Phase-contrast imaging in the EM

For cryo-EM imaging of biological molecules the most important imaging mode is phase contrast. The phase of an electron wave is shifted by its passage near the nucleus of each atom. The positive potential it experiences (a few tens of volts over a path length on the order of an angstrom) results in a small phase advance on the order of a milliradian.



Figure 1. An electron wave is phase-shifted when it passes near an atomic nucleus and is "elastically scattered". The phase shift is wildly exaggerated in this picture.

The average inner potential of water (or vitreous ice) is about 5 volts; for lipid bilayers, about 6.5V; for protein, about 7.5V. It is the contrast between protein and ice that we are usually imaging. Below is shown the calculated inner potential of a small protein in water. The map has been smoothed by filtering it to 10 Å resolution.



Figure 2. Upper panels: electrostatic potential inside a protein, smoothed by a 10 Å filter. Lower panels: simulated images of the protein with 0.5 and $2\mu m$ defocus.

The amount of phase shift the electron-wave encounters is proportional to the integral of the electrostatic potential along its path. The proportionality constant is called σ and has

the numerical values of 0.86, 0.73 and 0.65 mrad/VÅ for 120, 200 and 300 keV electrons, respectively. (Those units are milliradians per volt-angstrom.)

As electrons pass through different regions the phase difference gives rise to essentially no intensity difference in images with a microscope that is exactly in focus. This is why we use "phase-contrast" and "interference-contrast" devices on light microscopes to see cultured cells, which are also phase objects. In the EM the primitive way that people traditionally get contrast is to defocus the objective lens, as illustrated in the lower panels of Fig. 2. The process by which this phase-contrast image is formed is a bit complicated. One way do derive it is by use of the "Fresnel Propagator", the integral that shows how wavefronts develop in free space. The Fourier transform of the Fresnel Propagator yields the contrast transfer function (CTF) directly, and I'll put that theory into the notes for a later lecture. Another way to derive it is based on diffracted electron waves, and that's what we'll consider here.

1. An overview

Here's how we'll proceed in the next two sections to describe phase contrast by defocusing the microscope. We'll make use of a model object that has just a periodic variation in phase—imagine ripples in a thin film at the object plane with periodicity *d*. We can use an object like this to obtain a completely general result, because using Fourier transforms we can decompose any 2D distribution of density into ripples of varying spacing. Just below a specimen like this there is no contrast, because a variation of phase doesn't change the intensity of the beam. However, we'll show that the object's effect on the propagating electron beam is to produce two diffracted beams, with diffraction angles depending on the periodicity of the ripples. If we look at a distance *z* below the specimen the diffracted electron waves, because they are traveling at angles, travel farther and have delayed phase shift compared to the main, undiffracted wave. If we pick particular values of *z* the discrepancy in phase shift is a multiple of 90 degrees, and there will be an optimum amount of contrast. The particular optimal *z* values depend on the periodicity of the ripples. If we set the objective lens to focus an optimal distance *z* away from the specimen, we will have optimum contrast in the magnified image as well.



Figure 3. A periodic phase specimen gives rise to diffracted waves.

Part A of Fig. 3 above shows a diffracted wave (orange) produced by periodic scatterers (red). In part B, the two symmetric diffracted waves interfere to produce positive contrast at one distance (red circle) and produce negative contrast (blue circle) farther from the specimen.



Figure 4. A numerical simulation of the interference of 200 keV electron waves below periodic specimens with d=10 Å(left) and 7Å (right). Note that the vertical dimension of these pictures is compressed roughly 50-fold.

In this derivation I make use of two basic mathematical tools. One is the Taylor expansion for the exponential function

$$e^{y} = 1 + y + \frac{y^{2}}{2} + \frac{y^{3}}{6} + \cdots$$

and similar approximations for squares and square roots; we will be ignoring terms of second order or larger; and the other is Euler's formula for the complex exponential,

$$e^{iy} = \cos(y) + i\sin(y).$$

We also make use of one result from quantum mechanics, which says that the intensity of a wave (or the probability of detecting a particle) is the squared magnitude of the wavefunction, typically written $|\Psi|^2$.

