

Ptychography

John Rodenburg and Andy Maiden

Abstract:

Ptychography is a computational imaging technique. A detector records an extensive data set consisting of many inference patterns obtained as an object is displaced to various positions relative to an illumination field. A computer algorithm of some type is then used to invert this data into an image. It has three key advantages: it does not depend upon a good-quality lens, or indeed on using any lens at all; it can obtain the image wave in phase as well as in intensity; and it can self-calibrate in the sense that errors that arise in the experimental set-up can be accounted for and their effects removed. Its transfer function is in theory perfect, with resolution being wavelength-limited. Although the main concepts of ptychography were developed many years ago, it has only recently (over the last ten years) become widely adopted. This chapter surveys visible light, X-ray, electron, and EUV ptychography as applied to microscopic imaging. It describes the principal experimental arrangements used at these various wavelengths. It reviews the most common inversion algorithms that are nowadays employed, giving examples of meta code to implement these. It describes, for those new to the field, how to avoid the most common pitfalls in obtaining good quality reconstructions. It also discusses more advanced techniques such as modal decomposition and strategies to cope with 3D multiple scattering.

1) Introduction

We have an object, possibly a very small object, and we want to make a magnified image of it. There are various strategies open to us. For many years the answer was to make one or more lenses as accurately as possible, and arrange them as accurately as possible in a microscope column (Figure 1a). New possibilities opened up with the advent of computers. If the image has minor systematic faults arising from the physical hardware, we can process it in the computer to improve upon it (Figure 1b). Alternatively we can build flexible adaptable optics that are controlled by a feedback loop, via detailed quantification of the distortion in the image. One of the biggest breakthroughs in transmission electron imaging relies on computationally measuring lens aberrations, and then compensating for them using complicated non-round lenses driven by dozens of variable currents (Figure 1c and see Chapter ****EDITOR**** in this volume). Similar compensation strategies have been employed in astronomical telescopes and in many other fields of optics.

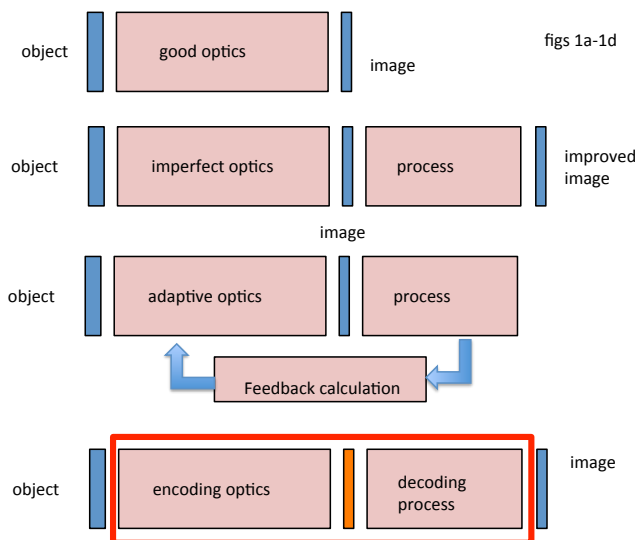


Figure 1: The computational imaging paradigm. a) The conventional microscope. b) Errors in aberrated or imperfect optics are corrected by post-processing. c) Errors in the optics are measured in a detector plane. Variable optics are then adjusted to improve the image via a feedback computation. d) Ptychography: the detector measures something rich in information, but not the image. Decoding computation is used to form the final image. The system is a type of transmission line.

A more radical conceptual leap is to realise that perhaps we should not worry about our image at all. All we need is a detector lying somewhere in an optical system that records something rich in information (Figure 1d), whether an

image or not. Now what we need is optics that *encodes* information about our object as efficiently as possible before it reaches the detector. Once we have those data, we can *decode* them (assuming we know something about the encoding optics) and computationally generate our image. This imaging paradigm is nowadays widely called ‘computational imaging.’ A communication channel replaces the optical transfer function of the traditional lens: structural information is transferred via three steps: physical encryption; detection; and finally a decoding algorithm.

Ptychography is a method of computational imaging. It employs a source of radiation (light or matter waves of arbitrary wavelength), an object that scatters that radiation, and a detector. We will see that we can have all sorts of optical elements between the source and the object, and between the object and the detector: the variety of modern implementations of ptychography is enormous. But it has five defining properties:

- (1) There must be an optical component, which is usually, but not always, the object itself, which can move *laterally* relative to whatever is illuminating that optical component.
- (2) The detector must be in an optical plane in which the radiation scattered from the optical component has intermixed to create an *interference pattern*, usually a diffraction pattern, but more generally any interference pattern, possibly even an image.
- (3) The detector collects at least two interference patterns, arising from *a minimum of at least two known different lateral physical offsets* of the illumination with respect to the object or other relevant optical component (modern implementations can use 100s or 1000s of lateral offsets). The offsets must be arranged so that adjacent areas of illumination overlap with one another.
- (4) The source of radiation must be substantially (but not necessarily wholly) coherent.
- (5) The image of the object is generated by a computer algorithm, which solves for the phase of the waves that arrived at the detector, even though the detector itself was only able to measure the intensity (flux or power) of the radiation that impinged upon it. If we were designing a normal communication channel, say a telephone transmission line, the very last thing we would ever choose to do is to have this catastrophic disposal of phase information right in the middle of the system. But that’s what happens with light, X-ray and electron detectors. An important strength of ptychography is that it can handle this regrettable ‘phase problem’ with no effort at all.

One thing is clear. Unlike the immediacy of a conventional microscope, ptychography puts a huge obstruction between the microscopist and the image. First, we must wait while at least the two interference patterns are recorded: the experiment takes time. Second, we have to rely on the computer to reconstruct the image from the data. The data usually looks nothing at all like the object of

interest: we must wholly trust a computer algorithm to deliver our results, something that unnerves quite a lot of scientists.

So, why would anyone want to use such a roundabout way of creating an image? There are three key benefits of the technique.

(1) It does *not need a lens*. Most implementations of ptychography do in fact use a lens, but the lens can be very poor quality: it will not affect the final (high) resolution of the ptychographic image. This has been the driving motive for X-ray ptychography, where high-resolution lenses are hard to make and very costly. X-ray ptychography nowadays routinely improves upon lens-based X-ray imaging resolution by a factor of between 3 and 10. Ironically, the original motive of ptychography was to overcome the resolution limit of electron lenses, but aberration correctors (Chapter **EDITOR**) now provide such high resolution – fractions of an atomic diameter – that extra ptychographic resolution has little to offer, at least at the time of writing. Of course, at visible light wavelengths, lens optics is a mature field, already offering wavelength-limited resolution.

(2) It produces a *complex image* (that is, both the modulus and phase of the exit wavefield at the back of the object). Image phase is an elusive thing, which is nevertheless crucial for imaging transparent objects like live biological cells that do not absorb radiation, but do introduce a phase change in the wave that passes through them. Consequently all sorts of optical methods have been developed over the last century for expressing the phase in an image, for example Zernike contrast, differential phase contrast, holography or processing through-focal series. However, it is a matter of fact that ptychography produces extraordinarily good phase images: the transfer function of the technique is, at least under most circumstances, almost perfect. This phase signal remains a pressing need over all wavelengths, and is the main motive for ptychography at visible light and electron wavelengths. It is also the key to the success of high-resolution X-ray ptycho-tomography (see section 6.1).

(3) We have stated that all sorts of optical components can be used in a ptychographical experiment. One would suppose that the characteristics of these components would have to be known exactly, or at least very well. After all, methods like holography need exquisite optical alignment, and even then various calibration steps must be undertaken to characterise the reference wave inhomogeneities. But a remarkable peculiarity of ptychography is that the method *self-calibrates*. It blindly characterises the optical components in the experimental set-up. It computationally provides a map of all the aberrations in any lens which is being used in the system, including apertures and slits. It can measure (and remove the effects of) any partial coherence in the source. It can find and correct for errors in the lateral displacements that are themselves the central source of the ptychographical information. It can infer the physical position of the detector. It can even correctly estimate the intensity of thousands of pixels that are inoperable in the detector, and even infer the intensity that would have been measured outside the edges of detector had the detector been larger.

What is the secret of this remarkable technique? There are many inverse computational imaging methods that solve for extra information, say the phase of an image, using multiple images collected as a function of some variable or other. More images mean more measurements, and more measurements usually mean more overall diversity in the entire data set. It happens that the source of diversity in ptychography – lateral shift – is easy to implement experimentally; unlike, say, a through-focal series, ptychographical data can be collected in endless abundance; and the diversity of these data is large. In other words, the communication channel of ptychography (Figure 1d) has a very wide bandwidth (see Section 4). Because most of this bandwidth is redundant, any errors in the encoding system can be corrected: it is very hard for the message (the image) to be lost or corrupted by instrument noise (of course, the fundamental limits of counting statistics will always apply).

Any computational imaging strategy must have its decoding algorithm. Ptychography's involves a procedure that must solve the 'phase problem', which has historically been seen as extremely difficult. Wave amplitudes add linearly, their intensities do not, and so the solution space is highly non-linear. We might suppose that the decoding algorithm in ptychography must be very complicated and very ill-conditioned. This is not the case. Perhaps another key reason for its success is that the most popular reconstruction algorithms available today are both intuitive and very easy to code. They also invariably work without too much tweaking or insider knowledge. It is astonishing that any one of the core algorithms can be used for any of the very diverse range of ptychographical set ups, or for any of wavelength – photon or electron. The only exception to this is that the "WDD" inversion method (see section 10) must have very densely sampled data: conversely, any of the iterative algorithms works for densely sampled data.

1.1 Nomenclature

Ptychography/cCDI: History dictates that in certain communities ptychography is seen as a type of coherent diffractive imaging (CDI). Recent developments, especially Fourier ptychography which records images but not diffraction patterns, perhaps renders this classification out-dated. Furthermore, CDI is inextricably linked with the term 'oversampling', which is not a fundamental constraint in ptychography. Here we will therefore reserve the term 'conventional CDI' (cCDI) for methods that recover structure from a single diffraction pattern (see Chapter **EDITOR** this volume which is dedicated to this subject): ptychography always uses data from more than one interference pattern and is probably best thought of in terms of Figure 1d.

Illumination/probe-object/specimen. The illumination function is very often called a probe, because the illumination is often made using a lens that converges a conical beam onto the specimen. We will use 'probe' interchangeably with 'illumination' depending on context. The same applies to 'object' and 'specimen'.

Exit wave or transmission function. There has been some confusion about whether ptychography solves for the exit wave of the object or the transmission function of the object. Very early work on ptychography sometimes used ψ_e to represent the exit wave from a specimen illuminated by a plane wave. This is only valid if the object function is indeed identical to the exit wave under plane wave illumination, which is true if the object is infinitively thin. However, as soon as we introduce depth effects, say by solving for multiple layers of a specimen (see Section 6.2), then the only interpretation of the object function is as a physical transmission function. The actual exit wave, from one probe position, bears no obvious relation to any of the layers in a 3D object, and there is no single two-dimensional function that can account for all the different exit waves that occur at every probe position. Propagation can also lead to features in the exit wave having higher amplitude than any part of the incoming wave. In short, we solve for transmission functions (and sometimes multiple layered transmission functions), not the exit wave. We do however have to solve for each different exit wave at each probe position.

Object functions as propagating waves: In some configurations, ptychography solves for a wave while an aperture of some type acts mathematically in the role of the illumination. In Fourier ptychography (Section 5.2), the aperture lies in the back focal plane of the objective lens and the 'object' function is the complex-valued diffraction pattern lying in the same plane. In selected area ptychography (SAP – Section 5.3) the aperture lies in an image plane and now the 'object' is the complex-valued image formed by the objective lens. The important point is that the mathematics of ptychography applies to any complex-valued function moved across another complex-valued function. Which one of these scatters (object/aperture) and which illuminates (probe/image or diffraction pattern) is inconsequential.

Fourier/detector projection: Historically, the projection in the diffraction plane of an iterative reconstruction algorithm has been called the 'Fourier constraint'. Because in ptychography this constraint can occur in the Fresnel near field or in the image (in the case of Fourier ptychography), we will call it here the 'detector constraint'.

Bright- and dark-field data: We may occasionally use the term 'dark-field intensity'. If the illumination is of the form of a convergent beam, then in the far-field the aperture in the probe-forming optics appears as a disc. The intensity in the disc is what is used for bright-field imaging in scanning transmission microscopy (the intensity there is collected as a function of a continuous scan of the probe). The intensity outside the disc is then the dark-field intensity, which is invariably much weaker than the bright-field disc. By reciprocity, in Fourier ptychography dark field intensity is called the same as in conventional microscopy: i.e. when recording a dark-field image the incident beam has been tilted far enough so that it is blocked off by the objective aperture. The resolution of a perfect lens can only be improved by ptychography if dark-field data is processed.

The ‘Fat-H’ and the ‘Trotters’: We will later introduce these two informal terms that are widely adopted by the community, but are not recorded in any published paper. We think this is timely because the subject to which they relate (Wigner Distribution Deconvolution (WDD) – see section 10) has had a recent resurgence. They are compact terms for complicated data structures and are now regularly used in discussions at conferences, etc.

STEM/STXM: We will call both the scanning transmission electron microscope (STEM) configuration and the scanning transmission X-ray microscope (STXM) ‘STE/XM’. This is because, being optically equivalent to one another, ptychography treats them both identically.

Ronchigram: This term is used in the electron literature but rarely in the X-ray imaging literature. It refers to the unscattered beam created by a convergent focused beam in the far-field. This is usually circular in electron microscopy (being the shadow image cast by the condenser aperture). For an X-ray Fresnel zone plate lens it is usually doughnut-shaped because of the central stop required to block undiffracted intensity. If Kirkpatrick-Baez (KB) mirrors are used, it is rectangular. When the probe is defocused from the object plane, it is equivalent to a Gabor in-line hologram.

Circles: As a warning to the reader, we remark that the science of ptychography involves lots of diagrams of circles. The probe function (in real space) is often circular, or represented by a circle. Diffraction disks (in reciprocal space) from a crystalline object are circular when a focused lens with an aperture in its back focal plane is used to form the probe. The Fat-H and the trotters are made out of parts of circles. Fourier ptychography and SAP ptychography have their own circular apertures. The modulus constraint in the complex plane is circular. Know which circle is which: they are not all the same!

2) A Brief History

This Chapter is not an historical review. However, for the benefit of those new to the subject, we now make one or two non-essential observations about its history.

First – where did the name come from? Ptychography derives from the Greek ‘ptycho’, meaning to fold. Hoppe and Hegerl [1] introduced it to describe a method of calculating the phase of the Bragg reflections from a crystal [2-4]. If a localised spot of radiation illuminates a crystal, the Fraunhofer diffraction pattern is a convolution (or in German ‘Faltung’, folding) of the crystal Bragg reflections with the Fourier transform of the illumination function. The latter is wide, because the illumination is narrow, and so the Bragg peaks, which are usually perfect spots, are made to overlap one another. If the radiation is coherent, the overlaps interfere with one another (as can be seen in Figure 2c, which we discuss in detail in Section 10). Hoppe observed that this interference could be used to estimate the phase difference between any pair of overlapping discs, bar an ambiguity of a complex conjugate. He realised that the ambiguity

could be resolved by shifting the illumination to a second position. He also proposed that the method could be extended to generalised non-periodic objects.

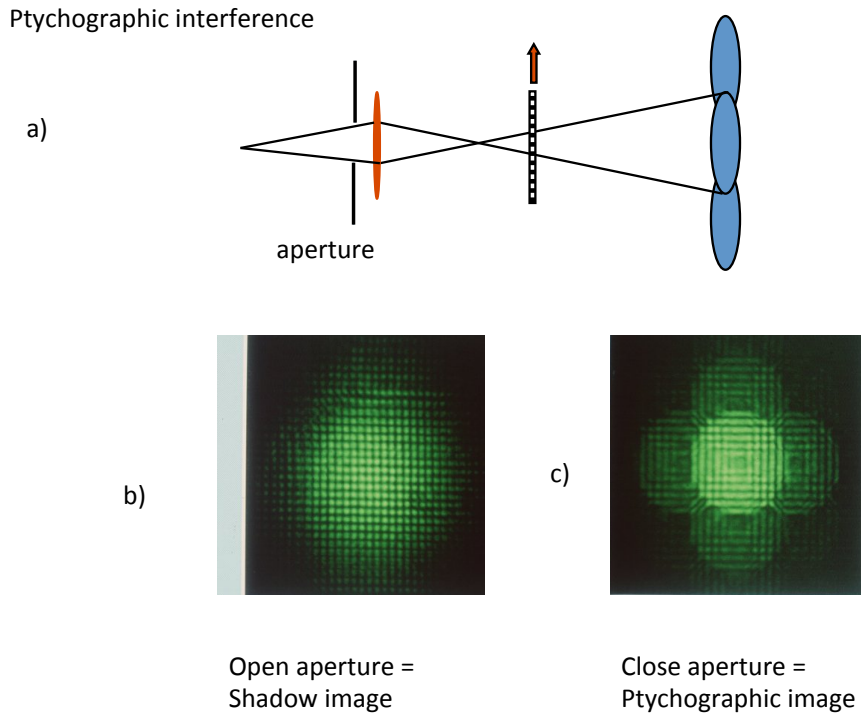


Figure 2: a) A lens brings a beam to a cross-over in front of a periodic object. b) In the far-field we see a shadow image (experimental data using a laser incident on a TEM grid). c) When the lens in (a) is stopped down by an aperture, we see explicit diffraction orders which interfere with one another.

For an explicit description of this phenomenon, see [5]. Modern ptychography has very little in common with this original concept, but the name has stuck. In fact it is still true to say that the word captures the essence of a diffraction pattern determined by a convolution, and a shift of illumination relative to the object. Note that in the context of near-field or full-field ptychography, which we will discuss in Section 5.4, the Fresnel integral can also be formulated as a convolution, but in this case there is no requirement for the illumination to be localised. In Fourier ptychography the measured data is the convolution of the image (the Fourier transform of the object diffraction pattern) with the Fourier transform of the lens aperture, i.e. the impulse response function of the lens.

After having had one research student (Hegerl) work on the technique, Hoppe abandoned it, as he describes in [6] written at the time of his retirement in 1982. There was some work done on ptychography by a small group in Cambridge led by one of the authors during the 1990s (reviewed in [5]). Nearly all of this was based on non-iterative ptychographic inversion algorithms, except for the work of Landauer, who developed iterative algorithms for Fourier ptychography [7]. Chapman successfully applied one of these techniques to soft X-ray ptychographic data[8], but this class of direct inversion algorithm required very

large quantities of data, which could not easily be handled by the computers available at the time. The X-ray data took a long time to collect because in those days source brightness was relatively low. Certainly the electron detectors available then were utterly dismal. Now that technology has moved on, there is a resurgence of interest in these techniques [9, 10], which will be discussed in Section 10, and which may yet prove to be very powerful.

The big explosion of interest in ptychography began in 2007, starting in the X-ray synchrotron microscopy community [11-13]. We can identify four reasons for this:

First, the development of third generation synchrotrons supplied very bright, spatially coherent sources, suitable for conventional coherent diffractive imaging (cCDI).

Second, following the first “Coherence” conference in 2001 [14], there developed a large community of scientists interested in both the physical implementation and iterative solutions of the cCDI X-ray phase problem as it relates to single diffraction patterns from finite objects [15, 16]. This meant that when the first real-space iterative solution to the ptychographical phase problem was demonstrated experimentally [11, 17], there were many workers who could immediately implement it on existing beamlines and instrumentation. It helped that the simplest experimental set up required only an aperture and a stepper stage, and that the associated iterative solution [18], although very quickly superseded by more comprehensive approaches, was very simple to code. Furthermore, because of the diversity of ptychographical data, all reconstruction algorithms for it are relatively robust, at least compared with those used for cCDI.

Third, there was a strong demand for higher resolution in X-ray microscopy that could not be easily satisfied by improved optics, but which could be delivered easily by ptychography [12, 19]. Although ptychography was originally developed to overcome the electron lens resolution problem, by the time it came to maturity aberration-corrected lenses could provide all the resolution one could usefully employ, although its ability to recover image phase accurately is still in demand.

Finally, the phase sensitivity of ptychographical micrographs, measuring quantitatively and linearly the projected optical potential, meant it could be very effectively used for tomographic imaging [20, 21], which has become one of its most scientifically significant applications (Section 6.1).

These benefits were already established in the literature by 2010. Since then there have been numerous developments over many different wavelengths and in many different optical configurations. We will try to cover most of the important trends in this Chapter, but as the rate of progress in the field accelerates, much of what we write here will quickly become out of date. Treat this chapter as an elementary introduction to the field.

3) How Ptychography Solves the Phase Problem

Chapter ****EDITOR**** of this volume is dedicated to single pattern diffractive imaging. In many situations of experimental importance, such as the strategy of ‘diffract and destroy’, we can only record one diffraction pattern because after one exposure the object of interest has been damaged or completely destroyed. However, we can use some of the concepts in cCDI to work our way towards an understanding of ptychography, especially in how it solves the phase problem. There is therefore inevitably some small overlap with Chapter ****EDITOR**** in what follows.

3.1) The Phase Problem

Let us consider a very simple version of ptychography, as shown in Figure 3. A source of illumination passes through an aperture and then the object. The resulting exit wave from the object propagates to the far-field Fraunhofer diffraction plane, where it is recorded on a detector with $N \times N$ pixels. For simplicity, we assume the aperture is so close to the object that there is no diffractive spreading of the beam between the aperture and the object. Figure 4 shows two $N \times N$ arrays of complex (real and imaginary) numbers, related to one another by Fourier transformation. The leftmost array shows an estimate of our object where it has been illuminated by the round aperture. The rightmost array represents the pixels in our detector: this array is the data we measure. Because the Fraunhofer integral is a linear, invertible Fourier transform, we should be able to back Fourier transform the pixels in the detector array and then discover the exit wave coming from the object.

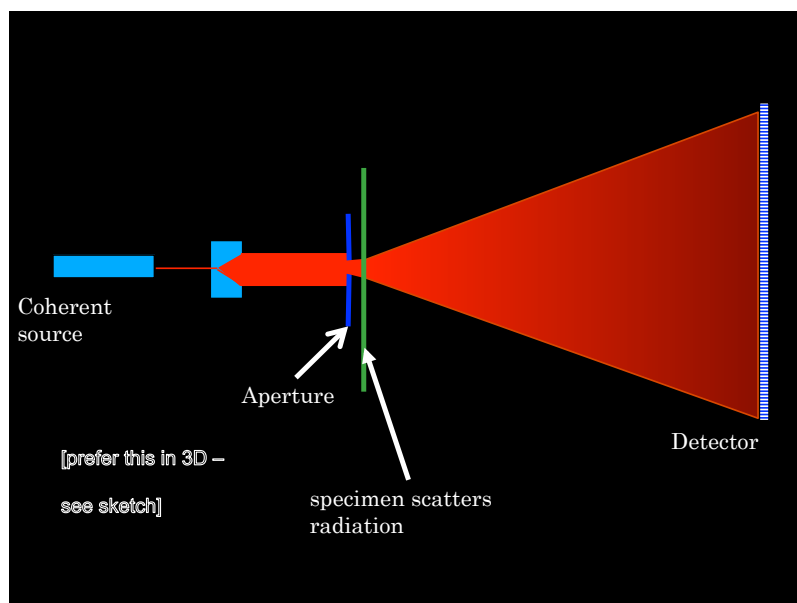


Figure 3: The simplest type of diffractive imaging experiment. An aperture, which we will start by assuming lies right against the specimen, i.e. there is no propagation spreading of the wavefield between it and the specimen. A detector lies in the far-field Fraunhofer diffraction plane.

The exit wave is also a complex function, intimately related to the structure of the object. Roughly speaking, its modulus represents the object's transmittance (1 equals total transmittance, which is free space, 0 being total absorption), and its phase is the accumulated phase difference, relative to free space, caused by the real part of the refractive index of the object as the wave passes through its thickness. Ideally, we want to measure both the modulus and the phase of the exit wave to find out the most about the object. In general, all microscopes suffer from the image phase problem, although there are many ways to express phase in an image, at least approximately. Ptychography is a way of achieving a very clean exit wave phase estimate.

For some radiations, for example electromagnetic radio waves, it is easy to build an $N \times N$ detector that can measure both the modulus and the phase of the wave arriving at it. This is how aperture synthesis radio astronomy works. In our case, however, all the radiations that have short enough wavelengths to be useful as high resolution microscopes oscillate at correspondingly high frequencies: no detectors can sense the phase of these oscillations.

What we have here is a classic inverse problem. If we know the object wave, calculating its diffraction pattern – the ‘forward calculation’ – is easy: it requires one two-dimensional Fourier transform from the LHS to RHS of Figure 4. But the backward, or inverse, calculation – inferring the object from the recorded data – appears to be profoundly intractable: we can assign any phase at all to each pixel of the diffraction patterns, but how can we select the single set of phase assignments that correspond to the actual phases lost in the experiment?

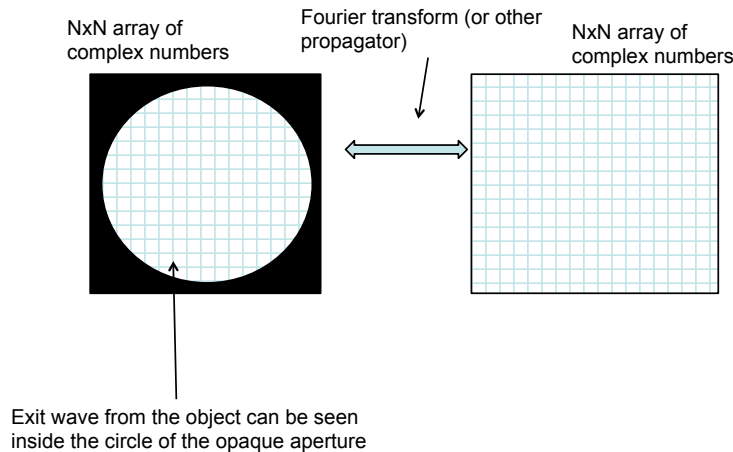


Figure 4: The Fourier relationship between a real-space object delineated by an aperture and its complex-valued Fourier transform lying in the far-field diffraction plane.

3.2 Iterative solution methods

Like all inverse problems, we proceed by applying ‘constraints’, i.e. knowledge from the experiment (the data) and also a priori information, which we know

about the object independently of the measurements we have made. In the two-dimensional image phase problem, the most powerful of these is if we know the object is both two-dimensional and exists wholly within a delineated, finite, area [22, 23]. This area is called the ‘support’ of the object. In practice everything outside this region is either free space or must be blocked off by an aperture, like in our experiment in Figures 3 and 4.

Each pixel in the diffraction pattern corresponds to a single Fourier component in the object exit wave. Changing the phase of that pixel has the effect of shifting laterally the Fourier wave component corresponding to that frequency in the object function. The phase of all the Fourier components in the object must therefore be such that they add up to zero outside the support. One can imagine that if all the phases are correct except one, then it is bound to give an amplitude contribution to the object wave outside the support, because it will not cancel out all the other correct amplitude contributions, also lying outside the support. The set of possible phases is now fantastically reduced, although it is still not obvious that there is only one unique combination of phases that gives rise to the localisation. In fact, it turns out that this single constraint can very often imply a unique object wave solution[23], except for several unavoidable ambiguities such as a shift of the whole object function, or that the object is its complex conjugate centrally inverted.

Even if there is a unique solution, this does not mean that we can construct a solution algorithm that will always find it. A key breakthrough in the generalised 2D image phase problem occurred when Fienup[24, 25] modified a solution strategy originally pioneered by Gerchberg and Saxton [26, 27]. The method is intimately related to the iterative methods used in ptychography, so it is worth explaining it conceptually in some detail. With reference to Figure 5, we set up an iterative computational loop. On the left hand side we have our computational array representing our estimate of the object function, and also our known aperture, which selects part of our object function. On the right hand side we have a computational array representing an estimate of the modulus and phase of our measured data. We also have the measured data itself (its modulus) in an array of identical size.

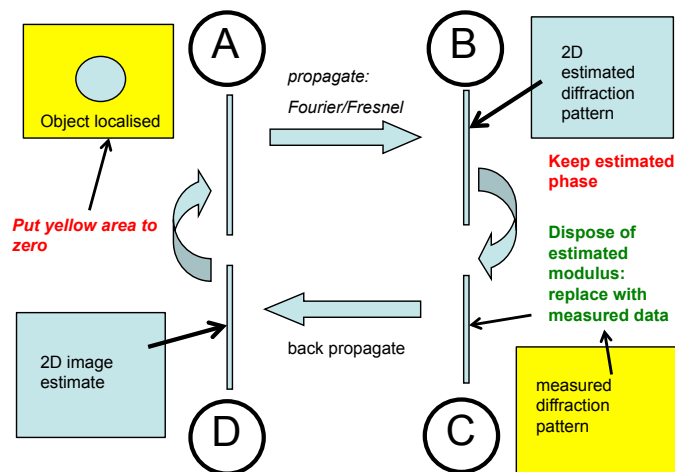


Figure 5: Computation process for projection onto constraint sets. Real space is on the left hand side, where the aperture constraint is applied; reciprocal space on the right, where the modulus of the measured diffraction pattern is applied, but where phase remains untouched.

In Figure 5, A to B and C to D are forward and backward propagation transforms, respectively. These model the relationship between our object plane and our diffraction plane: normally via a Fourier transform or a Fresnel integral of some type. Between B to C, we enforce our knowledge of what we have measured in the detector plane: that in this plane the modulus of the wavefunction must be the square root of the intensity of our data. Between D and A we enforce our aperture constraint: there must be no amplitude lying outside the extent of our aperture. Where we start the iteration is a matter of taste, although if we know anything else about the object (more a priori knowledge) – say that it's likely to be mostly transparent, which would be true for many specimens of interest – then we could start at D with an image field made up of 1s (total transparency) in every image pixel.

Now we apply our aperture constraint, moving from D to A. We put all the values of our object function to zero everywhere outside the aperture (mathematically, this process is called a projection). We computationally propagate the result to B. The calculated estimate of the detector wave at B is complex, but its modulus will (most likely) bear no relationship to the measured modulus at the detector. Moving from B to C we perform another projection, this time applying the constraint of the measured modulus (we replace the modulus that arrived at B with the measured modulus): but we don't touch the phase that came out of B. Now the modulus of the data at C is correct, but the phase is almost certainly wrong.

When we back transform from C to D. The wrong phase from C will almost certainly result in the image having some amplitude over all the field of view, including outside the region of the aperture where we know there should be none. We get rid of this wrong result by simply reapplying the aperture constraint (D to A), forcing all those wrong pixels to zero. And then off we go around the iteration again, perhaps for as many as 10,000 times. Sometimes, but certainly not always, this strategy will converge upon a reasonable estimate of the object. There are dozens of variations on this approach, some of which we will discuss later.

Crucial to what follows is how many data we measure relative to the number of variables we are attempting to solve for. The Fourier transform of the object wave maps one-to-one to the number of pixels in the diffraction plane. Once we have lost the phases of the diffraction pattern pixels, we still need at least enough measurements – more than twice the number of pixels in our object wave – in order to give the two numbers we require for the real and imaginary components of the object wave pixels.

The number of pixels in the detector fixes the size of our calculation box in the object space, which must be able to contain our aperture. Of course, as far as the computer is concerned, the two arrays – in the object space and the detector

space – are only arrays of numbers, which are always of the same dimension. In reality, the detector pitch has a physical size corresponding to the angle subtended from the specimen. This size is inversely proportional to the field of view of our calculation box in the object space, such that, for small angle scattering

$$\Delta\theta = \frac{\lambda}{D}, \quad (1)$$

where $\Delta\theta$ is the angular dimension of a detector pixel, D is the field of view in the object space, and λ is the wavelength of our radiation.

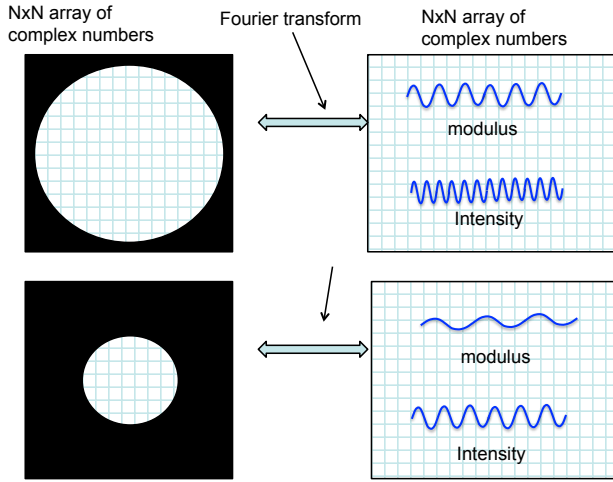


Figure 6: Intensity components of the Fourier transform of an object have half the periodicity of amplitude components. When the aperture fills the real-space object estimate (top left), its Fourier transform (top right) is undersampled by a factor of two. Halving the size of the aperture (bottom) means that now the intensity of the diffraction pattern can be sampled adequately, at the Nyquist periodicity in reciprocal space.

With reference to Figure 6, the upper sine waves in both detector arrays represent a modulus component of the highest frequency that can occur in the diffraction pattern determined by the corresponding physical width of the aperture in the object plane. The lower sine waves are the intensities of these modulus components, which have twice the periodicity of the underlying modulus (the periodicity of \sin^2 , say, is twice that of \sin). Clearly the sampling condition of the *intensity* is not the same as the sampling in our original *complex function*. For the same sized object (or in our case aperture), the Fourier transform array of intensity will be under-sampled by a factor of two.

If we want to measure intensity properly, we have a choice. We can buy a new detector with four times as many pixels ($2N \times 2N$) in order to fulfil the Nyquist sampling in the detector plane, or stick with the same detector and make the (physical) diameter of the aperture less than half the width of the calculation box. Let's do the latter, as shown in Figure 6. We have now halved the periodicity of all intensity components in the diffraction pattern so that the detector pixel size can indeed record all the information in it. (We could also put

the detector twice as far away from the object, but the resolution of our object pixel will then worsen by a factor of two.)

The requirement for the object aperture to be half the lateral size of the calculation box in real space means that that the majority of the unknowns in real space – the empty pixels – are in fact known: they are all zero. There are now more than enough numbers measured in the diffraction pattern to solve for the real and imaginary parts of the object within the smaller aperture area. Note that making the aperture smaller still will not give us any more information because this has the effect of sampling the diffracted intensity at more than the Nyquist frequency. By definition, the Nyquist condition has already captured all the information there is; a qualification is that higher sampling may help if the modulation transfer function (MTF) of the detector falls off quickly.

3.3 Ptychography: multiple diffraction patterns

Now comes the trick. Our detector fixes the size of the calculation box surrounding our aperture function. But there is nothing to stop us declaring an indefinitely large array in our computer in order to describe a much bigger object. We match the pixel size of this large array with that of our detector-defined calculation box in the object plane. We move the aperture from one position to the next over the object (or move the object with respect to the aperture) as shown in Figure 7, and collect a diffraction pattern from each aperture position. We now run our iterative loop in Figure 5 on each of these areas, one after the other.

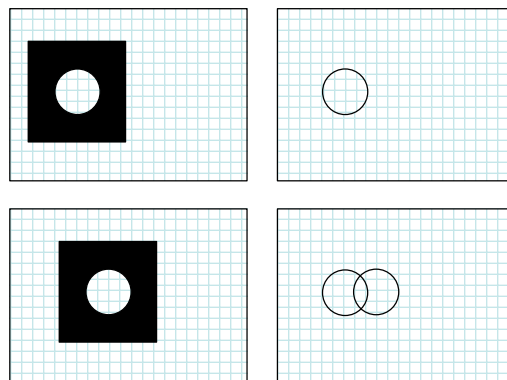


Figure 7: If we move the aperture over a larger field of view, we can collect a diffraction pattern from each aperture position. The constraints are still performed as before (Figure 4), but an on-going estimate is maintained over the whole field of view (right hand side). An estimate of the object function from the first aperture position is fed into the object estimate for the second position within the area of overlap.

The first published example of such a calculation is shown in Figure 8[28]. Four calculations are run simultaneously, but the areas covered by any one aperture do not overlap with any of the others: the calculations are completely independent of one another, and show some of the usual ambiguities inherent to the phase problem. Most noticeably, the cormorant in the phase part of the

image has a centro-symmetric inversion also present, with its phase reversed. This is the complex conjugate ambiguity arising from the fact that the Fourier intensity is the same for both these centro-symmetric functions; these two solutions ‘fight’ with one another because they are both equally valid given the recorded data.

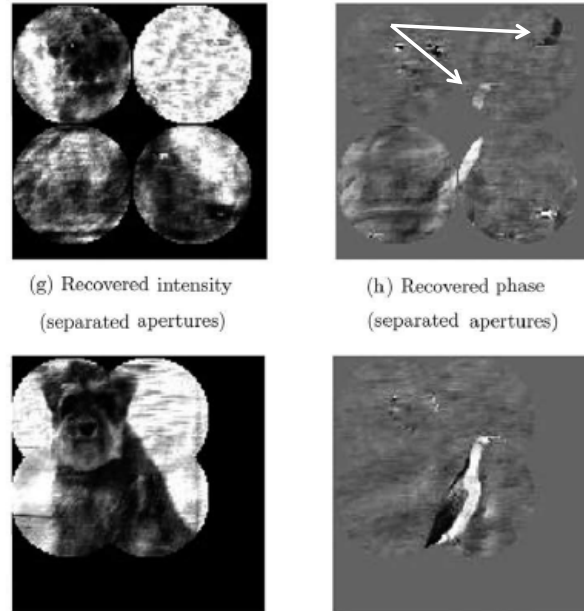


Figure 8: When separate reconstructions are undertaken (via the method in Figure 4), each using a single diffraction pattern from four entirely different areas of an object (top), the usual ambiguities of the phase problem arise. Images on the left are the modulus, and on the right are the phase, of the reconstruction. In the top right, the cormorant appears twice, one reflected and opposite phase (i.e. its complex conjugate). When the four calculations are undertaken simultaneously with overlapping areas (bottom) constrained to be identical, the reconstruction loses the ambiguities. Taken from [28].

Now consider the lower half of Figure 8. There is one continuously updated estimate of the whole area of the object. At each iteration, a circle of the object corresponding to one of the aperture areas is removed and imbedded a separate aperture box, as with the single diffraction pattern iteration. Once the object has been updated (a whole cycle from A to A in Figure 5), this circular area is replaced from where it was removed. Now when we begin our iteration for the adjacent aperture area, we already have a first estimate of the object in the overlap area. This information is fed into the second iterative loop, thus forcing the object solution to be consistent with all the diffraction patterns.

We see that the overlapping update of the object very quickly delivers a much better reconstruction than when we were processing each aperture area independently. The picture appears after just a few iterations. Ambiguities are destroyed. Centro-symmetric ambiguities cannot exist in adjacent aperture areas: there has to be only one value for both functions in the area of overlap, so the ambiguous polarity of both object estimates is forced to resolve itself. This is the power of ptychography. The degree of overlap between these simple aperture functions can be really very small, yet still the solution is forced to be

unique. Ptychography provides a new prior: knowledge of the illumination positions, or at least their relative positions. It also provides more measurements than unknowns because some of the unknowns (object pixels) are expressed in more than one diffraction pattern. The subset of object functions that are consistent with two diffraction patterns - and with the exact known illumination positions and their precise area of overlap, where the object wave must be identical for both diffraction patterns - is drastically reduced.

Hoppe's original formulations of ptychography reached a similar conclusion, although by a rather different route. He thought about the solution strategy in reciprocal space, in terms of interfering diffraction beams [2, 5]. Sampling intensity between any two spots makes it possible to estimate their relative phase within an ambiguity of a complex conjugate. Changing that interference condition, also by shifting the illumination to a new position, can obtain a second estimate of relative phase in order to resolve the ambiguity. The real space picture shown in Figure 8 is probably much easier to understand.

So ptychography can solve the phase problem easily because it folds together information from more than one diffraction (or scattering) pattern. Remember, the support constraint cCCD problem is generally soluble with just one diffraction pattern, except for a few ambiguities; a little extra information from the illumination overlap constraint is a disproportionately powerful way to remove these ambiguities and improve the likelihood of finding a correct and unique solution.

Anything more than this – any extra information in our data over and above the need to solve the phase problem – can now be used for all sorts of different things. In section 4 we discuss how it can be used to account for experimental errors and unknowns. Sections 5 and 7 will describe other uses for diversity: multi-slice volumetric imaging and multi-modal decomposition of incoherent states in the illumination and/or object or detector. Describing ptychography as a solution of the phase problem is perhaps therefore an understatement. Yes, it solves the phase problem, but that is only the first step, and a tiny first step, of what it can achieve.

3.4 An example ptychographic algorithm: update for a spatially soft illumination function

Unlike cCDI, real-space ptychography rarely has a sharp support function. Having an aperture right up against an object is impractical. (Although Fourier ptychography and SAP do indeed employ sharp apertures.) In real-space ptychography, the illumination is not sharply defined, but is 'soft' in the sense of an extended, slowly decaying or ringing amplitude, like an Airy disc or a wave propagated from an aperture to the object, which gives rise to Fresnel fringes. In this section we discuss how an iterative reconstruction can cope with this type of soft illumination.

When we do the reconstruction for a 'top hat' sharply defined illumination function, we can cut out the current estimate of the object and put it into a

separate calculation box. After going around our iterative loop (Figure 5), we then ‘paste’ the new function back into the image from where it came. Of course, we only paste the area defined by the probe, not the whole calculation box function, most of which will contain zero amplitude. We do not touch the area of the object that was not illuminated at this probe position. The whole process is called ‘the object update’.

When we have a soft-edged illumination function, the update has to be subtler. Now we have to copy a box in the object that is big enough to contain most of the illumination. We multiple this copy of the small area of the object by the probe function (D to A) to get the exit wave function, ψ_e . Then we go round the iterative loop. What comes out of C to D, is a new estimate of the exit wave, which we can call a corrected version of ψ_e , namely ψ_c . It is corrected because the experimental data has been fed into the loop (B to C). ψ_c will usually look substantially like ψ_e , certainly after the iteration has run over all the probe positions many times.

However, unlike the sharp aperture, we cannot just cut out a part of this function and paste it back into the image estimate, because it is unevenly modulated by the probe amplitude. Instead we use the new estimate of the soft exit wave to alter, but not replace, the existing running estimate of the object. For example, there may be points within the illumination function (say the rings of an Airy disc function) that are zero. No photons or electrons went through those pixels of the object, so it is unreasonable to change our estimate at those pixels based on whatever we measured in the diffraction plane at that probe position: we just leave them alone. Conversely, areas that were strongly illuminated by the probe scattered most information into the diffraction pattern, so it makes sense to weight the alterations we make in the object estimate more heavily in those areas, and less so in weakly illuminated areas.

How can we do this in a consistent reliable way for a complicated probe? We can develop a heuristic algorithm as follows [18]. A more formal treatment can show that this update approximates to Newton’s method [29]: it is a very efficient and effective search algorithm, although many more complicated, but computationally more intensive algorithms, can improve upon it.

The two-dimensional exit wave is given by

$$\psi_e = a \cdot q, \quad (2)$$

where a is a two dimensional illumination function and q is a small area of our two-dimensional object function, located around the probe position. For brevity, we do not include the x, y coordinates of the functions. If these were 2D arrays in MATLAB, for example, the multiplication would be pixel by pixel, coded as

Exitwave=Illumination.*Specimen; [COPY EDITOR – (SIC), including ‘;’]

All the arrays have the same pixel size, but the size of the box our probe is imbedded within is usually much smaller than the total object size.

We go round the right hand side of our iterative loop, A to B to C to D, thus applying the detector projection constraint. The back propagation C to D gives us a new exit wave, which also corresponds to a new estimate of the object function, such that

$$\psi_{NEW} = a \cdot q_{NEW}. \quad (3)$$

We want to alter q in the light of ψ_{NEW} , to give a better estimate of it, q_{NEW} . ψ_{NEW} should be an improved estimate on ψ_e because we have injected known experimental data during the detector projection constraint. Subtracting the equations and rearranging, we have

$$q_{NEW} = q + \frac{1}{a}(\psi_{NEW} - \psi_e). \quad (4)$$

The trouble with this equation is that when a is small or zero – which it certainly will be in places if it was something like an Airy disc – the second term will tend to infinity. A common way of dealing with this is via a Wiener filter. If we multiply top and bottom by the conjugate of a , a^* , the denominator is then real, so we can add a small real number, ε , to avoid this catastrophe, giving

$$q_{NEW} = q + \frac{a^*}{(|a|^2 + \varepsilon)}(\psi_{NEW} - \psi_e). \quad (5)$$

However, we are still giving the same credence to the change we are going to make to q at any point spanned by the illumination. It would seem logical to change it most where the amplitude of a is large, as we postulated above. The simplest scheme is to multiply the second term by the magnitude of a , scaled so that its maximum is unity. That is to say we put

$$q_{NEW} = q + \frac{|a|}{|a|_{MAX}} \frac{a^*}{(|a|^2 + \varepsilon)}(\psi_{NEW} - \psi_e), \quad (6)$$

where $|a|_{MAX}$ is a single number which is the value of the maximum modulus of the probe. All the other terms are 2D functions, with the subtraction, multiplication and addition being pixel by pixel. Now we are completely changing the object with the new estimate at the point where the probe has maximum modulus, and all other points are only being changed in proportion to the modulus of the probe incident at that point. Points not illuminated are not changed at all. A little thought will show that when a is the sharp aperture we first described, this update has an identical effect as the cut-out-and-paste strategy. When the solution is correct, the object is not altered: an elementary requirement of any search algorithm.

Once the update has been applied at one probe position, it must be applied at all other probe positions spanning the desired field of view, continuously updating the same object function. The whole process is repeated, perhaps 50 times – i.e. 5,000 updates for a 10x10 array of probe positions – always refining the same estimate of the object. The algorithm is called the ptychographical iterative

engine (PIE) [17]; a name that also playfully teases an eminent scientist who, in the early 1990s, described ptychography as ‘Pie in the sky’. It can be altered in all sorts of ways by introducing various constants or raising the scaling factor to some power. We will discuss these changes further in Section 9.

3.5 A survey of ptychographic algorithms

Since the arrival of the PIE algorithm in 2004 [18], a growing list of alternatives have been demonstrated, such that today (early 2017) it is a difficult task to keep abreast of all the developments; to ease the burden somewhat, this section provides a brief historical survey.

Our survey will split ptychographic algorithms into two kinds. Class 1 are those that invert the standard ptychographic data set, where the illumination is coherent, no account is taken of noise, the specimen shifts are accurately known and the multiplicative approximation is satisfied. (Apart from the original PIE, all of the algorithms in this category also solve for the probe.) Class 2 are those algorithms that loosen one or other of the standard assumptions – for example by accommodating partial coherence or allowing for thick (non-multiplicative) probe/specimen interactions.

