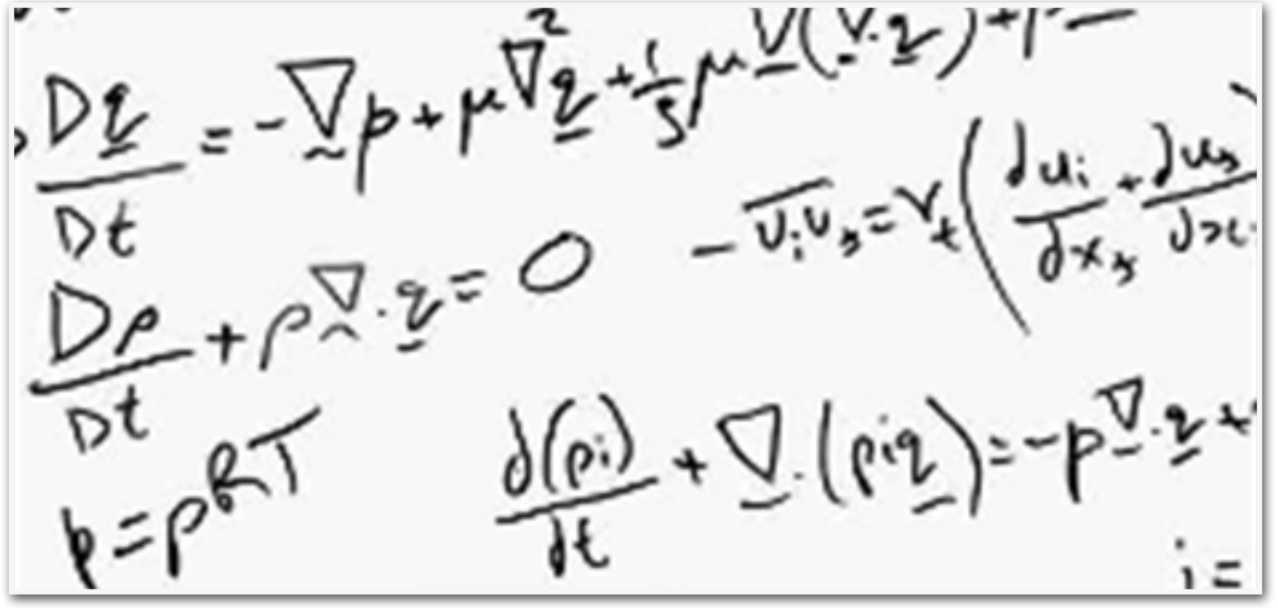
Science and Technology Facilities Council

Developing Improved Optimization Algorithms for Ptychographic Image Reconstruction

Jaroslav Fowkes
STFC Computational Mathematics Group
Visiting Research Fellow at Oxford University