2. A sinusoidal phase object.

Suppose we have a specimen that produces a small phase shift of the incoming electron wave. Assume the phase shift varies across the specimen in the manner of a cosine function,

$$\phi(x) = \varepsilon \cos(2\pi x / d). \tag{1}$$

This "sinusoidal grating" is oriented along the *x* axis and has a period *d*. Its magnitude is given by ε , much smaller than 1. Meanwhile we let the incident electron(s) have the time-independent wavefunction

$$\Psi = e^{ikz} \tag{2}$$

where the propagation constant $k = 2\pi / \lambda$. Eqn. (2) says that the electron waves have wavelength λ and are propagating in the *z* direction. Just beyond the specimen the wavefunction reflects the phase shift of the specimen as

$$\Psi = e^{i[kz+\phi(x)]}$$
$$= e^{i\phi(x)} \text{ at } z=0^+$$

We can approximate this by the first terms of the Taylor expansion of the exponential,

$$e^{i\phi(x)} \approx 1 + i\phi(x) \tag{3}$$

which relies on the assumption that $\phi(x)$ is very small. This important approximation is called the *Weak Phase Approximation*. It says that the effect of introducing the specimen— in this case, a periodic change in the electron wave's phase—is equivalent to saying that the specimen introduces a whole new set of waves with small amplitude, superimposed on the incident wave. The incident wave is unchanged in amplitude, to a first approximation. The fact that the new wave amplitude is an imaginary number is a shorthand for saying that it is advanced in phase by 90°.

3. Representation as diffracted waves

The second term in (3) represents a periodic set of sources of new waves. The new waves are diffracted waves of amplitude $i\epsilon/2$ that propagate at angles $\pm \theta$. The angles satisfy

$$\sin\theta = \frac{\lambda}{d}.\tag{4}$$

(Because the electron wavelength is small, in practical situations θ is tiny, at most a few degrees.) At a distance *z* below the specimen the diffracted waves have a phase shift retarded by a factor $\cos\theta$ relative to the phase shift *kz* that the undiffracted wave has. At a distance *z* below the specimen we have the overall wavefunction

$$\Psi(x,z) = e^{ikz} + i\phi(x)e^{ikz\cos\theta}.$$
(5)

Here I've taken advantage of the fact that the two diffracted waves combine to give a periodicity in *x* just like the original $\phi(x)$. Factoring out e^{ikz} we have

$$\Psi(x,z) = e^{ikz} \left(1 + i\phi(x)e^{ikz(\cos\theta - 1)} \right)$$
(6)

What interests us is the intensity of the electron waves, which is what is magnified and transferred to the detector. Quantum mechanics tells us that the intensity is given by the magnitude squared of the wavefunction. The magnitude squared of the wavefunction is equal to the sum of squares of the real part and the imaginary part. The magnitude squared of e^{ikz} is always equal to 1 so we can ignore that factor in eqn. (6), and the magnitude squared of the rest is

$$|\Psi|^2 = \left[\operatorname{Re}\left(1 + i\phi(x)e^{ikz(\cos\theta - 1)}\right)\right]^2 + \left[\operatorname{Im}\left(1 + i\phi(x)e^{ikz(\cos\theta - 1)}\right)\right]^2$$

It turns out that the imaginary part is very small, because only the second term involving ϕ contributes to it. That term's magnitude is at most ε and its square therefore so tiny it can be ignored. The real part however is on the order of unity, and its square is (using Euler's formula)

$$|\Psi|^2 \approx (1 - \phi(x) \sin[kz(\cos \theta - 1)])^2.$$
 (7)

Then, again neglecting a term of order ε^2 after expanding the square the intensity is

$$|\Psi|^2 \approx 1 - 2\phi(x) \sin[kz(\cos\theta - 1)].$$

Finally, we can make use of the fact that the angle θ is very small. To second order in θ ,

$$\cos\theta - 1 \approx -\frac{1}{2}\sin^2\theta = -\frac{1}{2}\lambda^2/d^2$$

And replacing $k = 2\pi / \lambda$ and noting that $-\sin(-y) = \sin(y)$ the intensity is given by

$$|\Psi|^2 \approx 1 + 2\phi(x)\sin(\pi z\lambda/d^2)$$
(8)