The first algorithm to appear in class 1, after PIE, was the conjugate gradient approach suggested by Manuel Guizar-Sicairos [30]. This was also the first algorithm to solve for the probe and the first to employ a global, rather than a step-by-step approach to ptychographic reconstruction. Figure 9 explains this important distinction; most class 1 algorithms adopt the global update strategy, since in this way the many well-tested non-linear optimisation routines are readily adapted to the ptychographic problem.

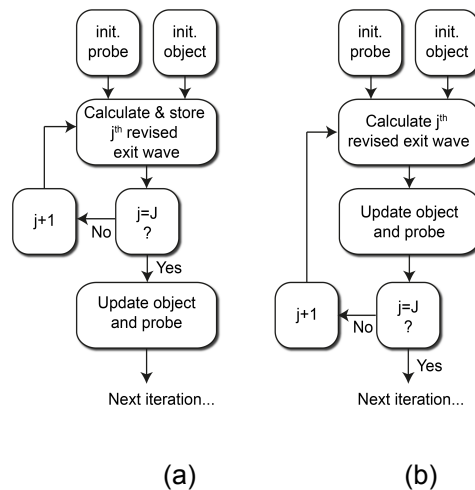


Figure 9: There are two strategies that iterative algorithms take to recover an image from a ptychographic data set. In (a), a whole collection of updated exit waves are calculated in parallel, one for each of the diffraction patterns in the data set. This collection is then used to perform one batch update of the probe and the object. Popular algorithms such as the Difference Map and Conjugate Gradient method take this approach. In (b), updated exit waves are calculated serially, one-by-one, with each

update being fed into a corresponding update to the object and probe. This is the tack taken by the 'PIE' family of algorithms.

Next, a key paper by Thibault and colleagues in 2008 [12] conclusively demonstrated the power of simultaneously solving for the probe. Thibault's team repeated the original X-ray experiment by Rodenburg et al. by imaging a zone plate using hard X-rays, but used the probe-solving ability of the Difference Map (DM) algorithm [31] to realise a significant improvement in image quality and resolution over the earlier work.

The Authors' ePIE algorithm [32], published in 2009, extended the PIE scheme to solve for the probe. Optical bench experiments were used in the original paper, but shortly after, Schropp and colleagues [33] used ePIE in the X-ray regime to characterise the X-ray beam's focus (perhaps the first important real-world application of ptychography), and in the same year (2010), ePIE was shown to work with electrons [34]. Along with DM, ePIE has become the most widely used reconstruction method, so Section 9 will look in detail at the mechanics of these algorithms, and how they are coded.

The ptychographic inversion problem lends itself well to a variety of non-linear optimisation strategies, as Marchesini and colleagues showed in a wide-ranging survey in 2010 [35]. The survey covered conjugate gradient and Newton-type 2nd order optimisation, as well as set projection approaches, in particular the Relaxed Average Alternating Reflections (RAAR) method popular amongst the cCDI community. The survey paper began a series of studies by Marchesini's group, which continued with papers on alternating direction minimisation [36] and the idea of phase synchronisation to accelerate algorithm convergence [37], as well as 'class 2' algorithms to combat diffraction pattern noise. RAAR itself has gone on to form the basis of the SHARP ptychography system at the Advanced Light Source, where ptychographic images can now be obtained in close to real time [38].

Almost all of the work on ptychography up until the start of 2014 concerned X-ray microscopy. However, around this time the emergence of Fourier ptychography and further demonstrations of electron ptychography began to broaden the appeal of the technique, and so spurred further interest in new algorithms. One example was the 'GPILRUFT' scheme used to reconstruct atomic scale images of cerium dioxide at Oxford [39]. GPILRUFT tackled the reconstruction by linearising the inversion problem, and so was the first to go some way toward provable convergence results, although the significant practical difficulties with electron ptychography that the Authors faced seemed to outweigh any benefits from the new algorithm. Fourier ptychography (FP) used ePIE at the outset [40], but the very different nature of the data in FP – combined with the fresh eyes of newly-interested research groups – quickly resulted in alternatives; rather than give a full run down here, the Reader is directed to a comprehensive review by Yeh et al., in particular for the comparison there between the step-by-step and global approaches [29].

The most recent work on class 1 algorithms, at least that the Authors are aware of, come from two papers. A 2015 paper by Hesse [41]– working with D. Russell Luke, inventor of RAAR – presented the ‘PHeBIE’ proximal gradient algorithm, together with a welcome rigorous look at the convergence properties of ePIE and DM. This year (2017), a paper by one of the Authors [Maiden, Optica, in press] re-examined and improved ePIE, with changes to the probe and object update steps (see Section 9) and the introduction of ‘momentum’, an idea borrowed from the machine learning community.

Of the algorithms in class 2 (those allowing a relaxation of the assumptions in the standard ptychographic model), most have attempted to deal with noisy data, and most of these have assumed that noise arises from counting statistics and so is governed by the Poisson distribution (see Section 4.7). Quite early on, Thibault and Guizar-Sicairos took this tack with their maximum likelihood algorithm [42]; since then, ePIE has been adapted to accommodate Poisson noise [43], and a variety of schemes have been used for FP to the same end [29, 44]. Another major source of noise, camera readout, was combatted by Marchesini by adapting the Fourier constraint in RAAR [45], and by the Authors with an adaptation of ePIE in the electron [46] and optical [47] regimes.

Arguably, the most important class 2 advance came with the advent of mixed-state ptychography [48] (see Section 8). The mixed state forward model can quite readily be applied to any of the conventional algorithms. Apart from dealing with partial coherence in the X-ray [48], electron [49] or optical [50] regimes, one or other mixed-state algorithm has since been employed to deblur diffraction patterns in ‘fly-scan’ ptychography, where the probe rapidly scans across the specimen without stopping [51]; multi-wavelength ptychography [52]; ptychography with a vibrating specimen [53]; and in the previously mentioned ‘probe relaxation’ algorithm to handle a probe that fluctuates during the experiment [54].

Another popular grouping of class 2 algorithms correct errors in measurement of specimen translations (see Section 4.4). That this is possible was first shown by Manuel Guizar-Sicairos in his early conjugate gradient paper [30]. Later an annealing algorithm that randomly agitated the measured specimen positions during the reconstruction showed that position correction could be effective in optical and electron ptychography [55], and a cross-correlation based add-on to ePIE gave excellent results in the X-ray regime [56]. A refined conjugate gradient search also solved the position error problem effectively [57].

Last in our survey are the class 2 algorithms that relax the thin (or multiplicative) specimen assumption. Two approaches have been reported: multi-slice ptychography (see Section 6.2 and [58]) and diffraction tomographic ptychography [59]. This is an exciting area for further research, although the hugely enlarged object space for volumetric imaging makes the reconstruction task immensely more demanding.

The following Sections provide further details of the myriad ways in which ptychography can be implemented, improved and expanded; we will revisit ptychographic algorithms in more detail in Section 9.

4) Sampling and removal of artefacts in images

We are going to use a very old result to illustrate some important concepts about information content in ptychography. Figure 10 was published by Bunk et al [60], almost immediately after the first experimental demonstrations of visible light and X-ray iterative phase-retrieval ptychography[11, 17]. It used the PIE reconstruction algorithm described in the last section.

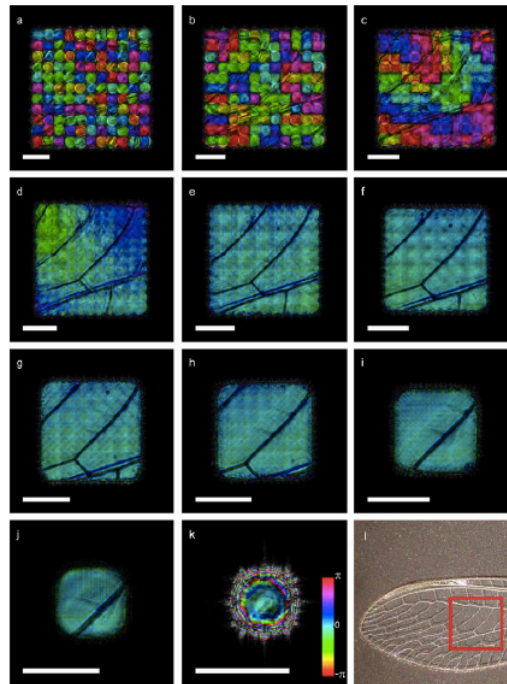


Figure 10: An early very example of a visible-light ptychographic reconstruction obtained using iterative solution methods, collected in the simple aperture configuration (section 5.10), illustrating the improvement in the reconstruction as the degree of overlap between adjacent illumination positions increases. Taken from [60]. Artefacts can now be removed by various strategies (section 4.8).

If you are completely new to ptychography, you might be disappointed that these reconstructions seem to be full of artefacts, especially in view of what has been said in the previous sections. The first thing to stress is that their lack of quality is absolutely nothing to do with the capabilities of the authors of the paper. When these results were published, they were cutting edge, and certainly no worse than the first proof-of-principles [refs]. But at that time really the only thing that was known about ptychography was that it could solve the phase problem for an indefinite field of view, as discussed in the previous section. All the many developments that have taken place since then mean that now very high resolution, artefact-free reconstructions can be obtained with almost total reliability. For example, see Figure 11, where a modern visible light ptychograph is compared with traditional contrast techniques. However, we start our

experimental narrative here because (a) the work in Figure 10 represents the first experimental exploration into the effect of ‘extra’ information in ptychography (beyond the solution of the phase problem), and (b) because we think it will be useful to illustrate to any new-comer to the field what sort of things can go wrong if you don’t know the tricks of the trade. We will re-assess the artefacts in these pictures in section 4.8.

In this section we are going to consider the width and nature of the communication channels illustrated in Figure 1d in the context of ptychography. First we consider sampling of our data. A great emphasis in the early days of X-ray cCDI was on the sampling condition in the diffraction plane, which was called over-sampling [61]. In ptychography the intensities measured at the pixels in the detector change as we scan the illumination or object. If we move the illumination in very small steps, the changes are small and incremental: changes are much larger for large step sizes. A more general view of sampling in ptychography is therefore to examine not only the sampling in the diffraction pattern in reciprocal space, but also the sampling in real space: the grid over which we scan the illumination. We need to consider the sampling over a four-dimensional cube made up of 2D diffraction patterns collected from an array of 2D probe positions. (Or, in the case of Fourier ptychography, the sampling of the illumination beams in angle space and the pixel sampling in the image plane.)

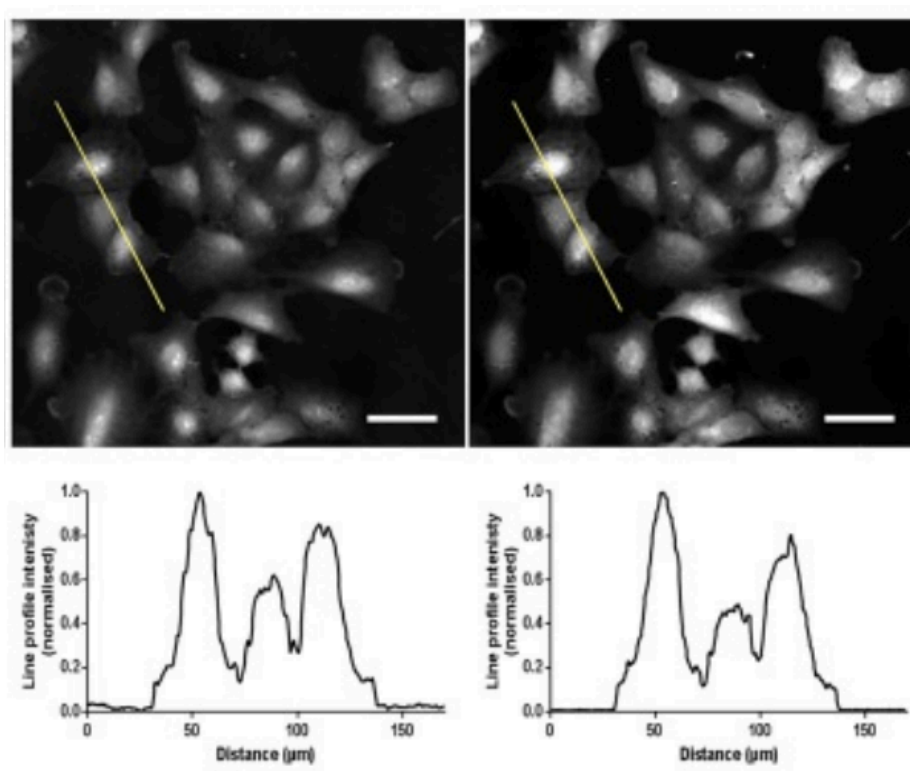


Figure 11: Example of a modern visible-light ptychograph of cells (left) – compare Figure 10. The right hand image is the conventional fluorescence image. Ptychography does not need fluorescence signals, so the cellular structure can be imaged directly without affecting the cells in any way, e.g. for screening live embryos. Taken from [62].

There are two ways of thinking about sampling in real space. One way is simply to state the periodicity of the probe movements in real space. More commonly, workers in the field often talk about the 'overlap parameter'. This is the ratio of step size through which the illumination is moved in relation to the width of the illumination. Because the illumination is invariably roughly circular, both these definitions are imprecise: the step size has to be about 30% before all the gaps between circular areas have been covered just once.

Let's look at Figure 8 in detail. In this visible light optical experiment, an 11x11 grid of illumination positions is fed into the PIE serial iterative reconstruction method [18]. In the first frame, the probes do not overlap with one another, at least as defined by the aperture diameter. In subsequent frames the probes are made larger so that the overlap between them increases in steps of 10% up to 100%. In fact, some structure comes out of the 0% overlap data set because PIE can account for diffraction effects caused by the small propagation distance from the aperture to the sample, thus allowing some information to 'seep' between probes. Clearly, 100% overlap contains no ptychographic (probe shift) information whatsoever. The result is worse than an error reduction support constraint algorithm because the PIE constraint in real space has soft edges arising from the broadening of the probe.

Clearly, as the overlap increases, the quality of the reconstruction becomes better and better, at least until the 100% overlap catastrophe. Is this what we expect? Because of the geometry of the gaps between circular apertures, as discussed above, the overlap must be at least 30% before every pixel of the object is illuminated even once. This accounts for the sudden jump in the quality of the image between Figures 10c and 10d. As the overlap increases further, we are making more and more measurements for a smaller and smaller field of view of the specimen: the ratio of measured data points to unknowns is increasing. Remember, a single diffraction pattern contains enough numbers to solve for an isolated object (any one of these illuminated areas). Once we have enough overlap to suppress the few ambiguous solutions that can arise in cCDI, it is not obvious why having any further extra data – often called 'redundant data' – should necessarily make the reconstruction better. We will find that a key application of this redundant ptychographic data is to suppress the artefacts present in these early results. (Of course, if redundant data is employed usefully, the word 'redundant' becomes a misnomer.)

A key requirement for cCDI is that the sampling in the diffraction plane must become smaller (more dense) as the size of the object increases. This follows from a simple analysis of the scattering geometry – that beams scattered from the edges of the object will become out of phase more quickly as a function of scattering angle if the size of the object is large: i.e. the detector pixels lying in angle space must be smaller to pick up all the relevant interference information. As we have seen, when we measure intensity in the far-field, the calculation box over which we solve for the object must have dimensions of roughly twice the size of the object itself. One might suppose that this same condition must hold in ptychography. Indeed, most ptychographic reconstructions are undertaken with the probe imbedded in a similar calculation box.

Surprisingly, the minimum sampling condition for ptychography is not constrained by the probe size. Rather than think of overlap as a measure of redundancy, it is more informative to think of the probe movement defining a grid of real-space sampling. The fundamental minimum sampling condition in ptychography must take into account both real- and reciprocal-space sampling. Strangely the size of probe is independent of the sampling requirement, quite unlike in conventional cCDI, provided that for a given real-space sampling the probe is big enough so that adjacent illumination areas overlap somewhat and span the entire field of view.

If we simplify the illumination shape as a square, so that we do not have to handle the awkward geometry of overlapping circles, it can be shown using simple physical arguments [63] that the minimum ptychographic sampling condition is

$$\Delta R = \frac{1}{2\Delta u} \quad (7)$$

where ΔR is the sampling interval in real space and Δu is the sampling interval in reciprocal space. The same conclusion can be reached by a more formal derivation [64].

We see that we can exchange sampling between real and reciprocal space as we wish: it is as if we have a dial that can, in a continuous manner, reduce sampling in one plane and increase it in the other, while still preserving the necessary quantity of information to reconstruct the specimen. If the probe is large, but the sampling in real space is very fine, this formula implies that the pixel size in the detector can be large, even though the structure of the diffraction pattern is very fine (a large probe in real space implies small features in reciprocal space). This is quite contrary to anything that follows from cCDI. However, it transpires that we can recover unmeasured small pixels (that do satisfy the conventional diffraction sampling condition), from the large pixel data – see section 8.5. Very dense sampling in real space is normally associated with a very small probe (see Sections 5.1 and 10 below), so that features in the diffraction plane are anyway very large and can be captured by only a few large detector pixels. This type of data, although subject to the same sampling condition, is better processed by non-iterative means (see Section 10.4)

It should be emphasised that the fundamental sampling condition relates only to Fourier domain ptychography where the scan is over an infinite field of view, and where we know the probe function. It is also the minimum sampling required to solve the phase problem. In any practical ptychography experiment the sampling in diffraction space is high and there is considerable probe overlap in real space. So we generally have much more information than we need. Now we discuss the things we can do with these extra data in order to improve image quality.

4.1) Probe recovery:

One of the most important breakthroughs in iterative phase retrieval ptychography was to discover that it is possible to solve for both the object function and the illumination function [12, 65]. The two functions express themselves equivalently in the mathematics, so perhaps this is not quite so surprising. It was known some time ago that the WDD method (section 10) could be used to solve for both object and illumination, but experimental tests on the optical bench were not particularly convincing [66]. On the contrary, iterative methods to retrieve the probe work very well. The two most popular algorithms for this simultaneous recovery involve either projections over the whole data set at once (DM) or a serial update process (ePIE), which were briefly introduced in Section 3.4 and will be discussed in detail in section 9.

An immediate unintended consequence of this development was that workers in the X-ray field began to use ptychography not to make images of an object, but solely to characterise and reconstruct the illumination function. There are now many examples in the literature. Because the full complex field is recovered, it can be back-propagated to the lens aperture, thus elegantly displaying any phase aberrations in the optics. This is enormously more informative than a simple resolution test, say by scanning the focus of the beam across a knife edge. In the particular example shown in Figure 12, [67], the probe calculation from a refractive aberrated optic has been used to make a perturbing phase plate that corrects for the aberrations. This is an example of how ptychography can enhance the technology of its lens-based imaging cousins in order to improve the very fine probes used for analytical STX/EM. Similarly, Figure 13 shows a cross-section through an electron probe recovered from ptychographic data in the scanning electron microscope [68]. The explicit map of the complex wavefield of the probe in both of these examples is not available by any other means.

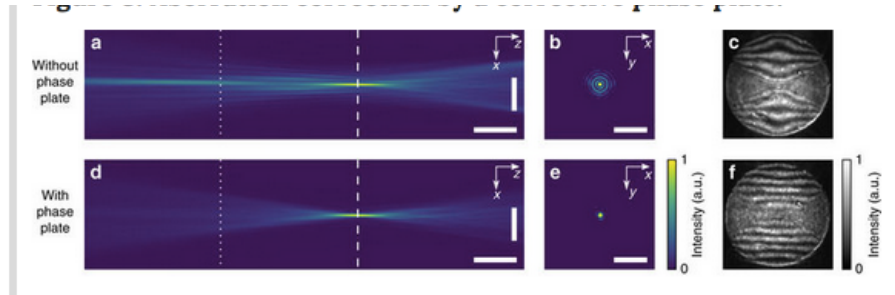


Figure 12: An example of a cross-section through a focused X-ray beam, calculated via ptychography, taken from [67]. The beam is calculated at one level of defocus where the object is positioned, and then propagated computationally to produce the cross-section. The optics are imperfect (top), generating a large crossover. In the lower picture, the optics have been corrected by inverting the aberrations in the lens measured from the top cross-over. The inference fringes of the right (caused by a diffraction grating in the beam) are like those in Figure 2: when they are straight, there is only defocus present and no higher-order aberrations.

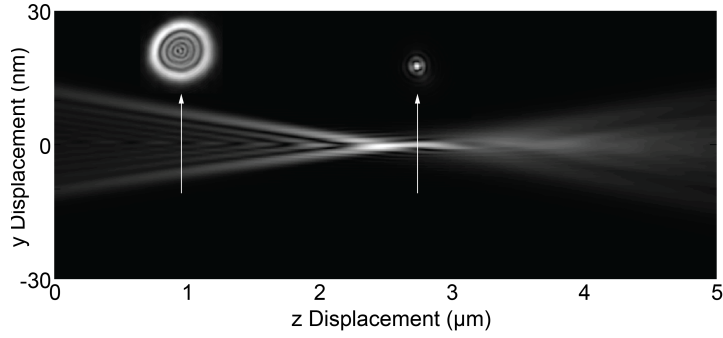


Figure 13: As Figure 13, but for an electron probe in a scanning electron microscope (SEM), taken from [68].

No matter how well the optics within a ptychography experiment are calibrated, all reconstructions nowadays solve for both the object and the probe. Of course, for a set up that remains constant from one day to the next, it is logical to start the reconstruction with the last known probe solution.

An interesting and possibly very important development in probe recovery has recently been demonstrate by Odstrcil [54]. In the context of EUV ptychography, experimental constraints dictate that every single probe is different and unknown. We might suppose that absolutely no progress could be made in such a situation. The whole technique of ptychography depends on the probe and the object remaining constant. The premise of his reconstruction technique is that though all the probes are different, each probe can be described as a sum of a few (5-10) fundamental probes, all of which are orthogonal to one another. There are still innumerable possible probes, but each one is described by a few numbers, instead of the 1,000s of pixels needed to describe a completely general probe.

A little thought will suggest that this is a very reasonable assumption. After all, the optical components remain same. In this case, each shot for the EUV source has a different structure, but each probe will be perturbed by a set of possible variables that can change in the experiment, and these variables may be rather few. If each probe were completely different from every other, we would indeed have an impossible problem.

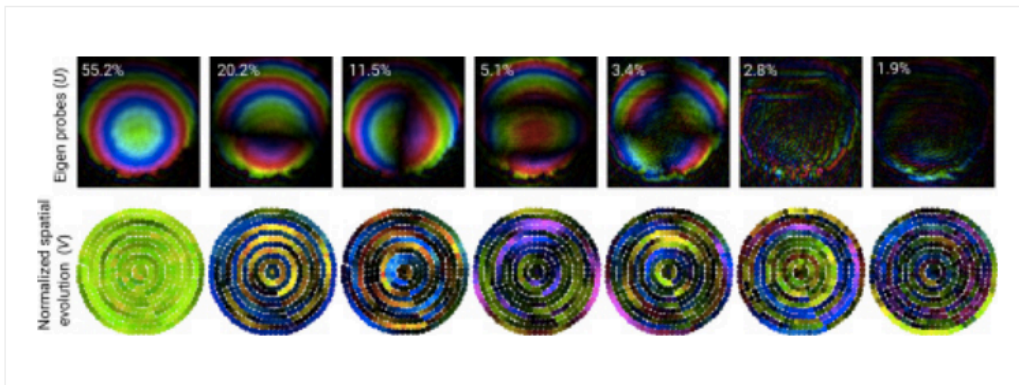


Figure 14: A set of orthogonal probe functions that can be used to compose a probe function that varies from one position to the next. See text for more details: taken from [54].

If all the real probes are already known, finding the underlying fundamental probes can be done by the standard techniques of principal component analysis. But at the start of the reconstruction, they are not known. The reconstruction starts by assuming all the probes are the same and does a normal reconstruction. The resulting probe and object are very poor approximations of their real counterparts. However, the object function can be used in the iterative update to make a new estimate of each of the probes for all the positions. These are now used to find principal components. They are not the actual principal components because the first estimate of all the probes is bad. Furthermore, none of the wrong probes can be fully described by the small number of wrong principal components. The updated probes are projected onto the first estimate of the principal components, thus making a new set of probes that are now just described by the first estimate of the principle components. These new probes are used to update the object. Then they are updated themselves. The second iteration of probes creates a second iteration of principle components, and so on and so forth. The algorithm convergences on the actual principal components and hence the actual probes. Of course, because each probe is only described by a handful of numbers, we only need a fraction of the diversity in the ptychographical data set to solve for them all. Some example results are shown in Figure 14.

4.2) Some pathological instances where Ptychography struggles

We have said that ptychography suppresses all of the ambiguities that arise in cCDI. This is not quite true: it does, rarely, suffer from its own special ambiguities. Of course, now that we are solving for two complex functions, object and probe, we can expect that the sampling condition will become twice as demanding. That is true, but other factors must also be taken into account when solving for both functions. The specimen and probe functions can never be completely and unambiguously separated from one another. A simple example is that the probe can increase in amplitude, while the specimen reduces in amplitude (appears more opaque), but the product of the two maintains the total measured flux on the detector. This is not serious as far as observing the structure of the object, but it needs to be handled carefully if quantitative absorption data is required, say by calibrating the total flux in the probe in an area of free space around the object. During the reconstruction the probe can be periodically propagated to the detector plane (without the influence of the object) where it can be constrained by the correct free-space intensity. Indeed, it is always advisable to scale the first estimate of the probe by the integrated intensity in the detector plane. If there is a large disparity between the intensity of the physical probe and the first guess of the estimated probe, many reconstruction algorithms find it very hard to recover. If the edge of the field of view of the reconstruction is very bright or very dark, you have probably made this mistake.

More profound questions arise when we consider the information content of the object and the illumination. To get any diffracted information, the specimen and probe must have structure. If the object structure is sparse, consisting of a very

few simple features separated by large areas of constant phase or modulus, then we might suppose that the probe is very poorly constrained. Consider a largely non-transparent object with only a few empty features. If the probe is scanned with large step size, but over a small field of view, only a few of these object features will intersect with it: there may be a large subset of areas within the probe are never transmitted through to the detector, yet alone solved for by any algorithm!

Even when the object function has a lot of structure, there are certain types of probe which are difficult to solve for, one example being a defocused convergent beam, which we will discuss further in Sections 5.5 and 10.6.6. Another example occurs in visible light ptychography, where it is common to use a fixed of some type to create a probe with complicated phase and modulus structure [69]. Counter intuitively, convergence is poor when both the object and illumination is highly structured. However, once a good estimate of a complicated probe is known (and can be used to seed the reconstruction), then convergence onto the object function is much better than using a probe with little structure. The WDD formulation can be used to suggest probe structures, which are more likely to improve the convergence of the reconstruction (see Section 5.6).

There are various very unusual combinations of object function and illumination structure and/or shift positions where ptychography provides no extra phase information at all. A trivial example is if all the illumination positions are identical so that the overlap between them is perfect (see Figure 10). Obviously all the diffraction patterns are identical and therefore lend no extra diversity. Similarly, if the object is periodic, and the scan of the illumination has the same periodicity (or any factor times the object periodicity), then all the diffraction patterns will also be identical. Certain illumination functions can also cause the obliteration of diversity, for example a convergent beam of finite angular extent when incident on a high-frequency periodic structure can mean that there is no overlap in the diffraction orders in the far-field (see caption of Figure 2), in which case no phase information can be expressed.

If the entire field of view is free space, then clearly we cannot find any sort of sensible solution. If the reconstruction starts with the assumption that the object is free space, and with a known probe function (which is now the only information expressed in the diffraction pattern, but without its phase), then in theory the object function should not depart from free space. We find that in general if a significant area of the field of view has some sort of object structure, then areas that are free space will be reconstructed correctly, although residual errors in the probe reconstruction arising from the limited field of view occupied by the object may express themselves in free space at the probe position locations. The free space problem is clearly a condition for which the conventional microscope is vastly superior: it will show blank free space. Luckily, not many microscopists want to look at free space.

If the object is unknown, and especially if it is likely to be sparse or weakly scattering, it is always better to use a probe that has more structure within it. This can be shown using arguments based on the WDD method [70, 71],

although whether these are directly applicable to iterative methods has yet to be proved. Figure 15 shows an X-ray example of how making the probe (in this case formed by a zone plate lens) much more complex by the introduction of a pinhole improves the reconstruction quality.

4.3) Non-periodic scan

Although we have formulated the sampling condition in terms of a periodic scan over the object, it was realised quite early that periodic scans are not optimal [65]. As we noted in the previous section, ptychography offers no information if the probe is shifted across a periodic object, at the periodicity of that object, because each diffraction pattern is identical and contains no phase information. We can reverse the argument. A probe scanned periodically over a specimen will not contain any ptychographical information for a Fourier component in the object that matches multiples of that periodicity. A periodic scan will always tend to produce image artefacts at that periodicity. However when we solve for both the probe and the object, the problem creates the so-called ‘raster scan pathology’, first pointed out by Thibault et al. [21]. Either the probe or the object can develop structure at the scan periodicity, causing a further source of ambiguity.

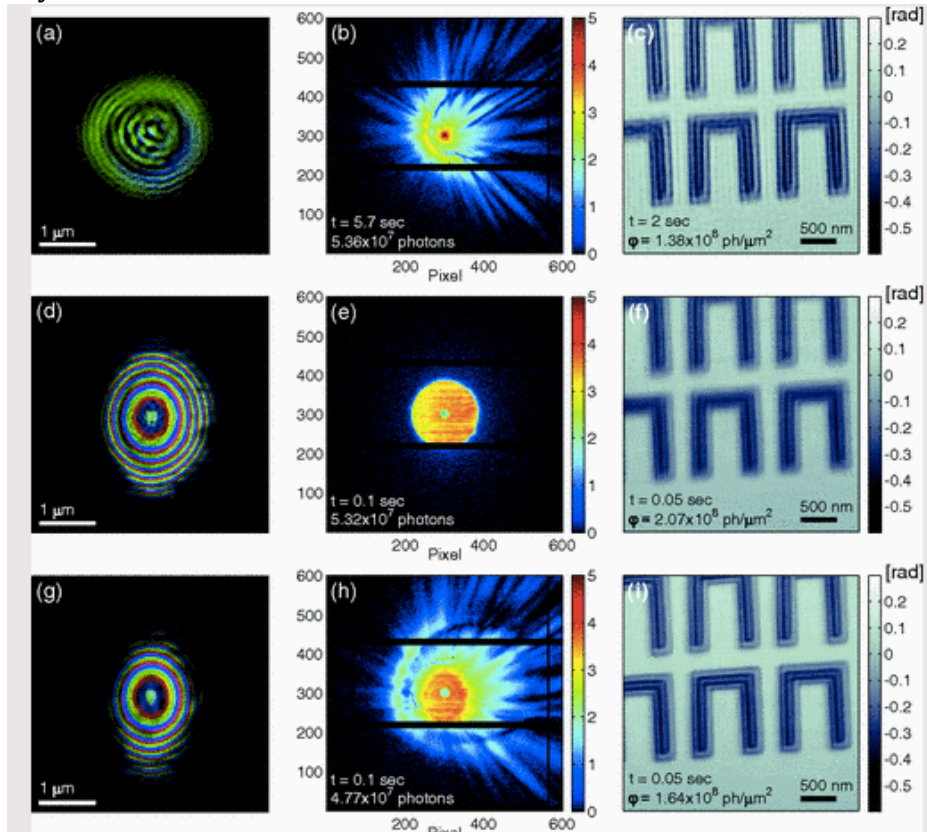


Figure 15: Effect of having a more structured illumination, taken from [70]. (a)-(c) show the real-space probe reconstruction, the diffraction pattern from the probe when there is no specimen present, and an example reconstruction, respectively, for when a simple aperture is used to form the illumination. (d)-(f) as above, but for convergent illumination. (g)-(i) as above, but for convergent probe clipped by an aperture, which clearly extends the probe function in reciprocal space. The quality of the reconstruction

improves from top to bottom. This is shown quantitatively in original publication using Fourier ring correlation [70].

The solution is to deliberately introduce non-periodicity into the scan. One common way of doing this is via a spiral scan, starting from the centre of the field of view [12, 20, 65]: a technique that is now used very widely in the synchrotron X-ray world. Alternatively, a broadly periodic scan can have small random offsets added to each probe position. There are situations where neither of these strategies is easy to implement, for example when a STX/EM configured for smooth rectilinear scans is modified to collect ptychographical data. In fact, if the early iterations in the reconstruction uses computationally perturbed probe positions, then periodic artefacts from a regular scan can be suppressed at the cost of resolution. Final polishing of the solution can then use the real regular probe positions [72].

4.4) Refining probe positions

We can suppose that our knowledge of the probe positions relative to the specimen is the key a priori constraint in ptychography, replacing the real space object support constraint in cCDI, especially when we are solving for both the specimen and the probe. However a densely-sampled data set can allow refinement of the probe positions after the experiment has been completed. This has proved important for electron ptychography (at least when real-space step sizes are large). A STEM scan is designed to be periodic, but when random position offsets are added to these (see previous section) hysteresis in the scan coils does not always move the probe to the assumed positions.

If we know our scan positions but think there might be distortions from specimen drift, stretching, or rotation of the scan, then these can be parametrized using only a few variables, which become a few more variables in our search space. Guizar-Sicairos and Fienup were the first to investigate the search for unknown probe positions using conjugate gradient methods [30], but this is computationally intensive. The two most commonly used techniques have low computational overhead, increasing the cost of the whole reconstruction by only a factor of 3 or so. Both look for perturbations in the position of every scan point one at a time, which reduces the computational overhead, but they use quite different mechanisms: annealing and cross-correlation. In both cases, an initial reconstruction is obtained, making no account of position errors.

In the annealing algorithm [55], each probe position then has a number (say 5) of random offsets applied to it, but only up to a given maximum. Using the existing estimate of the object and probe, a diffraction pattern is calculated from all 5 positions. One of these will most closely match the measured diffraction pattern from that point, i.e. it will have the lowest error metric. This position is now chosen as the 'correct' position, and then the object estimate is updated using that probe position. (Note that we are describing this in terms of a serial update algorithm, like the aperture serial iterative update described in Section 9.1, but it can be incorporated into the parallel methods.) The same process is applied to all probe positions. For the next iteration the new altered probe

positions are the starting point, but once again random offsets are added to these. Thus it is possible for a probe position to wander quite far from its original putative position. However, as the calculation proceeds, the random maximum distance added to the current probe position estimates is slowly reduced. This forces an estimated probe position to settle on a single point, in the meantime gradually stopping it from jumping large distances from a good quality estimate. Figure 16 shows the improvement that can be obtained using the method, in this case for electron ptychography data. The effects of a drastic period of drift, starting half way through the experiment, are entirely removed.

The correlation algorithm [56] starts by storing a copy of the current estimate of the object function. It then updates the original object function at just one probe position. The current (pre-updated) estimate of the object has previously been reconstructed in this particular probe area using lots of diffraction patterns from all the overlaps occurring within it. In calculating the next update, this ‘good’ image (averaged from lots of data) is fed into the reconstruction algorithm at A (figure 4), where the exit wave estimate is generated. It is the detector update, B to C, that impresses the wrong position information for this probe position, but most the original image data (determined mostly by the phase of the diffraction pattern, which is not changed) will survive and still be present in the new estimate of the exit wave at D. In other words, the updated object function should look like the previous object estimate, but with the newly updated area being a copy of the image shifted to the wrong position. This is true at least to first approximation.

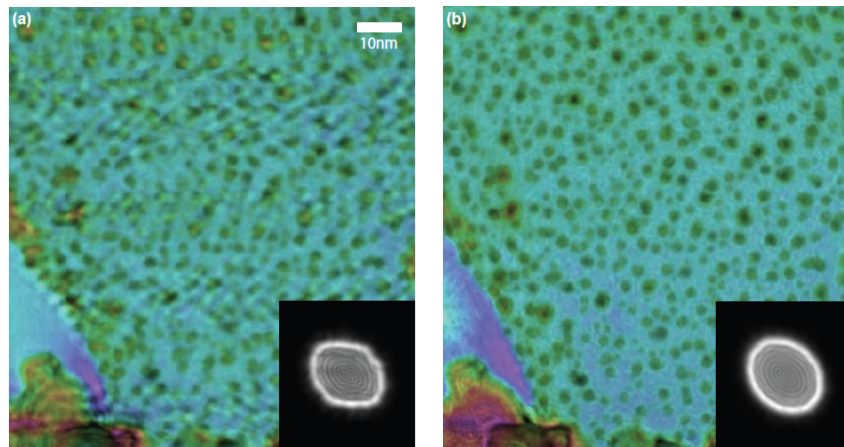


Figure 16: Example of the improvement in the object and probe reconstruction when distortions (present in the left hand side reconstruction) are removed by probe-position refinement. These electron data were seriously damaged by an unpredictable specimen drift [55].

We have a copy of the pre-updated estimate of the object, and the updated estimate, with part of it shifted. Cross-correlating these two should give a peak that is displaced from the origin. The magnitude of this displacement is very small, because the cross-correlation is dominated by the areas of the two images that are mostly identical. However, the peak will lie in a certain direction from the origin in the two-dimensional plane of the cross-correlation. This can be used to steer the next estimate of the probe position. The length of the vector from the

origin to the peak has to be multiplied by some factor to define an actual new position for that probe.

New algorithms for probe position refinement continue to appear. For example, Tripathi et al [57] have combined conjugate gradient methods with the conventional DM and ePIE core algorithms, giving excellent results (Figure 17). Needless to say, researchers tend to be quite conservative, using algorithms that they have confidence in. Most algorithms have free parameters that can be tweaked, and so a lot depends on experience; the optical set up being used also impacts their efficacy. Consequently, it is hard to compare them objectively. Groups choose one, develop the requisite knowledge to optimise it, and then tend to stick with it. Probe positions are just another set of dimensions in the solution space, so there are undoubtedly much more comprehensive and efficient ways for solving for them yet to be found.

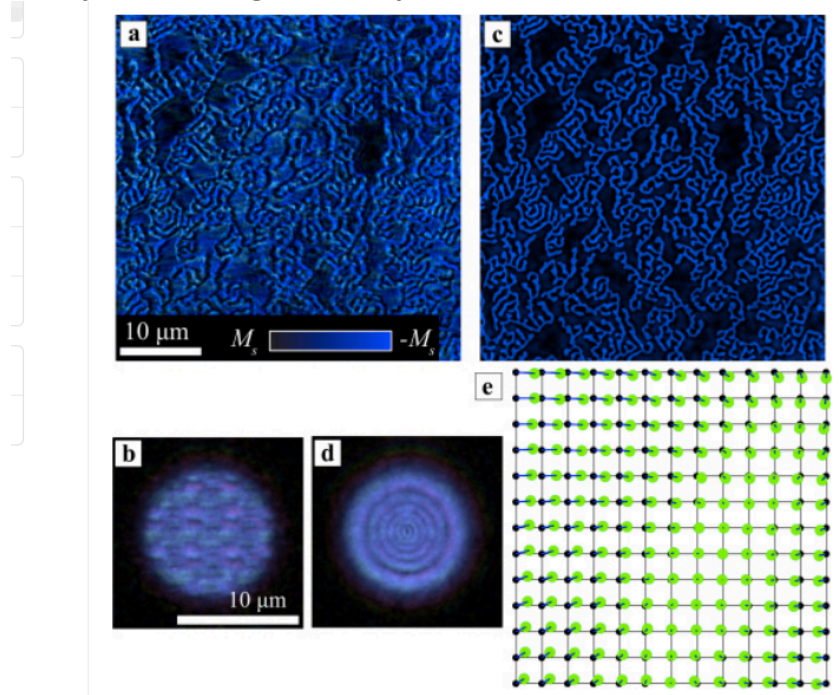


Figure 17: An example of probe-position correction in X-ray ptychography, taken from [57]. (a) and (b) are the uncorrected object and probe reconstructions. (c) and (d) are after probe position refinement. (e) compares putative and actual probe positions calculated from the probe-refinement procedure. See [57] for more details.

4.5) Field of view

In addition to sampling *per se*, another important variable is the field of view of the whole scan. When the sampling in real space has high periodicity but the probe is large (or in other words, the overlap is very large), the centre of the field of view will be illuminated many times, whereas the edge of the field of view is only ever illuminated once. With reference to Figure 18, the ratio of the probe size, the step size and scan size (4x4) is such that only the very centre of the object is illuminated 12 times (the corner probe positions do not overlap in this

area). Extending the scan to 6x6, and the area illuminated 12 times increases in area by a factor of 9. In other words, at low scan sizes, the constraints on the data (generated by multiple sampling the same object area) very rapidly increase as the scan is enlarged. This accounts for the (perhaps surprising) fact that the bigger the field of view (the more numbers we have to solve for) the easier it is to solve for those numbers. Early attempts at iterative ptychography, especially electron ptychography [34, 73, 74], often used very small scan patterns, which may account for the fact that only the very central regions of such reconstructions were of reasonable quality. Using a small field of view also makes solution of the probe function much more difficult, because we need extra diversity to solve for it. In general, a 10x10 scan with a 70% overlap parameter is a safe minimum requirement: anything less than this should be avoided. Larger fields of view are always more desirable.

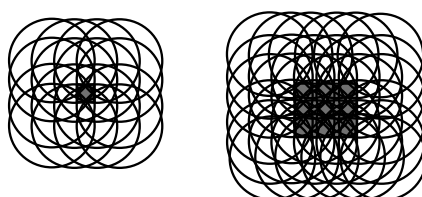


Figure 18: As the field of view is increased, defined in terms of the number of illumination areas, the region where the object has been illuminated most often increases in size quickly.

4.6) Missing data and data truncation

In practice, most ptychographic data sets fulfil the fundamental minimum sampling criterion in Equation 7 by many factors. This means that astonishingly large quantities of data can be discarded, or simply not measured, without affecting the quality of the final reconstruction. One way of thinking of this is via Hoppe's 'ptycho' convolution. If in a conventional real-space ptychography experiment the illumination is parallel (which of course means it has no localisation at the specimen or convolution in the far-field), then a particular scattering vector will arrive at just one pixel on the detector. If data from this pixel is lost – say that detector pixel is faulty – the Fourier component in the object relating to that scattering vector is also irredeemably lost. However, in ptychography we have a localised probe, which means the diffraction pattern is convolved with the scattered amplitude. (Note that in Fresnel full field ptychography, where there is no localisation in the illumination, there is still a convolution in the Fresnel integral.) This means that information relating to any one scattering vector is expressed in a number – often a very large number – of pixels surrounding the faulty pixel. We can therefore happily dispose of the signal from a detector pixel on the understanding that information expressed around it will fill the gap left by it. We do this in an iterative reconstruction algorithm by what is called 'floating' the dead pixel. When the exit wave is propagated to the detector, the missing pixel assumes the modulus and phase of the forward calculation. This is well determined by the existing estimates of the

object and probe (propagated to the missing pixel via the exit wave) which have been generated by all the ‘good’ detector pixels in all the many diffraction patterns.

A more radical manifestation of this phenomenon is accounting for intensity data that is scattered outside the detector, and then using redundancy in the ptychographic sampling to recover intensity that would have been measured had the detector been large enough [75]. This sounds improbable, but it does work. As an example we refer to a visible light optical demonstration in Figure 19. The diffraction pattern (Figure 19a) has been recorded as usual. The well-developed speckle arises from the fact that the illumination in this particular experiment is highly structured and there is a wide range of angles in the incident radiation. Clearly, at the edges of the detector the intensity is still strong and we can reasonably infer it extends beyond the edges of the recorded data.

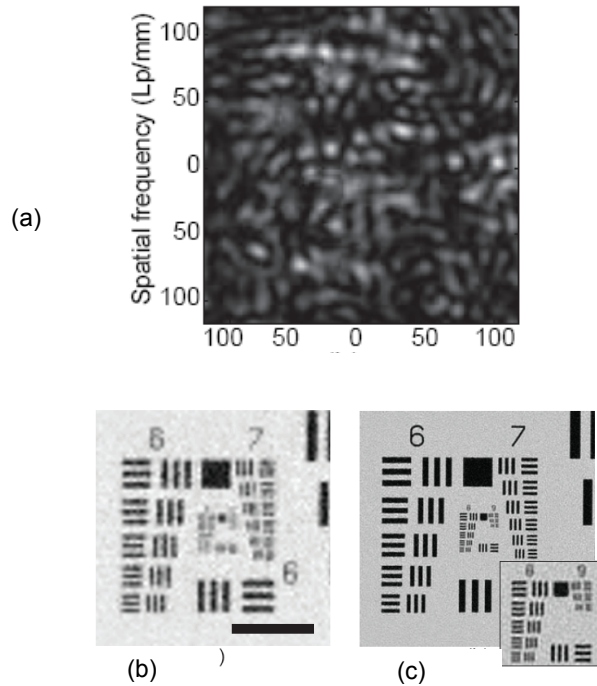


Figure 19: An example of reconstructing data that has not been measured, taken from [75]. (a) the diffraction pattern, which is clearly larger than the detector size. (b) Reconstruction using the measured data. (c) Reconstruction with a much larger diffraction pattern, but with the same data. Pixels outside the region of measured data are left to ‘float’. See main text.

In Figure 19c we see two reconstructions. The low-resolution image has been reconstructed as usual, using only the recorded data. The high-resolution image has used a much larger computational array for the detector, with all pixels outside the measured region being ‘floated’ during the reconstruction. This method is not giving us anything for free or breaking the laws of physics in any way. The lost data has to have a certain value in order to be consistent with the convolution of the object diffraction pattern with the angular distribution of the illumination function (i.e. the Fourier transform of the illumination function). In

this case, the latter is very wide in diffraction space. We end up solving for a region of reciprocal space the width of the illumination (in reciprocal space) convolved with the width of the detector. This is exactly the same as the transfer function of an ordinary optical microscope, which is the convolution of the condenser lens aperture size with the objective lens aperture size.

Recovery of lost diffraction data does have some practical applications. Many high-performance X-ray detectors are arranged in tiles, with gaps between each tile. Rather than interpolating the data in these regions, or simply ignoring them, ptychography can recover accurately the measurement that should have been made there. In the context of scattering outside the whole detector, it must be emphasised that if the diffraction patterns do indeed spill over the extent of the detector (this should not happen if the experiment has been designed properly), there can never be a consistent solution for the inverse calculation. Truncated data should ideally therefore always be padded with floating pixels.

One may ask – ‘how many pixels can I ignore, and how does that number relate to the necessary minimum sampling condition?’ We leave this as a computational exercise for the reader, who may be surprised at the vast quantity of pixels that can be ignored and ‘floated’. Two hints: choose the pixels randomly, not in any sort of systematic array; and remember that just because an image looks ‘OK’ that does not mean to say you have actually recovered all the information that was in the object in first place. Although sparse objects can be hard to reconstruct, because of what we discussed in section 4.2, objects that are moderately sparse (for example resolution test specimens) contain rather low information, and so seem to reconstruct well even if the sampling condition is not reached.