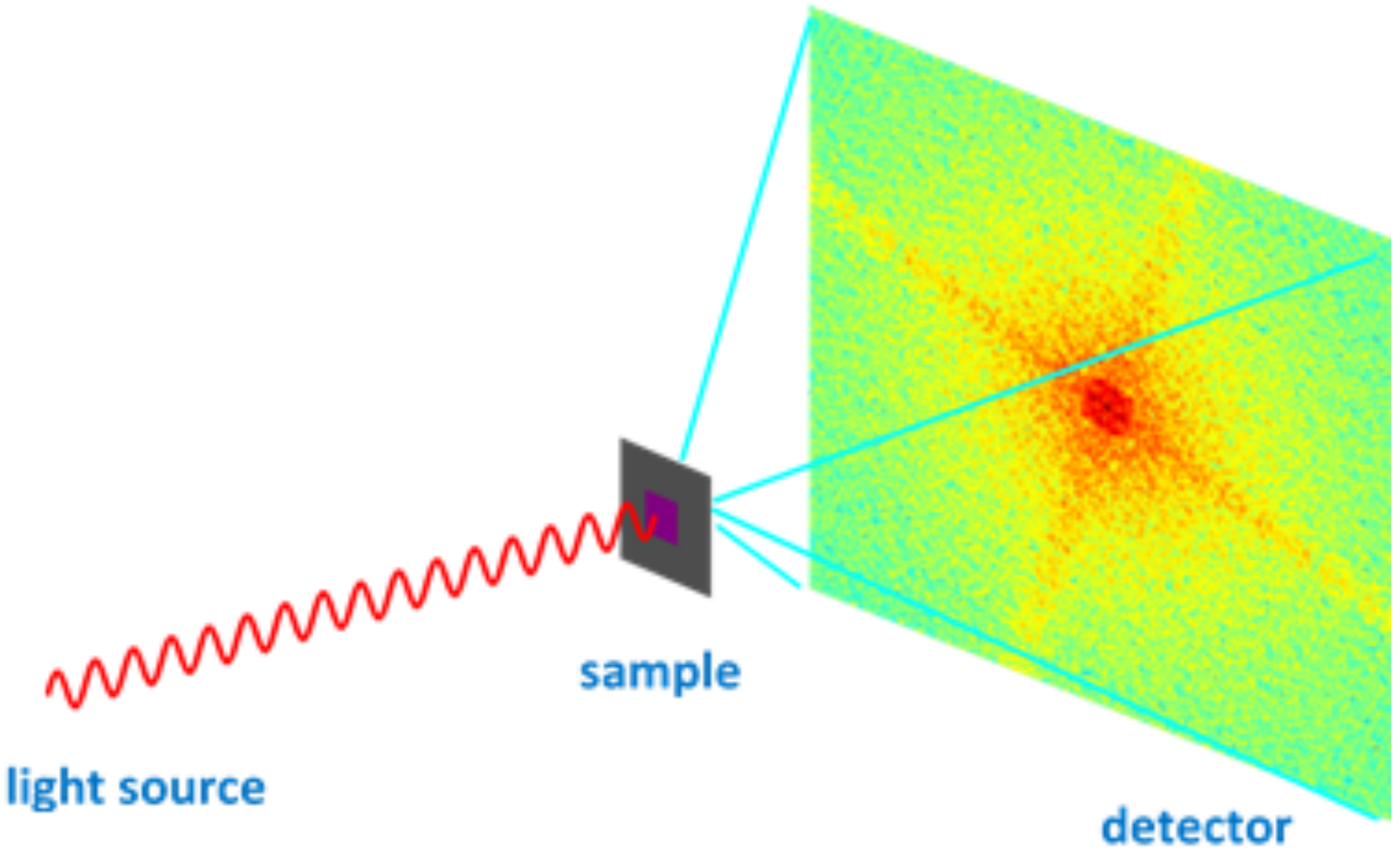
STFC Rutherford Appleton Laboratory

STFC manages the UK's major science facilities, at the Harwell campus we have:

- Diamond Light Source
- ISIS Neutron Source
- Central Laser Facility
- RAL Space

