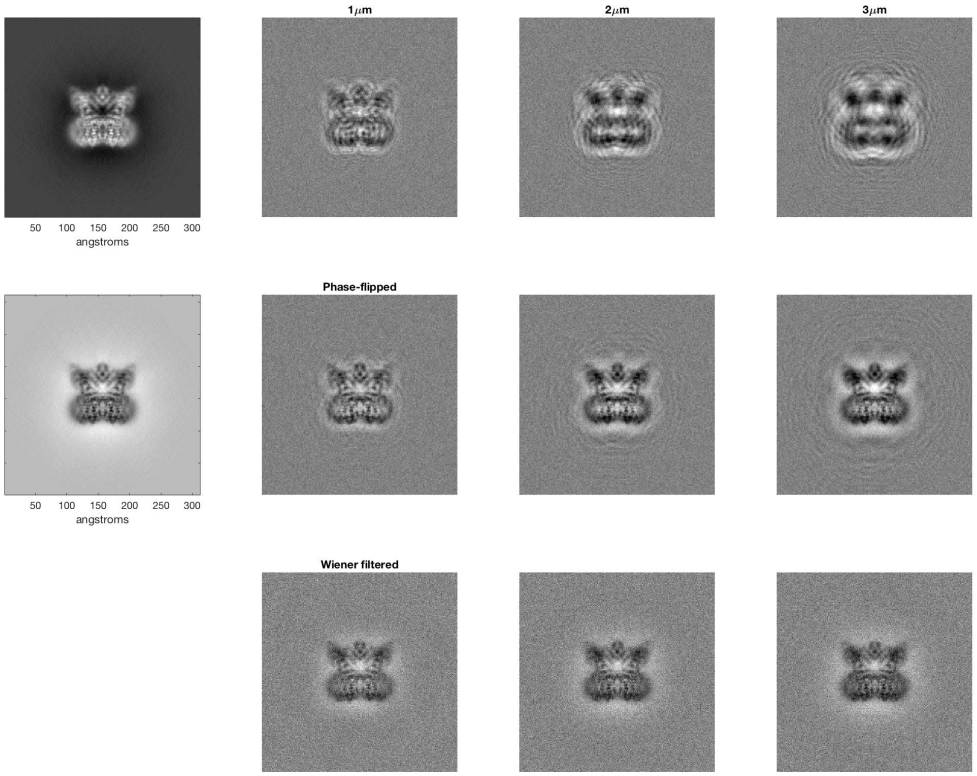
Figure 1. CTF and sgn(CTF)

So how do you "correct" an acquired image for the CTF? The answer is that you can't. There is information missing at frequencies where the CTF is zero; and near the zeroes, where there *is* information, it is nevertheless often unusable because its amplitude is so low. There are tricks that you can do, however.

1. Phase flipping, as we've just discussed.
2. Combine data from multiple images, obtained at various defocus values. This way the zeros from one image are filled in by data from others. This is the most powerful

method, which is used in electron crystallography and advanced single-particle reconstruction. We'll be talking about this strategy later.

3. Perform an "inverse filtering" operation, for example with a Wiener filter. The Wiener filter consists of a set of Fourier weights that are designed to correct for the shape and polarity of the CTF function without unduly magnifying the noise near the zeros. It is optimum in the sense of giving a minimum squared error in reconstructing the original image.



Defocus-contrast images at different defocus values (top row) are "corrected" by phase-flipping (second row) or a Wiener filter (third row).

The Wiener filter was discovered by Norbert Wiener[1] (1912 PhD in mathematics at Harvard, at age 17). In our image processing it serves as a sort of deconvolution device. Let's model the imaging system in the Fourier domain. We start with the original object $O(u,v)$ and pass it through a filter (the CTF) with frequency response $C(u,v)$ to yield the image

$$X(u,v) = O(u,v)C(u,v) + \text{noise} \qquad (2.2)$$

[1] From Wikipedia: The Wiener filter is a filter proposed by Wiener during the 1940s and published in 1942 as a classified document. Its purpose is to reduce the amount of noise present in a signal by comparison with an estimate of the desired noiseless signal. Wiener developed the filter at the Radiation Laboratory at MIT to predict the position of German bombers from radar reflections.... The unmanned V1's were particularly easy to model, and on a good day, American guns fitted with Wiener filters would shoot down 99 out of 100 V1's as they entered Britain from the English channel, on their way to London.

So, why don't we just do the following to recover the original object's structure:

$$O'(u,v) = \frac{X(u,v)}{C(u,v)} \; ?$$

We can't because there are zeros in *C*.