The contrast transfer function (CTF) is defined to be the scaling of the intensity change relative to the original phase shift in the specimen. We write

$$|\Psi(x,z)|^2 = \text{const} + 2 \times \phi(x) \times \text{CTF}$$

where the CTF is

$$CTF = \sin(\pi z \lambda / d^2)$$
(9)

This function is a very important result. Given a phase object that has a sinusoidal variation with period *d*, the amount of contrast we observe in the image depends on the plane that we have focused on. There is zero contrast when *z*=0 and also at *z* values where $\lambda z/d^2$ has an integer value; there are alternating positive and negative contrast maxima in between. <u>Please note</u>, as we'll explain below, in practice we always focus *above* the specimen so the CTF as described in the literature has the opposite polarity!

4. How does this work?

The basic problem of EM phase-contrast imaging is that right at the specimen, at $z=0^+$, there is no contrast. What the analysis above has shown, however, is that beyond the specimen, at a distance *z* displaced from it, there is substantial contrast. The contrast comes from interference between the undiffracted and diffracted beams. We have a situation like that in holography. There is a reference beam (in our case, the undiffracted beam) and there are diffracted beams from the object that interfere with it to produce a hologram. So in the microscope we don't record an image of our specimen, we record an "in-line hologram" of it.

In practice the necessary values of *z* are remarkably large (see Fig. 4). Suppose we have a specimen with a periodicity of *d*=10Å that we wish to image with 200 keV electrons ($\lambda = .025$ Å). The diffracted beams will make an angle $\theta = \sin^{-1}(\lambda/d)$, only 2.5 milliradians, because $\lambda/d = .0025$. The contrast is proportional to $\sin(\pi\lambda z/d^2)$; the first maximum occurs when $\lambda z/d^2 = 1/2$, which works out to *z*=2000Å. In order to optimally image this specimen, we would change the objective lens current so that its focus is on the plane at *z*=2000Å away from the specimen. This is very far away from the object, some 80,000 wavelengths! The reason the distance is so great is that the angle is so small, and so the difference in path lengths is tiny.

| | | - |
|--------------|--------|---|
| Energy (keV) | λ (Å) | |
| 100 | .03701 | |
| 120 | .03349 | |
| 200 | .02508 | |
| 300 | .01969 | |
| 300 | .01969 | |

Table 1. Electron wavelength at popular accelerating voltages.

5. The Contrast Transfer Function, embellished

Usually the formula (9) is written in terms of the "spatial frequency" of the object f=1/d. Also in practice people like to actually focus on a plane *above* the specimen. The quantity $\delta = -z$ is called the "defocus" value, that is the distance above the specimen that the objective lens is focused. This works fine, because as far as the objective lens knows, it is collecting electron waves that could have started above the specimen!

The reason for focusing above the specimen is that one gets stronger contrast at low spatial frequencies. There the phase contrast is negative, the same polarity as the "amplitude contrast" that arises from the loss of electrons scattered at high angles from dense parts of

the specimen. It also works well with the extra effects of spherical aberration of the objective lens as we'll describe below. The contrast mechanisms work together and we get the strongest image.

To try to avoid confusion I will henceforth use δ as the "defocus" or "underfocus" value, which is normally given as a *positive* number. "Underfocus" means turning *down* the current in the magnetic objective lens, making it weaker and therefore focusing above the specimen.

We can explicitly include in the contrast transfer function (CTF) the amplitude contrast, by giving it a mixing angle α that is typically around .05 radians for biological specimens.