4.7) Shot noise

All data have noise. In the case of electrons and X-rays, specimen damage is always a concern, meaning that minimal dosage is always desirable, and so our preferred data will always suffer from a degree of Poisson noise, even if the detector is perfect. The lower the dose, the lower the specimen damage; but the fewer the counts the higher the noise. This is a leading issue in all imaging science, especially electron microscopy of soft matter like biological tissue or polymers.

Something widely misunderstood is that damage must be a serious weakness of ptychography because each area of the object must be illuminated many times. But this is only true if we worry about the noise in any one diffraction pattern. We must remember that each pixel of the object scatters photons or electrons into several diffraction patterns: these scattering events, and the information they contain, are not lost; they are simply distributed over a number of detector pixels that just happen to lie in different diffraction patterns. Provided our reconstruction algorithm can put together all these counts, the noise in the reconstruction is, like conventional imaging, determined by the total number of counts that passed through each pixel element of the object.

This phenomenon of dose fractionation occurs in many fields, for example tomographic reconstruction [76, 77]. Unfortunately it only works well if our detector is perfect. Background or readout noise mean that we need to minimise the number of times we readout the detector. If there were only one count per diffraction pattern scattered from the object, then this would be drowned out by just a few false counts arising from the detector noise. Hard X-ray detectors and very modern electron detectors do nowadays achieve virtually perfect event counting, so dose fractionation in ptychography can now be fully exploited. When we come to discuss the Wigner Distribution Deconvolution method later in Section 10, we will see that extraordinarily low counting statistics can be tolerated in each diffraction pattern.

A low count in each diffraction pattern has consequences for sampling in real space. Suppose our illumination area is large and we make exposures that are so short that on average only one photon or electron arrives in each diffraction pattern. Clearly, if we are going to get enough flux to pass through any one pixel of the object in order to form a reasonable image of it, then we cannot move our illumination in large step sizes because most image pixels will not scatter even a single photon or electron. The optimum step size will then depend on the characteristics of the detector: some single-event counting detectors can handle very few counts per pixels, so the step size must be very small. The smallest meaningful step size depends on the frequency spectrum of the probe. Moving it less than the periodicity of its highest frequency Fourier component will not alter the diffracted intensity (or rather the probability distribution of the intensity) to produce new, independent information, because we are sampling in real space at periodicity of less than the Nyquist condition.

With low count rates we must be careful about how we reconstruct the data. In any inverse problem, noise masks the minimum in the error metric, and can create many local minima and a false global minimum. Finding the minimum without getting stuck in local minima is much harder, and if we do find the global minimum, it will not be a perfect representation of the object function. After all, a perfect reconstruction implies we know the diffraction pattern perfectly, which we clearly do not, because the low counts are distributed stochastically, albeit with a probability determined by the underlying wavefunction. For noisy data, it is preferable to use a conventional algorithm (DM, ePIE etc.) initially, then when close to the solution, refine with Maximum Likelihood (ML) [42]. A formal study of the convergence properties of this approach has not as yet been undertaken, but the results are impressive. See Figure 20.

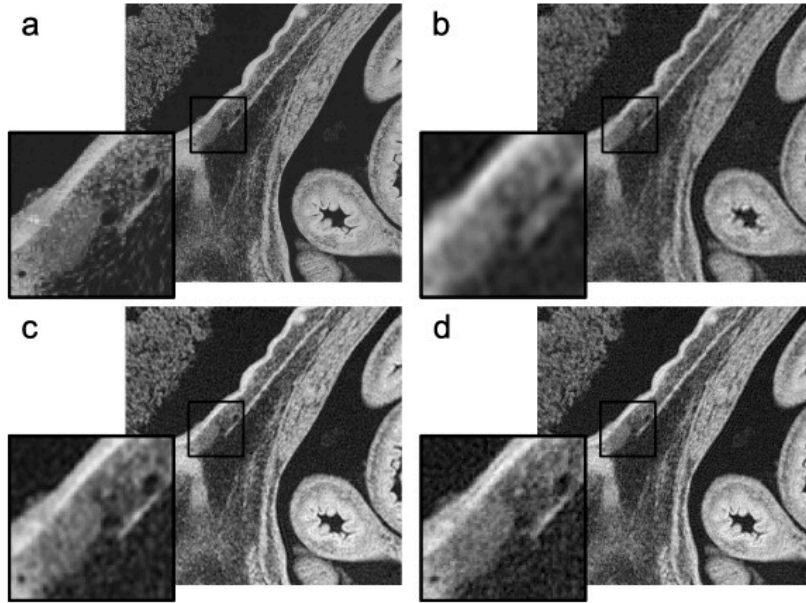


Figure 20: Demonstration of the maximum likelihood (ML) method by Thibault and Guizar-Sicairos, from [42]. (a) Original image. (b) Difference Map (DM) reconstruction. (c) ML reconstruction assuming Gaussian statistics. (d) ML reconstruction assuming Poisson statistics.

4.8) Artefacts in Figure 10

As we emphasised at the beginning of this section, Figure 10 was one of the best ptychographical reconstruction obtained by 2008 [60]. By now, we hope a reader new to the field will appreciate the developments that have occurred since then, so that if they were faced with a similar reconstruction they would know how to improve upon it. These are the issues:

- i) The object and the probe are sparse. At the centre of the field of view, the dominant feature in the object is just a line. Data transfer in ptychography is structure dependent. The probe is also a simple propagated aperture that does not have much diversity. Nevertheless, there is no reason why this shouldn't reconstruct, and where the real-space sampling is finest it does. But at the outset, the combination of probe and specimen means that this experiment is demanding.
- ii) There are periodic structures, because of the regular scan, which we now know cannot easily transfer certain frequencies (even with a known probe). An irregular or circular scan would immediately solve this problem.
- iii) The algorithm employed, PIE does not solve for the probe. The probe has been estimated from knowledge of the aperture and a computational propagation to the specimen, using a physically measured distance between the aperture and the object. Solving for the actual probe will certainly improve the solution.

iv) The scanning stage may not be perfect. Nowadays if there is any doubt about hysteresis or backlash, adopting one of the probe position refinement algorithms could well also improve the reconstruction.

v) Perhaps most important of all (although this has not been a subject covered in this section), is that we now know that this very simple aperture-only set up is one of the worst ways of doing ptychography. See Section 5.10.

vi) The data may also benefit from a modal decomposition (covered in Section 8), not because the laser source is incoherent, although it could have more than one mode within it, but because a ‘throw away’ mode can take out any detector noise which may well be present.

Finally, we remark that dose-fractionation properties of ptychography were not realised when these results were published. For this reason, the authors concluded that the optimum overlap condition is not the largest possible (very dense real-space sampling). They balance the overlap parameter with a consideration of total dose, assuming each diffraction pattern must have the same number of counts, and not that the counts can be fractionated between them. If the detector is not perfect, then of course their analysis still applies. If the detector is perfect, we now know that having as much overlap as possible is optimal, although this generates huge quantities of data, where each diffraction pattern may only contain rather few counts.

Section 5: Experimental configurations

Ptychography is very versatile. The ways it can be undertaken are diverse. Most of the optical set ups that have so far been explored are used with more than one type of radiation, although for good reasons rarely with all types. For example, the simple aperture configuration is easily implemented using visible light or X-rays, but would be fiendishly difficult to do with electrons. Making an aperture small enough at electron wavelengths, and opaque enough outside the aperture, would imply an extraordinarily large aspect ratio for the hole: very hard to make, and which would contaminate almost instantly. Similarly, Fourier ptychography is perfect for visible light and possible for electrons (where it has historically been called tilt-series reconstruction). But it is virtually impossible for synchrotron X-ray ptychography where the beamline direction is fixed: all the optics and the detector would have to be scanned around the object, an impossibly demanding experiment with little to recommend it. However, these examples are exceptions. The benefits and limitations of most aspects of any particular ptychographical optical set up are usually the same, independent of radiation type. In what follows, we will therefore categorise ptychography by optical configuration.

We first make a few general comments. In all that we have said so far we have assumed that the detector lies in the Fourier domain of the object function. In fact there is no requirement for this to be true, as long as we know the form of the propagator between the object and the detector, which, when the detector is

some distance from the object but not far enough to satisfy the Fraunhofer condition, will in general be a Fresnel propagator. All the reconstruction algorithms can equally well apply the detector intensity constraint at any plane downstream of the object.

So far we have mostly discussed an illumination field (a complex-valued wave) being incident upon a scattering object (a complex-valued transmission function): i.e. real-space ptychography. Remember that these can be exchanged with one another. We can instead have an aperture or stop of some type, analogous to the illumination, which multiplies a wavefield. We can then move the aperture or wavefield relative to one another in order to solve for both functions. This is the principle of Fourier ptychography, although the same situation occurs in other configurations we will discuss. The wavefield can be an image or a diffraction pattern and is usually formed by a lens.

In this section we assume that the multiplicative approximation applies (that the exit wave is the illumination function times the transmission function) and that the source of radiation is perfectly coherent. We will explore how to circumvent these approximations in sections 6 and 8 respectively.

5.1 Focused probe ptychography

With reference to Figure 21, we have a coherent source and a lens that focuses a tight beam cross-over through the plane of the object. In X-ray synchrotron ptychography the lens is very far from the source (many 10s of metres), so the radiation incident on the lens is parallel and the coherence width is roughly the size of the lens. The cross-over is then at the focal length of the lens. In scanning transmission electron microscopy (STEM) there are a number of lenses between the source and the final focussing lens, but the effect of these is to demagnify the source so that it appears (when looking back from the focussing lens) to be distant, thus ensuring good spatial coherence. Note that the spatial coherence width across the lens in this configuration (for all types of substantially monochromatic radiation) is approximately the inverse of the angular size of the source as seen when looking back to the source from the plane of lens.

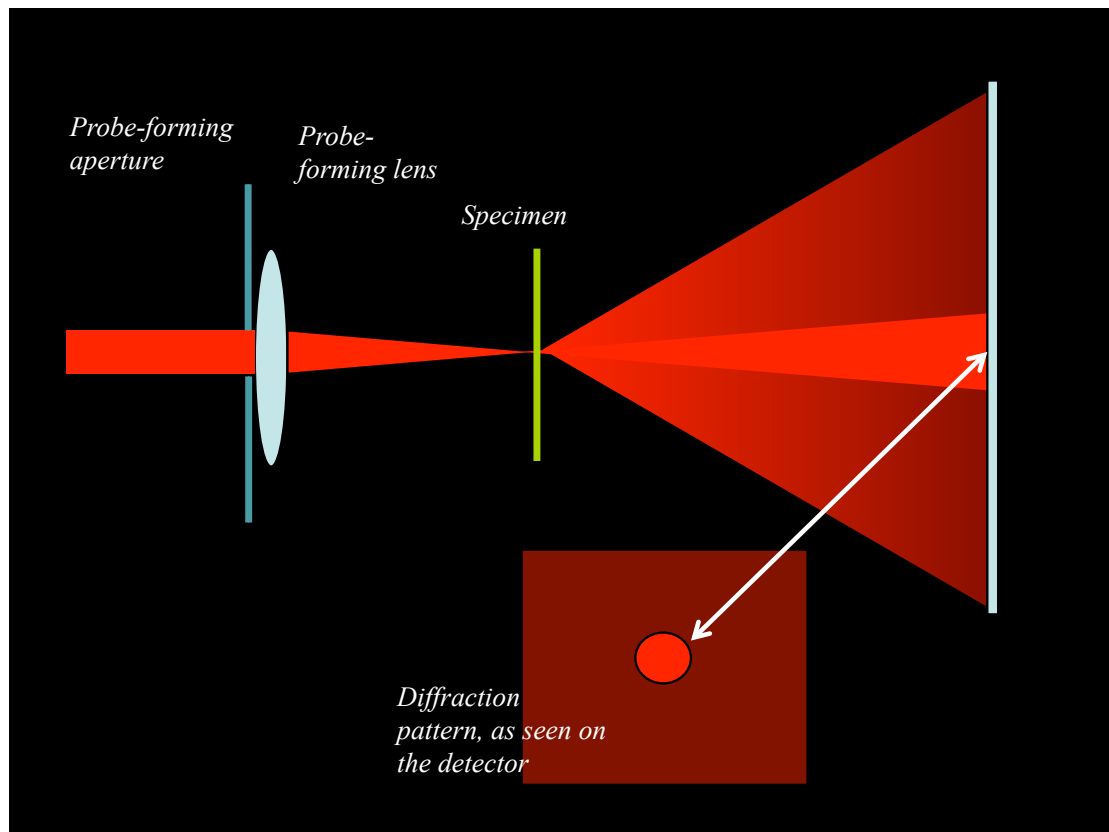


Figure 21: The focused probe geometry. A lens forms a beam-crossover in the plane of the object. In the far field, the diffraction pattern has a bright region (called the Ronchigram in electron microscopy), which is a shadow image of the lens pupil. Weak dark-field diffraction occurs outside this bright area.

If there is a circular sharp aperture within the plane of the lens, then in the absence of the specimen there appears a round disc of illumination on the detector (Figure 22). This is always the case in electron microscopy, but X-ray microscopy often uses a Fresnel zone plate to focus the beam, which requires a central stop [ref], and so the far-field pattern appears as a doughnut shape (also in Figure 22). If Kirkpatrick-Baez (KB) mirrors are used, which they often are because they do not absorb and waste any useable X-ray flux, then there is a rectangular box in the far-field. For the present discussion, we will only discuss the use of a circular aperture.

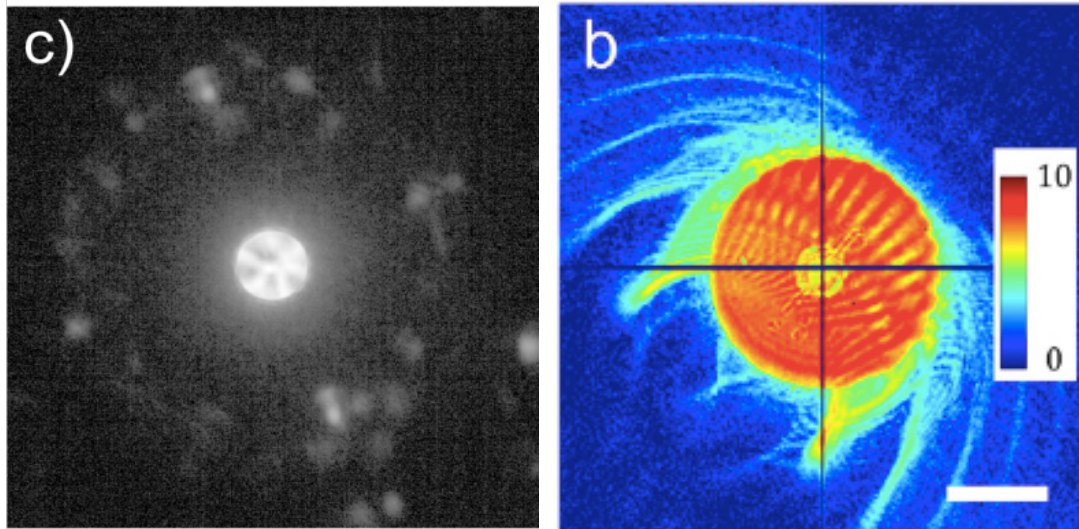


Figure 22: Diffraction patterns in the focussed probe geometry: left, for electrons in a scanning electron microscope (from [68]); right, for hard X-rays using a Fresnel zone lens (from [78]).

Compared to the difficulties of the simple aperture configuration (Section 5.10), one benefit of using the lens means that most of the unscattered counts are spread over a relatively large area, which avoids saturation of the detector, although there is still a large dynamic range between the central disk and the high-angle diffracted dark-field intensity. In all imaging configurations, there is a direct relationship between counts per unit area and obtainable resolution given a certain image contrast. Poisson statistics dictate that detectable contrast depends on \sqrt{N} of the total number of counts passing through a pixel. If we half the pixel size in x and y, we need four times the flux per unit area to be sensitive to the same contrast. For this reason, high-resolution ptychography generally employs a focused beam wherever a small field of view can be tolerated (for example [19, 79]).

A focused beam implies that the probe is very small and so the sampling in the diffraction plane can be very large. If we only had one pixel in the detector plane positioned right in the middle of the far-field disc, we would have created a conventional STE/XM in the bright-field mode. The output of this pixel as a function of the probe position, which is scanned on a very tight grid across the specimen, would be the conventional bright field image. Is this ptychography with a single diffraction pixel? It certainly represents the limit of low sampling in diffraction space and very dense sampling in real space, but it certainly is not ptychography, if only because it has not solved the phase problem: like all bright field images, the phase of the image has been lost.

When we double the information (in each x-y coordinate) by splitting the detector into 4 pixels, or at least 4 quadrants of a circle, as shown in Figure 23, we are now on the first step towards ptychography, sampling in reciprocal space on an extraordinarily coarse grid, and on a very fine grid in real space. However, these four pixels mean that now we have (in principle) enough information to solve the phase problem. We have 2 numbers in each x-y direction that can be

used to calculate the real and imaginary components of each real space image pixel. In fact, for this to be true we have to make some strong assumptions: (a) that the object is weakly scattering, (b) the illumination optics are perfect, and that includes not having any defocus, (c) we must accept that the reconstruction can only process data lying within the central disc of the diffraction pattern, so we rely on all the resolution coming from the lens (not from high-angle dark-field intensity). The only (non-negligible) gain is then the recovery of the image phase. To bring to bear the full power of ptychography to remove lens aberrations in the STE/XM configuration, to process the dark-field high-resolution scattering and to be able to cope with strongly scattering specimens, we must still sample reciprocal space on a fine grid (see section 10).

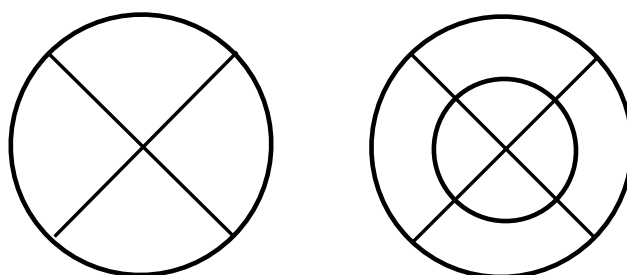


Figure 23: Sector detectors. The simplest configuration (left) can have its transfer characteristics improved by further subdivisions (right). See also Figure 76.

The focused probe arrangement has one very important advantage; analytical signals, like X-fluorescence spectroscopy, can still be simultaneously collected at the resolution of the probe cross-over. This is true for both X-ray and electron microscopy. Certainly the main rationale for aberration-corrected STEM is that elemental composition and bonding information can be obtained at atomic resolution, whether by X-ray spectroscopy or electron energy loss spectroscopy (EELS). The incoherent annular dark field (ADF) image also has several benefits (see Chapter **EDITOR**) that STEM microscopists are loath to lose. With a focussed probe geometry, X-ray and ADF data, as well as some less common signals like secondary electrons, Auger electrons and cathodoluminescence, can be collected simultaneously with ptychographic data. The EELS detector must be on the optic axis and so short of drilling a hole in the diffraction pattern detector, EELS cannot be collected simultaneously with ptychographical data.

Figures 24 and 25 show examples of electron ptychographs collected simultaneously with the ADF signal. Ptychography produces an excellent phase signal, which is sensitive to both heavy and light atoms. The ADF signal is sensitive to the atomic mass of the atoms. The principle advantage of ADF imaging is that the contrast is incoherent and it monotonically increases with the projected mass of the atoms. It therefore has higher resolution than the bright-field image, is approximately quantitative, and it does not suffer from coherent artefacts. However, a consequence of the mass dependence is that it is difficult or impossible to image light atoms within a matrix of heavy atoms.

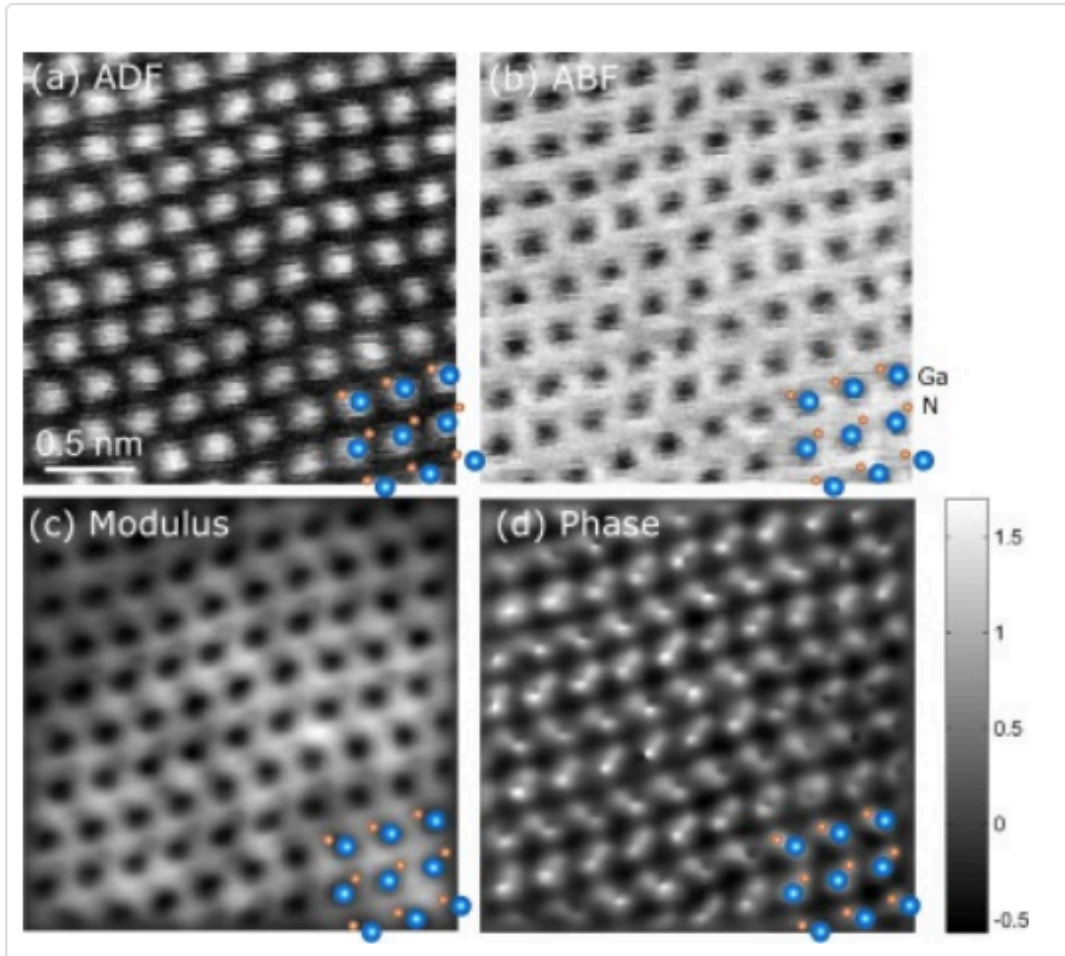


Figure 24: Image of GaN recorded by conventional electron contrast methods: (a) ADF and (b) and ABF images. (c) and (d), modulus and phase of the ptychographic reconstruction. Only the latter can clearly image the very light nitrogen atoms (a few are marked orange) between the heavy Ga atoms (blue). Reproduced from[9].

Figure 24 compares an ADF image with its ptychographical counterpart. The light oxygen atoms, easily visible by ptychography, are entirely absent in the ADF image. Similarly, Figure 25 shows a very light structure (carbon C_{60} inside carbon nanotubes), imaged with high phase contrast via ptychography, together with the ADF picking out a few heavy atoms. The ptychographic phase is also shown to have higher contrast than other less comprehensive sector-detector type phase imaging methods. The combination of ptychography with ADF imaging may well prove to be the most effective use of electron ptychography. Note that both Figures 24 and 25 were reconstructed using the WDD inversion method (Section 10).

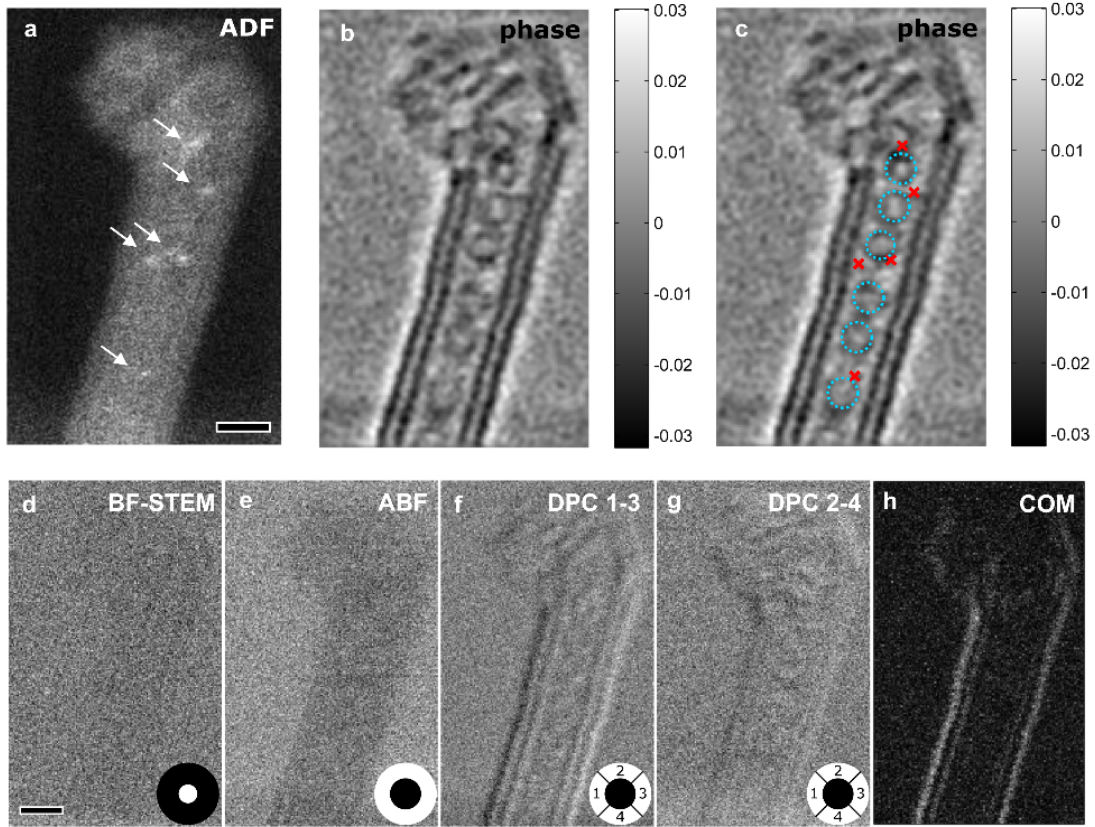


Figure 25: (a) ADF image of a carbon nanotube with C_{60} balls fitting inside it. A few heavy atoms are also picked out in the image. (b) and (c): Phase of the ptychographic image, the latter with the positions of the C_{60} balls and heavy atoms high-lighted. (d)-(g) Contrast from various configurations of sector detectors, all of which are weaker or noisier than the ptychographic reconstruction. Reproduced from [10], where full details can be found.

5.2 Fourier Ptychography

As we have seen in the previous section, a scanning transmission microscope employs a lens to focus the image of a small bright source onto the object: the image is constructed by scanning this tightly focused spot across the object while recording the transmitted intensity which falls on a detector downstream of the object. The optical set up in a conventional transmission microscope would at first appear to be very different. The object is illuminated by a plane wave and the resulting exit wave is brought to a focus at an image plane by a lens lying downstream of the object. During the late 1960s, when the first STEM instruments were developed, there was some confusion in the community when it was realised experimentally that the bright field STEM image has identical features, such as Fresnel fringes and limited contrast transfer, to the TEM image, despite the fact that they are formed in such a completely different way. It was Cowley who first suggested that the well-known principle of reciprocity could account for this equivalence [80]. This states that if we have a source of radiation at a point A which has an intensity I_A , and we record an intensity I_B due to this source at another point B somewhere else in the optical system, then the reverse of this experiment will give the same result: if B radiates with intensity I_A , the signal at A due to that source will be I_B . With reference to Figure 26, we can now

see that our two types of microscope – scanning transmission and conventional transmission – encode identical information. All we have to do is reverse the directions of the rays in the ray diagrams of the two machines.

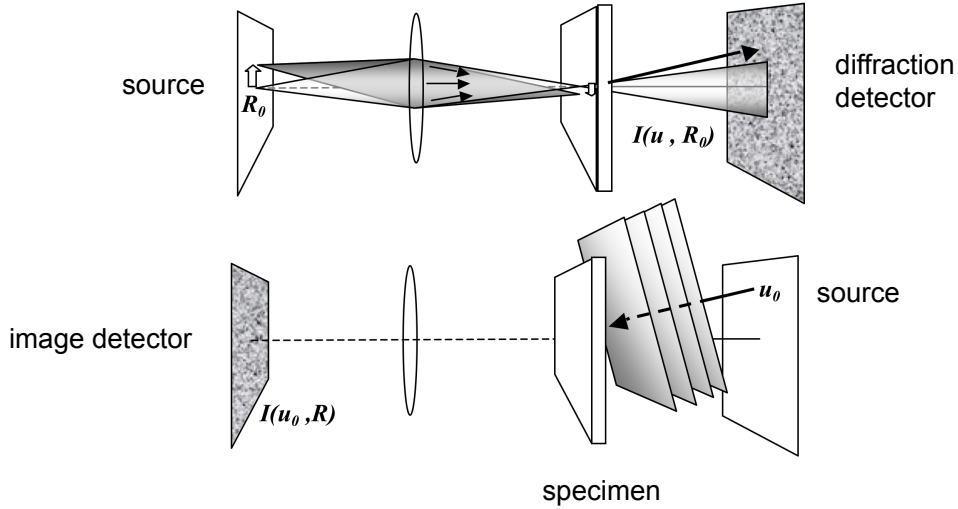


Figure 26: Two very different configurations of transmission microscope. At the top the STE/XM, at the bottom a conventional microscope with tilted illumination. Via the principle of reciprocity, both set ups can collect the same information.

Consider a single pixel in the detector plane of scanning transmission microscope. Keeping all the optical components the same, we now replace that with a source of radiation and we place a detector at the point originally occupied by the source. Remember, our scanning mode detector was positioned in the Fraunhofer diffraction plane a long way away from the specimen, so its coordinates are a function of angle. When replaced by a source, the incident radiation bathes the whole specimen with a tilted plane wave, as illustrated in the lower half of Figure 26. In the conventional transmission microscope we do not need to scan a probe, because the image arrives simultaneously over the whole image plane. Rather, each image pixel is, via reciprocity, like a different probe position, because the effect of moving the source in a scanning transmission machine is to move the probe. In short, we have a four-dimensional data set that can be collected in 2 ways: a set of diffraction patterns recorded as a function of probe position, or a set of images recorded as a function of plane wave illumination angle. It stands to reason that we can therefore use this reciprocal configuration to do everything that conventional ptychography can do. The method is nowadays called Fourier ptychography. It was first proposed by Hoppe, shortly after his work on ptychography [81].

Consider a conventional microscope in which the illumination is a coherent plane wave travelling parallel to the optic axis. In the back focal plane of the lens, we see the conventional parallel beam diffraction pattern. If the specimen does not scatter too strongly, this will consist of a bright spot on the optic axis with weaker diffraction amplitude from the specimen lying around it, as shown in Figure 27. Now when we tilt the beam, the bright central spot will move laterally

and, provided the specimen is not too thick, the diffracted amplitude will shift with it by the same amount. If we place an aperture also in the back focal plane, then we have constructed a sort of ptychographic experiment. The shifting diffraction pattern is like the object wave we want to solve for – except in this case it happens to be a diffraction pattern. The aperture is like the conventional illumination function. Our data are recorded in the Fourier domain of these two functions, which in this case is the image plane. Now the folding (convolution, or ‘ptycho’) of the wave intermixture is the convolution of the impulse response function of the lens/aperture and the exit wave of the object, which gives the convolved image recorded in intensity. All the general principles of ptychography apply. If we are going to call this technique Fourier ptychography, we have to rename conventional ptychography as ‘real-space’ ptychography.

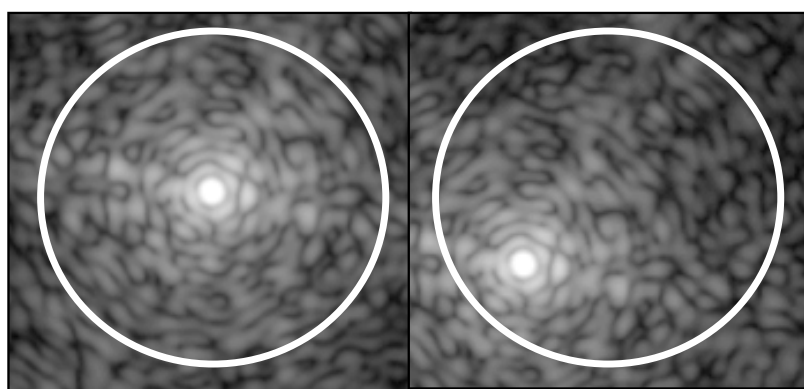


Figure 27: An illustration of the diffraction amplitude at the back focal plane of a conventional microscope. As the illumination is tilted, different parts of the diffraction pattern are steered through the lens. In Fourier ptychography, the aperture is treated as an illumination function, the diffraction pattern as the object. By shifting the angle of illumination a wide area of the diffraction plane can be reconstructed.

In visible light Fourier ptychography the imaging system often has a low numerical aperture, meaning that every image recorded in the image plane has very poor resolution. Once a large diffraction pattern has been calculated using the ptychographical methods, we can transform back and obtain a very high resolution picture. Why would anyone want to do this? After all, optical lenses are nowadays very good indeed. One obvious reason is that we end up with both the modulus and phase of the image, which is very important for imaging transparent objects such as biological cells. However, another key advantage is very high resolution combined with very large field of view. Supposing we have a CCD camera with 1000x1000 pixels. If we deliberately stop down the imaging lens so it has very poor resolution for each image we record, we can demagnify the image on the CCD and capture a very large field of view. If we now step the diffraction pattern in the back focal plane through enough incident beam tilt angles to extend the field of view of the diffraction pattern (not to be confused by the field of view in the image plane, which determines the pixel pitch of the diffraction pattern) say by a factor of 10, and then back transform to the image plane, we have a high resolution image of a wide field of view – 100,000x100,000

pixels. This wide field of view is vital for things like counting abnormal cells in a cell culture, where statistics from huge numbers of cells are key.

The physical set up of visible light Fourier ptychography also has some significant advantages over its real space counterpart. The different angles of illumination can be generated by an array of light emitting diodes (LEDs), so neither the illumination function nor the object function has to be moved. Moving the illumination in any optical set up, say by using deflection coils in an electron microscope or by using mirrors or prisms in the case of visible light invariably changes the shape of the probe via the introduction of aberrations or phase gradients, so that one of the principal constraints of ptychography is lost. There are ways around position-dependent probe variations (section 4.4), but it is preferable to avoid this complication. The disadvantage of moving the specimen is that it takes time, and there is invariably hysteresis in mechanical stages, although we have already described how computational refinement of probe positions is possible Maiden [55]. In contrast, a Fourier ptychographic microscope can be made as a fixed structure with no moving parts, with a fast readout camera easily synchronised with the switching of each illumination source.



Figure 28: A Fourier ptychography microscope, shown schematically in (a). (b) A conventional microscope has been modified using a Lego framework so that an array of photodiodes can illuminate the object at different angles (see bottom Figure 26). (c) A

typical image collected at low resolution (a single illumination tilt). (d) Magnified section of (c). (e) Ptychographically reconstructed image at same magnification as (d).

Figure 28 shows an image of an optical microscope, modified with an LED array mounted on a Lego structure, which was used to generate the first published visible light ptychographic image [40], together with a demonstration of the resolution improvement over the raw data. Figure 29 shows the reconstruction process, in both real space and reciprocal space. Figure 30 shows an example of a biological structure imaged using the approach. The final reconstruction here is composed of 0.9 gigapixels.

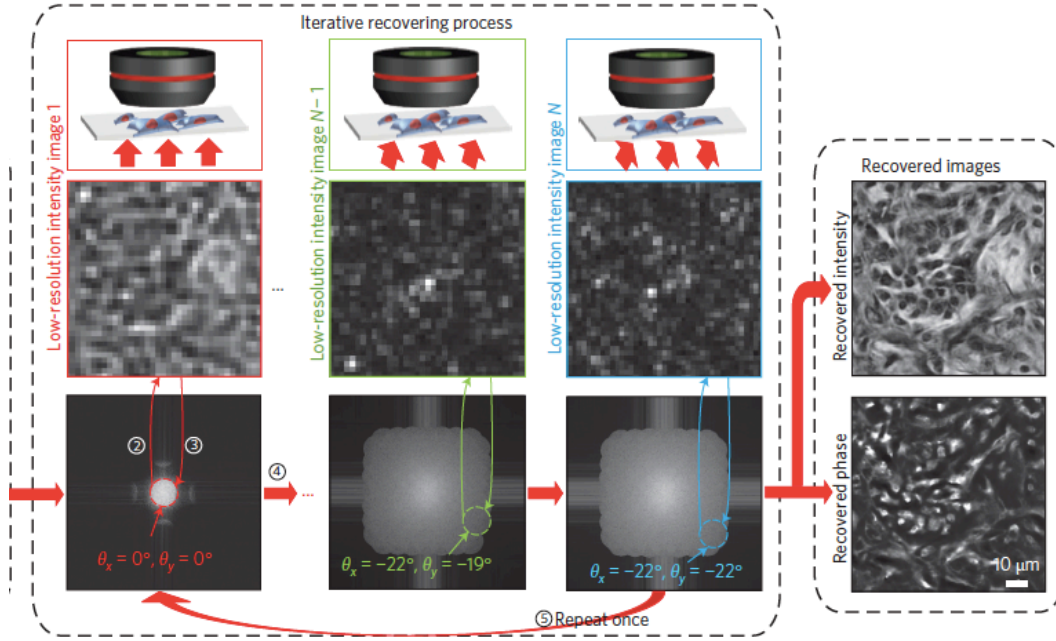


Figure 29: Iterative reconstruction for Fourier ptychography. Top three images showing different illumination angles. Below these the raw data for a bright-field image and two typical dark-field images. Bottom three central frames show the back-focal plane reconstruction developing. Far right is the recovered image intensity (top) and phase (bottom). Reproduced from [40].

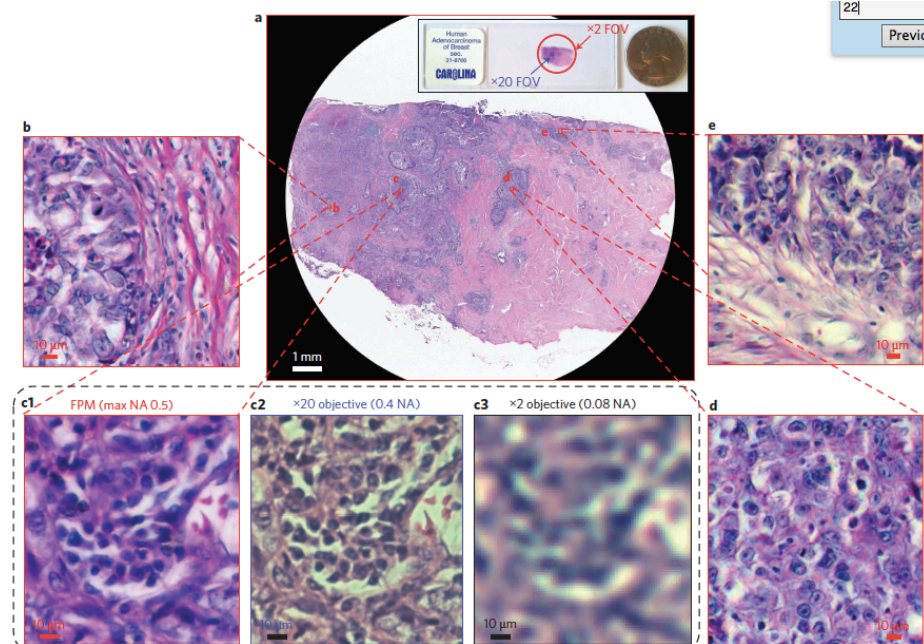


Figure 30: Top middle: example of a very wide field of view, high-resolution, 0.9 Gigapixel image reconstructed by Fourier ptychography [40]. The scale bar is 1mm. On all other images the scale bar is 10 μ m. (c2) and (c3) are conventional images taken with x20 and x2 object lenses, respectively. The other images are taken from various small areas of the reconstruction.

Fourier ptychography has also been undertaken in the electron microscope, the original concept pre-dating the recent interest in the visible light version by 40 years[82]. It is generally called ‘tilt series reconstruction’ in electron microscopy, and has shown the ability to improve resolution over and above that of a good electron lens [83]. It is impractical to have an array of electron sources, as is used with visible light, so the single illuminating beam must be scanned through a range of discrete incident angles by double deflection coils, which also have a habit of suffering from hysteresis. Figure 31 shows an example of the raw data acquired in an electron microscope as a function of illumination angle, and the corresponding reconstruction in the back focal plane. Figure 32 illustrates the gain in resolution in real space over and above a conventional through-focal series reconstruction, which uses only one normally incident beam.

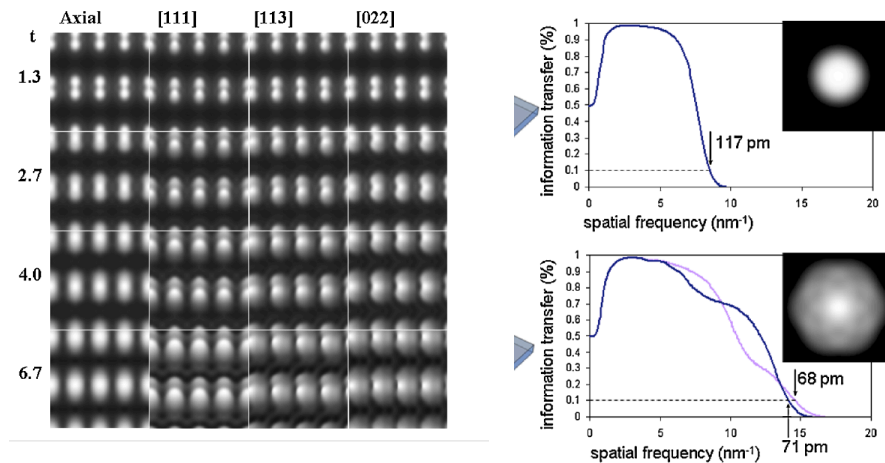


Figure 31: Fourier ptychography in the electron microscope, where it is usually called tilt-series reconstruction. Left: raw data from various tilt angles. The specimen is silicon orientated on the $\langle 112 \rangle$ zone axis. Right top: the region of reciprocal space passing through to the conventional image. Right bottom: the region of reciprocal space reconstructed. Taken from [83].

Since 2013, when visible light Fourier ptychography was first demonstrated, there has been a great deal of research undertaken on it and the field is expanding very quickly. All the inversion algorithms developed for real-space ptychography apply equally well, with one or two minor alterations, to Fourier ptychography. Indeed, all the key developments in real-space ptychography have been reproduced in Fourier ptychography, and many have been superseded: see for example [29, 59, 84-91]. If you understand reciprocity, everything we have discussed in Section 4 with respect to sampling, diversity and reconstruction refinement still applies, as do most of the methods we discuss in sections 6 and 8. The field is also making significant contributions to the theory of the inverse problem in ptychography, which we discuss in Section 9.

In the visible light domain, the Fourier configuration of ptychography is extremely promising. For more information, the interested reader is directed towards the book by Zheng that has recently been published on the subject [92].

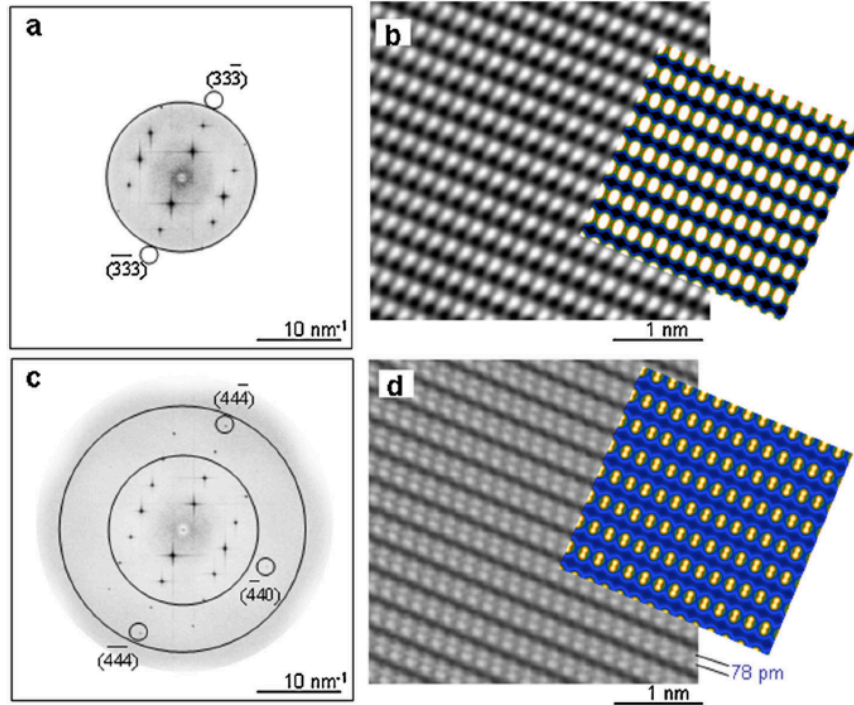


Figure 32: Data as in Figure 31. Top left: diffractogram (modulus of the Fourier transform of the image) for the conventional image, which is shown on the right. Bottom right, diffractogram of the reconstructed image, clearly extending much further into reciprocal space. Bottom right is the final high-resolution reconstruction, including the expected calculated exit wave in colour [83].

5.3) Selected Area Ptychography (SAP)

Another configuration where the reconstruction is of a wave-field instead of a physical object, in this case an image formed by a lens, is called selected area ptychography, or SAP. With reference to Figure 33, a conventional microscope with the specimen illuminated by coherent radiation is used to form a conventional image. An aperture is placed in the plane of the image and the resulting diffraction pattern is recorded some distance downstream of the aperture. The specimen is physically moved laterally so that the image wavefield moves across the aperture. We treat the image as our object function and the aperture as our illumination. Once again, everything we have said about real-space ptychography as far as reconstruction algorithms applies. All electron microscopes have a 'selected area (SA) aperture' in the first image plane in order to select one area of an object from which to obtain a diffraction pattern. This is used to characterise small areas of a specimen that may be composed of very small crystal grains or small isolated objects. Hence SA Ptychography: SAP. The configuration has been shown to work in the electron microscope in a proof-of-

principle experiment [46](Figure 34). Unlike real-space electron ptychography, it can image a very large field of view, and may well compete with conventional electron holography, say for mapping electric or magnetic fields.