UKRI Science and Technology Facilities Council

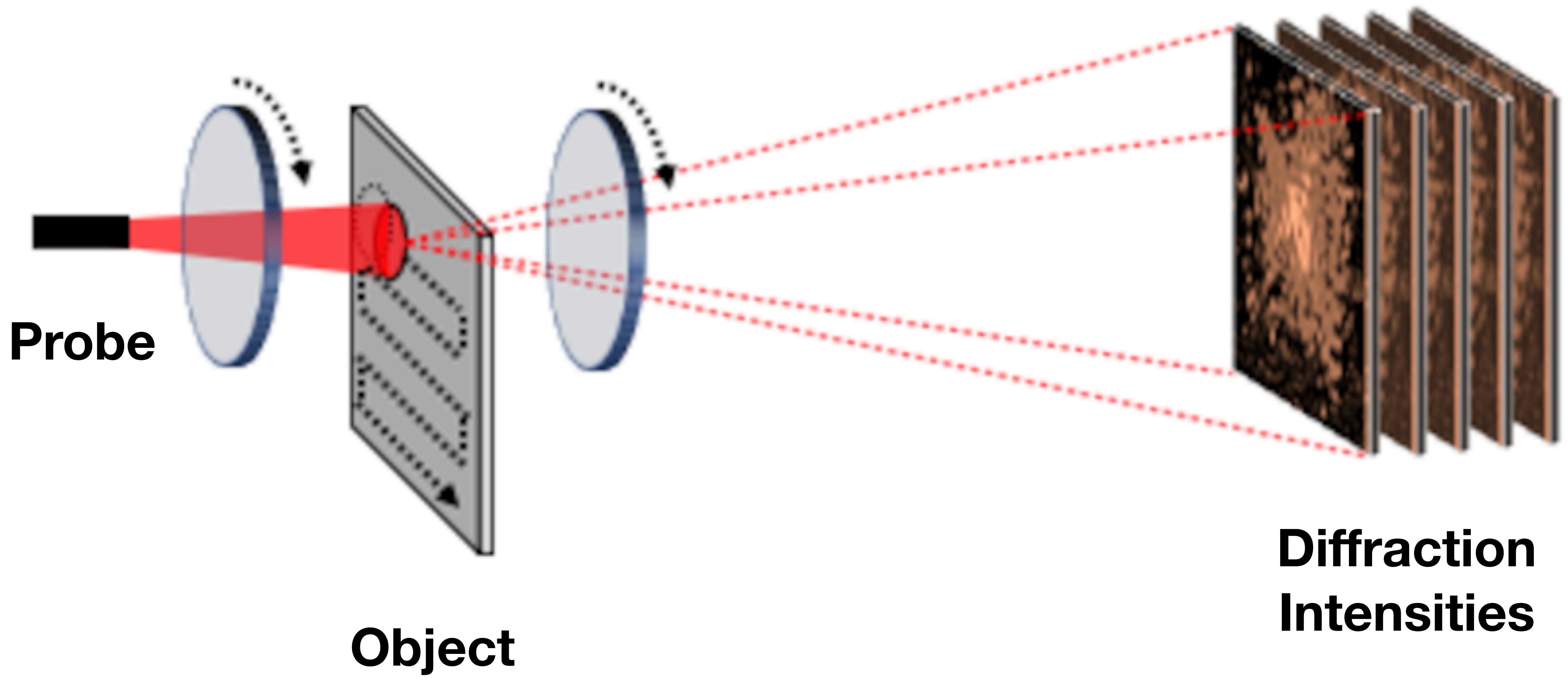
Coherent Diffractive Imaging



“CDI is a ‘lensless’ technique that allows imaging of matter at a spatial resolution not limited by lens