**Partial Coherence: The role of source size**

If we have two waves, exp(2i**u**1.**r**) and exp(2i**u**2.**r**+i), when we interfere the two the intensity will be

I(**r**) = 2+2cos(2[**u1**-**u2**].**r**+)

The result we get depends upon whether the phase  has a fixed value, or is statistically random. To see how we get coherence from incoherence, a classic example is to consider emission from a small source, as sketched in the Figure below.



R

When electrons are emitted, each point on the source emits a spherical wave with a specific energy, i.e. the phase at any position dependes only upon the distance from the point in the source. Let us assume (normally valid, although may be suspect with a very small sources, for instance a carbon nanotube) that the electrons are emitted incoherently from each point – i.e. there is no specific phase relationship (in a statistica sensen) between the electrons emitted from different points.

To define the geometry, we define an origin in the source O, and refer to all points in the source with respect to this. We are going to be interested in what we have at some “observation plane” far from the source, a (normal) distance R. Any point in this source will be refered to in terms of a distance from a second origin ***O***. We want to understand what is the relationship between the wave at two points u1 and u2. (For instance, we might interfere these two later, and we want to understand what will be the amplitude of the fringes.) To make life a little simpler (we don’t have to), let us say that both the source and the plane of observation are at right angles to the seperation vector R between the two.

Let us start with some point p at the source. The phase of the wave at u1 (and u2) will depend upon the distance it has traveled, which we will call R1. Dealing with everything as vectors,

**R**1 = **R** + **u**1 – **p**

If we say that the phase of the wave when it is emitted is zero, the phase at u1 will be

= 2|**R**1|/

where  is the wavelength. We can write a similar equation for the phase of the wave at **R**­­2. The phase difference between the waves at u1 and u2 will therefore be:

= 2|**R**1|-|**R**2|}/

We now want to convert from R1 and R2 to something a little more manageable. We can write (exact, no approximations)

|**R**1| = {R2 + |u1-p|2}

If R is large, we can expand this as a Taylor series and neglect terms with smaller values

|R1| = R{1 + |u1-p|2/R2}  R{1+|u1-p|2/2R2}  R+|u1-p|2/2R

Therefore, substituting these in and calculating the phase difference

  2/R+|u1-p|2/2R - R-|u2-p|2/2R]

 /|u1-p|2/R - |u2-p|2/R]

 /(u1)2 - (u2)2}/R - 2{u1-u2}.p/R]

The interference depends not on the difference between the phases, rather exp(i[]) which we can write as

exp(i[]) = exp(i/(u1)2 - (u2)2}/R - 2{u1-u2}p/R]\_

To simplify this, let’s introduce a new variable s=p/(R), and convert to vectors for s, u1 and u2 so that

exp(i[]) = exp(i(u1)2 - (u2)2 }/(R - 2i{**u**1-**u**2}**.s**)

Now, the last stage. Every point in the source will emit, but they do not all have to emit the same number of electrons per second. Let us describe the emission from each point by some function S(**s**) (using **s** as a vector rather than p). The amplitude of (for instance) fringes due to interference later down the column from the waves at u1 and u2 will be determined by summing the contributions from all the points in the source, i.e.

<exp(i[])> =  exp(i(u1)2 - (u2)2}/R - 2i{**u**1-**u**2}**.s**)S(**s**)d**s**

or

<exp(i[])> = exp(i(u1)2 - (u2)2/R}) exp(-2i{**u**1-**u**2}**.s**)S(**s**)d**s**

The first term on the right does not really matter – it’s amplitude is 1 so we can ignore it in many cases. Introducing a new term to describe our final result,

<exp(i[])> = (**u**1,**u**2) =  exp(-2i{**u**1-**u**2}**.s**)S(**s**)d**s**

The term (**u**1,**u**2) is called the “mutual coherence” between the two points **u**1 and **u**2. If it was zero, the two are completely incoherent and cannot interfere. If it is one, the two are completely coherent. The smaller the source, the more “coherence” there is between the two different waves – hence small sources are in general better.

Going back to the original example used at the top of this section, in general we can write:

I(**r**) = 2+2(**u**1,**u**2) cos(2[**u**1-**u**2].**r**+)

where is the average phase (which will include the term that we dropped earlier).

The idea that something small increases the coherence occurs in many places when you use any type of microscope. For instance, suppose we have a very large source, e.g. a tungsten filament or an LaB6 tip where the region from which there is emission can be several microns in size. In this case we can consider the condensor aperture to be incoherently illuminated (only points exceedingly close togethor are coherent). In this case two points at the object have a mutual coherence which is determined by the size of the condensor aperture – the smaller it is, the more coherence there is.

An important final point. When we create an image, we are really only doing this (in almost all cases) with the *coherent* part of the wave. We Fourier transform waves from one plane to another, e.g. from the object to the diffraction plane (objective aperture plane) by summing contributions – which we can only do if the contributions are coherent. As an example, the image of the source on the sample is formed from the electrons which are coherent across the condenser aperture. The size of the source (illuminated region approximately) is going to be determined largely by how much coherence we have. If we demagnify the source, we are expanding the region of coherence across the condenser aperture, thereby reducing our illumination intensity (longer exposures). A small source (e.g. a FEG) will save us a lot of effort, giving more electrons in a small region which is better for STEM operation. A FEG source does have other problems (stability, cost) so it is not always best.