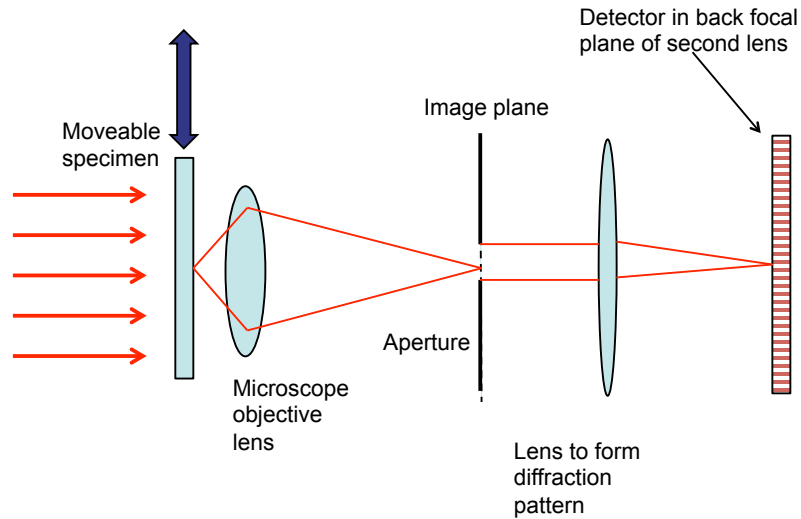


Figure 33: The SAP configuration. The object is moved, causing its image to move relative to a selected area aperture in the first image plane of the objective lens. The detector lies in the Fraunhofer plane (or more normally the Fresnel diffraction plane) of the aperture.

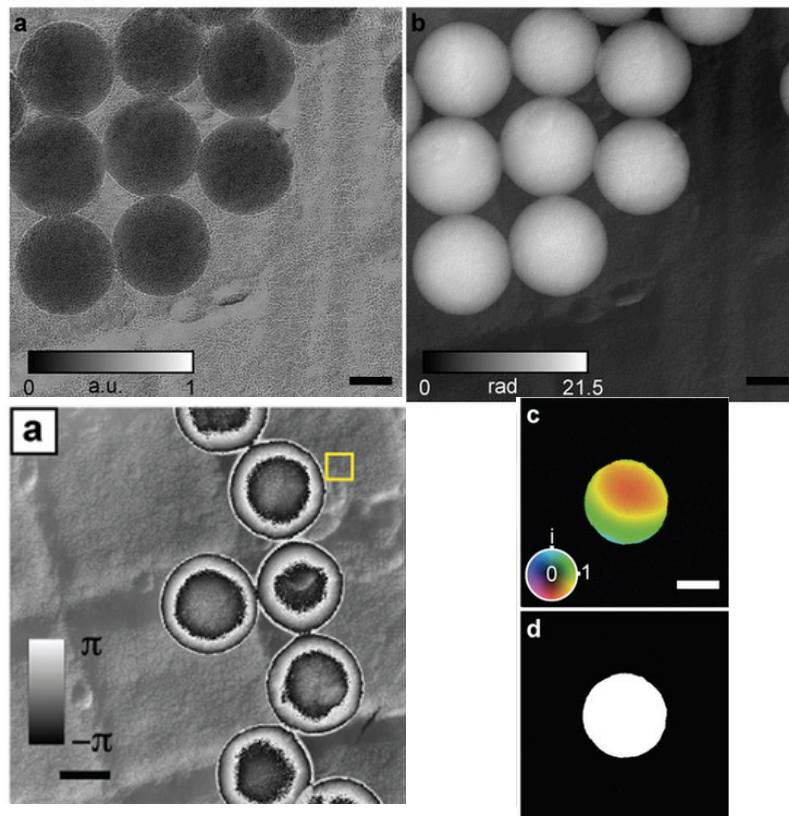


Figure 34: Example of electron SAP, taken from [46]. Top left is the conventional bright-field image. The spherical latex balls appear with flat contrast because the contrast mechanism relies on weak phase. In the unwrapped ptychographic phase, top right, the strong phase is rendered perfectly. Bottom left shows the phase wraps, clearly confirming that these objects are very strong phase. The fractured phase wrap in the balls is because of the gold structure underlying the balls. Bottom right: phase and modulus of the aperture function.

To date, the most extensive use of SAP has been at visible light wavelengths, where it is commercially available as a means of characterising biological cell life cycle [62, 93]. The main advantage of the technique is that the full coherent resolution capability of the optical objective lens can be exploited, giving high-resolution and extremely high quality images, together with a clean phase image. The latter is crucial for imaging live, unstained or unlabelled cells. Resolution can be increased further by arranging for the illumination to include a range of incident angles within it. We showed an example image using this technique in Figure 11. Because the free space background phase signal is so flat and there are no ringing effects in the image caused by distortion of low frequencies (as happens, say, with Zernike phase contrast), even very weakly scattering transparent objects appear with high contrast. This means that segmentation of the image is easy and accurate, allowing for reliable cell counting statistics and the measurement of other biologically important parameters such as reproduction rates, motility, cell volume, etc. (Figure 35). Very long experiments (several days) can be performed (in a suitable cell incubator) without the need for refocusing the image, which can be achieved computationally post data acquisition.

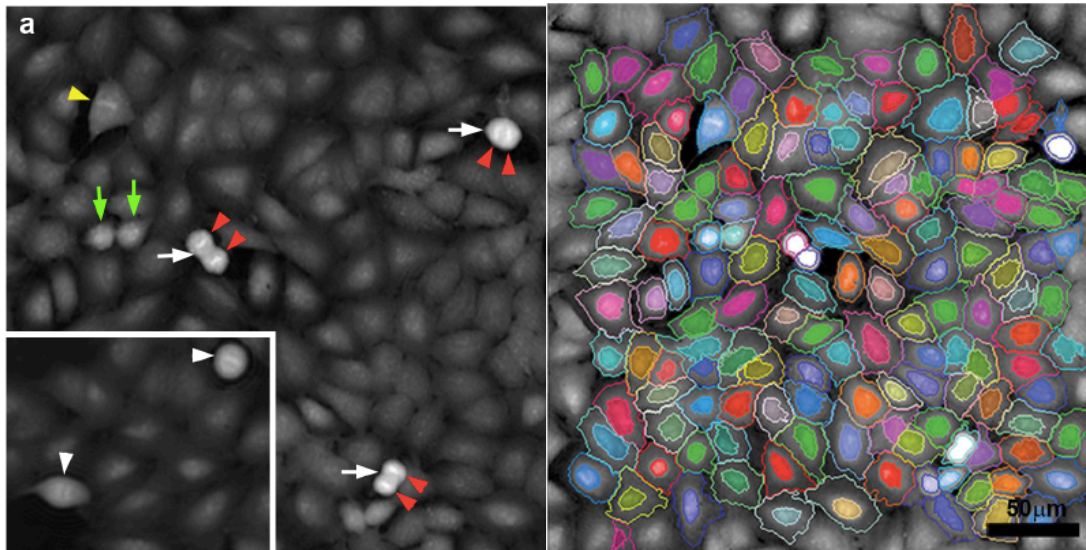


Figure 35: Visible light SAP (see also Figure 11), taken from [93]. One of the great advantages of the technique is that cells do not need to be stained or labelled, and so can be observed over days reproducing and moving. This image shows some cells that are at various stages in the process of division. See [93] for details. The phase image is particularly amenable to precise segmentation, as shown on the right.

5.4) Fresnel full-field ptychography

An early definition of ptychography suggested that a prerequisite for the method is that the illumination function is localised, so that the convolution in the Fourier domain allows diffraction components to interfere with one another. This has nowadays proved to be over-prescriptive. Consider the two experiments shown in Figure 36. Figure 36a consists of a corrugated wave front (i.e. the surfaces of constant phase depart significantly from plane surfaces) incident upon the object. Behind the object, but relatively close, is the detector. To work out the intensity of the radiation at the detector plane, we add up a sum of Huygen's elementary spherical waves, each centred on one point of the exit wave function, and having the modulus and phase of the exit wave function at that point. The real part of the impulse response of any one of these waves looks something like the graph on the right of Figure 36a.

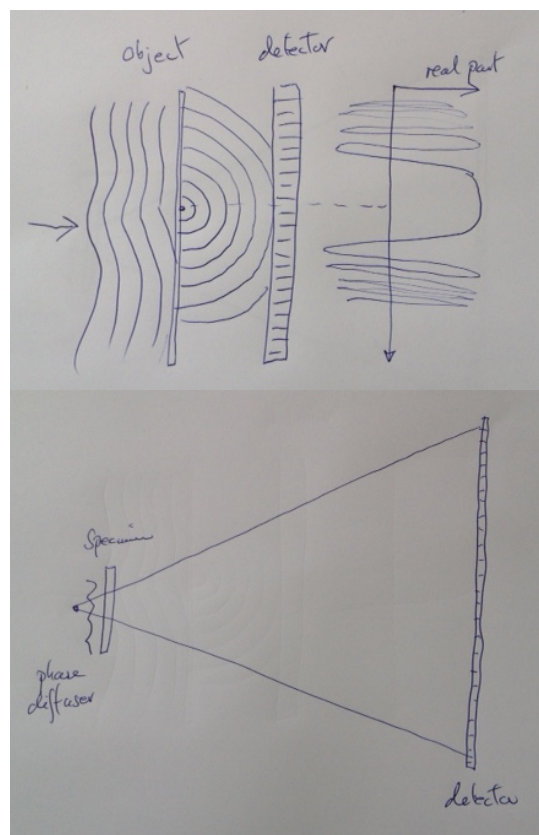


Figure 36: Full field Fresnel ptychography. The incident wave must have structure in order to provide the ptychographical diversity. Wavelets scattered from the object have most influence on the detector pixels directly downstream of them. This is because of the stationary phase effect of the Fresnel integral (top right). Bottom: Shadow imaging to increase the magnification of the technique.

This intermixture of the waves has a similar effect as the convolution in Fourier domain ptychography, although how the wave components add together is rather different. The Fourier integral involves the whole object at once, adding all rays that head off to the detector at a particular angle. In the near field, the propagation integral also adds rays arriving at any one detector pixel, but their path lengths and angles vary considerably from one position on the object to the

next. Like the Fourier integral, it seems as if all elements of the object contribute wave amplitude to each detector pixel. However, the pixel size of the detector means that the intensity of any one pixel is only affected by a rather localised region of the object exit wave. Over the surface of the detector, the elementary spherical wave has beyond a certain width, very rapidly varying oscillations. This is a stationary phase effect. An elementary spherical wave from one element of the object gives a large contribution to the scattering integral over an area of the detector where its phase is substantially flat, i.e. of roughly constant value. Detector pixels laterally displaced from the source of the elementary wavelet experience a quickly changing phase. Beyond a lateral displacement of more than the radius of the first Fresnel zone (the area in which the phase changes by less than π , the phase begins to change very rapidly, roughly as a function of the lateral displacement squared. There will very quickly come a point where the size of the pixel is such that it spans many phase cycles, so that the contribution from the wavelet integrates to zero. Partial coherence in the illuminating beam (i.e. a finite source size) exacerbates this effect. So, in Fresnel ptychography we do not need a localised source: the Fresnel integral itself defines a localised area that contributes to any one detector pixel, although the local area so defined is different for each detector pixel.

Fresnel ptychography requires us to move the object laterally with respect to the illumination. Of course, if the illumination is a simple plane wave, the out of focus image on the detector will just move laterally without changing at all, thus not giving us any information. That is why the illumination must have diversity – the wavefronts must be distorted or uneven. Henceforth, iterative solution of the ptychographic phase problem can proceed as before, using any of the standard algorithms, the only difference being a change from the Fourier propagator to one that models the physical propagation from the object to the detector. Note also the comments in Section 5.11.

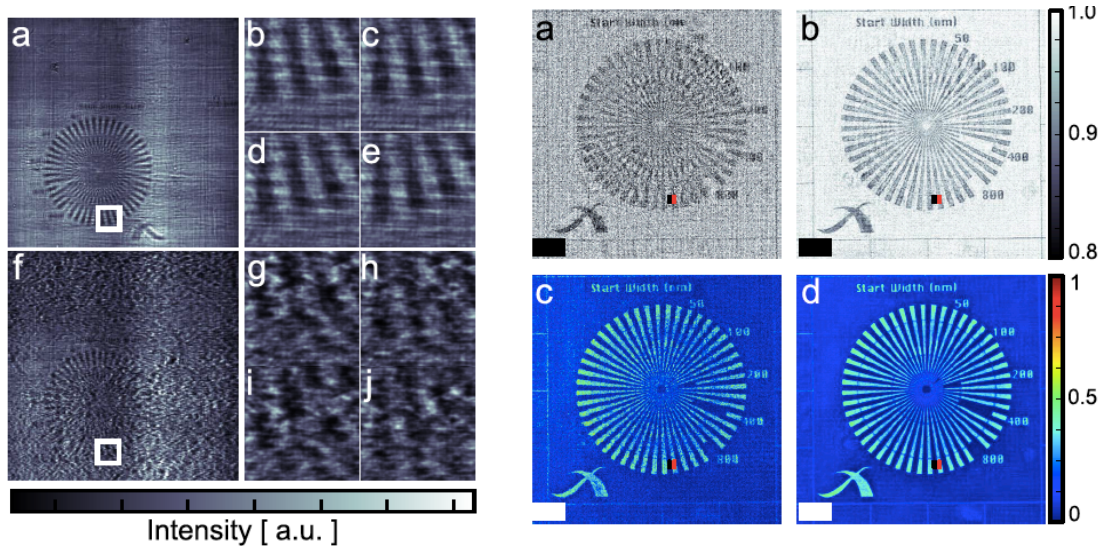


Figure 37: Example of hard X-ray near-field ptychography, from[94]. The frame on the left shows raw data without (top) and with (bottom) a diffuser. The smaller panels show the raw data as it is scanned. The data using the diffuser varies more rapidly, implying

greater diversity. Panel on the right are the reconstructions. (a) and (c) modulus and phase for no diffuser. (b) and (d) modulus and phase with the diffuser. Scale bar is $2\mu\text{m}$.

The experiment shown in Figure 36 will not give us any magnification of the object: the reconstruction has the same resolution as the detector pixel pitch (although the phase is recovered). X-ray near-field ptychography has therefore been undertaken in the X-ray shadow image microscopy mode that was pioneered by Cosslett [95] in the 1950s, see Figure 36b. The ratio of the distances from the source to the detector and source to the object determine the magnification constant. Figure 37 shows a conventional shadow image taken in this configuration using 16.9 keV hard X-rays [94]. The source was generated by the focused beam cross-over created by two Kirkpatrick-Baez (KB) mirrors. Figure 37a shows raw data taken in this configuration, with and without a fixed diffuser in the beam path. Figure 37b shows the ptychographic reconstructions. Interestingly, even without the diffuser, the beam line optics, which of course always introduces some minor imperfections in the wavefield, have introduced enough diversity in the incident wavefield for the reconstruction to work. However, when with the diffuser is in place, the reconstruction is very much better. This is an example of diversity improving ptychographic data.

Near field ptychography has several advantages. The field of view is large, even when only a few specimen positions are used. Strictly speaking, only four scan positions are needed to recover the complex components of each pixel in the x- and y-directions. (This is similar to the need for four sector detectors discussed in Section 5.1). Of course, more specimen positions are beneficial because they further constrain the solution: the results shown in Figure 37 nevertheless used only 16 positions. In real-space ptychography, the diffraction pattern always has a high-dynamic range, especially between the bright unscattered beam and dark-field features lying at high scattering angles. This can make it very difficult to choose an appropriate exposure time: what is correct for the unscattered beam is far too short for the scattered beams. Conversely, in near-field ptychography the whole detector is evenly illuminated, which makes setting the optimal exposure time easier.

5.5) Defocused probe ptychography

With reference to Figure 38, we can use a lens to form a convergent beam but rather than place the object at the exact focus of the probe, we can defocus it somewhat, or equivalently move the specimen up or downstream of the beam focus. This configuration combines near-field ptychography (Figure 36b) and the focused probe scanning transmission microscopy STE/XM. The difference between near-field ptychography is that the degree of defocus is relatively small, so that the field of view within the central disc is small, and that diffracted data lying at high angles outside the bright disc are also processed. It is therefore a complicated mixture of Fresnel-type interference and Fourier domain diffraction. Of course, the reconstruction process remains the same, the only difference being the probe structure is dominated by curved wavefronts. If that curvature is included correctly, the far-field pattern is just the Fourier transform of the exit wave. We do not have to use the Fresnel integral for the central disc,

because the pre-multiplier of the curved phase distribution, followed by a Fourier transform, is itself a way of constructing the Fresnel integral.

This type of defocused probe is very commonly used in synchrotron X-ray ptychography, because by moving the specimen forwards and backwards away from the beam cross-over the diameter of the probe can be changed at will. Thus the probe size and step size can be matched to the field of view (see, for example, in [96]). Another benefit is that it helps keep the number of exposures small, which is important when the duty cycle of the camera readout and/or the settling time of the stage make up a significant proportion of the total elapsed time of the whole experiment. This is especially true of ptycho-tomographic scans that can take many hours, or even days.

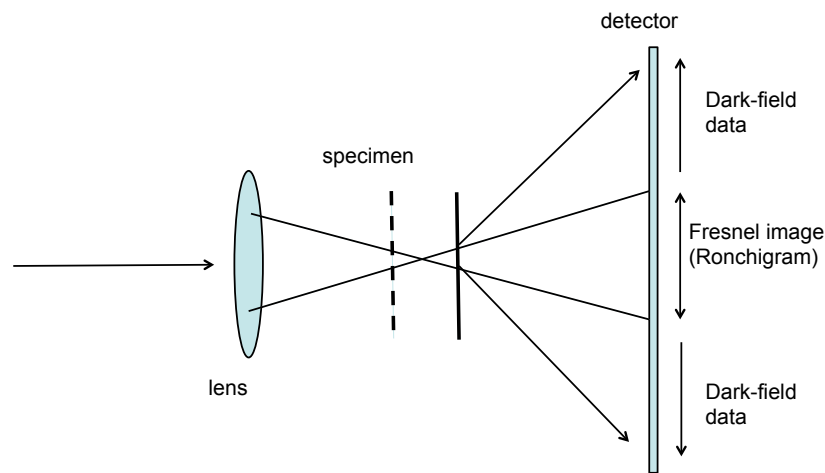


Figure 38: Defocused probe ptychography. The far-field is both a magnified Fresnel shadow image (Figure 36), but also has high-angle dark-field intensity.

Yet another benefit of having a large probe is to limit dose-rate specimen damage effects, at least for electron ptychography where damage can be severe.. As we have emphasised, ptychography is a dose-fractionation method: moving the illumination by large step sizes or small step sizes does not affect the total dose that needs to go through the sample in order to produce an image with adequate signal to noise in each reconstruction image pixel. However, there is some evidence that dose *rate* can be as important as *total dose*, for example in the time-dependence of damage observed by electron energy loss spectroscopy [97]. It is possible that ions displaced by knock on damage can relax back into their original location if there is sufficient time to do so before the next knock on event occurs. That means that a low dose rate per unit area may induce less damage for the same amount of total dose. Using a large probe achieves exactly this. It is also possible that a large area of illumination will ameliorate other annoying problems that arise in electron microscopy, such as the build up of contamination (sometimes exacerbated by a focused probe) or local charging of the specimen, which can lead to uncontrolled and sudden specimen movement.

Figure 39 shows an example of a defocused probe electron ptychograph obtained from a scanning electron microscope. The was not a STEM operating at

high accelerating voltage, but a conventional SEM (an FEI Quanta 600) operating at 30keV, with a two-dimensional detector mounted below the specimen stage. The stage had also been modified to accommodate a transmission specimen, which in this case was a standard TEM resolution test specimen consisting of small gold particles sitting on a thin amorphous carbon support film. Figure 22a, discussed previously, shows an example of the raw data. Although it is hard to see it, the central disc, which in electron microscopy is called the Ronchigram, has some structure within it, which is essentially the same as the Fresnel near-field image in the equivalent X-ray experiment shown in Figure 37. The difference here is that the range of illumination angles in the beam is small and the experiment is also going to process the dark-field diffraction peaks lying well outside the central discs – indeed, this is where all the high resolution information in the experiment comes from.

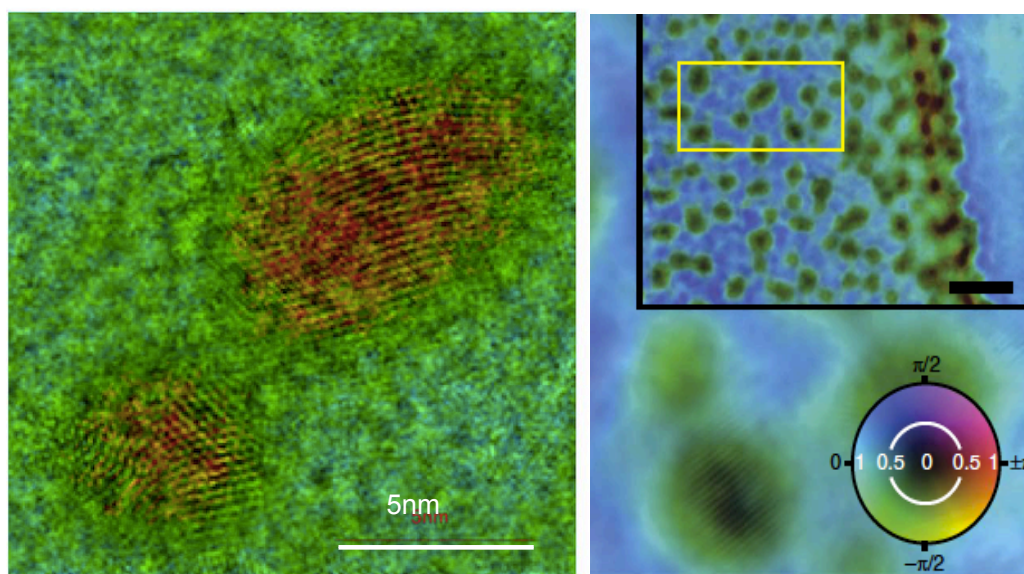


Figure 39: Electron ptychography with a defocused probe in a scanning electron microscope (SEM), taken from[68]. The specimen is a standard test specimen consisting of gold particles on amorphous carbon. Phase is represented by colour, brightness modulus. The scale bar on the large field of view top right is 15nm. The enlarge image has been contrast enhanced for reproduction; scale bar 5nm.

Atomic fringes are visible in some of the gold particles – those that are orientated on a zone axis. The smallest fringes visible are separated by 0.23 nm, corresponding to an increase in resolution over the lens capability by a factor of about 5. These results imply we could dispose of conventional TEMs, at least for imaging (as opposed to focused probe analysis), and use instead a rather less costly SEM fitted with a transmission detector. In fact, careful inspection of Figure 39 shows that some of the atomic fringes are delocalised from the gold particles – a similar problem that arises in defocused TEM images, and one that is fatal for accurately determining the exact position of atomic columns. Delocalisation is particularly sensitive to any intensity pedestal or read-out noise in the detector. The comprehensive solution would be to have a single electron detector.

From the point of view of the ease of reconstruction, using a defocused probe has both strengths and weaknesses. The central disc is essentially a Gabor hologram. In any iterative reconstruction this means that the first low-resolution image of the object should be a pretty good holographic estimate. Indeed, it is known that cCDI reconstructions are improved in this Fresnel mode [98]. If there are errors in the probe positions, this is very obvious in the Ronchigram, and can be used to coarsely adjust probe position errors, provided only defocus (and not higher order aberrations) are present[99]: in this case the Ronchigram is an undistorted near-field image of the object.

Unfortunately, reconstructing data from curved wave illumination also has several hazards, which can quickly lead to stagnation of the reconstruction process, or simply give a completely wrong solution. Defocus corresponds to adding an extra curved phase to the transfer function of the lens. This curvature will also appear across the Ronchigram disc in the far-field. In real space in the sample plane, there is also a corresponding curved phase over the illumination. In an iterative reconstruction, the phase of the correct curvature must be seeded in the first estimate of the probe in real space.

Suppose that the probe is physically defocused so it has a diameter D . There is only one phase curvature over this probe that will give rise to a disc in the far-field of the correct, recorded diameter. As an extreme example, suppose our first guess of the probe has no phase across it at all. Propagating this to the far-field will give us an Airy disc – an intensity distribution only a fraction of the width of the measured far-field disc. When we apply the Fourier constraint, creating a bright disc of modulus, and back Fourier transform we have a function that is nothing like our real probe function. In fact it will probably be so small that it will not even overlap with the adjacent probes that were used to create the data. Recovering from such a remote position in the solution space is virtually impossible.

A further problem is that the magnification of the Ronchigram image is also a function of defocus (as is obvious from Figure 38). This means that if the step size in real space is poorly calibrated, the only way the reconstruction algorithm can reconcile the conflicting data is to both increase (or decrease) the magnification of the image and increase (or decrease) the size of the illumination, achieved by changing the estimated defocus of the illumination. The result is a reconstruction that looks out of focus, but it cannot be put back into focus simply by re-propagating to the correct plane, because the reconstruction does not relate to any actual physical plane within the wave disturbance; it is just the best estimate of the object given by the conflicting data. The same effect occurs if the object to detector distance is not measured accurately, something which is always poorly calibrated in the conventional transmission microscope where the intermediate lenses are used to form the diffraction pattern.

In X-ray ptychography the object to detector distance is fixed, and can be measured very accurately. The stepper motors used to scan the object are also usually well calibrated, as is the focal length of a zone plate lens. It is also easy to measure the distance the object has been moved out of the beam crossover,

again by using a stepper motor in the z-direction. Apart from inputting the correctly defocussed probe function at the start of the reconstruction algorithm, which can be immediately calculated from these experimental parameters, the problems described above rarely apply.

5.6) Diffusers

As we remarked in Section 4.2, the bandwidth of ptychography in the sense of the transmission line in Figure 1d is a function of both the probe structure and object structure. An object that has broad flat features, i.e. one that has low entropy, is in general more difficult to reconstruct. The probe and the object appear equivalently in the mathematics of ptychography, except the probe contributes to every diffraction pattern, whereas each region of the object is only expressed in a few diffraction patterns (an exception to this is near-field ptychography). It is incontrovertible that having no structure in the illumination – a flat plane wave covering the whole of the object plane – cannot possibly give us any ptychographical information at all. It would seem logical therefore, that having a probe function that has lots of structure can greatly reduce the likelihood of encountering a probe-specimen combination that cannot easily reconstruct.

Think of a probe composed of random phase and modulus. The randomness appears in both real and reciprocal space, in the latter appearing as a well-developed speckle pattern. As we discussed in Section 4.6, this sort of pattern is excellent for reconstructing gaps in the data that have not been recorded in the diffraction plane, either because of missing pixels, or because the detector is too small so that intensity has fallen outside it. At the other extreme, a simple large aperture with flat phase has a tiny Airy disc response in the far-field, so the ptychographical convolution at any one pixel is only substantially affected by a few pixels around it. The diffuse probe would seem to constrain the data set much more effectively than a simple probe.

The choice of probe positions also matters. If we integrate all the flux that has passed through the object, summing up the intensity that arrived at it from all the probe positions, it would be unfortunate to find that some areas had not been illuminated. We could never possibly reconstruct the object at those points. A probe with strongly varying random modulus would be likely to span the object with a relatively even total flux. Sharp features in the probe also make the intensity at each detector pixel change more rapidly as a function of probe position, which would seem to put more information into the recorded data. Another issue is the bit depth of the detector. An even speckle pattern is more likely to optimise the total information content read out from the entire diffraction pattern, especially if the detector is imperfect in any way, because each pixel has made the most of its available dynamic range.

The question of the optimal probe has not been fully resolved. Some insight can be offered by the WDD method, which requires a division by a function that depends on the probe. If the probe is made in such a way that this division is stable (i.e. the Wigner distribution relating to it – see equation 32 - has few

minima), then there is some evidence that the reconstruction is more stable, noise robust and accurate[70, 71]. Suffice it to say that diffusers, placed at one position or another in the optical path, generally improve the reconstruction [optical useful resolution].

5.7) Bragg ptychography

One of the principle applications for cCDI is a method of using a Bragg diffracted beam from a small crystalline particle [robinson]. The configuration has two principal advantages. By tilting the object through a small angle, many diffraction patterns can be recorded as the 3D Bragg reflection is scanned through the Ewald sphere, thus plotting out the intensity of the 3D Fourier transform of the object. Using the knowledge that the particle is finite, one can then use single shot Fienup-type iterative methods to recover the phase of the volumetric plot of the reflection and thus reconstruct the shape of the crystal. More interestingly, any departure in perfect crystallinity will alter the intensity and phase of parts of the reflection. The Bragg condition by definition assumes a fixed phase relationship between all the scattering points (atoms) in the object. If atoms become displaced, say by a strain field, then these relative phases change. Similarly, the real space reconstruction of the object will have internal phase shifts mapping the strain field[100].

Hruszkewycz et al. were the first workers to demonstrate experimentally that the same principle can be applied to ptychography [101]. They investigated the strain of an epitaxial SiGe layer grown on silicon-on-insulator (SOI) device. The geometry of the experiment is shown in Figure 40. A cross-sectional TEM image of the object and the measured strain maps are shown in Figure 41. The phase of the ptychographic reconstruction gives a direct measure of the displacement of the atom planes in the SiGe; the derivative of this gives the slope of the planes, which can be mapped out and compared with calculations. More recent work has demonstrated that the method can also be extended to mapping 3D strain in semiconductors[102, 103].

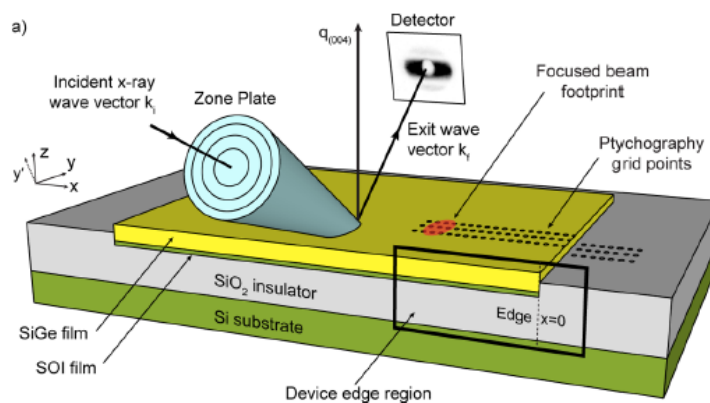


Figure 40: Geometric set up for Bragg ptychography, reproduced from [101].

Bragg ptychography has potentially very important applications in the semiconductor industry, where strain induced by epitaxy of materials with dissimilar unit cell size can be used to control the nature of the band-gap. Although local strain can be measured by electron microscopy, the need to prepare a thin sample leads to relaxation of the strain: inference of the original bulk strain is then difficult. X-ray Bragg ptychography can deal with bulk materials in their original state of strain, although there are limitations on the depth of penetration into the surface of the material and, relative to electron microscopy, the resolution of the technique.

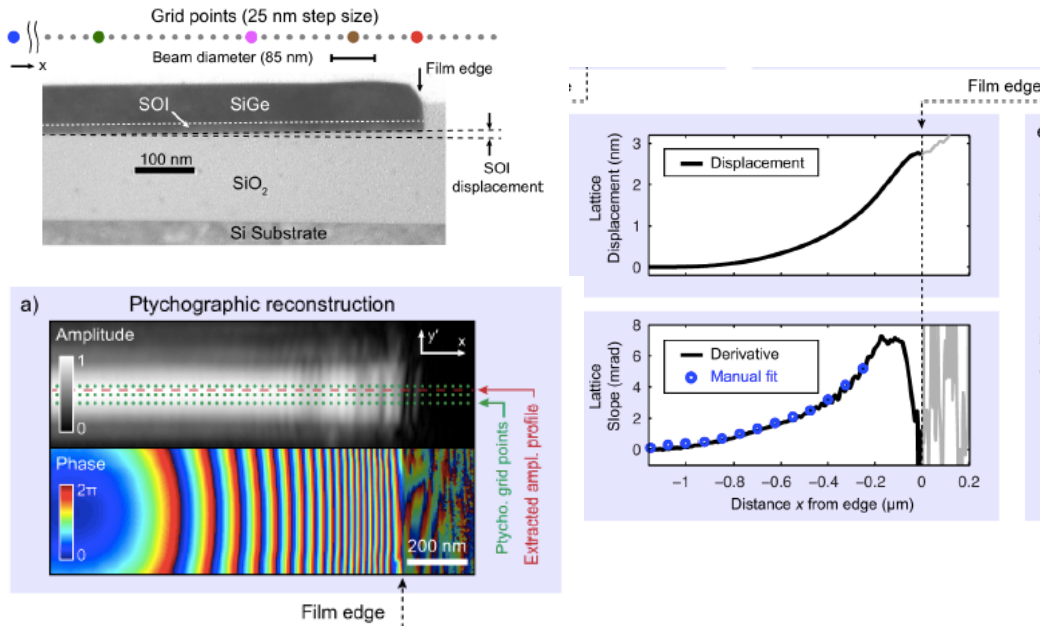


Figure 41: Example results from a Bragg ptychography experiment, reproduced from [101]. Top left is a cross-sectional TEM view of the sample. Bottom left is the ptychographical reconstruction in modulus and phase (colour coded). The phase is proportional to displacement from the unstrained condition. On the right, this, and its derivative, is plotted, the latter being proportional to the curvature of the atomic planes. The data is compared with calculation.

5.8) Visible light reflective ptychography

Visible light has been used to demonstrate ptychography in the reflective configuration, with both the illumination and detector normal to the surface of interest, as shown in Figure 42. Clearly, the phase of the reflected beam is sensitive to surface topology, and vertical sensitivity has been shown to be comparable with white light metrology [104]. The comparison is shown in Figure 43. There is a very wide array of competing surface topology measurement techniques, and so it is unlikely that visible light reflective ptychography will have wide application, even though these early result could be significantly improved upon. A complication is that to measure a structure with

vertical features larger than half the wavelength, multiple phase wraps abound in the image: a very serious problem when a large step change in height is encountered. A solution is to employ a second colour of light, in a second experiment, very close in wavelength to the first, thus generating a large artificial wavelength by forming the difference between the two phase images, as demonstrated in [104].

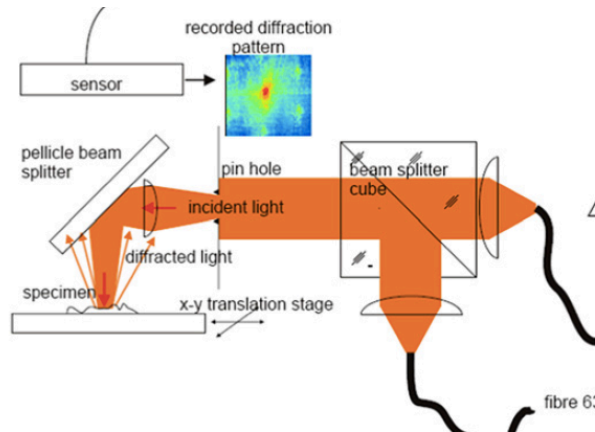


Figure 42: Set up for visible light ptychography in the normal incidence reflective mode. In this case, two sources, very close in wavelength, enter on the right. By switching between them, a long synthetic wavelength can be generated by combining two reconstructions. Taken from [104]. See main text.

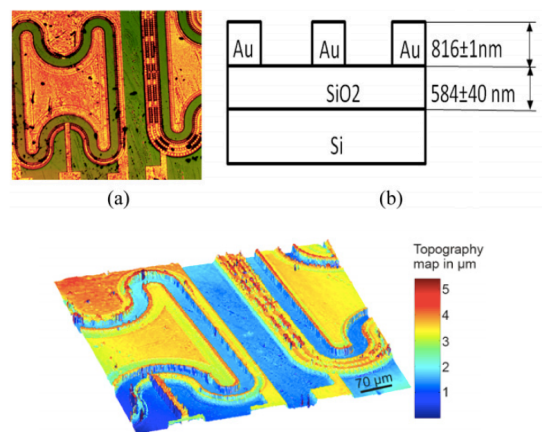


Figure 43: (a) White light interferometry of the test structure in (b). (c) The reflective ptychographic reconstruction from the same object. Reproduced from [104].

5.9) Transmission and reflection EUV ptychography

A considerable limitation of X-ray ptychography is the need for a synchrotron to obtain high flux and high coherence. Beam time is scarce, so experiments cannot be easily refined during a single scheduled run. A promising alternative is to use a higher harmonic source or laser produced plasma EUV sources in the ordinary laboratory environment. A coherent laser source can generate pulses which are very well controlled, both spatially and in time, using all the many optical techniques nowadays used for femto-second studies. Such pulses can be passed through a non-linear medium, such as a gas. In the pulse of intense electric field, electrons are almost dissociated from their respective nuclei, but accelerate and decelerate passing through the atomic potential, thus adding harmonics to the transmitted EM wave. In this way, EUV radiation is produced.

As far as ptychography is concerned, the huge benefit of this method is that the source of radiation is essentially fully coherent, unlike a synchrotron that relies on a large distance between source and optics (and thus the consequent loss of useful flux) to achieve spatial coherence.

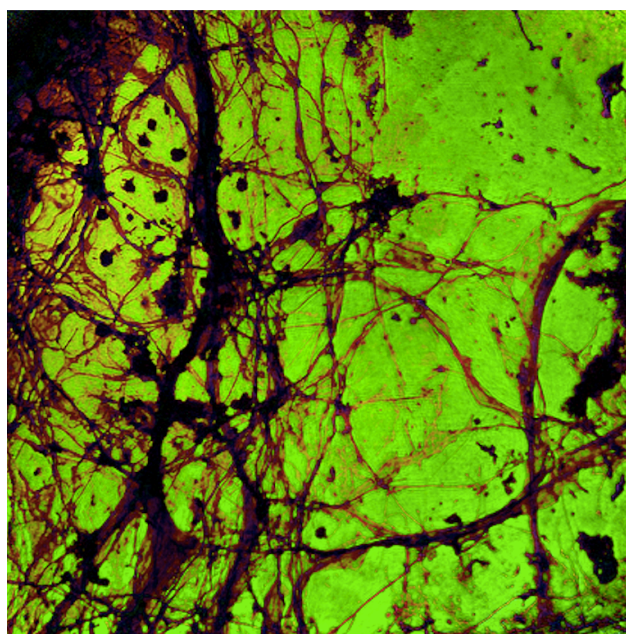


Figure 44: Transmission EUV ptychograph of rat's neurons. Colour and modulus coded as in colour wheel. Courtesy of Jo Bailey, John Chad, from the Centre for Biological Sciences, University of Sheffield, and Magdalena Miszczak, Michal Odstrčil, Peter Baksh, and Bill Brocklesby, from the Optoelectronics Research Centre, University of Southampton.

EUV ptychography can only image thin and weakly scattering transmission specimens. Figure 44 shows a ptychographic reconstruction of cells from a rat's brain. The resolution is about that of a visible light optical microscope. A disadvantage of EUV is that the specimen must be held in vacuum, which means that it is not possible for biological structures to be imaged wet, non-desiccated or live. However, there may be many other potential applications to very thin objects which otherwise do not scatter light strongly. Note that this

reconstruction, used the varying probe algorithm described in section 4.1. Another important application is to surface topology measurement using glancing angle reflection, an example of which is shown in Figure 45[105]. In this geometry, care must be taken to map the detector coordinates to the scattering configuration and the elongated probe shape and phase. The method has very promising applications in high-resolution semiconductor metrology.

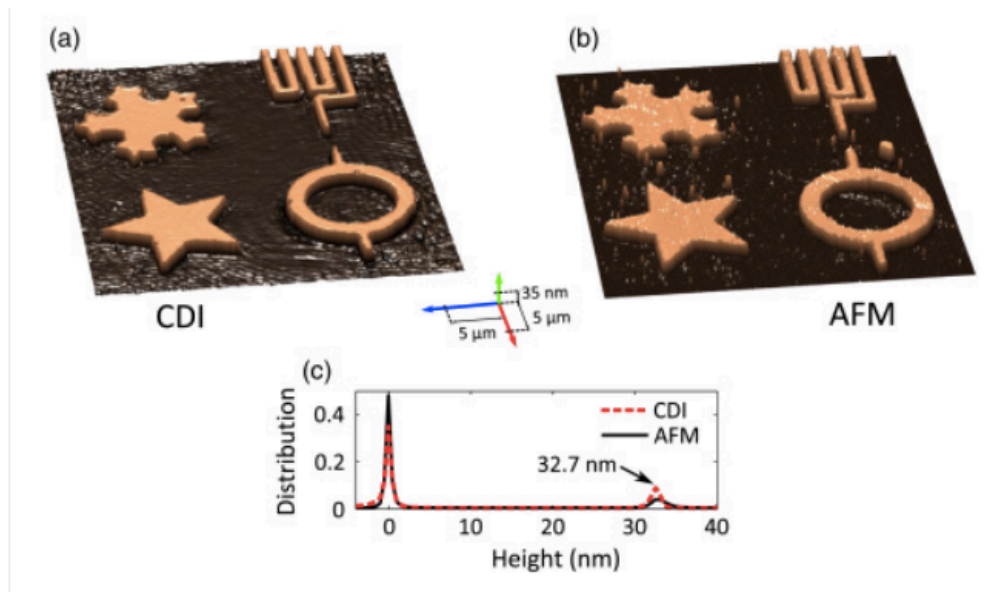


Figure 45: (a) Modulus and (b) phase of a reflective EUV ptychograph of a test object. The phase is essentially a topographical plot of the object. The reconstruction (from [106]) compares favourably with the AFM and SEM image (not shown).

5.10 The simple aperture

The simplest ptychographical set up imaginable comprises a source, an aperture, a moveable object and a detector, with no lenses and no other optical components. This was the original goal of totally lensless imaging, and at first ptychography seemed to liberate imaging from the need of any sort of interferometer or lens at all. But despite its simplicity, the aperture-only set up should be avoided if at all possible.

The biggest problem is the very bright central spot of the diffraction plane. All types of detector find this hard to handle. Single photon/electron devices are count-rate limited, so that the full flux of the source cannot be employed. Over-exposed CCD pixels bleed charge into adjacent pixels. A central stop can mitigate the problem, but the powers of ptychography to fill in missing pixels are at their weakest when the probe function in reciprocal space (in this case an Airy disc) is

so narrow, as we discussed in Section 5.6. Losing this low frequency information leads to unwelcome large-scale distortions in the image.

5.11 Probe reconstruction in the Fresnel configurations

Many of the configurations we have discussed have been described in terms of the detector lying in the Fourier domain. In fact, it is often convenient to place the detector, or its conjugate equivalent, nearer to the object. So for example in the case of SAP, the diffraction lens can be defocused to avoid the high intensity zero-order diffraction peak.

One may suppose that the Fresnel integral must be used in the reconstruction process, and that consequently the exact distance from the object to the detector must be known. In fact, if the reconstruction simply assumes the detector is in the Fraunhofer plane, the object function appears as usual. However the probe function will have a phase curvature over it, with a radius equal to the object to detector distance. Without deriving the reason for this formally, we observe that the phase has the effect of a computational lens, steering parallel beams (which correspond to the Fourier integral) to a focus on the detector. The approximation is only true for small scattering angles, but is another example of how ptychography can self-calibrate.

6) Volumetric imaging

A two-dimensional picture of an object is good, but imaging it in three-dimensions is greatly more informative. In the field of biological imaging, cell colonies grown on a flat piece of glass cannot possibly satisfactorily model their development in a natural three-dimensional tissue structure. There has therefore been a huge investment in developing reliable volumetric imaging methods, most notably in the visible light domain with confocal scanning microscopy. This is now the workhorse of many biological studies. The ability to label and map the distribution of individual proteins is a powerful component of technique, allowing detailed studies of how genetic information is expressed within different parts of a cell (see Chapter ****EDITOR****).

Material science also has a pressing need for three-dimensional information. One of the biggest weaknesses of electron microscopy has historically been the projection effect. All the three-dimensional information in the object is concentrated into a two-dimensional image, rather like a shadow image. Electron tomography ‘add-ons’ are now supplied by most electron microscope manufacturers for moderately low resolution reconstructions, and the most recent research is now demonstrating atomic resolution in 3D, which is as much as can ever be hoped for (see Chapter ****EDITOR**** of this volume).

There are two very different ways of undertaking 3D imaging via ptychography, which we discuss in the next two sections. The first is an extension of conventional tomography, which puts together many ptychographical images recorded at different object rotations: we call this ptycho-tomography. Alternatively, a single data set is used: the probe is scanned as usual but without

rotating the sample. The reconstruction procedure is then via a multislice update, wherein the propagated wavefront through layers of the object is reconstructed for every probe position and every layer in the object. Unlike ptycho-tomography, this ‘multi-slice’ method can account for multiple scattering in the object.

6.1) X-ray Ptycho-tomography

To date, ptychography has had its biggest scientific impact in volumetric imaging at high-resolution: X-ray ptycho-tomography. Its first application [20], see Figure 46, was at hard X-ray wavelengths for which it is ideally suited. Hard X-rays can penetrate thick objects, which is clearly good for tomography. They can also pass through air without creating too much unwanted scattering, unlike soft X-rays where the object must be in vacuum or close to very thin transparent windows upstream and downstream of the object, which themselves create unwanted scattering. But at high energies X-rays pass through an object often with very little absorption. It so happens that the real component of the refractive index of many materials at these energies (which induces a phase change in the X-ray beam) is much larger than the imaginary component (which determines absorption). The image phase of a ptychograph is therefore of a much higher contrast than the conventional absorption signal. Even better is the fact that phase accumulates linearly as a photon passes through an object, and the rate of that accumulation is related to the refractive index, which is material dependent. This means that the phase image really is the linear projection of the matter within the specimen. All this, combined with the much-enhanced resolution of ptychography over other X-ray methods, means that there is a huge application niche for the technique in both material science and biological science.