aberrations. This technique exploits the measured diffraction pattern of a coherent beam scattered by an object to retrieve spatial information”

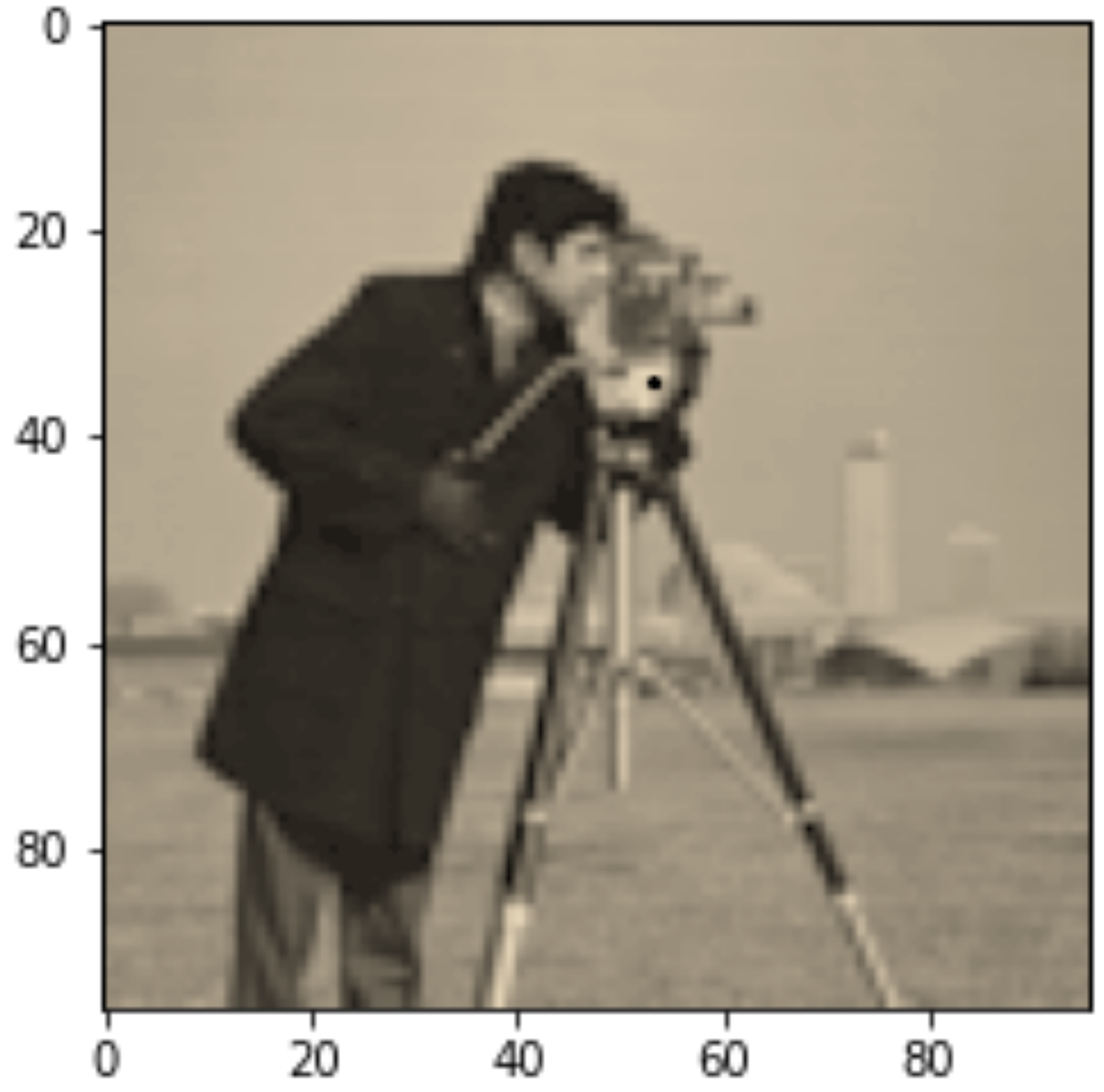
Science and
Technology
Facilities Council