Finally, a correction to the CTF comes from spherical aberration in the objective lens. Magnetic lenses are terrible, and one of their bad properties is that when beams arrive at the periphery of the lens they are bent more strongly than they ought to be. It is as if the focal length of the lens decreases for rays at high angles and therefore arising from high spatial frequency components of the object. If we focus above the specimen, the effect of the aberration is to, in effect, decrease the defocus for electrons arriving at steep angles. This is like the problem of glass lenses that have spherical rather than parabolic curvature: they bend light rays arriving at the periphery more strongly than they should. Magnetic lenses have nothing "spherical" about them, but the effect is approximated in the same way by what is called the spherical aberration coefficient C_s , which has units of length. In the two cryo-EMs in our facility C_s is about 2 mm. With the C_s term the contrast transfer function has this form

$$CTF = \sin(-\pi\lambda\delta f^2 + \frac{\pi}{2}C_s\lambda^3 f^4 - \alpha).$$
(10)

where the small quantity α describes the degree of amplitude contrast, that is contrast that is present with zero defocus. The C_s term only becomes important at high spatial frequencies, i.e. around 3 Å resolution.



Figure 5. The χ function and the contrast transfer function at 200 kV and defocus of about 0.2 µm, without and with spherical aberration C_s =2mm. With this choice of defocus you can see that the contrast-transfer function remains roughly constant over a range of spatial frequencies up to a resolution of about 3 Å (3.3 nm⁻¹) because the C_s term opposes the defocus term. This choice of defocus is called the "Scherzer defocus".

An alternative way the CTF is described in the literature is

$$CTF = sin(\chi - \alpha)$$

In this notation the "wave aberration function" $\,\chi\,$ is given as

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

The first term is the effect of a perfect lens; the second is the effect of spherical aberration. There are further terms that you could include, but they describe higher-order aberrations in the lens

that become significant only at resolutions beyond 1 Å. The top of Figure 5 shows a graph of χ as a function of spatial frequency.

6. Envelope function

A high defocus value improves the visibility of your protein, because it gives you more signal in the low-frequency range that shows the outline of the particle. But there is a cost in the resolution of the images. It can be difficult to undo computationally all the rapid oscillations in the CTF, for one thing. But a physical limitation is quite serious. A defocus of 1 μ m 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*.



Figure 6. CTFs at 0.5 and 1.5μ m defocus for 300 keV electrons, with decay at high resolution (*B*=20, 40) typical of a FEG microscope. Values are plotted up to 0.33 Å⁻¹, that is to a resolution of 3 Å.

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 arriving at an angle of 10^{-3} radians will be shifted by $10^{-3} \times 1$ µm = 1nm compared to the pattern imaged with zero-angle electrons. 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 guns (FEGs; found on Yale's Glacios and Krios microscopes) are so important: they allow the effective electron source size to be so small that angular spreads of 10^{-5} or 10^{-6} radians are attainable, which in turn allow high resolution at high defocus values. Still there is a decay at high resolution with an FEG, of the approximate magnitude illustrated in Fig. 6.

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, which goes by the name *envelope function*, the CTF looks like

$$CTF = \sin(\chi - \alpha)e^{-Bf^2/4}$$
(11)

and *B* has units of nm² or Å² and is called the "B-factor" or "envelope factor". Good cryo-EM images have *B* values of 50-100 Å², but even these values are not so ideal. At *B* = 100 Å² spatial frequencies of 5 Å are attenuated to 1/e of their original amplitude, and higher spatial frequencies are attenuated even more.

Figure 6 shows an example of the complete CTF at two defocus values. The zero-frequency value is slightly negative, due to amplitude contrast. The magnitude of the CTF oscillates but also decays with frequency, and the decay is faster at higher defocus due to the effects of spatial incoherence.

References

Almost every book on high-resolution EM talks briefly about the CTF, but few derivations are understandable; hence I've given my own here. An overview of the subject is given in: Wade, R. H. 1992. A brief look at imaging and contrast transfer. *Ultramicroscopy* 46: 145-156.

A description of the CTF and important practical issues in EM image formation can be found in chapters 11 and 12 of a book by Yale's own Peter B. Moore, *Visualizing the Invisible*, Oxford 2012, available online at http://site.ebrary.com/lib/yale/docDetail.action?docID=10560926

Not for the faint of heart, but the definitive book on TEM physics is: Reimer, Ludwig. 2008. *Transmission Electron Microscopy*, 5th Ed. Springer Verlag.