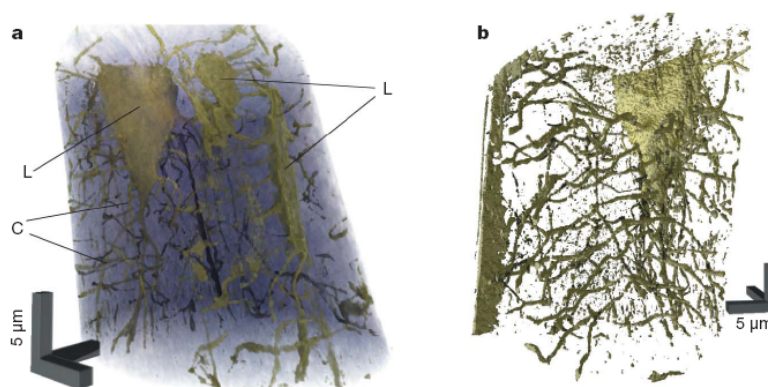


Figure 46: The first reported example of a 3D X-ray ptycho-tomographic reconstruction [20]. The sample is bone. It is the phase signal that allows for the high-contrast ptychographical imaging of biological structures, which otherwise do not absorb strongly at hard X-ray wavelengths.

The pioneering work at the Swiss Light Source in the Institut Paul Scherrer has refined the technique so that nowadays it is used as a routine method which can

analyse and image all sorts of materials: the list of publications dedicated to specific science problems is far too numerous to list here. One example is the rather nice series of tomographs showing the in-situ fracture of a microcomposite in Figure 47. Figure 48, showing a tomographic reconstruction of a significant volume of an Intel device, is a recent example at the time of writing. The extraordinary size, detail and resolution of the reconstruction is stunning. Figure 49 shows a detector device (of the same type used to collect the data). The experimental reconstruction is so good it looks almost like a CAD drawing.

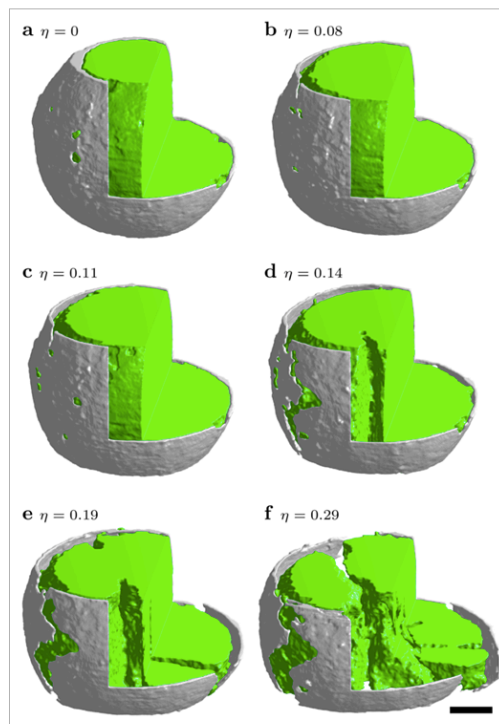


Figure 47: In-situ ptycho-tomography time series of the destruction under compression of a micro-composite. See [107] for more details.

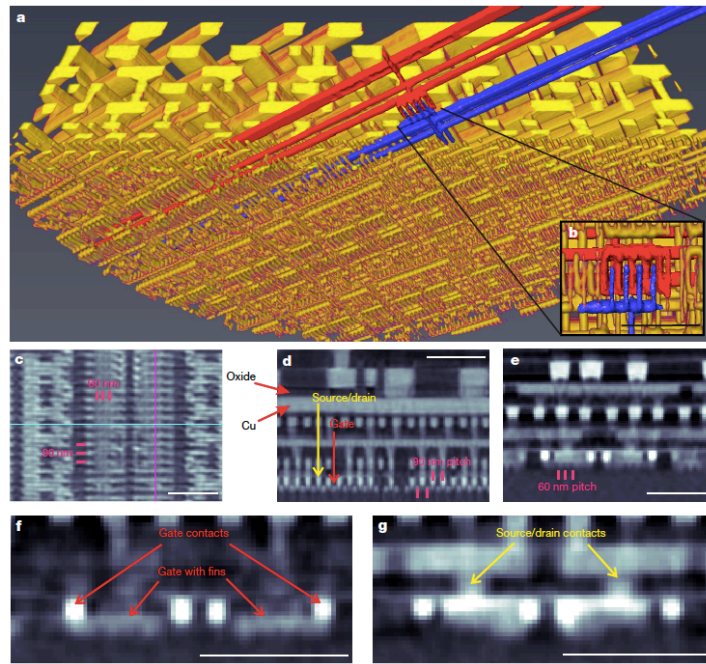


Figure 48: Ptycho-tomography of a volume of an Intel microprocessor. Scale bars are all 500nm. Reproduced from [96]

Ptycho-tomography encounters all the usual problems of tomography, such as registration of successive projections. Thermal instabilities inducing specimen drift during a long scan can mean that a poorly mounted specimen moves out of the field of view, etc. When very high resolution is required, these issues are best addressed by investing in very high quality stages with laser interferometric feedback. A computational complication is that a thick object will induce phase wraps in the image. This is a dramatic non-linearity within an otherwise excellent linear signal. Luckily, there are numerous ways of handling phase wraps: it is a very large field in its own right, but care must be taken.

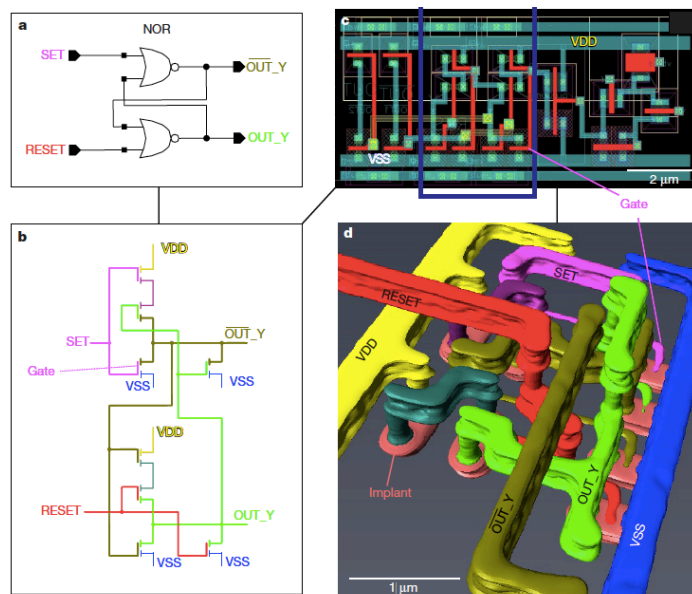


Figure 49: Ptycho-tomography of a detail of a solid-state hard X-ray detector, from [96]. The same type of detector was used to collect the data. The circuitry, shown

schematically on the left hand side in (a) and (b), is explicitly visible in the experimental reconstruction (d). (c) is the blueprint for the design used in the manufacturing process.

6.2) Multi-slice reconstruction

In everything we have discussed so far, the object function has been modelled as a two-dimensional transmission function. So for example in hard X-ray ptychotomography, any one ptychograph is treated as a projection of the electron density through the whole (thick) object onto a two-dimensional surface normal to the incident beam; an assumption that is implicit in the back-projection methods used in the tomographic reconstruction. Similarly, a thin, weakly scattering object in transmission electron microscopy is accurately approximated as a 2D projection, constituting an integral of the 3D atomic potential of the object along the direction of the optic axis. Indeed, for a long time the projection effect in TEM was one of the technique's key weaknesses, in that detailed understanding of an atomic arrangement, say occurring at the interface of two crystallites, could only be easily undertaken if the pertinent feature repeated itself along the beam direction. Atomic scale tomography (Chapter **EDITOR**) is nowadays making significant progress in tackling this problem.

The 2D approximation breaks down for two reasons. The first arises from the geometry of the rays scattered by features in the object that lie at the same x, y point in the 2D plane of the projection, but are separated in the z -direction, parallel to the optic axis. With reference to Figure 50, the 2D approximation assumes the diffraction pattern at a particular angle arises from the path difference (and hence phase difference) between any two points in the object plane (like points B and C). When these are separated along the optic axis (A and B), an extra path difference is introduced, shown as Δp , meaning that the 2D Fourier transform can no longer be used to calculate the diffraction pattern. The effect can also be thought of in terms of the curvature of the Ewald sphere in reciprocal space [108].

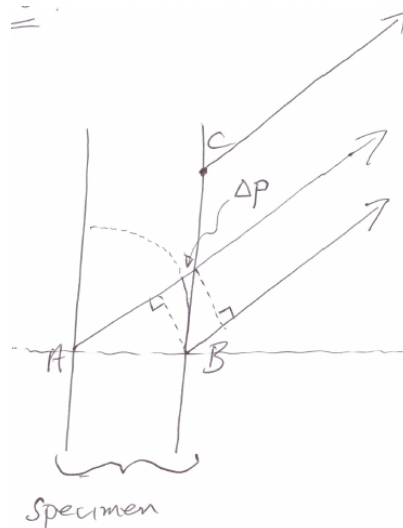


Figure 50: In the forming the Fourier integral, parallel rays from a single surface of the object are summed (e.g. points B and C). When the object is thick, rays from points in the same (x, y) position (A and B) have an extra path length introduced, indicated by Δp . The

geometry is best handled by computing the scattered amplitude of where the Ewald sphere cuts the 3D Fourier transform space of the whole object, at least in the first Born approximation.

A second effect is multiple scattering (or, in the parlance of electron microscopy, dynamical scattering). The mathematics of ptychography, which has no constraints on the form of the specimen function or the illumination function, can deal with an arbitrarily strong 2D object (i.e. one with very strong phase and modulus changes within it). Strong phase can be represented by a Taylor series expansion of $e^{i\phi}$, which leads to a diffraction pattern that can be formulated as multiple convolutions, equivalent to multiple scattering [109]. However, in practice, strong phase requires a substantially thick object. The geometric and multiple scattering 3D effects then become intermixed so that the exit wave bears little or no relation to the projection of the object. This is particularly problematic for electrons, which for many materials of interest scatter very strongly.

There are various ways of calculating the effects of thickness-induced phase changes and multiple scattering. One of the most common and flexible approaches used in electron microscopy is the multi-slice method originally proposed by Cowley and Moodie [110]. In this, the 3D object is represented by a series of 2D slices lying normal to the optic axis. The layers are assumed to be transmission functions (like those in everything we have discussed so far) that are thin enough to satisfy the two-dimensional approximation. The incident wave forms a product with the first layer in order to calculate an exit wave from that layer. The exit wave is then propagated, via the Fresnel integral, angular spectrum method or similar, to the second layer, where it forms a new incident wave. The exit wave from the second layer is the product of its transmission function with this new incident wave. The process – product of incident wave times transmission function, propagation, product of new incident wave on the next layer, etc. – is used through the whole specimen. The Fresnel propagations account for the geometric breakdown of the two-dimensional approximation, and the serial scattering from each layer accounts for multiple scattering.

However, this technique is only appropriate for a forward calculation: given a model object, we can use it to calculate what the exit wave will look like. In electron microscopy, much work has been spent altering specimen models in order to find a good match with the measured bright-field high-resolution image, which itself is an interference pattern altered by the additional effects of the transfer function of the lens. Even if the exit wave can be measured in modulus and phase, say via a through-focal series, no one image can be inverted directly to give the 3D object: a 2D image does not have enough measurements in it to solve for all the many layers of the 3D object.

The same does not apply to ptychography, where it is now well-established that the data collected in a single ptychographic scan can, surprisingly, solve for many 2D layers within the object, at the same time removing multiple scattering effects and calculating the evolution of the incident radiation as it propagates through the object [58, 111-113]. Once again, this is possible because of the enormous diversity in ptychographic data.

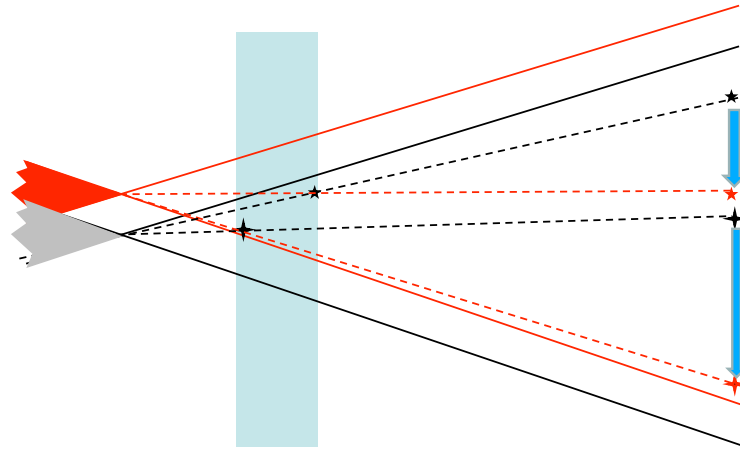


Figure 51: Simple ray diagram illustrating why ptychography (probe movement) encodes 3D information. For a convergent probe (or any localised probe) features in the object cross the shadow image at different rates (for a constant probe shift speed), according to their depth in the object. In reality, wave interference effects greatly complicate the diffracted information, but the latter is still encoded with similar information.

We note that the probe is localised and so is necessarily composed of a sum of incident plane waves, which have a significant range of incident angles (k -vectors). A simple ray diagram, illustrated in Figure 51, suggests that as a defocused STE/XM probe is scanned laterally, features in the object at different depths will appear to move over the shadow image at different rates. In reality, for finite wavelength, interference effects dominate the diffraction plane and in real-space the probe can have very complicated wave structure. But this mode illustrates that 3D information affects the recorded data, and so in principle can be extracted from it. Ptychographical translation diversity also means that we get a different exit wave for each probe position, unlike the single exit wave in conventional imaging. If the step size (sampling) in real space is small, there exists hundreds or thousands of exit waves to process: plenty of data to provide multiple slices in the object.

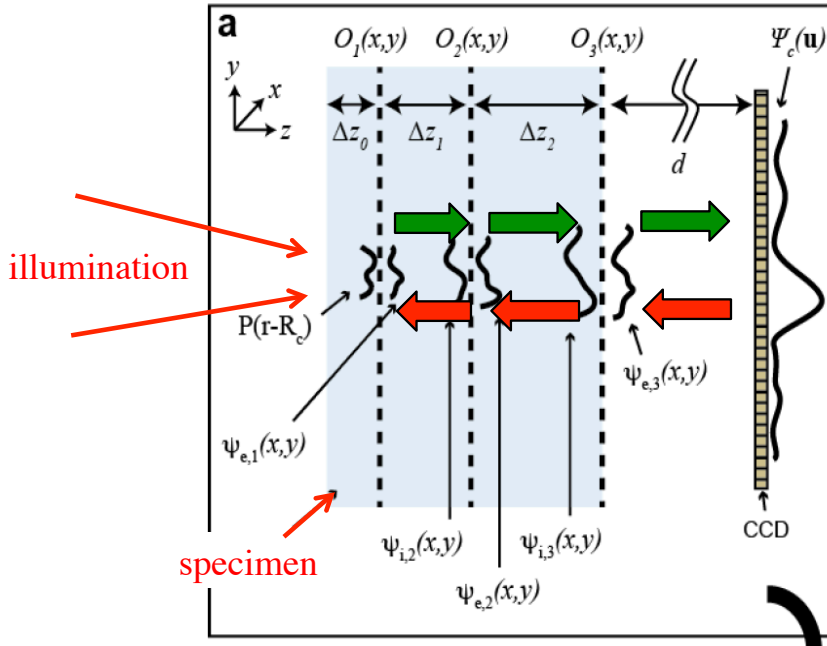


Figure 52: The inverse multi-slice method. At each layer of the specimen, the incident wave from the previous layer is treated in the same way as the probe in a 2D ptychographical reconstruction. The forward calculation (green pointers) proceeds as usual (see main text). The inverse calculation uses the normal update of object layer and incident wave, at each layer. The updated incident wave is back-propagated to be used in the update for the previous layer, etc.

The first algorithm to demonstrate multiple-layer reconstruction computationally reversed the forward multi-slice calculation [58], as shown in Figure 52. To start, the forward calculation is carried out, but each incident and exit wave from each layer is stored for later use. As usual, there is a running estimate of each layer of the object and also of the probe incident upon the first layer. After undertaking the forward calculation to give an estimate of the diffraction pattern, the detector modulus constraint is applied as usual. Back propagation gives us a new estimate of the exit wave from the last layer. The last layer of the object is then updated as usual for the two-dimensional case (equ 6 or 9), except the role of the probe is replaced by the incident wave at the last layer calculated from the forward calculation. This incident wavefunction is then also updated as usual as if it were the probe, and then back-propagated to the second from last layer, where the procedure is repeated using the stored incident and exit waves at that layer from the forward calculation, and so on and so forth. Finally, the actual probe function incident on the first layer is updated and used for the incident wave at the next probe position to be processed.

Figure 53 shows a visible light example of a 3D reconstruction through slices of a root. It compares favourably with the confocal microscopy image of the same object. The data reconstructed 34 layers of the object, each separated by 2μm. Only 5 of the reconstructed slices are shown. In generating such an image, the algorithm had to calculate two images (modulus and phase) for each layer, two images for the probe, and two images for the exit waves from each layer: i.e. 138 two-dimensional images from one ptychography experiment. (We note that the incident waves are uniquely defined by propagation from the previous exit wave,

so they do not constitute independent variables.) Ptychography is indeed a very information intensive technique. On the other hand, we know that two lenses in the confocal configuration can obtain all this information: ptychography just happens to do it in a computer. Similar results have been obtained in X-ray ptychography [114].

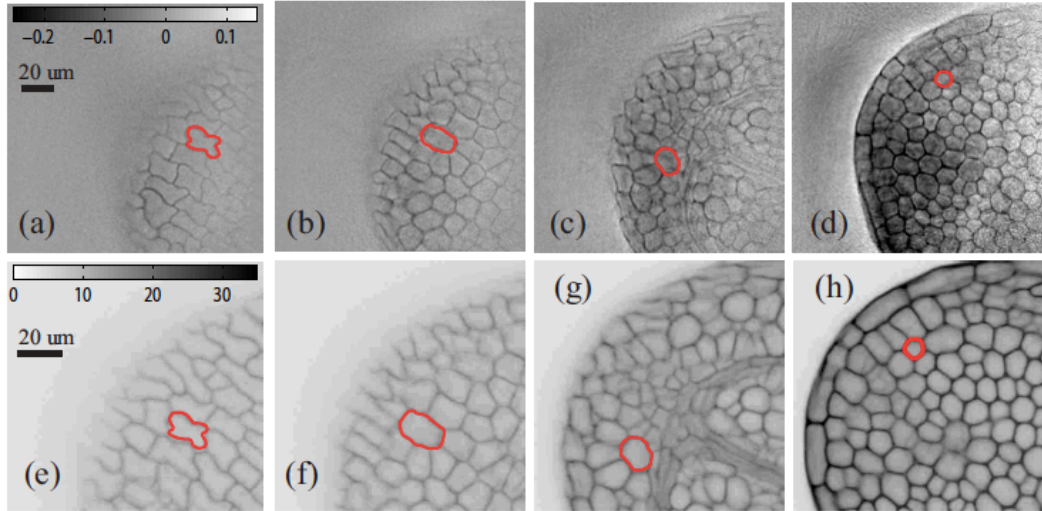


Figure 53: Top: selected slices from a ptychographical multi-slice reconstruction of an embryonic root tip [21]. Bottom: comparison with conventional confocal images of the same slices. Ptychography does not require the specimen to be labelled or stained.

Fourier ptychography (Section 5.2), which of course contains identical 3D information, is usually thought of as solving for the diffraction pattern lying in the back focal plane. The multiple layers cannot be solved for there because as the illumination is tilted, the Ewald sphere rolls through reciprocal space, so the diffraction pattern changes as it is moved. However, the lens and aperture transfer function can be regarded as a propagator between the exit surface of the object and the detector plane (the image). Diversity arises from the different incident waves angles so that the inverse propagation gives an equivalent result.

In what we have described, both the forward and back propagation depends on knowing, or estimating, the separation of the layers and the refractive index of the 'free space' between them. If either of these is wrong, the propagation integrals give the wrong wavefunctions, and so the reconstruction algorithm does not converge. However, these can just be put into the algorithm as another set of free variables, as shown by [115], and illustrated in Figure 54 in the case of multislice X-ray imaging.

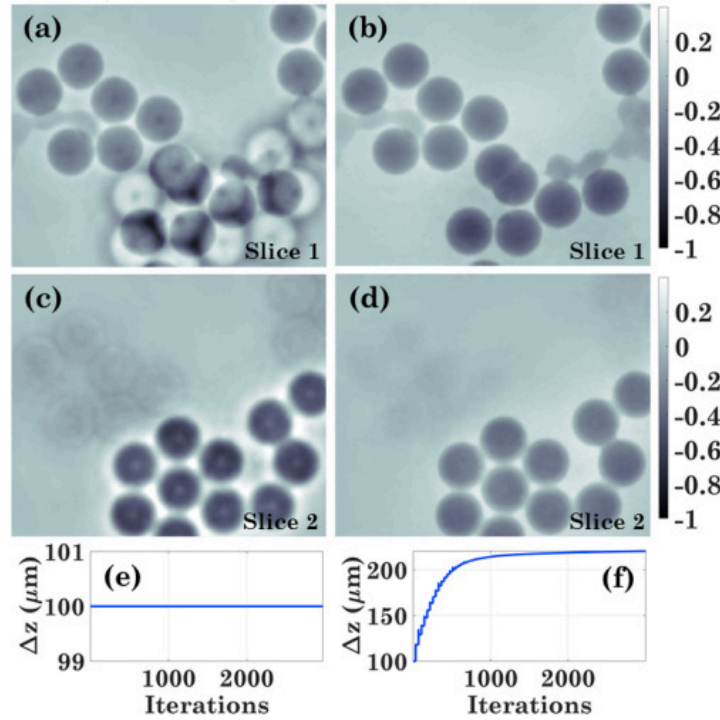


Figure 54: X-ray 3D multi-slice reconstruction. All are phase images. (a) and (c) reconstruct two layers, but their separation is assumed known and fixed. (e) plots the separation as a function of iteration (fixed). (b) (d) and (f) are similar, except here the separation is also recovered as a free variable, greatly improving the reconstruction [115].

This particular multi-slice formulation also does not account for backwardly propagating waves that have been reflected off the layers: forward-only scattering is a good approximation for the behaviour of high-energy electrons and X-rays, but not for visible light. Ever more comprehensive search algorithms within larger solution spaces may accommodate these issues.

The depth resolution of the technique clearly depends on the angles subtended at the specimen by the illumination pupil and the angular size of detector, but it is also affected by the strength of the scattering from one layer to the next. A strongly scattering layer increases the range of incident angles upon the next layer, and hence the potential lateral and depth resolution. A weakness of the approach is that because ptychography relies on coherent wave inference, the 3D transfer function in reciprocal space is doughnut shaped: at high or low resolution, the depth resolution is very small [112]. This performance compares poorly with the transfer characteristics of confocal microscopy, where the contrast mechanism arises from incoherent fluorescence. Intensity in real space means that the transfer function in reciprocal space is the autocorrelation of the coherent transfer function, which has the effect of filling in the missing low frequencies, enhancing both lateral and depth resolution. There has been work on incoherent optical Fourier ptychography using structured illumination [90], which could be a truly revolutionary development in the technique.

A potentially important application of multi-layer reconstruction using visible light is to image large biological cells, or clusters of cells, without having to kill or stain them: in ptychography strong contrast arises from the real part of the refractive index, which is expressed in the phase of the transmission function. This could be useful for, say, checking the viability of human embryos before implantation. X-ray imaging is less dependent on the breakdown of the projection approximation because the scattering angles involved are very small and so the depth of field is generally much larger than the thickness of the object. Reversing and removing multiple scattering effects in electron microscopy via ptychography could represent a major breakthrough, overcoming one of the biggest limitations of imaging with electrons, although whether this will be possible remains to be seen. We note that the WDD method can also extract depth information, but this has only been demonstrated for weakly scattering objects [9]; see section 10.6.4.

7) Spectroscopic imaging

One of the most common and useful ways of mapping elemental distributions in specimens is to collect the fluorescent X-ray spectrum from the object while it is being irradiated by a scanned focused probe of high-energy electrons or X-rays. As long as the incoming beam has sufficient energy, it can eject inner electrons from the specimen atoms. Electrons that then fall into the resulting empty core state can irradiate X-rays which have characteristic energies specific to the particular element, see Chapter ****EDITOR****. This fluorescent signal is incoherent and so it cannot be used in conventional ptychography (although see [90]).

However, we can plot the distribution of a given element using coherent ptychography if we take two images, one above and one below the absorption energy of the core state. Figure 55 shows an example taken of a fibroblast cell that has within it cobalt ferrite nanoparticles [72]. These are not visible in the image taken at 703eV, below the absorption edge, which is at 710eV, but are visible in the image taken above the absorption edge, at 712eV. Interestingly, the phase of the absorption can also be measured. Figure 56 shows an example of a very high resolution map of two separate iron compounds within a particle, scanned as a function of energy [79]. Each point in the image has a different spectral response. Because the shape of the absorption lines depends on the local bonding environment of the iron, the authors were able to map the relevant compounds using principal component analysis. The authors compare the resolution of the same type of analysis undertaken with a focused probe STXM with a 25nm optic. The resolution of the ptychographic chemical map is estimated to be 18nm, compared with 70nm for the STXM data.

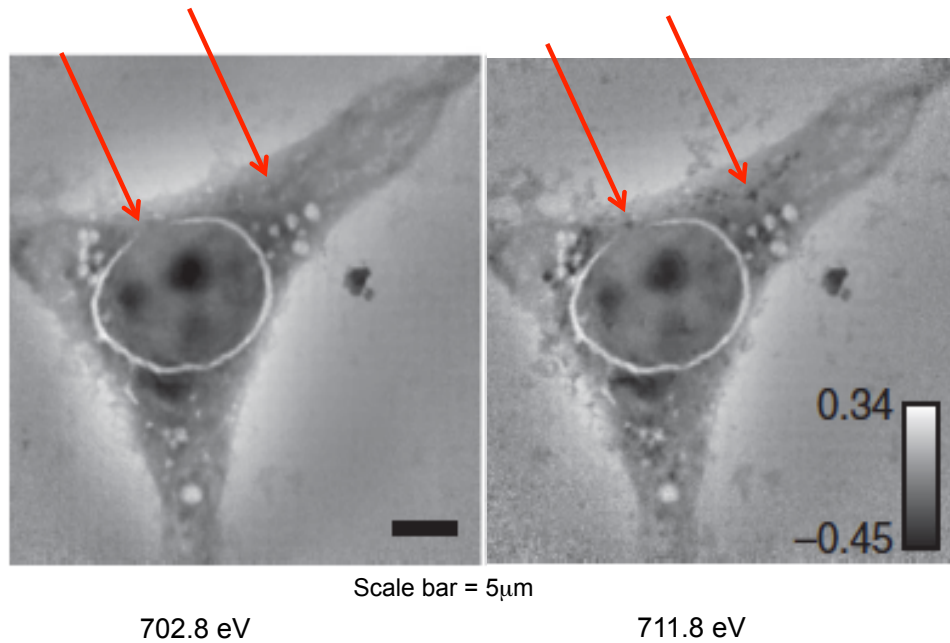


Figure 55: Soft X-ray phase ptychographs of a Balb/3T3 mouse fibroblast, marked by CoFe_2O_4 particles, taken below (left) and above the absorption edge of Fe [72]. The distribution of iron is clearly visible in the latter.

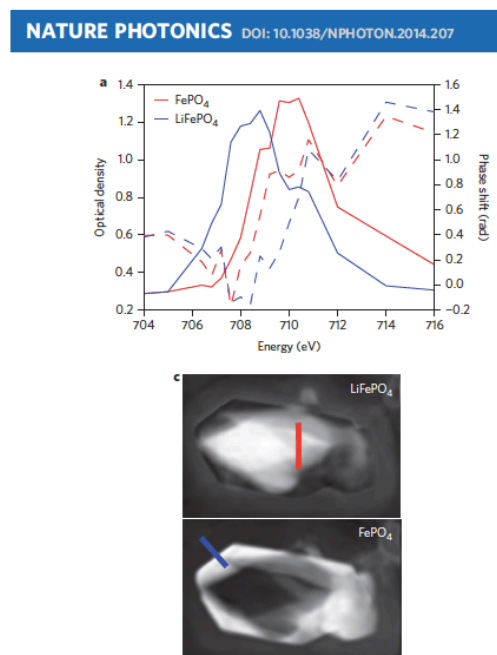


Figure 56: High-resolution X-ray ptychographical chemical mapping of FePO_4 and LiFePO_4 in a small particle [79]. By taking images at different energies, the loss peaks (top) can be used in a principle component analysis to map the two compounds (bottom).

8) Mixed State Decomposition and Handling Partial Coherence.

We saw in Section 4 that a typical ptychographical data set is extremely rich in diverse information. This can be used to correct automatically many imaging parameters. In Section 5.2 it was found we could extract even more information. Provided the sampling in both real and reciprocal space is dense so that the minimum sampling condition defined by Equation 7 is well surpassed, we have seen that we can solve for dozens of 2D layers through the object thickness.

Thibault and Menzel [48] proposed one of the most important extensions for the use of information diversity in ptychography. An assumption of the phase problem is that when we measure the intensity of a pixel it has associated with it one modulus and one lost phase. The pixel has to be small enough so the wave does not vary substantially across its width, i.e. the sampling condition is fulfilled. But what happens if two separate non-interfering waves (i.e. ones that are incoherent with respect to one another) are incident on the detector? We only measure one intensity, but now we have lost the two phases, and, even worse, the two moduli as well. We seem to have four unknowns where before we had only one unknown. In fact, in this case we have only three unknowns because we know the intensities of the two moduli must add up to the measured intensity, a piece of information that will be key.

There are many situations where this occurs in practice. X-ray and electron sources are mostly incoherent across their physical width in the plane of their emission. However, a long way from a small incoherent source, the wave becomes substantially spatially coherent. A star is a huge incoherent source, but seen from earth it twinkles coherently, a result of the Van Cittert-Zernike theorem. Good coherence requires the source to be a very long way from the experiment, but then flux per unit area is low, so we must balance our desire for as much spatial coherence as possible with the competing need for as much flux as possible. Inevitably, there will always a small degree of partial coherence in our wave experiments.

It is not only a diffracted wave that can be a source of incoherence. Vibrations in the specimen or any part of the instrumentation can be equally harmful. These are more generally called state mixtures. Our detector is sampling many different configurations of the experiment during the time it takes to make an exposure. This is equivalent to adding together (incoherently) the coherent waves that would have been scattered from all the different states in the system during the measurement time.

Coherence theory is a large subject area in its own right. One can consider any two points in a wavefield. Each oscillates in time. If they oscillate in perfect synchrony (though usually with different phase) then they are coherent. If there is no correlation between their disturbances, they are wholly incoherent with respect to one another. The general situation lies between these extremes: there is some statistical correlation, but it is not perfect. The coherence function describes the degree of correlation between pairs of points in the wavefield, but this can be an awkward way of analysing the effects of partial coherence. Wolf

[116] suggested a different approach, widely adopted in practical situations. The wavefield is decomposed into a set of modes, each of which is entirely incoherent with respect to any other mode. The modes do not interfere with one another, but can be treated separately, each propagating through the optical system independently. State mixtures in the object and the detector, or any part of the optical system can also be treated as modes.

An example would be modelling the effects of partial coherence caused by having a finite source. The source can be divided up into points, each of which is perfectly coherent. Each source wave (mode) is propagated through the whole optical system to the detector where its intensity is added to the intensity of the other waves that arrive at the detector from all the other point sources. This process might blur the intensity at the detector because the extended source has induced significant incoherence into the experiment. However, if we choose our points on the source to be very close to one another, and the whole source is small, the intensity at the detector from the different points might be, for all intents and purposes, identical. This means that all these different modes are so similar to one another they may as well be treated as one mode. In general, we can decompose a partially coherent wavefront into as many modes as we like, but this is not an optimal representation of its coherence properties. The modes we will talk about here have been orthogonalised with respect to one another. This can be thought of as a sort of principal component analysis, minimising the number of modes we need to describe the system completely.

In quantum mechanics, the density matrix is used to handle mixed states. In any particular representation, the usual single-state operators (for energy, position, momentum, etc) can operate on it. To find the expectation of a particular measurement, the trace of the resulting matrix is formed, which is simply a way of calculating the total probability (expectation value) of making a measurement when two or more states that are incoherent to one another are present in the same experiment. Diagonalising the density matrix is equivalent to finding the set of incoherent states that are orthogonal to one another. A pure state then has only one entry of unity in the density matrix. This type of analysis is now very common in the field of quantum computing, where the decoherence of a wavefunction limits the capability of a real-world quantum computer. Thibault and Menzel wrote their paper casting the ptychographic incoherence problem in these terms. In fact, actually undertaking a multi-modal decomposition in ptychography is computationally very easy and the process is quite intuitive, as we hope to show below, so a reader not familiar with quantum mechanics need not worry about understanding the process from this perspective.

8.1) Visible light model example

We start with a simple experiment using visible light. With reference to Figure 57, we undertake a ptychography experiment where three completely different wavelengths of light (green, blue and red) illuminate the object simultaneously. These three wavelengths are incontrovertibly incoherent with respect to one another. We have one specimen object, but the different colours of light will be absorbed differently in different regions of the specimen if it has any colour

differences within it. To ensure this is the case, the specimen is composed of an artificially manufactured projector slide that has been specially prepared; it consists of three superposed images each of a different colour (Figure 59). (Ideally, the pigments used for three colours would each absorb one, and only one, of the three incident light wavelengths, but this has not been achieved perfectly in this experiment.)

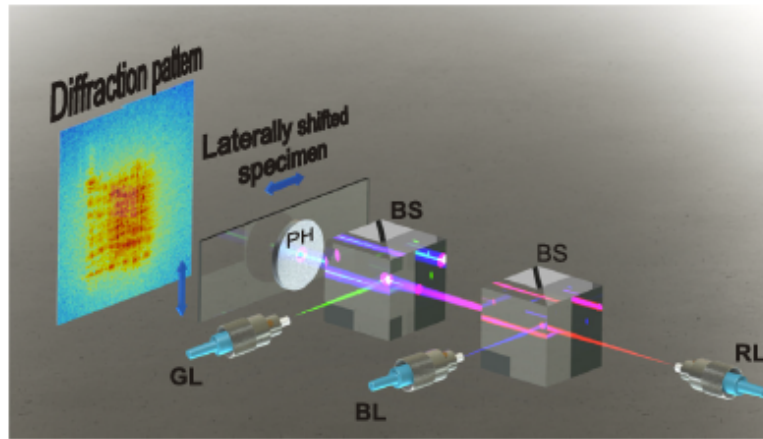


Figure 57: Example of ptychographic multiplexing. Three distinct wavelengths of light are incident simultaneously. The detector is only sensitive to the total summed intensity. From [52].



Figure 58: The test specimen used in Figure 57, from[52]; an old-fashioned projector slide consisting of three super-posed images, each of a different colour. The dyes do not absorb at the laser frequencies in Figure 57 perfectly, so there is cross-talk in the reconstructions in Figure 60.

So, we have three ptychographical experiments going on simultaneously. Each colour of light sees a different sample. The different colours of light also interact with the illumination-forming optics in different ways (diffracting by different amounts before they reach the sample), so that we also have a different probe function for each colour of light. But we can solve for all of these functions, three objects and three probes, using the diversity in the ptychographic data, despite the fact that the intensity from each experiment is collected on the same detector

all at the same time. (The detector is colour insensitive, it simply measures the total power of light incident upon it.)

At first solving for all six functions from this scrambled up data set sounds impossible. Surprisingly, we just have to make one minor change to any one of the common iterative reconstruction algorithms. First, we set up and run three reconstruction iterations simultaneously, each one solving for their respective object and probe functions. The only difference is when we come to applying the detector intensity constraint. We don't know the intensity (and hence modulus) of any one of the colour signals at a particular detector pixel, but we do know the total intensity that they all add up to.

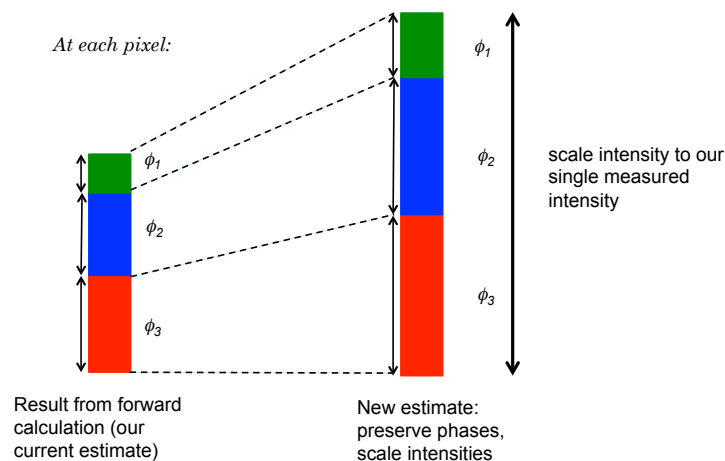


Figure 59: Graphical illustration of the detector intensity constraint when more than one mode is present in a ptychography experiment. See text for details.

In Figure 59, the height of the two columns represents intensity. The first column is the estimated intensity that has come out of our forward calculations (at B from A in Figure 5). The three simultaneous forward calculations have given us three estimated moduli, which have been squared and added together. Each forward calculation also gave us an estimated phase. The height of the column on the right hand side is the measured total intensity. To apply the constraint, we maintain the ratio of intensities of each colour in the estimated intensity, but scale them uniformly to fit the measured data. We now have three new moduli estimates, each the square root of their scaled intensity estimates, plus the three phases that came out of the separate colour iteration loops. These are fed back into their respective iterations at C in Figure 5. Amazingly, after running the iterations as usual, the three reconstructions appear from their respective iteration loops. It helps if the starting estimates of the probe or objects are slightly different so that they can diverge into the separate solutions, but we do not need to know whether those estimates have anything to do with the real functions: it is just an effective way to seed the three separate reconstructions. The form of the constraint being applied – that the sum of the calculated intensities must equal the measured intensity – just has to be true when the solution is correct. Diversity in the data (assuming there is enough) drives the algorithm to that solution.

Figure 60 shows the three object reconstructions. Note that there is some cross-talk between the images, but that is because the dyes in the object slide do not absorb at one wavelength exclusively, so some of the structure from one wavelength is expressed slightly in one of the other images. Even so, they are convincingly separated. Interestingly, each image and each probe reconstruction comes out a different size in their respective object arrays. This is because of the wavelength in Equation 1. The detector pixels are the same physical size for all the reconstructions, so the wavelength changes the magnification in the reconstruction array. This has been adjusted for in Figure 60.

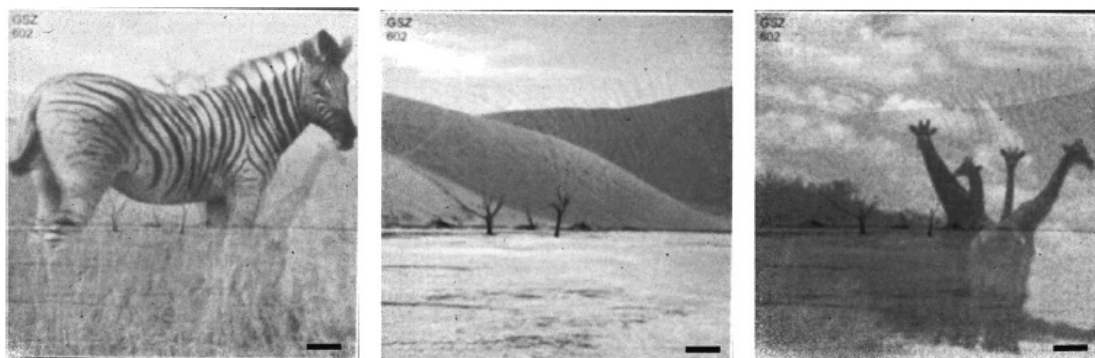


Figure 60: Reconstructions relating to the object in Figure 58. See text and [52] for more details.

8.2) X-ray Illumination modes

Most X-ray synchrotron beamlines have some partial coherence within them, no matter how carefully the optics is arranged. Even if the coherence width at the final slits lying upstream of the experimental set up is estimated to be entirely coherent according to the van Cittert-Zernike theorem, vibration in any intermediate optical element, for example the monochromator, can substantially reduce the effective coherence.

Unlike the light example given in the previous section, under most normal circumstances the object function is fixed, however the partial coherence is equivalent to having multiple modes in the illumination. In Figure 61 we show a multi-mode decomposition of an X-ray probe in the defocused condition (Section 5.5). The reconstruction proceeds in exactly the same way as before. In this case, 8 parallel iterations have been undertaken. There are an infinite number of ways that these modes can express themselves, each being a different representation made up of any linear combination of wavefunctions that might, or more likely, might not be orthogonal to one another. They can be orthogonalised using the standard Gram-Schmidt approach, but this is not fundamental insofar as we can choose any arbitrary vector with which to start the Gram-Schmidt process. Diagonalisation of the density matrix, which can be computationally undertaken with principal component decomposition, does give a unique and the most compact representation of the modes.

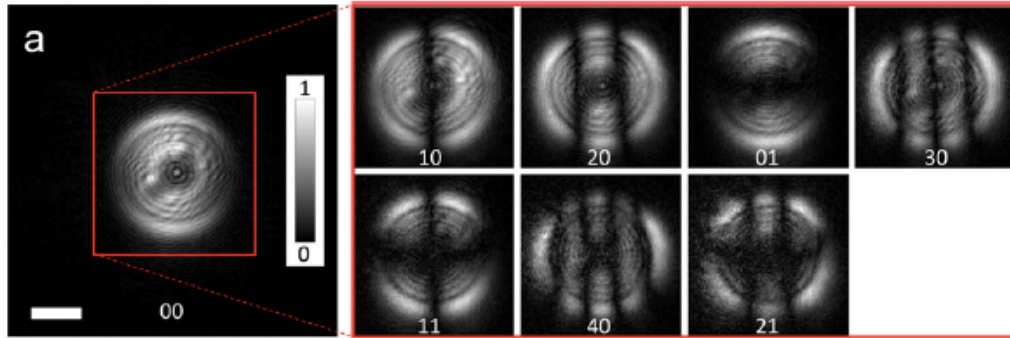


Figure 61: Orthogonal modal decomposition of partially coherent hard X-ray illumination [78].

As a consequence of the circular path of the high-energy electrons, the source in a synchrotron appears wider in the horizontal plane than in the vertical plane. As expected, we therefore see more lateral modes than vertical modes. Lateral incoherence appears as vertical fringes in the modal structure, because of the Fourier relationship between coherence and source width. In fact, the defocussed probe is not exactly in a Fourier relationship to the source, but the effect is the same. These results were obtained from a beamline that we had every reason to believe was fully coherent: the number of modes therefore came as quite a shock. It turned out that unbeknownst to anyone, the monochromator had a vibrational instability. The moral is: always perform a modal decomposition on all data that has any possibility of including partial coherence.

How many modes should you include in an illumination modal decomposition? You can declare in the computer as many modes as you like, but once they are orthogonalised only a few should have any significant power. Of course, you cannot solve for more modes than you have numbers in your data, so at some point the higher-order mode structures will disintegrate. Note that when you add up the intensity of all the modes they must be the same as the total intensity of the probe. If the underlying complex modes are normalised, then the diagonal terms of the density matrix represent the probabilities, or weights, of how the state mixture has been prepared. The sum (trace) of these is always unity because they are probabilities. The sum of the squares of the probabilities is a tidy measure of coherence: unity corresponds to total coherence (only one coherent state in the system); anything less is a measure of the degree of partial coherence.

Finally, note that the patterns in Figure 61 have no actual physical meaning, they are simply the lowest rank representation of the coherence of the experiment. However, to conserve computing power, one would not want to run more parallel probe estimates than are necessary in the reconstruction process, so in this sense the orthogonal representation has practical value.

8.3) Electron modes

Matter waves can also be decomposed into modes in exactly the same way, as shown in Figure 62. These data were collected on a TEM operating in the SAP

mode (Section 5.3). The source profile, as seen looking up the column from the detector plane, can be calculated by back propagating each complex mode, which are here lying in the image plane confined by a selected area aperture, back to the source plane via a back Fourier transform. The intensities of each mode are then added together to produce an estimate of the source, also shown in Figure 62. Adjusting the condenser alters the source size, an effect that can be seen in both the mode reconstructions and source shape reconstructions. Even though the column was aligned, we note the cold field-emission source is not perfectly round. It should be emphasised that these source intensity plots do not show the physical shape of the source, but rather its shape as seen through the selected area aperture. The image is diffraction limited because the total wavefunction relating to the source is truncated outside the aperture diameter.

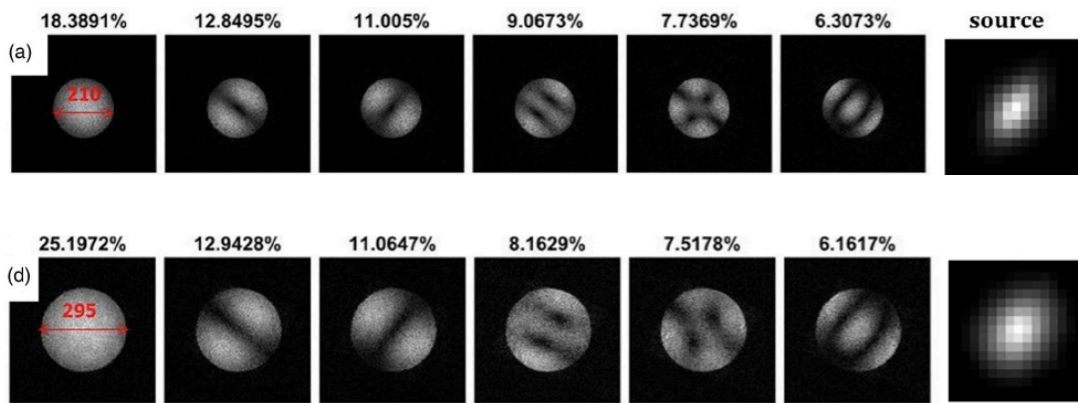


Figure 62: Modal decomposition of a propagating partially coherent electron wave, for two different spot sizes (apparent source size) [49]. The diffraction limited source, as seen backwards through the microscope, is shown on the right.

8.4) Mixed object state

Figure 63 is taken from the original Thibault and Menzel paper [48], demonstrating that this mixed state concept also applies to the object function. In this model calculation, each grey square represents a spin that can interact ferromagnetically or anti-ferromagnetically with its immediate neighbours. The system is in a temperature bath, enough to overcome the average bond energy so that the spins flip up and down randomly. The modelled ptychography experiment integrates data from all the oscillating spins over a longer time interval than it takes for them to flip. A phase change is expressed in the transmitted wave according to whether the individual spin is up or down.

After the mixed state decomposition, the principal modes can be extracted from the data showing that, at least on the scale of the probe diameter, the relative probabilities of the adjacent spins matches what would be expected at this temperature. This type of analysis is not dependent on the speed of the state changes relative to the integration time of the experiment, which means that in theory it could be applied to very high frequency phenomena, such as, in the case of electron ptychography, coupled bonding effects in an array of atoms. This may become a truly powerful experimental technique.