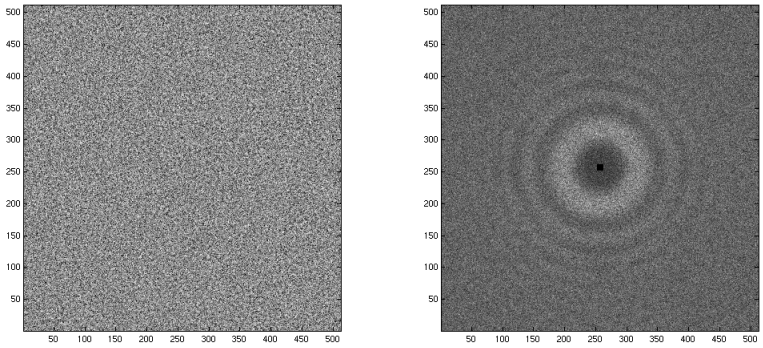
So, the next best thing (which gives a least-squared-error solution) is to instead compute

$$O' = \frac{XC}{w + C^2} \tag{2.3}$$

where the positive constant *w* keeps the result from blowing up at zeros in *C*. Optimally, *w* is chosen to be the inverse of the signal-to-noise ratio.

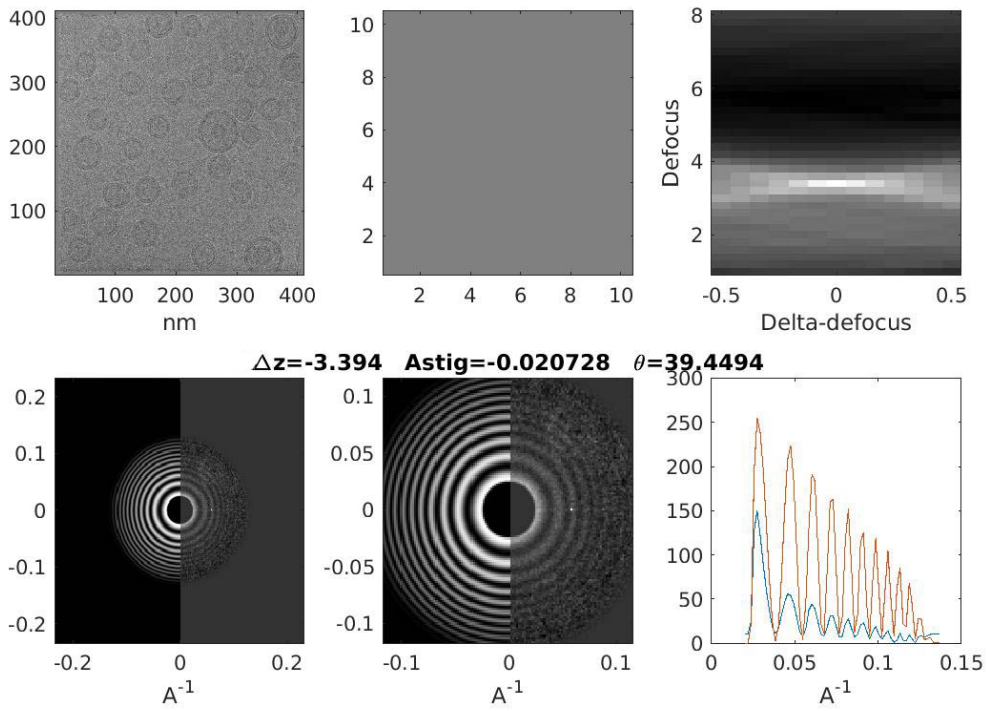
Determining the CTF

How can we know the exact value of defocus and other parameters, to be able to accurately model and correct for the CTF? Here is how it works in practice. Below left is part of an image, acquired with the CCD camera on our old F20 microscope. On the right is the power spectrum, obtained as the magnitude squared of the FT of the image on the left. The dark rings show where the zeros in the CTF are. The spectrum doesn't go to zero, but goes to minima set by the magnitude of the background shot noise. The positive and negative lobes of the CTF are not distinguished, due to the squaring operation.



You can fit the pattern of rings (called Thon rings) to very precisely determine the defocus and also get an idea of the *B* factor and other parameters of the CTF.

Below is the display of a CTF-fitting program. The circularly averaged power spectrum (lower right, blue trace) is fitted by the CTF^2 (orange trace). The fitting is actually performed in 2D, as



shown in the images in the bottom row of the figure. Once the defocus (here called Δz) is determined by fitting, one can reconstruct the CTF as shown left half of the bottom images.

If the Thon rings are not circular, you know that you have astigmatism. For high-resolution work it's essential that the CTF model includes astigmatism, so a CTF-fitting program will give you two defocus values, representing the major and minor axes of the elliptical rings, and also the astigmatism angle. Delta-defocus in μm, shown in upper right, is a measure of astigmatism and is very close to zero in this case.