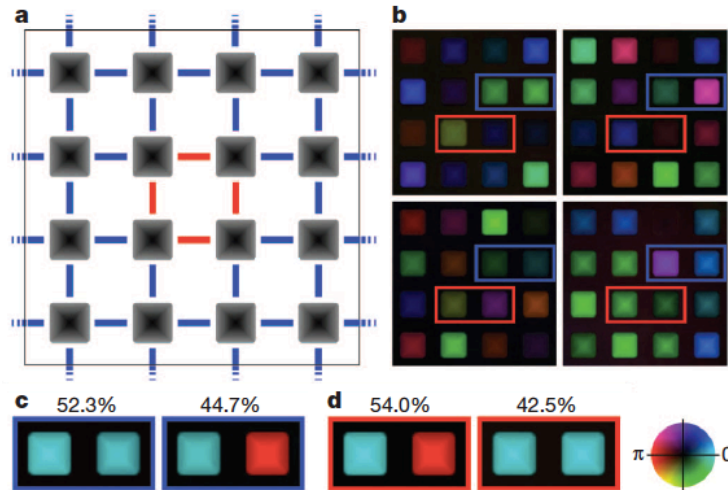


Figure 63: A model calculation showing how an object that has mixed states can reveal correlations in those states in a ptychographic reconstruction, despite them oscillating at a much higher frequency than the exposure time of each diffraction pattern [48]. Energy couplings for the red and blue bonds (top left) are in the opposite sense. The recovered images have probability distributions for the red and blue boxes (top right), as expected (bottom boxes).

8.5) Upsampling

One odd implication of Equation 7 is that the sampling condition in ptychography is not dependant on probe size. What does that mean if we have a large probe but only a few pixels in the diffraction plane? In Section 10.4, we will discuss direct ways of using very large pixels (sector detectors) to solve for the object, but these techniques rely on a highly focused (very small) probe, made by a perfect lens. The specimen must also be very weakly scattering. If the probe is large, large detector pixels cannot sample the rapid intensity variations that arise in the diffraction plane.

In order to exploit very dense sampling in real space, even though the detector pixels are larger than the features caused by a large probe, we need to ‘up-sample’ the big pixels [117]. During the reconstruction, this involves declaring an array size in the detector plane that would indeed satisfy the sampling condition given the size of the probe. Supposing we now have 3x3 pixels that fit into each big detector-sized pixel. We treat each computational pixel as a separate mode, running 9 concurrent reconstructions. The detector constraint is applied as before: after each forward calculation the modulus is changed according to the scaling of intensity illustrated in Figure 64. In this way, we reconstruct an artificial detector sampling that does fulfil the probe-size constraint. Doing this for data that is believed to be properly sampled can also be beneficial if there is a type of partial coherence in the beam that expresses itself as a convolution of the diffraction pattern. The method can also remove the MTF of the detector.

Although up-sampling works, it should be avoided if possible because there is bound to be a degradation in the results. The reconstruction of all the upsampled

pixels relies on the tiny changes of intensity in the big detector pixels that occur as the large probe is scanned in small steps across the object: data that is easily lost in noise or the finite bit depth of the detector. Figure 52 shows an example of up-sampling X-ray ptychography data.

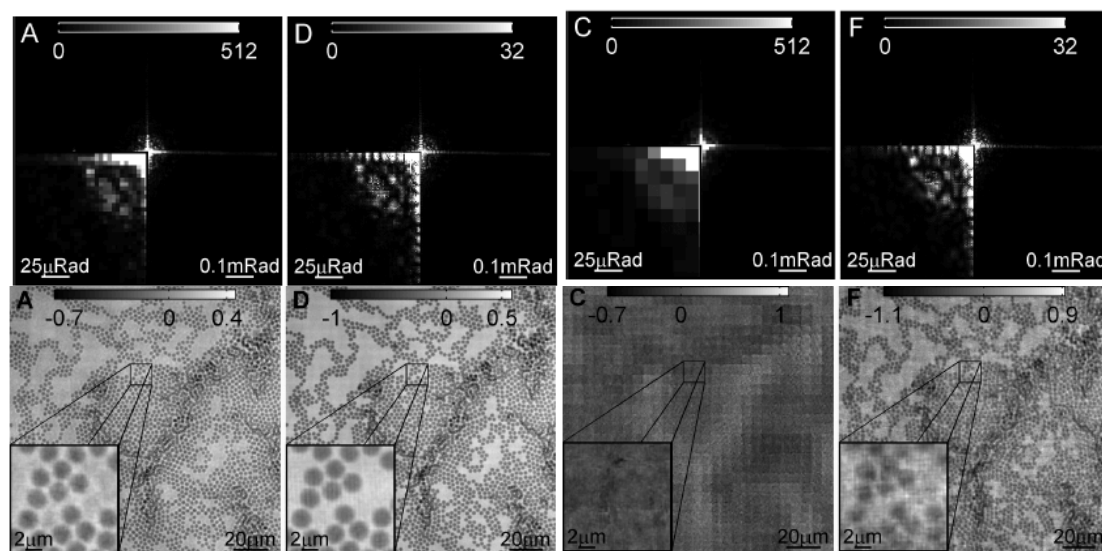


Figure 64: Example of up-sampling, taken from [117]. The diffraction patterns in the top row have had the lower left quadrant expanded, so as to show the process more clearly. Lower images correspond to reconstructions from the upper diffraction patterns. From left to right we have: the original data and its reconstruction; the original data up-sampled by 2x2; the original data binned by 3x3; the binned data to the left upsampled by 12x12. The best reconstruction comes from upsampling the raw data, possibly because there is some incoherence in the data.

8.6) Other uses: detector noise, diffuse scattering, continuous scan

Modal decomposition can be used for other things. A free mode (one that is not constrained by orthogonalisation) can be used to dump any intensity from the detector that is inconsistent over the whole data set. For example, if the detector has a pedestal – a constant offset or background noise – the inversion will try to put a delta function (the Fourier transform of a constant function) somewhere into the field of view of the object reconstruction. An incoherent mode can accommodate this. All the intensity that is inconsistent in any way with the forward calculation will be ‘dumped’ into it. Scattering by air in a hard X-ray ptychography experiment, or inelastic scattering in an electron experiment, will also be expressed in the mode, but in this case unevenly distributed over the detector. Dealing with this class of problem in a more controlled way, say by calibration or self-calibration of the detector, is probably a better approach.

A continuous line scan, in which data collection is speeded up by constantly taking exposures as the object is moved continuously across the probe, or vice versa, can also be handled by modal decomposition. Each exposure occurs over a blurred track of probe positions – i.e. a combination of several probe positions – each of which can be treated as an illumination mode [51].

9) Theory

In Section 3 we surveyed the wide and growing range of ptychographic algorithms reported in the Literature, and we have examined in Sections 4-8 the remarkable scope for exploiting the ‘redundancy’ in ptychographic data through clever expansions of the original ptychographic phase problem. In this Section, we will look in greater detail at how the most popular iterative ptychography algorithms work, and how they can be implemented on a computer.

9.1 The PIE family of algorithms.

Section 3.4 introduced the PIE algorithm, and explained its operation. It turns out that the reasoning behind that original formulation can be readily expanded to arrive at a whole class of algorithms that work in a similar manner. Returning to Eq. 6, reprinted below as Equ. 8 we saw that the core of the PIE algorithm was the object update function:

$$q_{NEW} = q + \frac{|a|}{|a|_{MAX}} \frac{a^*}{(|a|^2 + \varepsilon)} (\psi_{NEW} - \psi_e) = q + w\Delta q \quad (8)$$

Here, a new object estimate, q_{NEW} , is generated from the previous object estimate, q , by adding a specially weighted proportion of the old and new exit-waves, ψ_e and ψ_{NEW} , and dividing by the probe (with a fudge factor to avoid zero divisions). To make our discussion here clearer, we have rewritten this update in terms of a Δq – the exit-wave difference divided by the probe – and a weight function, w . It turns out that this weighting in the update function is only one of a whole host of possibilities that can be employed to reconstruct ptychographic data, as a very recent paper by one of the Authors explains [Maiden Optica in press].

For PIE, the weighting goes as the normalized probe modulus, $w = |a|/|a|_{MAX}$. This works well in practice and is often used by the Fourier ptychography community, where it has been re-derived as a second-order gradient descent [29]. The ePIE algorithm makes a very basic change, replacing the normalized probe modulus with the normalized probe intensity, $w = |a|^2/|a|_{MAX}^2$ which has the benefit of removing the need for the zero-division fudge factor since the $|a|^2$ term in w cancels the probe division in Δq . The result is an alternative update function:

$$q_{NEW} = q + \frac{a^*}{|a|_{MAX}^2} (\psi_{NEW} - \psi_e) = q + w\Delta q \quad (9)$$

We can plot the two weightings as a function of the probe modulus as shown in Figure 65. As is rather obvious, ePIE’s plot is a quadratic, meaning that where the probe is intense, there will be a large weight given to the Δq term in Eq. 9, whilst where the probe is dim, the weighting of Δq will be small and the object will change little in these regions. Equally obvious is the linear weighting for the PIE plot – again where the probe is intense Δq is strongly weighted, where it is dim

the weighting of Δq is smaller. What isn't obvious from these plots is whether either of these weighting functions is in any sense optimum, and this is an open question at the time of writing. What we can say is that there are further alternatives that also reconstruct ptychographic data very well, and which offer greater scope for tuning the reconstruction to accommodate a specific experiment – for example using different tuning parameters when the object is very weak or where the initial probe estimate is very poor. One weighting that we have demonstrated very recently to work extremely well takes the form $w = |a|^2 / (\alpha |a|_{MAX}^2 + (1 - \alpha) |a|^2)$, where α is a tuning parameter. The plots in Figure 65 give a couple of examples of how this function behaves for different α values – notice how the curve can be adjusted to give more or less weighting to dim parts of the probe, so the experimenter can adjust the algorithm to a lower weighting if data is very noisy, or to a higher weighting if the data is very clean. We have found in practice that an α value around 0.1 gives a considerable improvement in convergence rate over both PIE and ePIE. The object update function for this new form of weighting is:

We call this 'rPIE' because it can be expressed as a regularized version of the ePIE update.

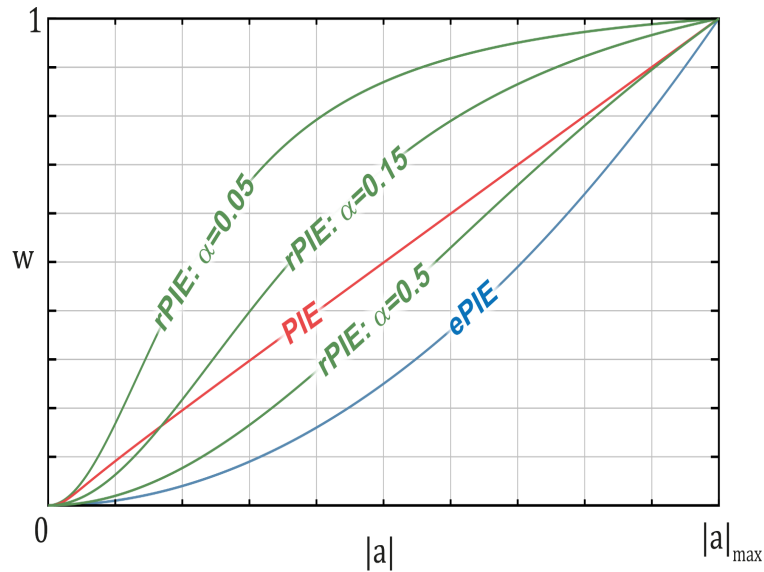


Figure 65: the way PIE-type ptychographic algorithms update the object estimate depends on the probe, a : for a given probe position, they update the object strongly where the probe is bright and only weakly where it is dim. The exact relationship between probe modulus and update strength (w), is shown in the Figure for three different algorithms. The new 'rPIE' update can be tuned to occupy different parts of the graph by varying its tuning parameter α .

The probe update

Although ePIE used a slightly different update function to PIE, the main advance it offered when it was first suggested was to solve for the probe as well as the

object. The implementation of this probe update is straightforward – simply interchange the appearance of a and q in any of the object update functions above to produce a probe update function, then apply this function after the object has been updated; for rPIE it helps to have a separate tuning parameter, β , replacing α as well.

9.2 Projection between sets methods.

Another way to think about the cCDI problem is illustrated in Figure 66. The plane of the Figure represents all the possible solution images that can exist. In reality, the dimensionality of the space is an enormous vector space, with each axis corresponding to the complex values of each image pixel. There are two sets of images lying within the coloured shapes. One set is consistent with the diffraction pattern intensity; the other is consistent with real-space priors that we may know about – for our discussion we can use the aperture constraint (where the object is known to be zero outside of a known support). The loop in Figure 5 alternately projects a current estimate of the solution onto the nearest point of the aperture constraint and then the detector (Fourier) constraint. It is the nearest point because the change we make in either domain is the minimum alteration we have to make to any pixel to get it to be consistent with its set.

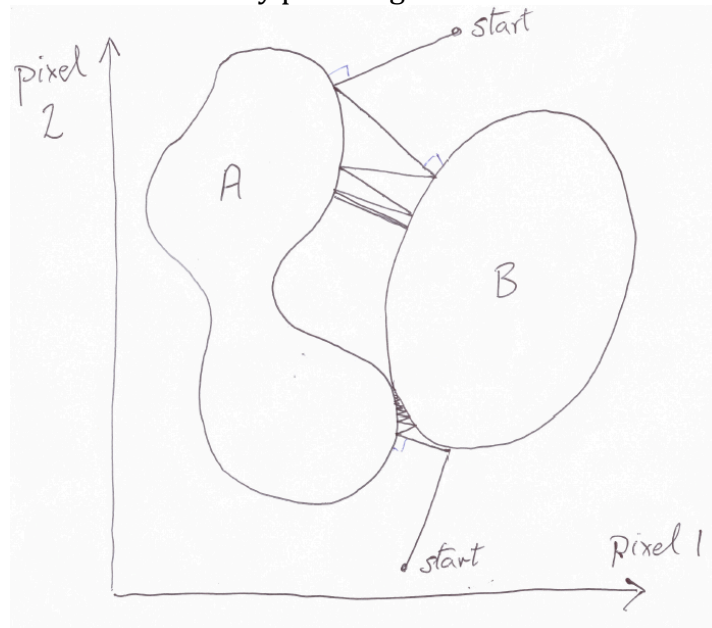


Figure 66: Graphical illustration of projection onto sets. One set is consistent with all possible images that satisfy the aperture constraint in real space, the other with images that have a Fourier transform whose modulus satisfies the detector constraint. See text for further details.

Even if there is only one unique solution image (the sets touch at one point – marked as ‘A’), there’s no guarantee our strategy will not get stuck jumping between the sets at the point ‘B’ in the diagram. Convergence to the correct solution is only guaranteed (and then only for perfect, noiseless data) if the two sets are convex, i.e. a line drawn between any two points within the set lies entirely within that set. Unfortunately the phase problem is non-convex: the steps B-C, revising the modulus of the detector wave estimates in Figure 5, can

be thought of as circles in the complex plane, and clearly a line between two points on the perimeter of a circle does not lie entirely on that circle.

Nevertheless, this ‘projection between sets’ concept is incredibly general and can be employed to any optimization problem – it was even used by Elser to solve Sudoku puzzles [31]. As a result, there is a large volume of literature devoted to set projection algorithms and analysis of their properties.

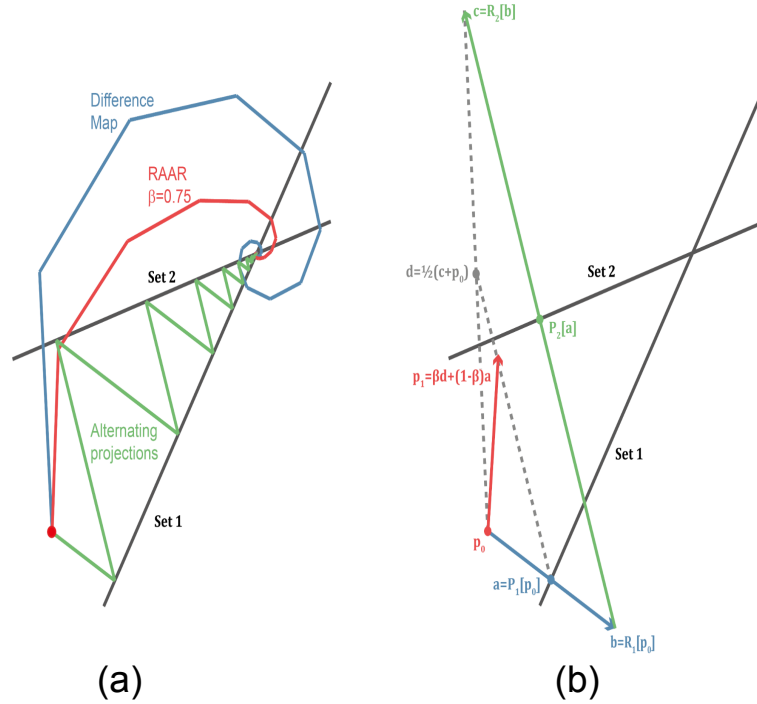


Figure 67: Parallel update algorithms, such as DM and RAAR, can be thought of in terms of projections between sets. Part (a) of this Figure shows schematically how the most simple set projection approach – alternate projections (green trace) – bounces between two constraint sets, and how more advanced methods spiral in to the intersection of the two sets. These advanced methods consist of a series of projections and reflections between the constraints, in the manner shown in part (b). A single iteration of DM starts at p_0 and steps through p_0 -b-c-d; a single iteration of RAAR goes p_0 -b-c-d- p_1 .

Consider next Figure 67. Here we will restrict attention to two convex sets, set 1 and set 2, represented by the two black lines in the Figure. We have already discussed one strategy to find the intersection between these two sets – our required solution – which is to alternately project between the two sets. The green trace in Figure 67a shows how this strategy bounces between the two constraint sets and staggers its way toward the intersection. Because the two constraints shown here are convex, this strategy is guaranteed to converge to the right answer, but it takes quite a large number of steps to do it, and as Figure 67a shows when the sets are non-convex this strategy can become stuck. The Difference Map (DM) [31] is one alternative to alternating projection, and the way it spirals towards the intersection, like water down a plug hole, is illustrated by the blue trace in Figure 67a. (Note that DM in its most general form has a tuning parameter, β , but this is usually held at 1 for ptychography, under which condition DM is equivalent to several other algorithms, e.g. the Douglas-Rachford algorithm and an algorithm called Averaged Successive Reflections (ASR).) Yet

another method – Relaxed Averaged Alternating Reflections (RAAR) [118] – is illustrated by the red trace. RAAR can ‘tighten’ the spiral behavior of DM with a parameter β . The spiralling action of these two algorithms accomplishes two things: it speeds convergence, by eliminating the zig-zagging of the alternating projections routine, and it widens the accessible search space, which for non-convex constraint sets means they can escape the local minima illustrated in Figure 67a.

DM and RAAR both employ *reflections* as well as projections between sets. Referring to Figure 67b, consider the point p_0 . The projection of this point onto set 1 is $(P_1[p_0])$ and it lies at a , the nearest point on the line to p_0 . The reflection of p_0 about set 1 is at b – it lies in the same direction as a , but is twice as far from p_0 : we can express this reflection as $R_1 = p_0 + 2(P_1[p_0] - p_0) = 2P_1[p_0] - p_0$. In terms of these projections and reflections, alternating projections can be easily summed up as $P_2[P_1[P_2[P_1[p_0]]]]$ etc...

DM follows this pattern: from the point p_0 , reflect about set 1, then reflect about set 2, then go halfway between p_0 and the result of these two reflections. In Figure 67b this is the path b to c to d . RAAR adds a final step: draw a line between the points a and d and travel a certain proportion, β , of the way along this line to find p_1 – this is how RAAR ‘tightens’ the spiral in Figure 67a. (Clearly, for $\beta=1$ RAAR and DM are equivalent.) We can only really skim the surface of this fascinating topic, so we refer the Reader to the extensive literature for more details.

9.3 Implementing ptychographic algorithms on the computer

It would require a book in itself to describe implementation details for all of the many algorithms for ptychography; instead, the following gives a framework that the coder can extend by reference to the literature. We will first set out processes that are common to all of the algorithms, namely initialising the object and probe, forming the exit wave and propagating it, and updating the exit wave at the detector to match the measured data. From these preliminaries, focus narrows to pseudocode examples of the three algorithms discussed above – ePIE, DM, and RAAR.

Initialisation of the reconstruction.

Some additional nomenclature, summarized in Figure 68, is needed to deal with the discrete nature of the algorithms (i.e. the unavoidable fact that the diffraction data, specimen and probe must all be represented by finite-sized matrices in the computer).

- The diffraction pattern are assumed square and of pixel dimensions $N \times N$
- The pixel pitch of the diffraction patterns (in metres), i.e. the detector pixel pitch is d_c
- The object reconstruction is of pixel dimensions $K \times L$
- The pixel pitch of the object reconstruction is d_o
- The pixel pitch of the probe is the same as the object, d_o
- The pixel dimensions of the probe are as for the diffraction patterns, $N \times N$

- Remember, there are J diffraction patterns in total, and the specimen shift when the jth diffraction pattern was recorded is $\mathbf{R}_j = (x_j, y_j)$

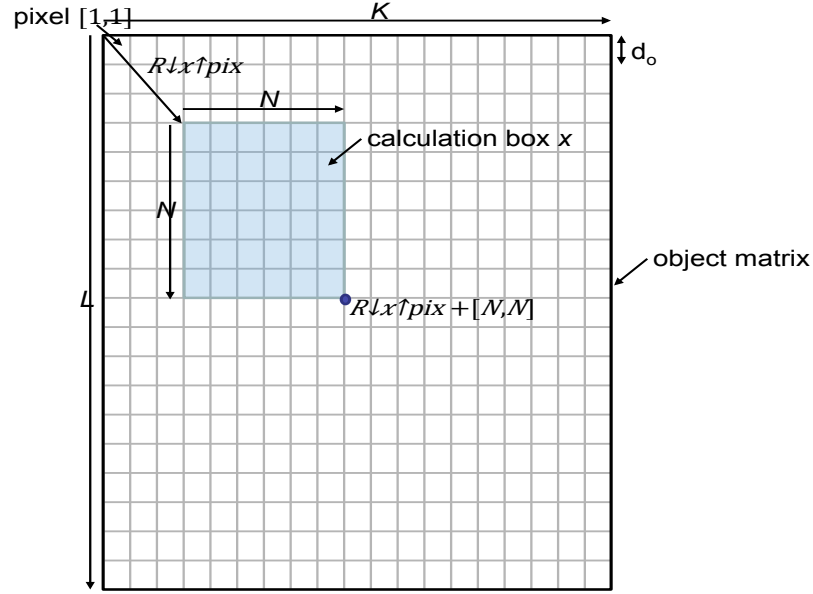


Figure 68: To explain how to implement ptychographic algorithms in code, we need to define some variables, as shown in this Figure. To digitally estimate the exit wave from a given specimen position x , a ‘calculation box’ with the same number of pixels as the detector must be extracted from the larger object matrix.

The first step in any of the algorithms is to decide the propagator, from which the pixel pitch of the object (d_o) follows. For the Fourier propagator the pixel pitch is:

$$d_o = \frac{\lambda z}{Nd_c} \quad (10)$$

Where z is the distance between the specimen and the detector, and λ is the illumination wavelength.

For angular spectrum and Fresnel propagation:

$$d_o = d_c. \quad (11)$$

Having established the pixel pitch, the object matrix can be initialised. Usually this is chosen as a matrix of 1s, representing free-space, whose size is governed by the extent of the specimen shifts. More exactly, K and L are chosen as:

$$(K, L) = \frac{\max(\mathbf{R}_j) - \min(\mathbf{R}_j)}{d_o} + (N, N) \quad (12)$$

Knowing the pixel pitch in the object matrix also allows conversion of the specimen shifts from the experiment into equivalent pixel shifts in the computer.

To do this we anchor the top left corner, pixel [1,1], as the origin and map the specimen shifts according to:

$$\mathbf{R}_j^{\text{pix}} = \text{ROUND} \left(\frac{\mathbf{R}_j - \min(\mathbf{R}_j)}{d_o} \right) \quad (13)$$

Initialisation of the probe is highly dependent on the experimental geometry used to collect diffraction data. The simplest case arises when the experiment uses an aperture to form the probe – here a disc of 1s embedded in a matrix of NxN 0s suffices, with the disc's diameter roughly matching that of the physical aperture (once scaled from pixels to metres via eqn. 10). In contrast, a convergent beam probe is easy to model, but hard to model accurately because of the difficulty measuring the exact amount of defocus. (A defocus error in the initial probe is one situation where many reconstruction algorithms struggle for convergence [119].) Assuming the defocus is known, the convergent beam probe is modelled by Fourier transforming an aperture multiplied by a quadratic phase profile. The size of the aperture should reflect the numerical aperture of the probe-forming optics, which itself can be determined from the brightfield disc in the diffraction pattern. If the diameter of the brightfield disc is D pixels, and the defocus is d_f metres, the initial probe can be calculated as:

$$\mathbf{F}^{-1} \left[\text{circ}(D) \cdot \exp \left(-i d_f d_c^2 \frac{n_1^2 + n_2^2}{\lambda z^2} \right) \right] \quad (14)$$

Where n_1 and n_2 index pixels in the diffraction pattern space (with the origin at the centre of the detector). For a probe with a diffuser in the beam path, the simplest strategy is to model an initial probe as above, disregarding the diffuser completely, relying on the reconstruction algorithm to untangle the diffuser's effect. Alternatively, if anything about the phase can be inferred (for example, a good approximation of the phase curvature at the detector plane), this approximate phase can be applied to a diffraction pattern, or an average of all of the diffraction patterns, and the result back-propagated to (hopefully) obtain a better initial probe estimate.

Regardless of how it is modelled, a useful final step in the probe initialization, as has been discussed, is to normalize the probe power to the diffraction data, by ensuring that the sum of the initial probe intensity over every pixel is equal to the pixel sum of the measured intensities in the brightest diffraction pattern. In what has become an indispensable final initialisation step, the diffraction data is transferred from computer memory onto a Graphics Processing Unit (GPU). To give an idea of the speed increase offered by GPU computing, a typical¹ Ptychographic reconstruction carried out with the authors' MATLAB version of ePIE takes 90 seconds to complete 100 iterations using an NVIDIA Titan GPU; the same reconstruction using an i7-4770 3.4GHZ CPU takes 868 seconds. Optimisation of the code and implementation in C gives even greater speed up.

¹ The data set in this case contained 400 diffraction patterns, each of 512x512 pixels.

Modelling the exit wave at the detector

To model the exit-wave leaving the specimen, for a particular specimen shift (say shift x), the ‘calculation box’ (see Section 3.3) must be extracted from the larger object matrix, as illustrated in Figure 68 – this equates to cutting out a region of pixels starting from R_x^{pix} and extending to $R_x^{\text{pix}} + [N, N]$. The final step in computing the exit-wave is to multiply the extracted reconstruction box, pixel for pixel, by the probe matrix.

Computer implementation of the propagators

Propagation of the exit wave to the detector plane can be via Fourier transform, the angular spectrum or Fresnel transform. MATLAB code that the Reader may use to implement these propagators digitally is given in Box (Fig 69). This code ignores multiplicative amplitude and phase constants, which do not have an effect on the ptychographic reconstruction; a more complete discussion of modelling wave propagation in MATLAB can be found in the book by Voelz [120].

Figure 69: **Matlab Code for propagators:**

```
function output = Propagate(input,propagator,dx,wavelength,z)
% Propagate a wavefront using a variety of methods
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% input:      the wavefront to propagate
% propagator: one of 'fourier', 'fresnel' or 'angular
spectrum'
% dx:         the pixel spacing of the input wavefront
% wavelength: the wavelength of the illumination
% z:         the distance to propagate
% output:     the propagated wavefront
% Setup matrices representing reciprocal space coordinates
[ysize,xsize] = size(input);
x = -xsize/2:xsize/2 - 1;
y = -ysize/2:ysize/2 - 1;
fx = x./(dx*xsize);
fy = y./(ysize*dx);
[fx,fy] = meshgrid(fx,fy);

switch propagator
    case 'fourier'
        if z>0
            output = fftshift(fft2(fftshift(input)));
        else
            output = ifftshift(ifft2(ifftshift(input)));
        end

    case 'angular spectrum'
        % Calculate phase distribution for each plane wave
        component
        w = sqrt(1/wavelength^2 - fx.^2 - fy.^2);

        % exclude evanescent waves
        notEvanescent = imag(w)==0;

        % Compute FFT of input
        F = fftshift(fft2(fftshift(input)));

        % multiply FFT by phase-shift and inverse transform
        output =
        ifftshift(ifft2(ifftshift(F.*exp(2i*pi*z*w).*notEvanescent)))
        ;

    case 'fresnel'
        % Calculate approx phase distribution for each plane
        wave component
        w = fx.^2 + fy.^2;

        % Compute FFT
        F = fftshift(fft2(fftshift(input)));

        % multiply by phase-shift and inverse transform
        output = ifftshift(ifft2(ifftshift(F.*exp(-
        1i*pi*z*wavelength*w))));
end
```

Revision of the exit wave

Replacing the modulus of the wavefront at the detector with the measured data is most efficiently achieved by dividing the propagated wavefront by its own modulus and multiplying by the square-root of the measured intensity. Care should be taken to avoid division by zero if this approach is adopted – for

example by adding a small number to the modulus before the division as in Box Fig 70 (eps is the smallest number MATLAB can represent).

Figure 70: **MATLAB code for propagators:**

```
Function: exitWaveNew = UpdateExitWave(exitWave, measurement, z)

detectorWave = Propagate(exitWave, z);
correctedWave = sqrt(measurement)*detectorWave/(abs(detectorWave)+eps);
exitWaveNew = Propagate(correctedWave, -z);
```

PIE-type algorithms

After the preliminaries given above, the PIE-type algorithms can be written in just a few lines of code. As an example, we give in Figure 71 pseudo-code for implementation of rPIE; changing the update function to realise ePIE or PIE is straightforward. One caveat: the code in Box (Fig 71) assumes a sequential order in which to address the diffraction patterns, in practice it is better to randomise this order, and re-randomise it after every iteration.

Fig 71: **MATLAB code for rPIE:**

```
Loop
For j from 1 to J

    reconBox = object(rPix(j):rPix(j)+[N,N]);
    exitWave = reconBox*probe;

    exitWaveNew = UpdateExitWave(exitWave, diffPatterns(j), z);

    tempProbe = probe;

    denomP = beta*max(abs(reconBox)^2) + (1-beta)*abs(reconBox)^2;
    probe = probe + conj(reconBox)*(exitWaveNew-exitWave)/denom;

    denomO = alpha*max(abs(tempProbe)^2) + (1-beta)*abs(tempProbe)^2;
    newReconBox = reconBox + conj(tempProbe)*(exitWaveNew-exitWave)/denomO;
    object(Rpix(j):rPix(j)+[N,N]) = newReconBox;

End

Until (converged)
```

Set projection algorithms

Looking back to Figure 67, in order to implement the set projection methods we need to define the two sets that represent the ptychographic problem, as well as the two projections onto these sets. The first set, set 1, represents the detector constraint we have already discussed: it is the set of all exit waves that have the

correct (measured) modulus at the plane of the detector. We project onto this set by taking the current estimates of the object and probe, and for each specimen shift forming the exit wave ($a.q$), propagating, correcting the modulus and propagating back. This is accomplished as shown in the pseudo-code of Box Fig 72.

Fig 72: **Pseudo code: parallel Fourier constraint:**

```

For j from 1 to J

  reconBox = object(rPix(j):rPix(j)+[N,N]);
  exitWave = probe*reconBox;
  exitWavesNew(j) = UpdateExitWave(exitWave,diffPatterns(j),z);

End

```

The second set, set 2 in Figure 67, represents the set of exit waves for which the probe and object are *consistent*. We illustrated this concept in Figure 8: the overlap between the regions of the object illuminated by the probe during the experiment links together the exit waves, because we know they must have been formed in the experiment by an unchanging object and a static probe. (Of course, we've seen that these assumptions can be somewhat relaxed in practice.) Projection onto this consistency set is via the probe and object update functions, which for the set projection methods take on a slightly different form to those of the PIE-type versions. Box (fig 73) and Box (fig 74) present pseudocode outlines of these updates, which are applied one after the other to implement the projection, with the updated object feeding into the probe update.

Figure 73: **Pseudo code description of the object update used by DM, RAAR and other 'batch' update algorithms:**

```

Function: object = BatchObjectUpdate(exitWaves,probe)

probeSum = zeros(K,L);
waveSum = zeros(K,L);
probeInt = abs(probe)^2;
conjProbe = conj(probe);

For j from 1 to J

  probeSum(rPix(j):rPix(j)+[N,N]) = probeSum(rPix(j):rPix(j)+[N,N]) +
  probeInt;
  waveSum(rPix(j):rPix(j)+[N,N]) = waveSum(rPix(j):rPix(j)+[N,N]) +
  conjProbe*exitWaves(j);

End

object = waveSum/(probeSum+eps);

```

Figure 74: **Pseudo code description of the probe update used by DM, RAAR and other 'batch' update algorithms**

```
Function: probe = BatchProbeUpdate(exitWaves,object)

objectSum = zeros(M,N);
waveSum = zeros(M,N);
objectInt = abs(object)^2;
conObject = con(object);

For j from 1 to J

    objectSum = objectSum + objectInt(rPix(j):rPix(j)+[N,N]);
    waveSum = waveSum + conObject(rPix(j):rPix(j)+[N,N])*exitWaves(j);

End

probe = waveSum/(objectSum+eps);
```

Alternating projections

Having defined the two projections, the most basic algorithm applies them alternately: project onto set 1, project onto set 2, project onto set 1... This is achieved in the fashion shown in Box (Fig 75).

Fig 75: **Pseudo code description of the simplest batch update method – alternating projections:**

```
For j from 1 to J
    reconBox = object(rPix(j):rPix(j)+[N,N]);
    exitWaves(j) = probe*reconBox;
End

Loop

For j from 1 to J

    reconBox = object(rPix(j):rPix(j)+[N,N]);
    exitWave = probe*reconBox;
    exitWavesNew(j) = UpdateExitWave(exitWave,diffPatterns(j),z);

End

    probe = BatchProbeUpdate(exitWavesNew,object);
    object = BatchObjectUpdate(exitWavesNew,probe);

Until(converged)
```

Projection onto
detector constraint set

Projection onto
'consistency'
constraint set

DM and RAAR

Implementation of DM and RAAR proceeds in a similar manner, the only complication being the slightly more involved way that the exit waves are updated when the detector constraint is applied. Both methods can be coded along the lines of Box (fig 76); setting β to 1 in this code gives the standard version of DM.

Figure 76: Pseudo code description of the RAAR algorithm – the standard implementation of DM is obtained by setting $\beta=1$:

```
For j from 1 to J
    reconBox = object(rPix(j):rPix(j)+[N,N]);
    exitWaves(j) = probe*reconBox;
End

Loop
For j from 1 to J

    reconBox = object(rPix(j):rPix(j)+[N,N]);

    waveToPropagate = 2*probe*reconBox-exitWaves(j);
    exitWaveNew = UpdateExitWave(waveToPropagate,diffPatterns(j),z);

    exitWaves(j) = beta*(exitWaves(j) + exitWaveNew) +
    (1-2*beta)*probe*reconBox;

End

probe = BatchProbeUpdate(exitWaves,object);
object = BatchObjectUpdate(exitWaves,probe);

Until (converged)
```

Implementation tips and tricks

Some general points that can be helpful:

- Algorithms can be accelerated by pre-computing the exponential phase terms in the propagators and by pre-square-rooting the diffraction patterns.
- Commonly implementation is via MATLAB – to avoid repeated use of “fftshift” in the reconstruction, the diffraction data can be fftshifted instead, as can the pre-computed exponents in the propagators – this can give a significant speed boost.
- Generally, single precision numbers are sufficient for excellent reconstruction accuracy – and offer another significant speed boost.

9.4 A basic comparison of algorithms

There has yet to be a comprehensive comparison of the different ptychographic algorithms with either real-world or simulated data, although Yang et al. have evaluated many of the batch-type algorithms (DM, RAAR, conjugate gradient) [35], and work by Waller and colleagues assessed a range of batch and serial approaches for Fourier ptychography [36]. The difficulty in performing such an evaluation comes from the huge range of real-world scenarios to which ptychography may be applied – even tests restricted to the X-ray regime would need to cover situations ranging from very weak phase objects illuminated with a high energy convergent probe to strong, optically thick samples illuminated by a diffused soft X-ray probe. Here, a brief comparison between the algorithms detailed in the previous section is provided – more as an example of their various traits than as any sort of assessment of their performance. It should also be said that the Authors have a great deal of experience with the ePIE-type algorithms, and much less knowledge of the tricks and short-cuts that might improve operation of the batch-type alternatives.

Our first comparison is by simulation. We used as a specimen a photograph of one of the Authors' daughters (Lucy), converted into a phase-only object with a phase range of $0-2\pi$ (so that the darkest parts of the photograph mapped to zero phase, and the brightest to a phase of 2π). As a probe, we simulated a convergent beam with a small defocus. After every iteration of each algorithm, we calculated the sum of the differences between the evolving object reconstructions and the true photograph – these error values are plotted in Figure 77. We also paused the algorithms at various points to take a snapshot of the phase reconstruction – these snapshots are shown in Figure 78.

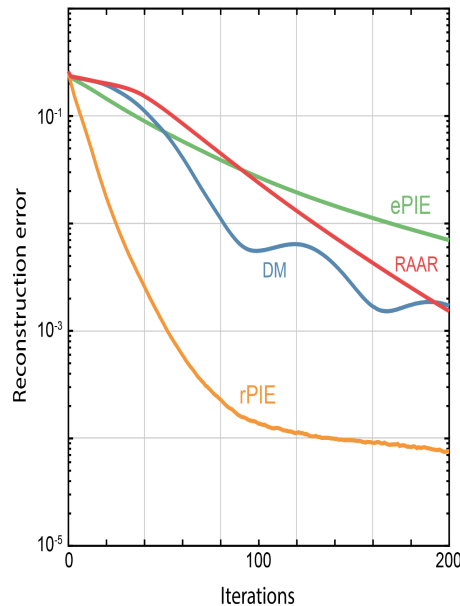


Figure 77: Progress of reconstructions using different algorithms in a simulated experiment. The graph plots an error metric that is the sum of the difference between the intensity of the exit-waves that the algorithms estimate and the measured intensities captured by the detector: note the link between the spiralling action of DM and RAAR in Figure 82 and this Figure.

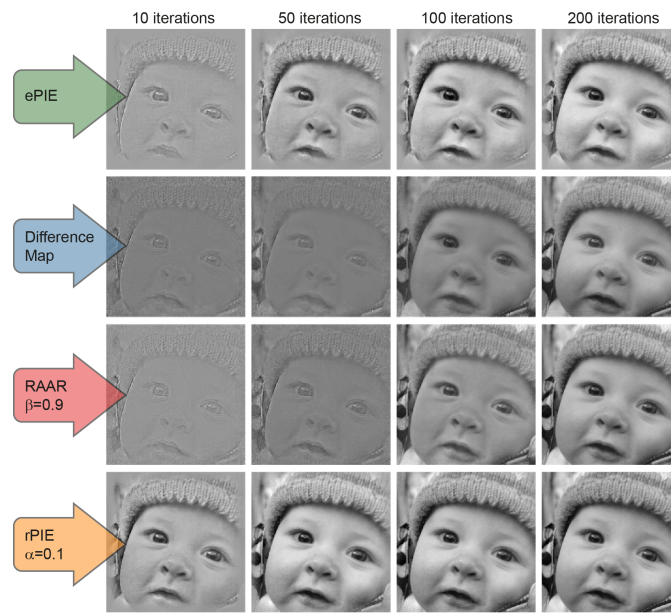


Figure 78: Progress of reconstructions using different algorithms in a simulated experiment. Here we have taken snapshots of the object estimate at various points during the reconstructions. Notice how the batch/parallel update algorithms, DM and RAAR, handle the centre and edges of the object quite equally, whilst the serial update algorithms, ePIE and rPIE, obtain the centre parts much more quickly but take time to reconstruct the edges.

We can note a couple of salient points here. First that the centre of the ePIE reconstruction evolves quite quickly, in the early iterations, whilst the outside part takes much longer to appear; second, that rPIE converges very quickly – at the centre and the edges – given this ideal, noise-free data set. The batch algorithms are much more balanced in the way they update the object, with the edges and the centre of the reconstructions evolving at an equal pace. DM converges quite quickly but tends to oscillate around the solution (like the spiralling action seen in Figure 67), whilst RAAR, although slightly slower initially, gives a very good convergence rate once it arrives near the solution.

For a second comparison, we collected data from an optical bench experiment. Our experiment used the simplest geometry of a probe formed by an aperture and a CCD camera placed in the far-field. As a specimen, this time we used a prepared microscope slide holding a section taken from a clam's gill (chosen only because it looks quite beautiful at high magnification).

After 100 iterations of each algorithm, the images in Figure 79 emerged (the amplitude part is shown). In this instance, DM does not fully converge and the result is a slightly speckled image. RAAR performs very well, with the resulting image displaying a good level of detail and good noise suppression. ePIE and rPIE also both produced good results (although perhaps RAAR just wins out).

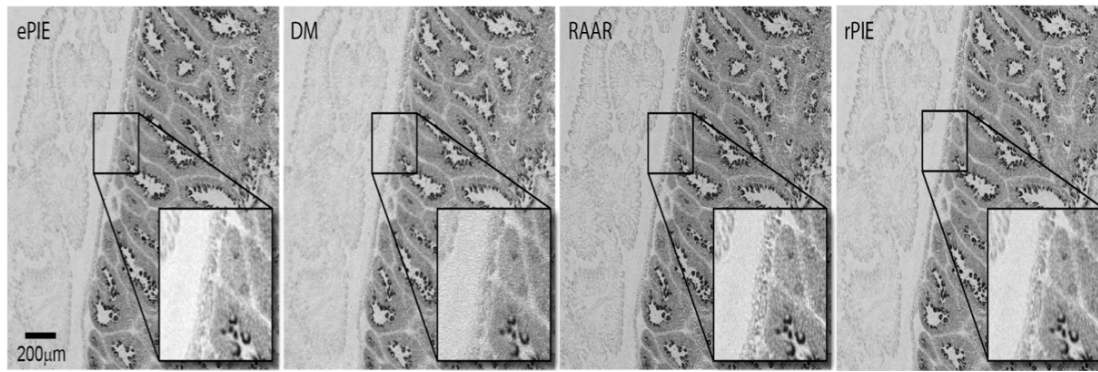


Figure 79: These amplitude images of a clam's gill were reconstructed using data gathered in an optical bench experiment. All of the algorithms work pretty well for this real-world data; the ptychographer must choose their own poison!

These simple comparisons mirror what is quite clear from the literature: that, given a carefully conducted experiment and reasonably clean data, the inversion problem at the heart of ptychography is well-conditioned, and amenable to solution by any optimisation technique – from simple gradient descent right through to the cutting-edge non-linear heavyweights.

10 Wigner distribution deconvolution (WDD) and its approximations

We are now going to discuss a class of direct, non-iterative solution methods for the ptychographic phase problem that can be used when the sampling in real space (i.e. the distance moved by the probe) also occurs at the Nyquist sampling frequency: this is determined by the rate at which intensity in the diffraction pattern changes as a function of probe position. The methods we will describe have their most practical implementation in the focused probe configuration (Section 5.1 and Figure 21). In this case, there is an aperture in the probe-forming lens so that, in the absence of aberrations, the probe is of the form of a bandwidth limited Airy disc function. The highest frequency in the illumination is determined by the diameter of the lens aperture, which means that as this probe is moved laterally, there is also a maximum spatial frequency at which the far-field intensity can alter. We can think of this via reciprocity, in the case of Fourier ptychography. The maximum spatial frequency that can arrive in a conventional image is also determined by the size of the aperture in the back focal plane of the lens employed, according to Abbe's theory. Clearly, there is no point in measuring the image (or moving the probe in real-space ptychography) at a higher spatial frequency (step size) than the maximum Fourier component of intensity in the image.

There are a couple of qualifications to this last statement. First, if we record the conventional image on a pixelated detector, say when undertaking Fourier ptychography, it is often advisable to sample at a higher frequency than Nyquist to avoid effects from the roll off of the spatial transfer function of the detector itself. Second, it must be remembered that the Nyquist frequency of the image is determined by the interference of beams passing by opposite sides of the lens aperture, i.e. separated by its *diameter*. This is twice the frequency of Fourier

components in the conventional *coherent* bright-field image obtained from a weakly scattering object, where the interference phenomenon is between a central unscattered beam on the optic axis and beams scattered up to the *radius* of the aperture. Later, we will see how this ‘double frequency/resolution’ issue manifests itself in ptychography.

Given equation 7, we might suppose that having maximal sampling in real space means that we can tolerate very low sampling in reciprocal space: we will find out below (section 10.4) that this is true, but only if the object is (i) weakly scattering, (ii) we are prepared to sacrifice our ability to remove even the most minor aberrations in the lens we are using to form the probe (including defocus), (iii) we accept that we cannot improve upon the resolution of the lens, as defined by its aperture, i.e. we cannot make use of dark-field scattering. Under these conditions we can nevertheless find the phase of the image more accurately than via the bright-field image.

To begin with we are going to start by thinking about the absolutely maximally sampled data set in both real and reciprocal space. We will define ‘maximal sampling’ in reciprocal space as the detector having a pixel size that is smaller than the reciprocal size of the *whole field of view* of the reconstructed image. This is much more demanding than simply being the inverse of the size of the probe (as is usual in ptychography), since the probe is invariably much smaller than the field of view. Matching reciprocal coordinates deriving from the field of view of the scan in real space and the total field of view as seen by the detector is only necessary if we want to use WDD to image strongly scattering specimens. Of course, these stringent conditions do not apply to iterative methods.

The experimental demands made by such a vast quantity of data are phenomenal – for a modest 512x512 pixel field of view, with a 512x512 diffraction pattern collected at every image pixel, we have nearly 69 billion measurements. If the bit depth of the detector is 16, you could only fit 8 of these data sets onto a Terabyte drive – and all this to solve for eight 512x512 pixel images! What is the advantage of all of this? One answer is that these extreme, very densely sampled data are the most we could ever hope to measure from a ptychography experiment, and so it must axiomatically be a ‘good thing’. Another answer is that very densely sampled data can be used to solve the ptychographic phase problem using a linear, closed form of inversion called Wigner Distribution Deconvolution (WDD). This was developed in the early 1990s [66, 108, 121-124], more than ten years before the modern iterative solutions for ptychography (for a review, see[5]). Given the agonising history of the phase problem during the 20th century, it is quite extraordinary that WDD solves an apparently non-linear and intractable inverse problem with a handful of Fourier transforms. It can also do almost everything else modern iterative methods can achieve: solve for the illumination, remove partial coherence effects and extract volumetric information from the object. Balanced against the absurd quantities of data it requires is the fact that it is computationally very fast. And anyway, in an age of ‘big data’ is this a problem? A domestic consumer can buy a terabyte disk for less than \$100; when the original work on WDD was done in the 1990s, the same money could buy 100Mbytes.

10.1 Notes on nomenclature

In this section we will be talking about four 4D data sets, each a function of two-dimensional variables, illustrated in Figure 80. Those in mixed real and reciprocal dimensions are called $I(u, R)$ and $H(r, U)$, where U is the reciprocal coordinates of R , and u is the reciprocal coordinate of r . We also have a function $G(u, U)$, which is a function of two reciprocal coordinates, and $L(r, R)$ which is a function of two real coordinates. Other conventions could be chosen (say that capital coordinates are the reciprocal of real coordinates). An advantage of the present scheme is that $G(u, U)$ is our most important function, and it is important to stress that it lies entirely in reciprocal space.

R is the probe position. r lies in the object space, and $I(u, R)$ is our detector intensity such that

$$I(u, R) = |\mathfrak{F}[a(r - R)q(r)]|^2 \quad (15)$$

where, as before, $a(r)$ is the probe and $q(r)$ is the specimen transmission function; it is important here that we keep track of coordinates, so we have now included the dependency on r explicitly for these functions. We will use $A(u)$ to denote the Fourier transform of the probe and $Q(u)$ to represent the Fourier transform of the specimen.

With reference to Figure 80, where each two-dimensional vector is represented by just one coordinate (so that the 4D data set is shown as a 2D function), the Fourier relationships between the data sets are as follows. Horizontal pointers represent a Fourier transform over just one variable, from u transformed to r , or vice versa. Vertical pointers are transforms also over one variable, R to U or vice versa. Diagonals represent Fourier transforms over all coordinates, (u, R) to (r, U) or vice versa, and (u, U) to (r, R) and vice versa.

(N.B: It has been realised that an earlier review that also adopted this convention included some mistakes [5]: Equ.s 94 of that reference, $L(r, U)$ should read $L(r, R)$, and in Equ.95, $D(u, r)$ should read $D(u, U)$, with the variables on the RHS substituted similarly. The same error is propagated through Equ.s 96-102, as well as in the text on p.122.)

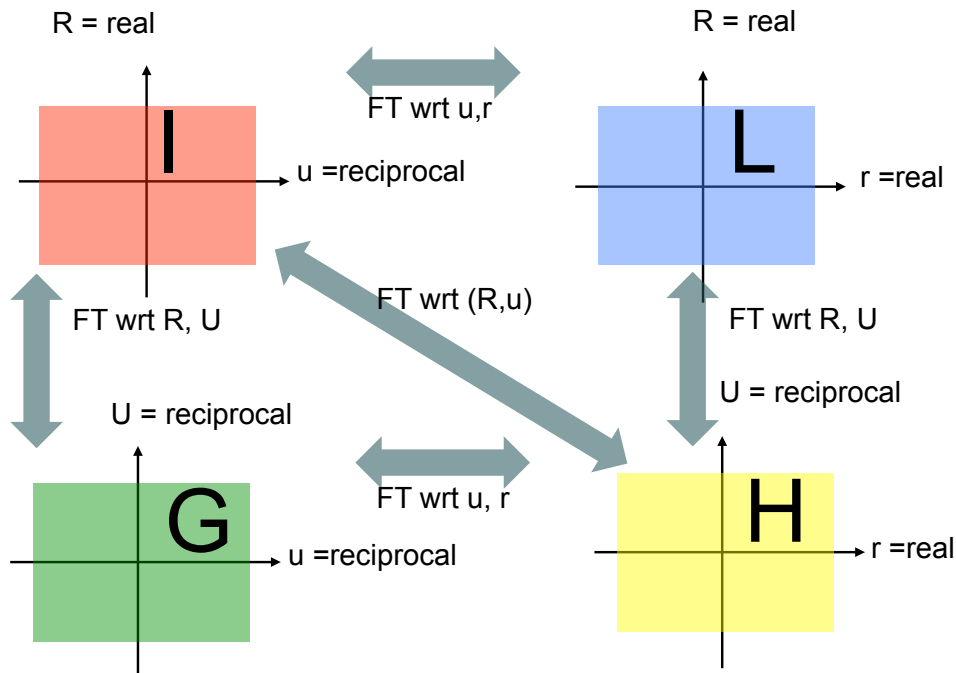


Figure 80: Schematic of the Fourier relationships between the recorded intensity, I , which is measured as a function of probe position, R , and detector coordinate, u , and the G - and H -sets. We do not discuss L , which is not relevant in the context of the main text.

10.2 Phase recovery from data sampled densely in real space

Rather than launching straight into the mathematics of WDD, we think it is important to give the reader a physical insight into the most important and mysterious aspect of the technique: how is phase information extracted from raw data which has only been recorded in intensity? A Fourier transform of a single diffraction pattern's intensity gives the autocorrelation function. Although this can be useful, especially if there is a strong reference signal in the original wave field, the object function is far from self-evident. Conversely, WDD relies on the principal strength of ptychography: probe movement. Let us see how this works.

Consider a focused probe that reaches its crossover some distance in front of a periodic grating, as shown in Figure 2 (we discussed this interference in relation to Hoppe's definition of ptychography). In the far field we would expect to see a shadow image of the object. (If you have an optical bench at your disposal, and you want to understand ptychography, this is an exceedingly informative experiment.) Figure 2b shows an example result. In this case, the object is a TEM grid illuminated by a laser beam focused by a single lens that has a variable aperture. In Figure 2c the aperture has been closed down. We now see discrete diffracted orders that are interfering with one another, giving fringes perpendicular to the scattering vector of the diffracted reflection, but with the same periodicity of the features that were cast in the shadow image of the object function. If the aperture is shut right down, the illumination is effectively parallel, so the disks become the usual diffraction spots and cannot interfere with one another. We see rather directly how interfering diffraction orders

evolve into a shadow image. Incidentally, if there are isolated features on the grating, like pieces of dust, they are not at all easily visible in the coherent shadow image because the interference of the diffraction orders dominate. If the source is partially coherent, resolution is lost but so are these very strong diffraction effects and so the isolated features become visible.

If we move the object (or probe) at a continuous speed, the shadow image and/or the interference fringes move across the diffraction plane at a rate that depends on the defocus of the probe. Greater defocus leads to less magnification in the shadow image, but this image appears to move laterally more slowly. The two effects cancel each other, so that at any one point in the shadow image, the variation of bright-dark-bright is determined only by the pitch of the grating, irrespective of defocus. This is exactly what would be expected given that the only change in the experiment is the object shift, so any change in intensity anywhere in the optical system must oscillate in synchrony with the periodic structure in the object. The degree of overlap between the diffracted disks is also directly related to the periodicity of the object, according to the usual diffraction grating equation.

Remember:

- (i) The position of the diffracted beam, and its overlap with other diffracted beams, is determined precisely by a specific periodicity in the sample.
- (ii) As the probe is moved, the intensity within the overlap areas changes periodically, at exactly the same periodicity within the specimen that defines the position of the diffracted beam. This intimate relation defines the structure of what we will call the 'Fat-H', as well as other characteristics of WDD. This principle is not confined to crystalline or periodic objects.

It is impossible to picture the full densely sampled 4D data set. We are therefore going to use one-dimensional lines, where a line represents a 2D image or a 2D diffraction pattern. The raw data set can then be represented as a 2D function, plotted as a function of probe position and diffraction pattern intensity, as shown in Figure 81. Horizontal lines correspond to diffraction patterns. Vertical lines are images (the signal detected at a diffraction pixel as a function of probe position). The data are the same as in a Fourier ptychography experiment, where vertical lines are images collected at a particular angle of illumination. We will also sometimes plot 2D functions that represent other slices through the 4D function. All these are *not* plots of one variable against another, but of a *function of two-variables*, where each point in the 2D plane will have a value that is not shown, although we could in principle show this using variable shading. Except for the raw data, all the functions are complex-valued.

The comprehensive diffraction/imaging 4D data set

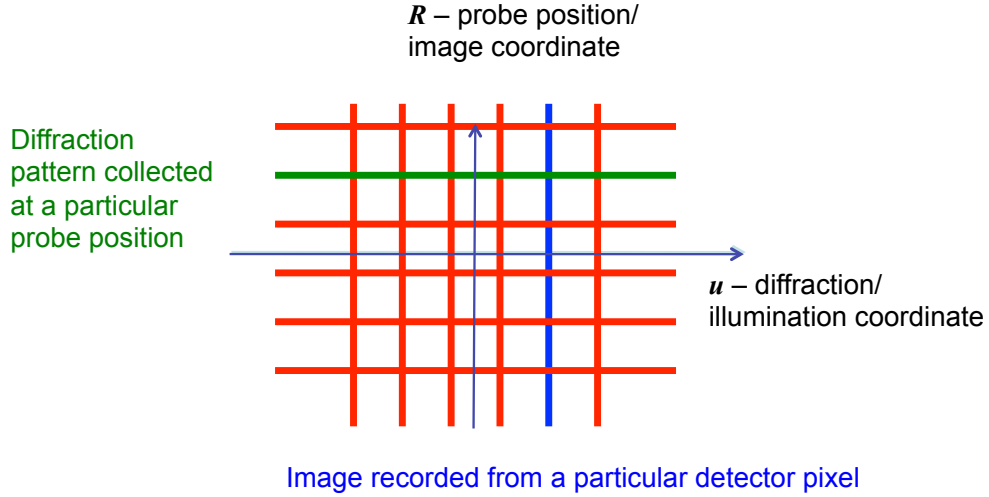


Figure 81: The maximally sampled data set. Along horizontal lines we have diffraction patterns, each from one probe position. Along vertical lines we have images, each recorded as a function of probe position using the signal collected from one diffraction pattern pixel. In Fourier ptychography, vertical lines are also images, collected from one illumination angle in u .

We are going to start by considering ptychographic phase retrieval for a periodic object, as in Figure 2. First we look at the raw intensity data, plotted as a function of R and u , namely $I(u, R)$ (shown top left in Figure 80), where R is the probe position coordinate and u is the diffraction pixel coordinate. If we have a periodic object, we have multiple strips, as shown in Figure 82: in this example, the strips (1D representations of the 2D diffracted disks in the full 4D data set) are not overlapping. When they do overlap (Figure 83), we now see interference, which periodically changes as a function of probe position. For simplicity the interference is shown as if the probe is perfectly focused. If it were defocused the interference fringes in this plane would be diagonal. As the probe is moved, the interference then shifts laterally across the overlap in the diffraction plane. For the focused probe, there is then no structure in the overlap of the disks, but the interference signal still changes as a function of probe position. The position of these fringes relative to one another will deliver the solution of the phase problem.

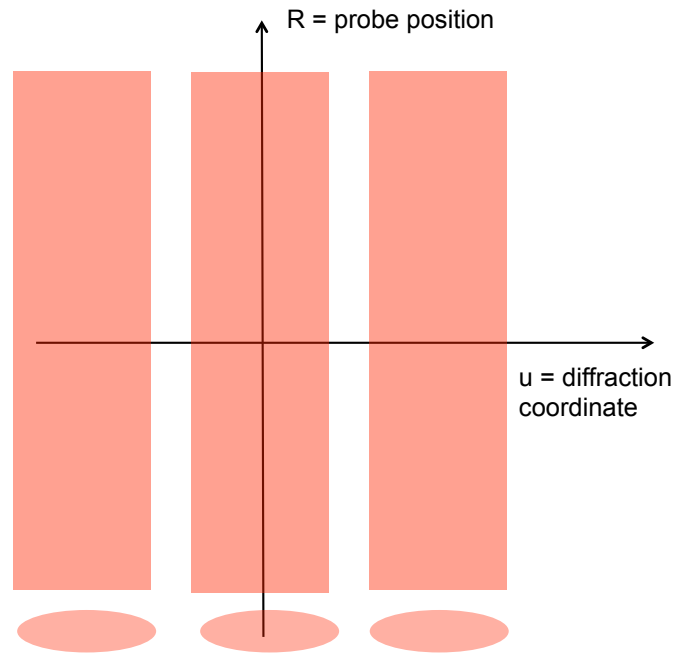


Figure 82: Shaded regions show where there is intensity in the recorded data when there are non-overlapping diffraction discs arising from a crystalline specimen. The discs at the bottom of the diagram show a perspective view of the two-dimensional diffraction pattern. In the main diagram, each diffraction pattern is a horizontal line, as in Figure 80.

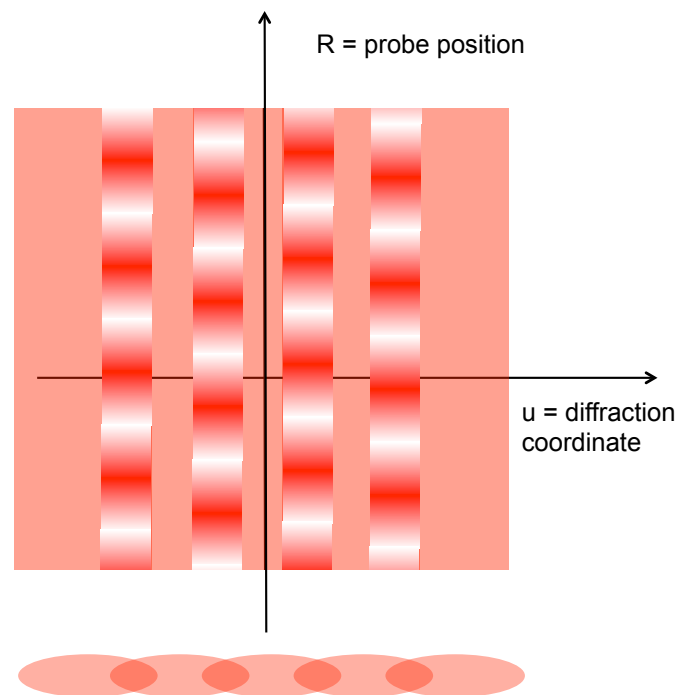


Figure 83: Raw data when discs from a crystalline specimen (as in Figure 2) overlap. The interference within overlap changes periodically. If there was defocus in the probe (Figure 2), these interference effects would be diagonal: think of a horizontal line moving down the Figure. Each pattern has fringes in the overlap region, which move laterally as the probe is moved. The position of these fringes solve the phase problem.

We now take a Fourier transform with respect to the probe shift coordinate, but *not* across the horizontal 1D diffraction pattern coordinate. This means we take out a vertical strip of pixels in our 2D data set, do a 1D Fourier transform on it, and then replace it in the same strip where we took it from, and so on for all such vertical lines. The result is shown in Figure 84, where now the vertical coordinate is a reciprocal coordinate of the probe position R labelled by U . We call this function the ‘G-set’ [122, 123]. The u coordinate remains unchanged, spanning the detector. Except at $U = 0$ (the zeroth component of the Fourier transform), there is no amplitude in the G-set in the vertical direction wherever the diffracted disks did not overlap – because these regions did not change as a function of R . However, we have lines of amplitude, each with the width of the aperture overlap, and positioned at the frequency of the structural element in the object that gave rise to the interference. When we do the mathematics, we will find that the phase of these features correspond directly to the phase difference between each pair of the diffracted discs, although we may have to deconvolve the influences of an aberrated or defocused probe. Once we have all such phase differences, we can construct the whole Fourier transform of the object: the phase problem is solved, once again by exploiting ptychographical probe-movement translational diversity.

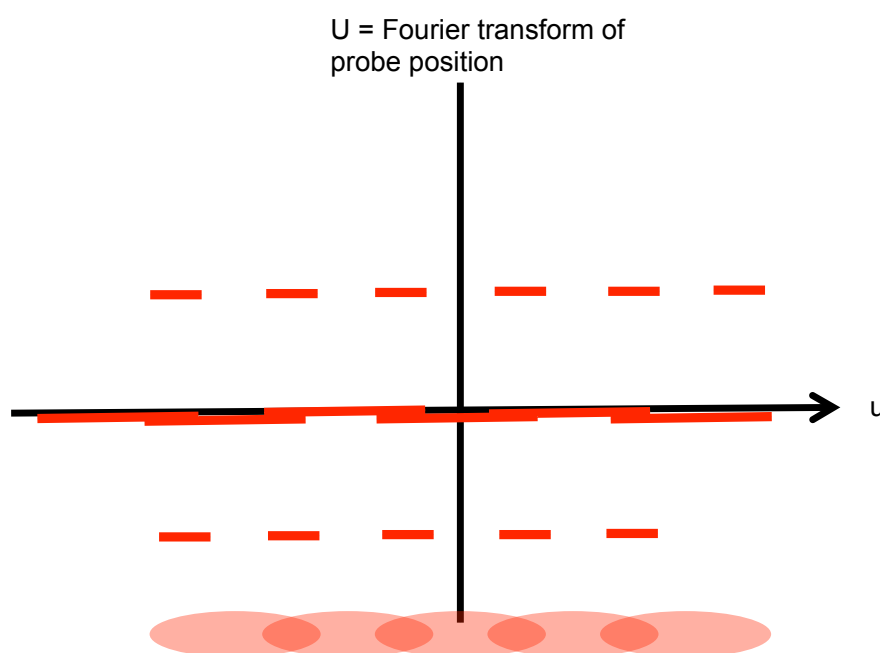


Figure 84: Once the Fourier transform is taken with respect to the probe position, the periodic features in Figure 83 appear at specific frequencies.

This particular focused probe experimental geometry was how Hoppe first formulated the concept of ptychography, at least as a gedanken experiment [2]. Instead of using all the probe positions, he proposed using just two positions, which just about provides adequate information to ‘unlock’ the phase problem if the object is indeed a perfect crystal and the diameter of the interfering discs are such that there is only one overlap occurring at any one point in the diffraction

pattern [5]. Moving over towards a general non-crystalline object there is a continuous spectrum of diffracted intensity, and so many diffraction disks, and their interferences, all overlap with one another inseparably. The advantage of collecting a whole field of view of probe positions, and Fourier transforming with respect to probe position coordinate, is that the interferences in the regions of overlap are teased apart.

But this is not the end of the story. It turns out (see below) we can quite easily form an image from a weakly scattering object using the G-set directly. More generally, if the probe is aberrated the lines (i.e. the aperture overlaps) in Figure 84 are complex variables with fine structure. In the case of defocus, the fringes cause any particular point in the diffraction pattern to come bright and dark at different times (the variation of intensity has different phases) as the probe position is scanned. Worse, if the object is strong, diffracted disks do not only interfere with the central disk, but with other, possibly strong, diffracted disks. This means that there can be multiple overlap areas at any one value of U , which can themselves overlap with one another! Our single Fourier transform has not perfectly separated all the ptychographic interferences. We will see that these more complicated effects can be deconvolved via the WDD method; in the meantime, we will explore more fully the weak phase object approximation in the case when the probe-forming optics are perfect.

10.3 Weak phase object approximation: the 'Fat H'

Before we can go further, we have to derive a mathematical definition of the G-set. We recall that the exit wave from our object function $q(r)$, with incident probe $a(r)$, can usually be approximated as a simple point-by-point product,

$$\psi(r) = a(r - R).q(r) . \quad (16)$$

The complex amplitude arriving at the detector is then

$$M(u) = \int \psi(r) e^{i2\pi u \cdot r} dr = \mathfrak{F}\psi(r) , \quad (17)$$

The intensity at the detector is now

$$I(u, R) = |M(u, R)|^2 = |\mathfrak{F}(a(r - R).q(r))|^2 . \quad (18)$$

which can alternatively be written as a convolution of the Fourier transforms of a and q , namely A and Q :

$$I(u, R) = \left| \left(A(u) e^{i2\pi R \cdot u} \right) \oplus Q(u) \right|^2 . \quad (19)$$

The exponential derives from the Fourier shift theorem, and can be thought of a phase ramp added to the aperture transfer function, which, like a thin prism, has the effect of shifting the probe in real space [5].

What we aim to do is to form G , namely

$$G(\mathbf{u}, U) = \int I(\mathbf{u}, \mathbf{R}) e^{i2\pi \mathbf{R} \cdot U} d\mathbf{R}, \quad (20)$$

the Fourier transform of our data with respect to just the \mathbf{R} (probe position coordinate) – but *not* with respect to the detector coordinate, \mathbf{u} . Equation 20 is not very helpful for doing this, so we have to rewrite it, making the convolution explicit so that

$$I(\mathbf{u}, \mathbf{R}) = \iint A(\mathbf{u}_a) Q(\mathbf{u} - \mathbf{u}_a) A^*(\mathbf{u}_b) Q^*(\mathbf{u} - \mathbf{u}_b) e^{i2\pi \mathbf{R} \cdot (\mathbf{u}_a - \mathbf{u}_b)} d\mathbf{u}_a d\mathbf{u}_b, \quad (21)$$

where \mathbf{u}_a and \mathbf{u}_b are dummy variables: the result of the integral does not depend upon them, although they are needed to compute the integral. Now we substitute into Equ 20, to get

$$G(\mathbf{u}, U) = \iiint A(\mathbf{u}_a) Q(\mathbf{u} - \mathbf{u}_a) A^*(\mathbf{u}_b) Q^*(\mathbf{u} - \mathbf{u}_b) e^{i2\pi \mathbf{R} \cdot (\mathbf{u}_a - \mathbf{u}_b + U)} d\mathbf{u}_a d\mathbf{u}_b d\mathbf{R}. \quad (22)$$

Note that A and Q have no dependence on \mathbf{R} , in the above. After all, in ptychography the illumination and the specimen stay the same, wherever the probe is moved. We can therefore integrate over \mathbf{R} to give

$$G(\mathbf{u}, U) = \iint A(\mathbf{u}_a) Q(\mathbf{u} - \mathbf{u}_a) A^*(\mathbf{u}_b) Q^*(\mathbf{u} - \mathbf{u}_b) \delta(\mathbf{u}_a - \mathbf{u}_b + U) d\mathbf{u}_a d\mathbf{u}_b, \quad (23)$$

where we have used the fact that the integral over the complex exponential function is zero everywhere except at $\mathbf{R} = \mathbf{0}$. This is only strictly true if we integrate over infinite limits – a fact that does have consequences when our field of view is finite, as will be discussed in Section 10.6.2. We integrate over \mathbf{u}_a (the choice of \mathbf{u}_a or \mathbf{u}_b is not essential) in which case the delta function in Equation 23 only has value at $\mathbf{u}_b = \mathbf{u}_a + U$, so

$$G(\mathbf{u}, U) = \int A(\mathbf{u}_a) Q(\mathbf{u} - \mathbf{u}_a) A^*(\mathbf{u}_a + U) Q^*(\mathbf{u} - \mathbf{u}_a - U) d\mathbf{u}_a, \quad (24)$$

or more conveniently for our discussion, we substitute $\mathbf{u}_c = \mathbf{u} - \mathbf{u}_a$ to give

$$G(\mathbf{u}, U) = \int Q(\mathbf{u}_c) Q^*(\mathbf{u}_c - U) A^*(\mathbf{u} - \mathbf{u}_c + U) A(\mathbf{u} - \mathbf{u}_c) d\mathbf{u}_c, \quad (25)$$

i.e the convolution

$$G(\mathbf{u}, U) = Q(\mathbf{u}) Q^*(\mathbf{u} - U) \oplus_u A(\mathbf{u}) A^*(\mathbf{u} + U). \quad (26)$$

The subscript u on the convolution is critically important: it denotes that we are only convolving the two functions in the u direction, which we will try to illustrate diagrammatically later.

Now let's try to pick this apart. The first thing to note is that when $U=0$, i.e. along the horizontal axis in Figure 84, we just have $|Q(u)|^2$ convolved with $|A(u)|^2$. This is happening at the zero component of the Fourier transform over R , so it is equivalent to the integral of the intensity of all our diffraction patterns. Physically, this is equivalent to an incoherent convergent beam electron diffraction (CBED) pattern collected using a wholly incoherent tungsten electron source (or an incoherent X-ray source). Each point in the (large) source gives rise to a displaced probe, and all the resulting diffraction patterns add together in the diffraction plane.

The next most important feature arises when we consider a weakly scattering object function. The Fourier transform of a weak specimen has a large spike at $u=0$, corresponding to the largely unscattered transmitted beam. At all other values of u , $Q(u)$ has very small amplitude. In Figure 85a we plot $Q(u)$ and $Q^*(u - U)$ on top of one another. The reader is asked to imagine what the product of these two functions will look like. Clearly, there is a massive spike at $u=U=0$, because this is where the two bright central features of $Q(0)$ meet up: the intensity of the transmitted beam.

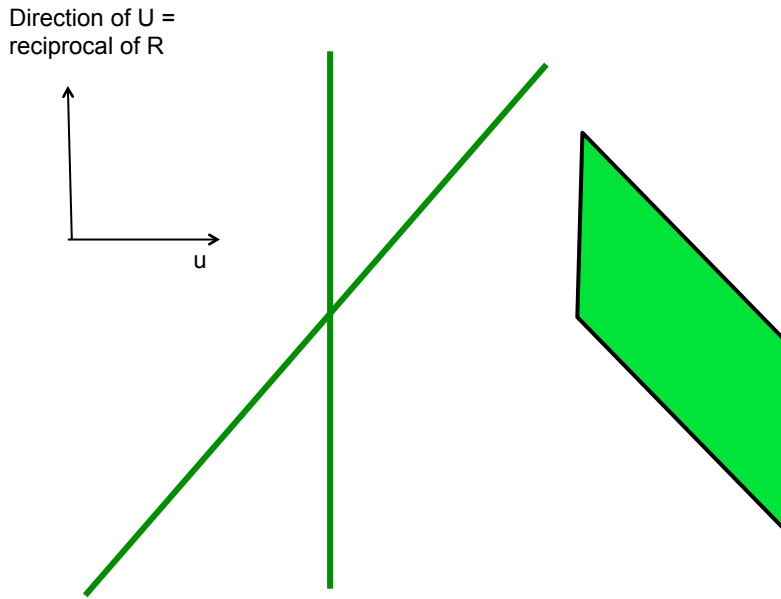


Figure 85: On the left, $Q(u)Q^*(u - U)$ in the G-set for a weakly scattering object. (The same function for a strong object is shown in Figure 94.) On the right we have $A(u)A^*(u + U)$ for a simple 'top hat' aperture function. Each point in the plane has a complex-valued function associated. The lines and shaded regions denote areas where amplitude can exist, though each point will have a complex value associated with it.

Now suppose $Q(0)$, the centre of the diffraction pattern, has an amplitude of unity. Along the vertical axis, $u=0$, we have

$$Q(0)Q^*(-U) = Q^*(-U). \quad (27)$$

That's easy: we can just take the data along this line, reverse it, take the complex conjugate, and Fourier transform to give the complex image! There is another line of high amplitude lying along the locus $u-v = 0$, where we have

$$Q(\mathbf{u})Q^*(0) = Q(u), \quad (28)$$

giving us another, even more direct estimate of Q . Everywhere else in this 2D G-set, at points not on these two lines, only very weakly scattered values of Q and Q^* meet up to form a product. In the weak object approximation, we ignore this second order amplitude. We cannot ignore it when the object is strong.

There is only one problem. To make this function manifest, we must have taken a Fourier transform with respect to R – but this weak phase object, with its central spike in reciprocal space, is, by inference, illuminated by a plane wave, so there has been no probe, no effects of probe movement and thus no ptychographical interference. The phase retrieval only works when we consider equation 26. It is the effect of the convolution of the aperture, leading to the sort of fringes we saw in Figure 2, that gives us the phase. Ironically, once we have done the experiment, we will deconvolve (via the WDD method) the aperture function, and hence obtain the function in equation 36 and Figure 85a in splendid isolation as we show below.

Our G-set is in fact given by equation 26. We first explore where data can arrive in the G-set for a weak object. Now we consider the aperture term in equation 26:

$$A(\mathbf{u})A^*(\mathbf{u} + \mathbf{U}) \quad (29)$$

In one dimension, the simplest aperture is just a top hat function of unity modulus, with no phase components. A little thought will show that in u, U space, equation 29 then describes a skewed parallelogram, as shown in Figure 85b.

Now consider the consequence of the convolution in Equation 26 for a weak specimen. Remember that we are *not* convolving Figure 85a with Figure 85b in the two dimensions like the blurring of a two-dimensional image; we are convolving only along the u direction. At some value of U , we have to take out two rows of pixels along horizontal lines from Figures 85a and 85b, convolve these two one-dimensional functions, and then put the resulting one-dimensional row of pixels back into the G-set at the same value of U . We then do this for all such 1D functions at all values of U .

One way to picture this is as follows. A one-dimensional convolution, say $g(x) \otimes h(x)$ can be achieved by 'flipping' one of the functions, say $g(x)$ becomes $g(-x)$, and then forming the correlation of the two by moving one past the other and observing the integral of the product of the two functions as a function of displacement. For our functions, we can flip the aperture parallelogram (as a function of u), and shift it laterally across Figure 85a, as shown in Figure 86. With

some further thought, it can be seen that we end up with a function that looks like Figure 87, which we will call 'the Fat-H'. Note that under all circumstances, $G(u, U) = G^*(u, -U)$. This is a property of all Fourier transforms of real functions: here we have the Fourier transform of the raw (real, intensity) data along the original R coordinate. In all that follows we can ignore either the top or bottom half of the G -set. When we are dealing with real data sets (which for this technique are enormous), this is an important thing to remember – you can throw away half of it.

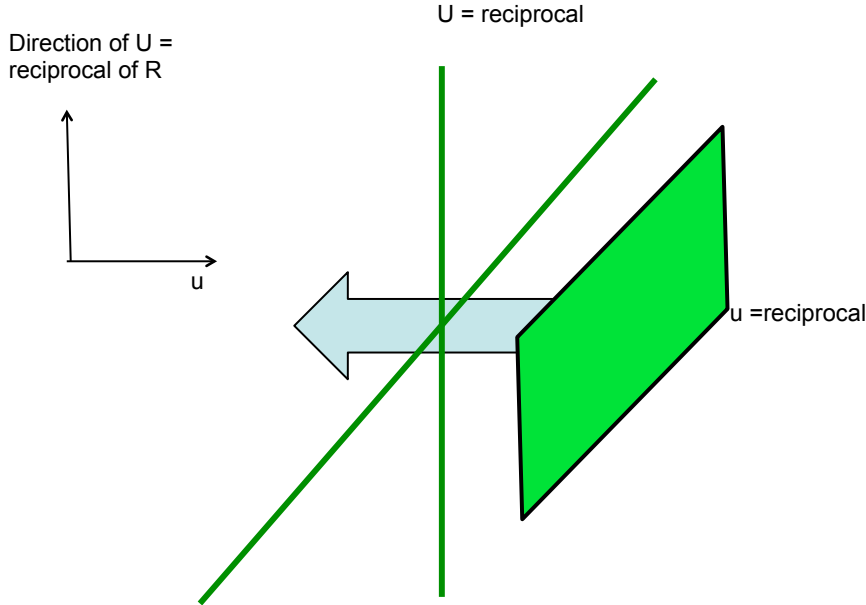


Figure 86: A way of picturing the convolution in Equation 26. For each separate value of U , we must form the integral of the two functions multiplied by one another as the parallelogram (a horizontally flipped version of $A(\mathbf{u})A^*(\mathbf{u} + \mathbf{U})$) is scanned across $Q(\mathbf{u})Q^*(\mathbf{u} - \mathbf{U})$.

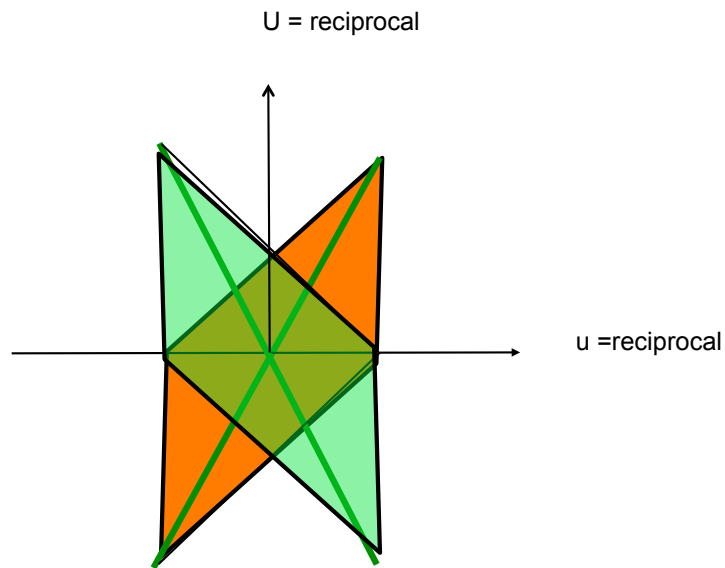


Figure 87: The result of the correlation in Figure 71 for a weak object function. We call this the 'Fat-H'. Lines drawn between the extreme tips of the structure represent symmetric scattering conditions (see section 10.6.5); in reality these are two-dimensional planes extracted from the 4D data set.

So far in this analysis all our 2D diffraction patterns have been represented as 1D lines, the other axis on our diagrams being reserved for the probe position or its Fourier transform. Next we consider what is happening in the 2D diffraction plane (which we label by coordinates u_x and u_y), but picked out at particular values of U , as shown in Figure 88. The five horizontal lines correspond to the five shapes shown in the 2D diffraction pattern plane shown on the right. As we go to higher and higher frequencies in U , which is the rate of change of intensity in the diffraction pattern as a function of probe position, the disks separate further and further – remember that the position of the diffracted disc in \mathbf{u} is itself determined by the periodicity in the object that gave rise to it – recall our discussion concerning Figure 2 and the movement of the shadow image fringes across it.

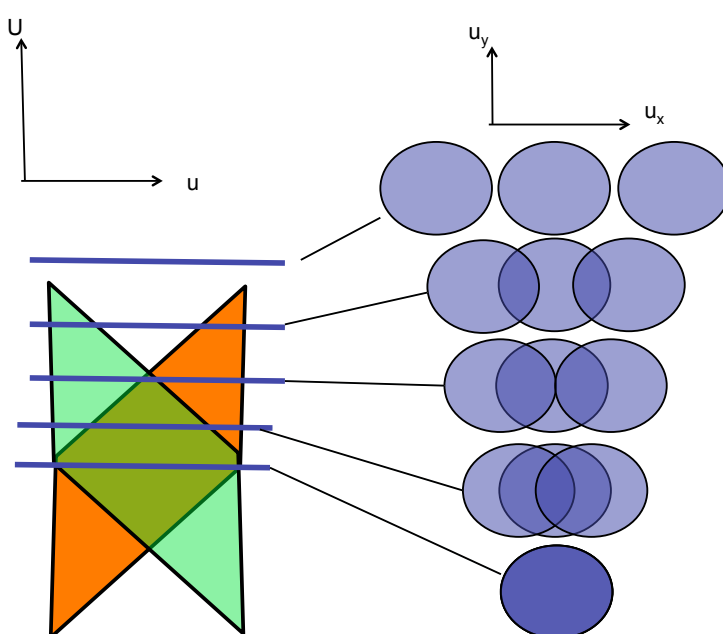


Figure 88: The Fat-H is drawn as a function of U and u , assuming both object and aperture are one-dimensional functions. In fact, every horizontal line in the Fat-H is a 2D plane plotted as u_x and u_y , shown on the right. At higher U (higher Fourier frequencies in the probe position coordinate), we see occluded aperture functions called the 'trotters'. See Figure 92 for an experimental example.

It is exceedingly important to understand that the presence of the occluded aperture shapes in Figure 88 (which below we will see are generally called 'the trotters') does not depend on the object being crystalline or periodic. Our experiment in Figure 2 used a periodic object as a simple demonstration. Any non-crystalline object is still made up of a set of Fourier components. Each one of these components lies at a particular value of U , and therefore can still only be expressed in the overlap areas defined by the aperture functions in the Fat-H. It is also important to appreciate that when we are dealing with the full 4D data set, we must take the Fourier transform with respect to the probe position over *both* its 2D coordinates in order to reveal the occluded aperture shapes (trotters) in Figure 88. Note that the multiply shaded areas are where there *can* be

amplitude for a weakly scattering object, but that does not mean there *is* amplitude at every such position. The presence of amplitude depends on structure in the specimen and the effects of phase aberrations in the aperture function.

As an aside, we will explain why those working in the field often call the occluded aperture functions ‘trotters’. During the 1990s, when this data set was first explored experimentally [122], Rodenburg built a cardboard 3D model of the 2D overlap regions as a function of just one of the coordinates in U . It looked somewhat like Figure 89. Cutting this object horizontally (i.e. at one value of U) gives the shape of the overlaps (Figure 88) in 2D, plotted as a function of u_x and u_y . Slicing vertically down the middle between the two pointed features (the points of smallest disc overlap), gives the Fat H (Figure 87), a function of u and U . This 3D object had an uncanny resemblance to an upside down pig’s trotter. Alternative names for the aperture overlap areas (e.g. the ‘aperture offset functions’) do not have the friendly and compact resonance of ‘trotters’. The name, always used in the plural even though the two occluded apertures are part of one 3D trotter, has stuck amongst the cognoscenti; in what follows we will use it in parentheses. In fact the pig’s trotter is a genuinely useful insight into the nature of the ptychographic data set, even if its name is flippant.

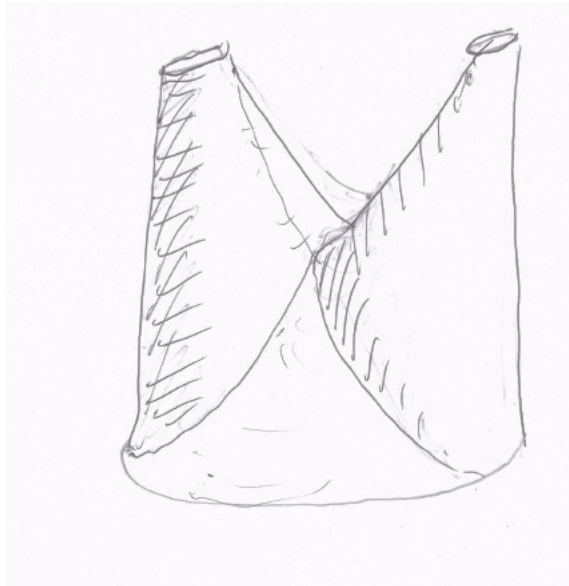


Figure 89: The pig’s trotter in 3D. One coordinate in U is plotted in the vertical direction. The other two coordinates are u_x and u_y . This is an alternative representation of Figure 88.

10.4 Sector detectors

In Section 5.1, we alluded to the fact that when the illumination is a perfectly focused probe, a ptychographic data set arising from a dense (Nyquist) sampling in real space can give us a phase sensitive reconstruction even if we only have four pixels in the detector plane. In fact, there are some very straightforward and direct ways of doing this. Indeed, so direct that the reader may become irritated that we have gone through all the shenanigans of constructing the G-set in order

to describe these techniques, although the G-set will become very important in later sections, when the probe is aberrated and/or of small numerical aperture and/or the specimen is strong.

Equations 27 and 28 tell us that when the object is weak, its Fourier transform is expressed directly into the G-set as a function of U . To get to here, i.e. to effect the ptychographic interference, we need a convergent probe, which consequently gives us the fat-H. If there are no phase components in the aperture, then because the convolution in equation 26 is taken only in the u coordinate, Q is unaffected in all the shaded areas in Figure 90. $Q(U)$ is expressed in every vertical line in the fat-H, lying at any u position within the central undiffracted disk. There is a problem in that the double overlap area, shown shaded in Figure 90, will have little or no amplitude if the object is weak phase. We will not labour through the theory here, but it derives from the fact that the image of a weak phase object has no contrast, and so its Fourier transform is zero. Where there is only one sideband present (unshaded regions in Fig 90) there is contrast in the image.

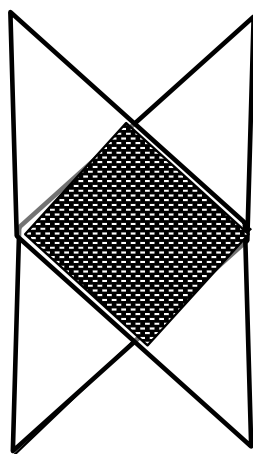


Figure 90: Amplitude in the shaded area of the Fat-H depends on the contrast transfer function of the lens. Unshaded areas are single sidebands, which always express contrast from the specimen, but are still affected by the complex transfer function of the lens. Sector detectors integrate vertical lines of these Fourier components.

So, thinking of the Fat-H, all we have to do is put two 1D detectors at $u>0$ and $u<0$. We do not even need to take the Fourier transform to form the G-set and then back transform; as the probe is scanned, the detectors pick up the original $q(R)$ as a complex number. In the two-dimensional detector plane, we have something that looks like the sector detector shown in Figure 23.

However there is a problem with the transfer characteristics of the images that come out of these detectors. At the centre of the detector we get no transfer at all. This is equivalent to a central bright-field detector in STE/XM: we see nothing if the object is weak phase except uniform brightness over the field of view. Both high frequencies and low frequency pass through the very edge of the detector – i.e. on the outer extremes of the Fat-H. In between we have partial transfer of different frequencies. However, this can be filtered out, at least approximately. Each sector gives an image. The Fourier transforms of these

images give diffractograms (the Fourier transform of the image intensity) – an integral over the area in u spanned by the detector, plotted as a function of U . It is possible to weight each point in each of the four diffractograms by dividing by the line length in the u -direction that intersects the shaded area in Figure 90. For the 4D data set, the division is by the area of the occluded apertures (trotters) at that particular value of U , see [125, 126]. In discussing sector detector transfer properties, the trotters are sometimes super-imposed on the sector detector, as shown in Figure 91. It then becomes clear that there can be benefits in using different diameters and annular divisions of the detector to improve the transfer characteristics of sector detectors.

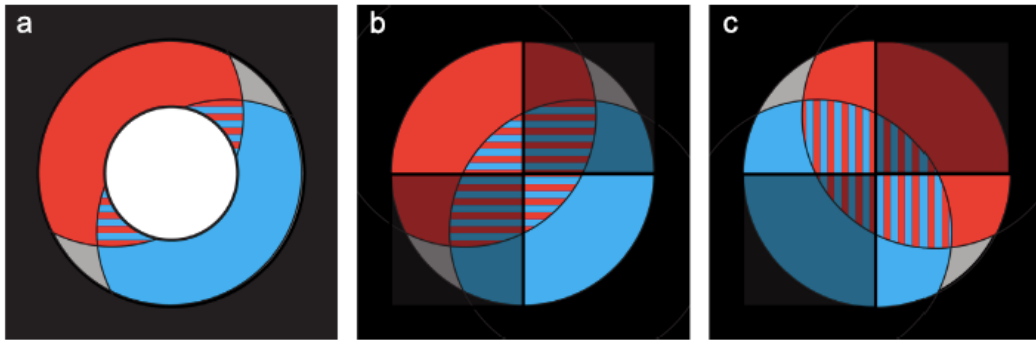


Figure 91: Sections through the trotters (Figure 88) superimposed on sector detectors [127]. This is a way of understanding the frequency transfer properties of sector detectors.

Sector detectors are nowadays commercially available in electron microscopes, although the processing done on the data is usually more approximate than what we have described above. For example, we can get an approximate estimate of the phase gradient in the object simply by taking the difference in the intensities measured at opposite sectors. This signal must then be integrated to give the absolute phase change induced by the specimen.

Note that at the extreme edge of the Fat-H, we get twice the resolution of the bright-field image, whose diffratogram lies along $u=0$: hence the title of the paper where the trotters were first observed [122]. This is nothing mysterious. The coherent bright-field image uses an incident plane wave that has a single incident k -vector. The maximum angle to which this can scatter is half the diameter of the aperture, which occurs at $u=0$ in the Fat-H. When we have a convergent probe, scattering can occur from one side of the aperture to the other, i.e. across its whole diameter. (Note that we are not talking about dark-field intensity, which is scattered outside the aperture disk in the focused probe configuration.) As we have said before, it is well known that conventional microscope resolution is defined by the inverse of the addition of the numerical apertures of both the condenser lens – the range of incident vectors illuminating the specimen – and the objective lens. The same applies here, except our ‘objective’ is the bright-field disk in our diffraction pattern, which we process computationally, not via another lens.

10.5 Dense sampling in real space *and* reciprocal space

Why bother to have an expensive 2D detector in the diffraction plane when a data cube that is densely sampled in real space can give an adequate phase image from a few sector detectors? After all, collecting a 2D diffraction pattern at every densely-sampled probe position massively increases the data we have to handle. The answer is that we can cope with aberrations in the optics, we can handle strong specimens, we can exploit dark-field intensity lying outside the central unscattered beam that contains higher-resolution information than the probe-forming optics, we can remove partial coherence effects, and we can choose to image specific layers in a three-dimensional specimen. The contrast in the final reconstruction is also much better [10]. Of course, iterative algorithms can do all of these things, and without having to have dense sampling in real space. But this section is about maximally sampled data – let us call it the ‘complete data set’ – and why we can invert it with a linear set of transforms. It would seem logical that if we have the data-handling capabilities necessary, the complete data set must be the most informative. Once we have the data, WDD is bound to give a faster reconstruction, but whether it is better than iterative methods remains to be seen.

Several workers have recently been obtaining this complete data set from the electron microscope using an ultra-fast (4,000 fps), single electron counting diffraction camera. Watching the data come out of this in real time as the probe is scanned is extraordinary. The central disc in the diffraction plane fills the entire camera. All that can be seen appears to be pure noise. But when a plane taken out of the G-set is displayed, the occluded apertures (trotters) are astonishingly clear, as shown in the example in Figure 92. This very powerfully demonstrates the dose fractionation property of ptychography. The noise statistics from all the many diffraction patterns has been re-assembled exactly as we expect, in this case by the Fourier transform integration over all probe positions.

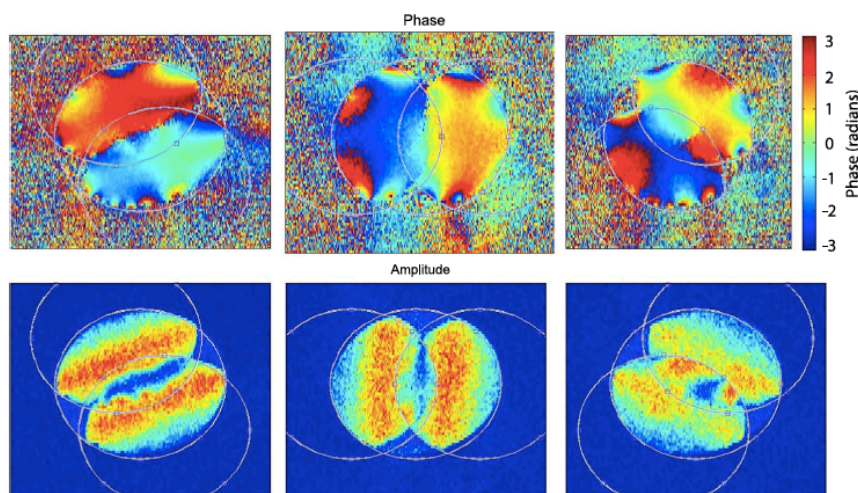


Figure 92: Experimental trotters in phase (top) and modulus (bottom), from [128]. Any aberrations in the lens are very sensitively expressed in the phase. These data were collected on a high-performance aberration-corrected machine, so the presence of phase distortions is surprising.

As we’ve hinted, we can do several things using this data to improve the fidelity of the reconstruction over and above that possible with sector detectors. Most

simply, we can avoid those parts of the Fat-H (shaded in Figure 90) where the sidebands of Q overlap with one another, say by integrating each slice in U only over the relevant trotter shapes. When there is any defocus or aberration in the probe, the shaded areas can have unwanted amplitude. After all, that is how we get contrast into a conventional bright-field image. By defocusing we pump amplitude into the diffractogram, which in the G-set lies along the vertical line, $u=0$. A sector detector unavoidably captures this unwanted region of data.

However, much more effective is to deconvolve the trotters from the data. We do this as usual by taking the Fourier transform across the plane of the trotters (i.e. over u , but *not* over U). In other words, we Fourier transform equation 26 with respect to u . By the convolution theorem this will give us the product of the Fourier transform of the functions depending of A and Q . The coordinate u was the reciprocal of the exit wave variable, r (equ 18). So we say this function depends on U and r , and we call it $H(U, r)$, where

$$H(U, r) = \left(\int Q(u) Q^*(u - U) e^{i2\pi r \cdot u} du \right) \times \left(\int A(u) A^*(u + U) e^{i2\pi r \cdot u} du \right) \quad (30)$$

or,

$$H(U, r) = \chi_Q(U, r) \cdot \chi_A(-U, r), \quad (31)$$

where for some general reciprocal function, F , we have

$$\chi_F(U, r) = \int F(u) F^*(u - U) e^{i2\pi r \cdot u} du, \quad (32)$$

which is our definition of a Wigner distribution, although in signal processing theory it is usually called an ambiguity function.

With reference to Figure 65, let us try to clarify all the steps we have taken, and also to describe the final steps we have to take in order to produce an image using the WDD method. At the top left we have our recorded data, I . This is a real function (intensity) recorded as a function of the diffraction plane coordinate u , and the probe shift coordinate, R . Below it is the G-set: a Fourier transform has been taken vertically over the probe position coordinate, R , transforming it to U ; the u coordinate remains untouched. When the specimen is weak, this is where we see the trotters and the Fat-H. However, the information relating to the specimen is still bound up in the G-set via the convolution in Equation 26. To remove the effects of the aperture, we now transform horizontally along the coordinate of the convolution, u , to the coordinate r , this time leaving the position of the rows of pixels in the G-set unchanged. We can alternatively take the obvious short-cut, which was how this theory was originally formulated [108, 121], by taking Fourier transforms over all the coordinates at once, jumping straight from I to H , as illustrated by the diagonal line. However, we then lose the ability to employ or understand the weak object approximations.

As we have described it, the model depends on reciprocal functions Q and A . The Reader is advised that most of the theoretical development of the WWD method in the literature used the real-space functions q and a in the definition of H and

χ_F . This has no impact on the key ideas, but we now think that understanding the convolution of A and Q in the G-set – and their deconvolution – might be an easier way to understand what are otherwise rather impenetrable equations. However, for the record, the equivalent definition of χ_F for a real space function, f , is

$$\chi_f(U, r) = \int f^*(R) f(R + r) e^{-i2\pi R \cdot U} du. \quad (33)$$

To proceed with the deconvolution we remove the aperture function (which for the time being we presume we know), so that we get

$$\chi_Q(U, r) = \frac{H(U, R)}{\chi_A(U, R)} \quad (34)$$

Like all deconvolutions, this division is highly unstable wherever χ_A is small or zero. Like the iterative update (Section 3.4), we use a Wiener filter, so that

$$\chi_Q(U, r) = \frac{\chi_A^* H(U, R)}{|\chi_A(U, R)|^2 + \varepsilon}. \quad (35)$$

Then all we have to do is back Fourier transform χ_Q with respect to r , to a function only dependent on Q . This is usually called the D-set. It exists in the same coordinate system as the G-set but now the aperture function has been removed. As we pointed out before, we need the aperture function to get the interference data in the first place, but it also places an important restriction on the D-set: there is no information beyond the extreme ends of the fat-H in the vertical direction.

The final step is to decide how we are going to handle the D-set, given by

$$D(u, U) = Q(u) Q^*(u - U) \quad (36)$$

It is bad enough thinking what this represents in a 2D plot, even worse to think about it in 4D! In Figure 93 we show our original interfering disk experiment next to the intensity of a diffraction pattern from a non-periodic object. For a simple periodic object, the disks give us the phase between the unscattered beam and the scattered diffraction orders, i.e. between two points in the diffraction pattern indicated by the white arrows. However, in general, when the object is non-periodic, the D-set gives us the phase difference between every single pixel in the diffraction pattern and every other single pixel. So for our 512x512 scan with 512x512 detector pixels, we have 69 billion pairs of relative phases: 6 are illustrated on Figure 78.

We should remember that there is a cut-off in the U -direction of D because of the finite width of the aperture, hence the finite height of the Fat-H, so only the relative phases between points separated by less than this distance in reciprocal space can be measured. Nevertheless, all pixels, over the whole diffraction plane (including all the dark-field data lying outside the central disk) can be reached by taking multiple steps from one pixel to the next, where each step is smaller than the cut off. In our optical crystalline example (Figure 93), this is like hopping

from one disk to the next. An experiment doing exactly this has been shown to work using electrons scattered from a silicon sample, thus obtaining an image (albeit only of a periodic crystal) at several times the intrinsic resolution of the lens used to form the focused probe [129]. An optical crystalline experiment, stepping much further out into reciprocal space, has also been demonstrated [130].

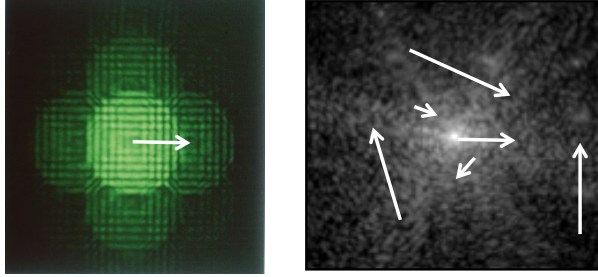


Figure 93: After deconvolution, the D-set ($Q(\mathbf{u})Q^*(\mathbf{u} - \mathbf{U})$) expresses phase difference between every pixel in the Fourier transform of the object and every other pixel. A single phase difference in a crystalline object (left) is easy to understand. All such differences in a non-crystalline object involves billions of phase differences, even for the Fourier transform of a modest field of view.

This process of ‘stepping out’ does not work well with non-periodic objects. The steps must be taken via features of high modulus to reduce the accumulation of phase errors, and thus the method can only use a fraction of the available data. A much more effective solution is to use a projection method [71], which repeatedly sums together phase differences in the 4D cube lying in planes of \mathbf{U} , at the same time working out along the \mathbf{u} -direction. This makes full use of the data, but is beyond the scope of this chapter. For more details, see [71].

Finally we remark – perhaps the most important observation of all – that when we fully deconvolve the data, there is no restriction on Q , and hence the object q , being weak. From the point of view of the mathematics it can be as radically strong as we like, incorporating massive and abrupt phase changes and wild variations in modulus. Of course, real specimens that are very strong tend also to have finite thickness. Sections 10.6.4 and 10.6.5 describe 3D imaging from the bright-field data, but there has been no work done on the influence of 3D scattering processes on dark-field WDD data, or whether 3D structure can be recovered using it. Q can extend as far out into reciprocal space as we like. Indeed, that was the original motive of WDD: to overcome the resolution limitation of the electron microscope lens. Figure 94 shows schematically how a strong object spans the D-set, and the associated cut-off due to the height of the Fat-H.

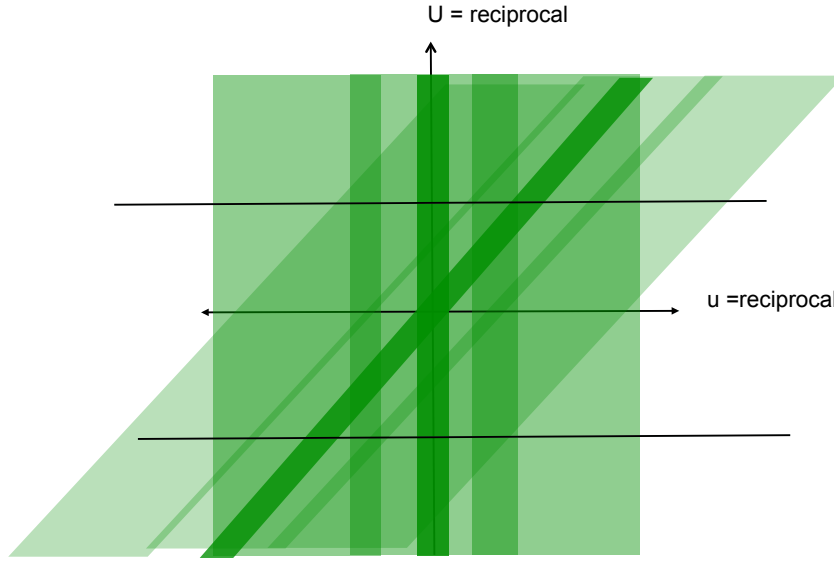


Figure 79: The D-set for a strong object. The reader is encouraged to imagine the product of the two functions $Q(\mathbf{u})Q^*(\mathbf{u} - \mathbf{U})$.

The WWD method was demonstrated with visible light during the 1990s [66, 123, 124]. There has also one proof-of-principle at soft X-ray wavelengths [8]. There is now some renewed interest at electron wavelengths [9, 10].

10.6 WDD: Miscellaneous

10.6.1 Partial coherence

The Wigner distribution, which includes a correlation-type relation, is known as a powerful tool for quantifying and understanding partial coherence, which is about statistical correlation. The same applies to WDD. Perhaps one of its most important characteristics is that, like modal decomposition in the iterative reconstruction methods (Section 8.2), it can remove the effects of partial coherence. This is not surprising – the data are the same, so the same information should exist within it. Many solutions of the phase problem start with the premise that the source and the interference processes are perfectly coherent. This is never quite true for short wavelength sources (X-rays or electrons), and so we must pay close attention to any retrieval strategy that can remove partial coherence.

When the source is of finite size, and every point of emission on it is incoherent with respect to every other point on it, then the mutual complex degree of coherence lying over the lens aperture (which lies in the Fourier plane relative to the source) can be derived via the Van Cittert-Zernike theorem, and is given by

$$\Gamma(u) = \Im s(r), \quad (37)$$

where $s(r)$ is the intensity distribution of the source. Incorporating this into our WDD schema is mathematically tedious [108], so we simply state the result. Now our final D-set is given by

$$D(U, u) = \Gamma(U) Q(\mathbf{u}) Q^*(\mathbf{u} - \mathbf{U}), \quad (38)$$

a surprisingly simple equation. If we think of the image obtained from any one detector pixel at position u as the probe is scanned, then a finite source will blur the coherent image, thus attenuating its high frequency components. The amplitude of the D-set is then attenuated by $\Gamma(U)$ in the U (and only the U) direction, because U is the Fourier transform coordinate of the probe position. In principle we can therefore divide $D(U, u)$ by gamma and restore a coherent data set. Other sources of incoherence like chromatic spread or detector pixel size and, in the case of electron microscopy, instability in the lens power supplies, can also be mapped in the G-set [131].

10.6.2 More about sampling and probe size

If the specimen is weak phase, we are by definition not interested in any scattered data lying outside the central undiffracted disk. The fat-H derives from the assumption that the second order cross terms between the scattered amplitude of Q and Q^* are negligible: only $Q(0)$ times $Q(u)$ has significant value. Equivalently, the D-set only has amplitude along the two lines $u=0$ and $u=U$. This means there is no opportunity for ‘stepping out’ or the projection strategy mentioned in Section 10.5. Under these circumstances the sampling in u only has to be sufficient to adequately deconvolve the occluded aperture function, equ 29 (the trotters). What is that sampling? It clearly must sample the trotters at a higher frequency than any modulus or phase structure within them. That is roughly the inverse of the probe size – i.e. the same sampling condition that applies to all other forms of ptychography. Actually, near the top of the fat-H, where the trotters are tending towards delta functions, their Fourier transform is somewhat wider. However, the deconvolution is only taking out aberrations and having the effect of performing an integration over the trotters, and so it does not need to be perfect.

Contrariwise, when we have a strong specimen, the whole plane of the D-set has significant amplitude. To cleanly undertake the deconvolution and then make use of all the phase differences in the D-set (at least when the object is non-periodic), the sampling in u must be the same as the sampling in U . The final result of the whole process, e.g. obtained via the projection method [71], is a single complex-valued diffraction pattern, plotted over u . Of course, the pitch of pixels in u must therefore be the inverse of the whole field of view (not just the size of the probe). Meanwhile, the weak phase object methods take all the reciprocal information from the U direction. This also has a pixel size that is the inverse of the field of view (as spanned by the probe), but having the flexibility to have so much lower sampling in u vastly reduces the demands on the size of the data set. There are possible solutions to this problem, say by tiling small fields of view, but at the time of writing we are not aware that such alternatives have been explored.

Finally we mention that the theory of WDD, at least for strong objects, depends on undertaking Fourier transforms over infinite limits, or periodically repeating objects. For a continuous image, the data must be attenuated at the edge of the field of view by a soft window function, and even more space must be left within the unit cell to accurately account for the probe function as it scans up to the edge of the field of view. All this is tractable, but a reader who wants to try to do WDD must be aware of this. If the probe is a focused cross-over it is very small, so this is not a significant problem.

10.6.3 Probe solution

The redundancy in the densely-sampled data set is extreme, and so it would be surprising if it were not possible to solve for the probe as well as the object function, as is routine when using iterative methods. Indeed there is such a solution [66] (there must be many others awaiting discovery). It combines elements of blind deconvolution techniques with WDD. In short, whenever we have an estimate of A , we can form the corresponding Wigner Distribution (equ 30) in the H -set. We divide as usual to solve for Q , and then transform, along the u coordinate to the G -set. We then estimate Q from data lying along the U coordinate. This is then used to form its Wigner Distribution. Now the data in the H -set is divided by this estimate, to give an estimate of A 's Wigner Distribution, and hence, after transforming back to the G -set, a new estimate of A ; and so on and so forth. The principle is that the convolution in the u -direction must be consistent with the function estimates taken along the U coordinate.

The method was demonstrated with an optical bench experiment, but given the dismal size of the data that could be gathered in 1993, the results were unimpressive.

10.6.4 3D Imaging

Nellist and co-workers have recently shown that applying WDD with probe functions constructed at different levels of defocus, slices can be selectively imaged from multiple layers of a thick, weak object [10]. This is not the same as solving for the image and then propagating to different defocii, in which case there would be Fresnel effects from out of focus layers. The method seems to pick out an actual plane within the object function. At the time of writing, the work is at a very early stage.

10.6.5 The Bragg-Brentano plane

It was recognised in the work that first described the weak object approximation of the G -set [122], that there exists two lines in it (two planes in the 4D data set) that have unique properties. They lie along $U=2u$ and $U=-2u$. They contain identical information because one is just the complex conjugate of the other. No matter what the aberrations in the aperture may be, if they are symmetric (which they often are), then the central value of the trotters, which lie along these line as illustrated in Figure 87 is always real and unity, because the

complex conjugate components of the symmetric aperture functions cancel each other. This is also true for defocus, which implies that an image formed from this data alone will have, in theory, infinite depth of field. It will be a projection of the object.

Another way of understanding this is that the centres of the spots arise from interference between an incident beam at, say, k_B and a scattered beam at $-k_B$. In conventional X-ray diffraction the specimen is often rotated at half the angular speed of the detector, so that the normal direction of Bragg planes remain parallel to a fixed direction within the specimen. In this way, a flat slice is taken out of 3D reciprocal space, instead of scattering over the curved Ewald sphere, which makes the analysis of the results much easier. A plane in 3D reciprocal space corresponds to a 2D projection in real space. The information along these special planes in the Fat-H is similarly symmetric, and so can also pick out a projection of the object. This projection phenomenon has been experimentally demonstrated on the optical bench [132]. Calculations using Bloch waves for crystalline specimens also indicated that this plane of data is relatively immune to dynamical (multiple) scattering effects, at least compared with the bright-field image [133].

10.6.6 Probe complexity and noise suppression

As mentioned in Section 5.6, the Wigner deconvolution can be used to explore optimal probes in ptychography. It would seem logical that if the χ_A function has few low modulus areas, then the deconvolution should be more stable. This would appear to be the case. Other noise suppression strategies can be employed to avoid low values of χ_A by using redundancy in the data. For more information on these issues, see [71].

11) Conclusions

This chapter has been intended as an elementary introduction to the subject of ptychography. We have also tried to give a flavour of recent developments in each of the many diverse areas of the subject. It is not complete: since the subject took off in 2007, there have been more than 600 papers published on the technique. We have necessarily been selective, reporting on what we think are the most significant aspects of the technique. Other authors would certainly take a different perspective. A previous review chapter was written only a few months after the first iterative phase retrieval ptychography images were published [5]. By the time it was in print it was already out of date. Ten years later the developments in ptychography, some astonishing, continue to pour out of research groups around the world. The literature is expanding exponentially.

Fourier ptychography is undoubtedly under-represented here. Since its appearance in its modern form in 2013, it quickly covered all the ground previously addressed in real-space ptychography, and is pushing ahead, creating an independent field. Several groups are very active as we write, publishing new

algorithms and new variants of the technique. We have also not had space to cover optical encryption with ptychography [134], non-linear ptychographical imaging [135], important developments in incoherent ptychography [90], and the many other refinements of experimental configuration and associated inverse algorithms.

Enabling technologies like microscopy usually follow a common development pattern. First the technology is invented and shown to work for simple test specimens; ptychography is well past this stage in visible light, EUV, X-ray and electron imaging. Next, the method is applied to solve a scientific problem that is ideally suited to the technique; this has been achieved in X-ray and electron ptychography. The method is then applied to answer scientific problems that can *only* be solved by the particular method; this is probably true in the cases of high-resolution X-ray ptycho-tomography, Bragg ptychography, and spectro-ptychography. Finally, the method becomes widely adopted as a standard part of wider scientific investigations, to the extent that its use is regarded a normal component of scientific investigation, fully exploiting its niche capabilities.

As yet, ptychography is not quite at that final stage of maturity. It is most advanced in X-ray imaging. However, so long as it remains confined to the synchrotrons, it can never be very widely used; there just isn't enough beamtime in the world, even though fourth generation synchrotrons will greatly speed up ptycho-tomography. The rapid advance of 'table-top' sources, some of which are very coherent, may bring about a step change in its usage at EUV or X-ray wavelengths in the ordinary laboratory. This may allow it to make a very big impact in all sorts of material and biological studies.

We can make one very reliable prediction. No one is going to throw away their aberration-corrected electron lenses, X-ray Fresnel lenses, KB mirrors or high-resolution optical lenses. There are many indispensable sources of image contrast that will never be delivered by even the cleverest computational optics. The most compelling use of a STEM aberration corrector is the ability to capture material specific signals, like X-ray spectra and electron energy loss spectra (see Chapter **EDITOR**). Modern machines can detect the elemental type of every single atom, at least in a two-dimensional, atomically thin structure [136]. The same applies in X-ray optics, where scanning focused probes can also resolve material-specific X-ray fluorescence, e.g. [137]. Material scientists crave for elemental and bonding information. They regard a scanning electron microscope (SEM) as virtually useless if it does not have an X-ray detector installed on it, despite the fact that modern SEMs can achieve sub-nanometer resolution with ease. Who wants just an image of a specimen when it is possible to know what element every bit of it is made from? Similarly, confocal visible-light microscopy is nowadays indispensable to vast areas of biological research, again relying on excellent lenses to focus a beam onto fluorescent dyes that can spatially resolve the active sites of specific proteins and other molecules. Lenses are here to stay.

But ptychography will find its niche, probably at all wavelengths, and it has many new things to look forward to. The ability to image state mixtures must have huge potential application, although where this will emerge most effectively is

hard to predict. 3D imaging of the refractive index of unstained biological objects must also be ripe for exploitation. We know also that many electron microscopists dream of a very simple electron ptychography microscope. This would comprise a source, one lens, a detector and a computer. However, there are difficulties. Electron ptychography is hard to do quantitatively without a good detector – preferably with single event counting. (The existence of such detectors at hard X-ray energies partly accounts for its success in that field.) But such electrons detectors are expensive (\$700k or so), which rather negates the idea of a low cost, high-resolution table-top TEM. But who knows – there may well be a market for such a machine as detector technology gets less expensive, which it inevitably will. Some X-ray ptychographers assert that it will eventually enable atomic resolution, at the same time overcoming the penetration limits of electron microscopy. We are sceptical: the information per damage event for X-rays is much lower than for electrons [138], but we would not discourage anyone from trying!

Finally we remark, again, that there remains one very fat and large elephant in the room. It has so far been impossible to *prove* mathematically that ptychography works. Despite its ability to skip over the phase problem with such nonchalant ease, it still relies on inverting a highly non-linear set of measurements. So yes, even the simplest heuristic algorithms give good pictures quickly and easily, but proving definitively why they do so is difficult. Even the most advanced algorithms have to make some assumptions. Luckily the applied mathematicians are slowly having their attention drawn to this rich and interesting field: ptychography needs them!

What next? Ptychography with neutrons? Surely the source size is far too incoherent and the interaction cross-section is far too small? But given the advances in the last ten years we have learnt not to discount anything...

References

1. Hegerl, R. and W. Hoppe, *DYNAMIC THEORY OF CRYSTALLINE STRUCTURE ANALYSIS BY ELECTRON DIFFRACTION IN INHOMOGENEOUS PRIMARY WAVE FIELD*. Berichte Der Bunsen-Gesellschaft Fur Physikalische Chemie, 1970. **74**(11): p. 1148-&.
2. Hoppe, W., *DIFFRACTION IN INHOMOGENEOUS PRIMARY WAVE FIELDS .1. PRINCIPLE OF PHASE DETERMINATION FROM ELECTRON DIFFRACTION INTERFERENCE*. Acta Crystallographica Section a-Crystal Physics Diffraction Theoretical and General Crystallography, 1969. **A 25**: p. 495-&.
3. Hoppe, W., *DIFFRACTION IN INHOMOGENEOUS PRIMARY WAVE FIELDS .3. AMPLITUDE AND PHASE DETERMINATION FOR NONPERIODIC OBJECTS*. Acta Crystallographica Section a-Crystal Physics Diffraction Theoretical and General Crystallography, 1969. **A 25**: p. 508-&.
4. Hoppe, W. and G. Strube, *DIFFRACTION IN INHOMOGENEOUS PRIMARY WAVE FIELDS .2. OPTICAL EXPERIMENTS FOR PHASE DETERMINATION OF LATTICE INTERFERENCES*. Acta Crystallographica Section a-Crystal Physics Diffraction Theoretical and General Crystallography, 1969. **A 25**: p. 502-&.
5. Rodenburg, J., *Ptychography and related diffractive imaging methods*. Advances in Imaging and Electron Physics, 2008. **150**: p. 87-184.
6. Hoppe, W., *TRACE STRUCTURE-ANALYSIS, PTYCHOGRAPHY, PHASE TOMOGRAPHY*. Ultramicroscopy, 1982. **10**(3): p. 187-198.
7. Landauer, M., *Indirect Modes of Coherent Imaging in High-Resolution Electron Microscopy*, in Department of Physics. 1996, University of Cambridge, UK: Cambridge UK.
8. Chapman, H.N., *Phase-retrieval X-ray microscopy by Wigner-distribution deconvolution*. Ultramicroscopy, 1996. **66**(3-4): p. 153-172.
9. Yang, H., et al., *Electron ptychographic phase imaging of light elements in crystalline materials using Wigner distribution deconvolution*. Ultramicroscopy, 2017.
10. Yang, H., et al., *Simultaneous atomic-resolution electron ptychography and Z-contrast imaging of light and heavy elements in complex nanostructures*. Nature Communications, 2016. **7**.
11. Rodenburg, J., et al., *Hard-x-ray lensless imaging of extended objects*. Physical review letters, 2007. **98**(3): p. 034801.
12. Thibault, P., et al., *High-resolution scanning x-ray diffraction microscopy*. Science, 2008. **321**(5887): p. 379-382.
13. Thibault, P., M. Guizar-Sicairos, and A. Menzel, *Coherent imaging at the diffraction limit*. Journal of Synchrotron Radiation, 2014. **21**: p. 1011-1018.
14. Spence, J.C.H., et al., *Lensless imaging: A workshop on "new approaches to the phase problem for non-periodic objects"*. Ultramicroscopy, 2001. **90**(1): p. 1-6.
15. Spence, J.C.H., U. Weierstall, and M. Howells, *Phase recovery and lensless imaging by iterative methods in optical, X-ray and electron diffraction*. Philosophical Transactions of the Royal Society of London Series a-

- Mathematical Physical and Engineering Sciences, 2002. **360**(1794): p. 875-895.
16. Nugent, K.A., *Coherent methods in the X-ray sciences*. Advances in Physics, 2010. **59**(1): p. 1-99.
 17. Rodenburg, J., A. Hurst, and A. Cullis, *Transmission microscopy without lenses for objects of unlimited size*. Ultramicroscopy, 2007. **107**(2): p. 227-231.
 18. Rodenburg, J.M. and H.M. Faulkner, *A phase retrieval algorithm for shifting illumination*. Applied physics letters, 2004. **85**(20): p. 4795-4797.
 19. Takahashi, Y., et al., *High-resolution and high-sensitivity phase-contrast imaging by focused hard x-ray ptychography with a spatial filter*. Applied Physics Letters, 2013. **102**(9).
 20. Dierolf, M., et al., *Ptychographic X-ray computed tomography at the nanoscale*. Nature, 2010. **467**(7314): p. 436-U82.
 21. Diaz, A., et al., *Quantitative x-ray phase nanotomography*. Physical Review B, 2012. **85**(2).
 22. Bruck, Y.M. and L. Sodin, *On the ambiguity of the image reconstruction problem*. Optics communications, 1979. **30**(3): p. 304-308.
 23. Bates, R., *Fourier phase problems are uniquely solvable in more than one dimension. I: Underlying theory*. Optik (Stuttgart), 1982. **61**: p. 247-262.
 24. Fienup, J.R., *RECONSTRUCTION OF AN OBJECT FROM MODULUS OF ITS FOURIER-TRANSFORM*. Optics Letters, 1978. **3**(1): p. 27-29.
 25. Fienup, J.R., *PHASE RETRIEVAL ALGORITHMS - A COMPARISON*. Applied Optics, 1982. **21**(15): p. 2758-2769.
 26. Gerchberg, R.W. and W. Saxton, *Phase determination from image and diffraction plane pictures in electron-microscope*. Optik, 1971. **34**(3): p. 275-+.
 27. Gerchberg, R.W., *A practical algorithm for the determination of phase from image and diffraction plane pictures*. Optik, 1972. **35**: p. 237.
 28. Faulkner, H. and J. Rodenburg, *Movable aperture lensless transmission microscopy: a novel phase retrieval algorithm*. Physical review letters, 2004. **93**(2): p. 023903.
 29. Yeh, L.-H., et al., *Experimental robustness of Fourier ptychography phase retrieval algorithms*. Optics express, 2015. **23**(26): p. 33214-33240.
 30. Guizar-Sicairos, M. and J.R. Fienup, *Phase retrieval with transverse translation diversity: a nonlinear optimization approach*. Optics Express, 2008. **16**(10): p. 7264-7278.
 31. Gravel, S. and V. Elser, *Divide and conquer: A general approach to constraint satisfaction*. Physical Review E, 2008. **78**(3).
 32. Maiden, A.M. and J.M. Rodenburg, *An improved ptychographical phase retrieval algorithm for diffractive imaging*. Ultramicroscopy, 2009. **109**(10): p. 1256-1262.
 33. Schropp, A., et al., *Hard x-ray nanobeam characterization by coherent diffraction microscopy*. Applied Physics Letters, 2010. **96**(9): p. 091102.
 34. Hübner, F., et al., *Wave-front phase retrieval in transmission electron microscopy via ptychography*. Physical Review B, 2010. **82**(12): p. 121415.
 35. Yang, C., et al., *Iterative algorithms for ptychographic phase retrieval*. arXiv preprint arXiv:1105.5628, 2011.

36. Wen, Z., et al., *Alternating direction methods for classical and ptychographic phase retrieval*. Inverse Problems, 2012. **28**(11): p. 115010.
37. Marchesini, S., Y.-C. Tu, and H.-t. Wu, *Alternating projection, ptychographic imaging and phase synchronization*. Applied and Computational Harmonic Analysis, 2016. **41**(3): p. 815-851.
38. Marchesini, S., et al., *SHARP: a distributed GPU-based ptychographic solver*. Journal of Applied Crystallography, 2016. **49**(4).
39. D'alfonso, A., et al., *Deterministic electron ptychography at atomic resolution*. Physical Review B, 2014. **89**(6): p. 064101.
40. Zheng, G., R. Horstmeyer, and C. Yang, *Wide-field, high-resolution Fourier ptychographic microscopy*. Nature photonics, 2013. **7**(9): p. 739-745.
41. Hesse, R., et al., *Proximal heterogeneous block implicit-explicit method and application to blind ptychographic diffraction imaging*. SIAM Journal on Imaging Sciences, 2015. **8**(1): p. 426-457.
42. Thibault, P. and M. Guizar-Sicairos, *Maximum-likelihood refinement for coherent diffractive imaging*. New Journal of Physics, 2012. **14**.
43. Godard, P., et al., *Noise models for low counting rate coherent diffraction imaging*. Optics express, 2012. **20**(23): p. 25914-25934.
44. Bian, L., et al., *Fourier ptychographic reconstruction using Poisson maximum likelihood and truncated Wirtinger gradient*. Scientific reports, 2016. **6**.
45. Marchesini, S., et al., *Augmented projections for ptychographic imaging*. Inverse Problems, 2013. **29**(11): p. 115009.
46. Maiden, A., et al., *Quantitative electron phase imaging with high sensitivity and an unlimited field of view*. Scientific reports, 2015. **5**: p. 14690.
47. McDermott, S., et al., *Characterizing a spatial light modulator using ptychography*. Optics Letters, 2017. **42**(3): p. 371-374.
48. Thibault, P. and A. Menzel, *Reconstructing state mixtures from diffraction measurements*. Nature, 2013. **494**(7435): p. 68-71.
49. Cao, S., et al., *Modal decomposition of a propagating matter wave via electron ptychography*. Physical Review A, 2016. **94**(6): p. 063621.
50. Li, P., et al., *Breaking ambiguities in mixed state ptychography*. Optics express, 2016. **24**(8): p. 9038-9052.
51. Huang, X., et al., *Fly-scan ptychography*. Scientific reports, 2015. **5**.
52. Batey, D.J., D. Claus, and J.M. Rodenburg, *Information multiplexing in ptychography*. Ultramicroscopy, 2014. **138**: p. 13-21.
53. Clark, J.N., et al., *Dynamic imaging using ptychography*. Physical review letters, 2014. **112**(11): p. 113901.
54. Odstrcil, M., et al., *Ptychographic coherent diffractive imaging with orthogonal probe relaxation*. Optics express, 2016. **24**(8): p. 8360-8369.
55. Maiden, A., et al., *An annealing algorithm to correct positioning errors in ptychography*. Ultramicroscopy, 2012. **120**: p. 64-72.
56. Zhang, F., et al., *Translation position determination in ptychographic coherent diffraction imaging*. Optics express, 2013. **21**(11): p. 13592-13606.
57. Tripathi, A., I. McNulty, and O.G. Shpyrko, *Ptychographic overlap constraint errors and the limits of their numerical recovery using conjugate gradient descent methods*. Optics express, 2014. **22**(2): p. 1452-1466.

58. Maiden, A.M., M.J. Humphry, and J. Rodenburg, *Ptychographic transmission microscopy in three dimensions using a multi-slice approach*. JOSA A, 2012. **29**(8): p. 1606-1614.
59. Horstmeyer, R., et al., *Diffraction tomography with Fourier ptychography*. Optica, 2016. **3**(8): p. 827-835.
60. Bunk, O., et al., *Influence of the overlap parameter on the convergence of the ptychographical iterative engine*. Ultramicroscopy, 2008. **108**(5): p. 481-487.
61. Miao, J.W., et al., *Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens*. Nature, 1999. **400**(6742): p. 342-344.
62. Kasprowicz, R., R. Suman, and P. O'Toole, *Characterising live cell behaviour: Traditional label-free and quantitative phase imaging approaches*. The international journal of biochemistry & cell biology, 2017. **84**: p. 89-95.
63. Edo, T., et al., *Sampling in x-ray ptychography*. Physical Review A, 2013. **87**(5): p. 053850.
64. da Silva, J.C. and A. Menzel, *Elementary signals in ptychography*. Optics express, 2015. **23**(26): p. 33812-33821.
65. Thibault, P., et al., *Probe retrieval in ptychographic coherent diffractive imaging*. Ultramicroscopy, 2009. **109**(4): p. 338-343.
66. McCallum, B. and J. Rodenburg, *Simultaneous reconstruction of object and aperture functions from multiple far-field intensity measurements*. JOSA A, 1993. **10**(2): p. 231-239.
67. Seiboth, F., et al., *Perfect X-ray focusing via fitting corrective glasses to aberrated optics*. Nature Communications, 2017. **8**: p. 14623.
68. Humphry, M., et al., *Ptychographic electron microscopy using high-angle dark-field scattering for sub-nanometre resolution imaging*. Nature communications, 2012. **3**: p. 730.
69. Maiden, A.M., J.M. Rodenburg, and M.J. Humphry, *Optical ptychography: a practical implementation with useful resolution*. Optics letters, 2010. **35**(15): p. 2585-2587.
70. Guizar-Sicairos, M., et al., *Role of the illumination spatial-frequency spectrum for ptychography*. Physical Review B, 2012. **86**(10): p. 100103.
71. Li, P., T.B. Edo, and J.M. Rodenburg, *Ptychographic inversion via Wigner distribution deconvolution: noise suppression and probe design*. Ultramicroscopy, 2014. **147**: p. 106-113.
72. Maiden, A., et al., *Soft X-ray spectromicroscopy using ptychography with randomly phased illumination*. Nature communications, 2013. **4**: p. 1669.
73. Putkunz, C.T., et al., *Atom-Scale Ptychographic Electron Diffractive Imaging of Boron Nitride Cones*. Physical Review Letters, 2012. **108**(6): p. 4.
74. Peterson, I., et al., *Nanoscale Fresnel coherent diffraction imaging tomography using ptychography*. Optics Express, 2012. **20**(22): p. 24678-24685.
75. Maiden, A.M., et al., *Superresolution imaging via ptychography*. JOSA A, 2011. **28**(4): p. 604-612.
76. Hegerl, R. and W. Hoppe, *INFLUENCE OF ELECTRON NOISE ON 3-DIMENSIONAL IMAGE-RECONSTRUCTION*. Zeitschrift Fur Naturforschung Section a-a Journal of Physical Sciences, 1976. **31**(12): p. 1717-1721.

77. Hoppe, W. and R. Hegerl, *Some remarks concerning the influence of electron noise on 3D reconstruction*. Ultramicroscopy, 1981. **6**(1): p. 205-206.
78. Li, P., et al., *Multiple mode x-ray ptychography using a lens and a fixed diffuser optic*. Journal of Optics, 2016. **18**(5): p. 054008.
79. Shapiro, D.A., et al., *Chemical composition mapping with nanometre resolution by soft X-ray microscopy*. Nature Photonics, 2014. **8**(10): p. 765-769.
80. Cowley, J.M., *IMAGE CONTRAST IN A TRANSMISSION SCANNING ELECTRON MICROSCOPE*. Applied Physics Letters, 1969. **15**(2): p. 58-&.
81. Hoppe, W., *IMAGE PROJECTION OF COMPLEX FUNCTIONS IN ELECTRON MICROSCOPY*. Zeitschrift Fur Naturforschung Part a-Astrophysik Physik Und Physikalische Chemie, 1971. **A 26**(7): p. 1155-&.
82. Hoppe, W., et al., *CONTRAST TRANSFER FOR BRIGHT FIELD IMAGE RECONSTRUCTION WITH TILTED ILLUMINATION IN ELECTRON-MICROSCOPY*. Optik, 1975. **42**(1): p. 43-56.
83. Haigh, S.J., H. Sawada, and A.I. Kirkland, *Atomic structure imaging beyond conventional resolution limits in the transmission electron microscope*. Physical review letters, 2009. **103**(12): p. 126101.
84. Dong, S., et al., *Spectral multiplexing and coherent-state decomposition in Fourier ptychographic imaging*. Biomedical optics express, 2014. **5**(6): p. 1757-1767.
85. Guo, K., S. Dong, and G. Zheng, *Fourier ptychography for brightfield, phase, darkfield, reflective, multi-slice, and fluorescence imaging*. IEEE Journal of Selected Topics in Quantum Electronics, 2016. **22**(4): p. 77-88.
86. Pacheco, S., G. Zheng, and R. Liang, *Reflective Fourier ptychography*. Journal of biomedical optics, 2016. **21**(2): p. 026010-026010.
87. Dong, S., et al., *Sparsely sampled Fourier ptychography*. Optics express, 2014. **22**(5): p. 5455-5464.
88. Tian, L., et al., *Multiplexed coded illumination for Fourier Ptychography with an LED array microscope*. Biomedical optics express, 2014. **5**(7): p. 2376-2389.
89. Zhang, Y., et al., *Self-learning based Fourier ptychographic microscopy*. Optics express, 2015. **23**(14): p. 18471-18486.
90. Dong, S., et al., *High-resolution fluorescence imaging via pattern-illuminated Fourier ptychography*. Optics express, 2014. **22**(17): p. 20856-20870.
91. Ou, X., G. Zheng, and C. Yang, *Embedded pupil function recovery for Fourier ptychographic microscopy*. Optics express, 2014. **22**(5): p. 4960-4972.
92. Zheng, G., *Fourier Ptychographic Imaging*, in *A MATLAB tutorial*. 2016, Morgan & Claypool Publishers.
93. Marrison, J., et al., *Ptychography—a label free, high-contrast imaging technique for live cells using quantitative phase information*. Scientific reports, 2013. **3**: p. 2369.
94. Stockmar, M., et al., *Near-field ptychography: phase retrieval for inline holography using a structured illumination*. Scientific Reports, 2013. **3**.
95. Cosslett, V.E. and W.C. Nixon, *An experimental X-ray shadow microscope*. Proc Roy Soc Ser B Biol Sci, 1952. **900**((140)): p. 422-431.

96. Holler, M., et al., *High-resolution non-destructive three-dimensional imaging of integrated circuits*. Nature, 2017. **543**(7645): p. 402-406.
97. Jiang, N. and J.C. Spence, *On the dose-rate threshold of beam damage in TEM*. Ultramicroscopy, 2012. **113**: p. 77-82.
98. Vine, D.J., et al., *Ptychographic Fresnel coherent diffractive imaging*. Physical Review A, 2009. **80**(6).
99. Hurst, A., et al. *Probe position recovery for ptychographical imaging*. in *Journal of Physics: Conference Series*. 2010. IOP Publishing.
100. Pfeifer, M.A., et al., *Three-dimensional mapping of a deformation field inside a nanocrystal*. Nature, 2006. **442**(7098): p. 63-66.
101. Hruszkewycz, S., et al., *Quantitative nanoscale imaging of lattice distortions in epitaxial semiconductor heterostructures using nanofocused X-ray Bragg projection ptychography*. Nano letters, 2012. **12**(10): p. 5148-5154.
102. Hruszkewycz, S., et al., *High-resolution three-dimensional structural microscopy by single-angle Bragg ptychography*. Nature Materials, 2016.
103. Chamard, V., et al., *Strain in a silicon-on-insulator nanostructure revealed by 3D x-ray Bragg ptychography*. Scientific reports, 2015. **5**: p. 9827.
104. Claus, D., et al., *Dual wavelength optical metrology using ptychography*. Journal of Optics, 2013. **15**(3): p. 035702.
105. Seaberg, M.D., et al., *Tabletop nanometer extreme ultraviolet imaging in an extended reflection mode using coherent Fresnel ptychography*. Optica, 2014. **1**(1): p. 39-44.
106. Zhang, B., et al., *High contrast 3D imaging of surfaces near the wavelength limit using tabletop EUV ptychography*. Ultramicroscopy, 2015. **158**: p. 98-104.
107. Bø Fløystad, J., et al., *Quantitative 3D X - ray Imaging of Densification, Delamination and Fracture in a Micro - Composite under Compression*. Advanced Engineering Materials, 2015. **17**(4): p. 545-553.
108. Rodenburg, J. and R. Bates, *The theory of super-resolution electron microscopy via Wigner-distribution deconvolution*. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 1992. **339**(1655): p. 521-553.
109. Cowley, J.M., *Diffraction physics*. 1995: Elsevier.
110. Cowley, J.M. and A.F. Moodie, *The scattering of electrons by atoms and crystals. I. A new theoretical approach*. Acta Crystallographica, 1957. **10**(10): p. 609-619.
111. Tian, L. and L. Waller, *3D intensity and phase imaging from light field measurements in an LED array microscope*. Optica, 2015. **2**(2): p. 104-111.
112. Li, P., et al., *Separation of three-dimensional scattering effects in tilt-series Fourier ptychography*. Ultramicroscopy, 2015. **158**: p. 1-7.
113. Godden, T., et al., *Ptychographic microscope for three-dimensional imaging*. Optics express, 2014. **22**(10): p. 12513-12523.
114. Suzuki, A., et al., *High-Resolution Multislice X-Ray Ptychography of Extended Thick Objects*. Physical Review Letters, 2014. **112**(5).
115. Tsai, E.H.R., et al., *X-ray ptychography with extended depth of field*. Optics Express, 2016. **24**(25): p. 29090-29109.

116. Wolf, E., *New theory of partial coherence in the space–frequency domain. Part I: spectra and cross spectra of steady-state sources*. JOSA, 1982. **72**(3): p. 343-351.
117. Batey, D., et al., *Reciprocal-space up-sampling from real-space oversampling in x-ray ptychography*. Physical Review A, 2014. **89**(4): p. 043812.
118. Luke, D.R., *Relaxed averaged alternating reflections for diffraction imaging*. Inverse Problems, 2004. **21**(1): p. 37.
119. Wang, S., D. Shapiro, and K. Kaznatcheev. *X-ray ptychography with highly-curved wavefront*. in *Journal of Physics: Conference Series*. 2013. IOP Publishing.
120. Voelz, D.G., *Computational fourier optics: a MATLAB tutorial*. 2011: Spie Press Bellingham, WA.
121. Bates, R. and J. Rodenburg, *Sub-Ångström transmission microscopy: a Fourier transform algorithm for microdiffraction plane intensity information*. Ultramicroscopy, 1989. **31**(3): p. 303-307.
122. Rodenburg, J., B. McCallum, and P. Nellist, *Experimental tests on double-resolution coherent imaging via STEM*. Ultramicroscopy, 1993. **48**(3): p. 304-314.
123. McCallum, B. and J. Rodenburg, *Two-dimensional demonstration of Wigner phase-retrieval microscopy in the STEM configuration*. Ultramicroscopy, 1992. **45**(3-4): p. 371-380.
124. Friedman, S.L. and J. Rodenburg, *Optical demonstration of a new principle of far-field microscopy*. Journal of Physics D: Applied Physics, 1992. **25**(2): p. 147.
125. Landauer, M., B. McCallum, and J. Rodenburg, *Double resolution imaging of weak phase specimens with quadrant detectors in the STEM*. Optik, 1995. **100**(1): p. 37-46.
126. Hornberger, B., M. Feser, and C. Jacobsen, *Quantitative amplitude and phase contrast imaging in a scanning transmission X-ray microscope*. Ultramicroscopy, 2007. **107**(8): p. 644-655.
127. Pennycook, T.J., et al., *Efficient phase contrast imaging in STEM using a pixelated detector. Part 1: Experimental demonstration at atomic resolution*. Ultramicroscopy, 2015. **151**: p. 160-167.
128. Nellist, P. and J. Rodenburg, *Electron ptychography. I. Experimental demonstration beyond the conventional resolution limits*. Acta Crystallographica Section A: Foundations of Crystallography, 1998. **54**(1): p. 49-60.
129. Nellist, P., B. McCallum, and J. Rodenburg, *Resolution beyond the 'information limit' in transmission electron microscopy*. Nature, 1995. **374**(6523): p. 630.
130. McCallum, B. and J. Rodenburg, *Error analysis of crystalline ptychography in the STEM mode*. Ultramicroscopy, 1993. **52**(1): p. 85-99.
131. Nellist, P. and J. Rodenburg, *Beyond the conventional information limit: the relevant coherence function*. Ultramicroscopy, 1994. **54**(1): p. 61-74.
132. Plamann, T. and J. Rodenburg, *Double resolution imaging with infinite depth of focus in single lens scanning microscopy*. Optik, 1994. **96**(1): p. 31-36.

133. Plamann, T. and J. Rodenburg, *Electron ptychography. II. Theory of three-dimensional propagation effects*. Acta Crystallographica Section A: Foundations of Crystallography, 1998. **54**(1): p. 61-73.
134. Shi, Y., et al., *Optical image encryption via ptychography*. Optics letters, 2013. **38**(9): p. 1425-1427.
135. Odstrcil, M., et al., *Nonlinear ptychographic coherent diffractive imaging*. Optics Express, 2016. **24**(18): p. 20245-20252.
136. Krivanek, O.L., et al., *Atom-by-atom structural and chemical analysis by annular dark-field electron microscopy*. Nature, 2010. **464**(7288): p. 571-574.
137. Yang, L., et al., *Imaging of the intracellular topography of copper with a fluorescent sensor and by synchrotron x-ray fluorescence microscopy*. Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(32): p. 11179-11184.
138. Henderson, R., *THE POTENTIAL AND LIMITATIONS OF NEUTRONS, ELECTRONS AND X-RAYS FOR ATOMIC-RESOLUTION MICROSCOPY OF UNSTAINED BIOLOGICAL MOLECULES*. Quarterly Reviews of Biophysics, 1995. **28**(2): p. 171-193.