Ptychography

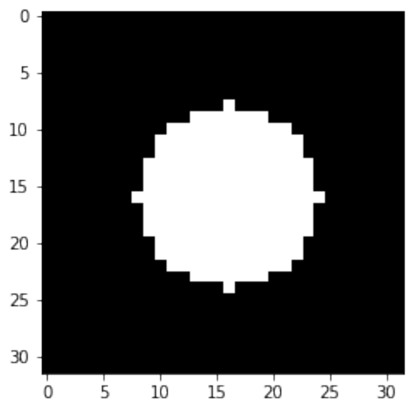


“Ptychography is an imaging technique in which a localized illumination scans overlapping regions of an object and generates a set of diffraction intensities used to computationally reconstruct its complex-valued transmission function”

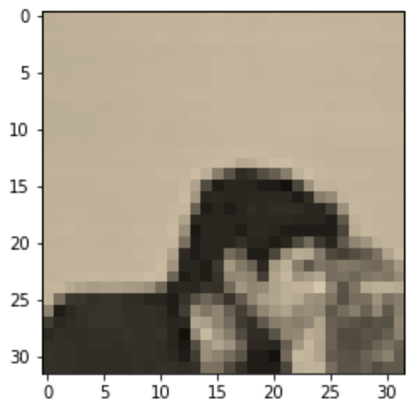
Exit Wave



$o \in \mathbb{C}^{m \cdot m}$



$p \in \mathbb{C}^{n \cdot n}$



$o_j \in \mathbb{C}^{n \cdot n}$

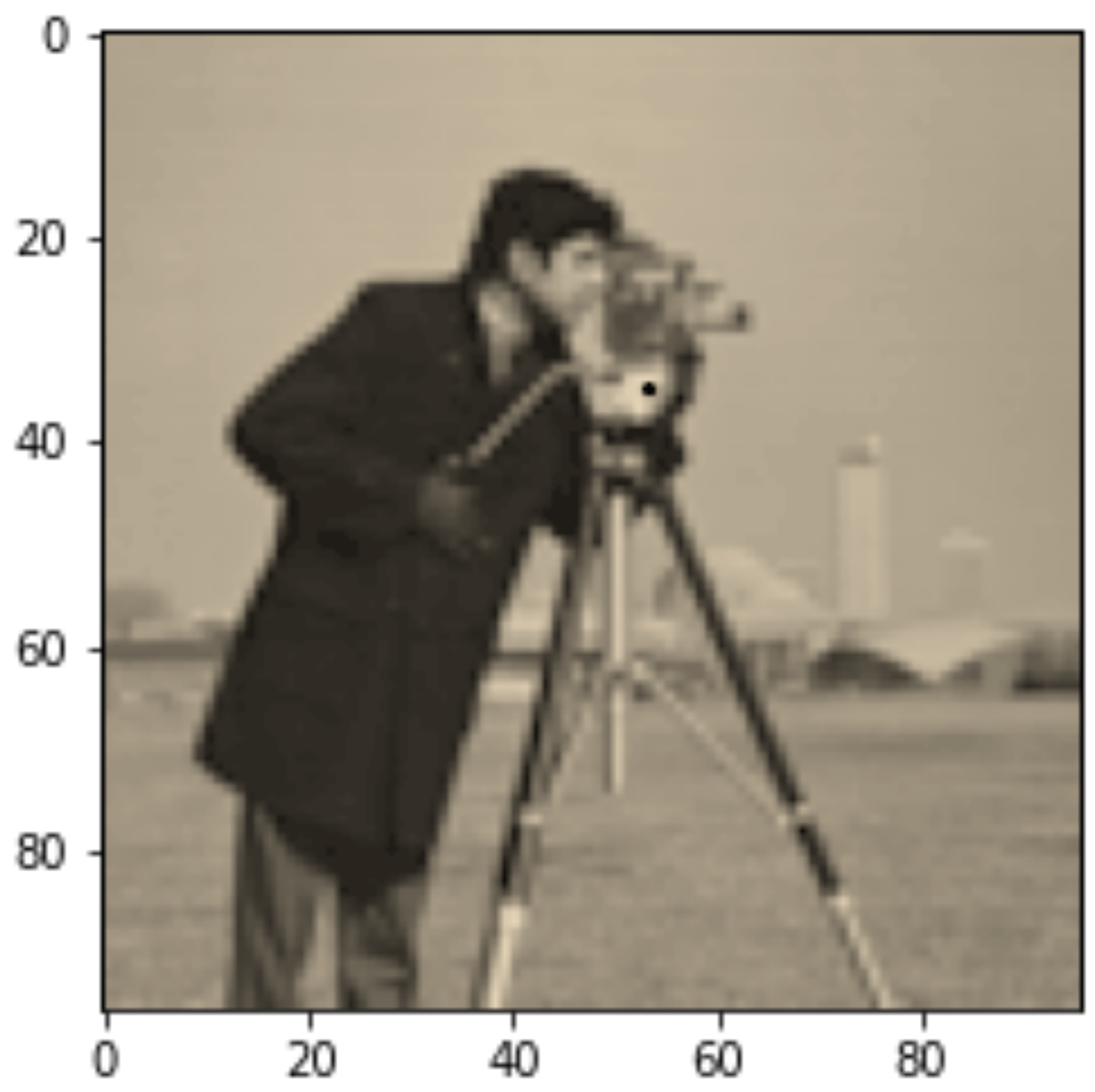
Exit Wave:

$$e_j = o_j \, p$$

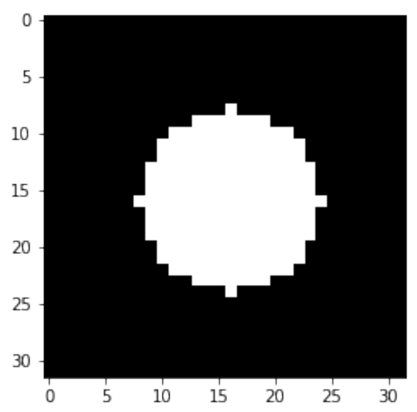
by product we mean elementwise multiplication

Science and Technology Facilities Council

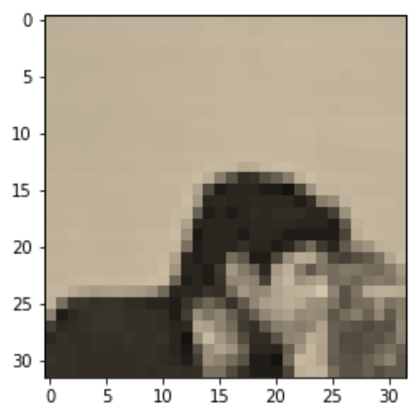
Wave Model



$o \in \mathbb{C}^{m \cdot m}$



$p \in \mathbb{C}^{n \cdot n}$



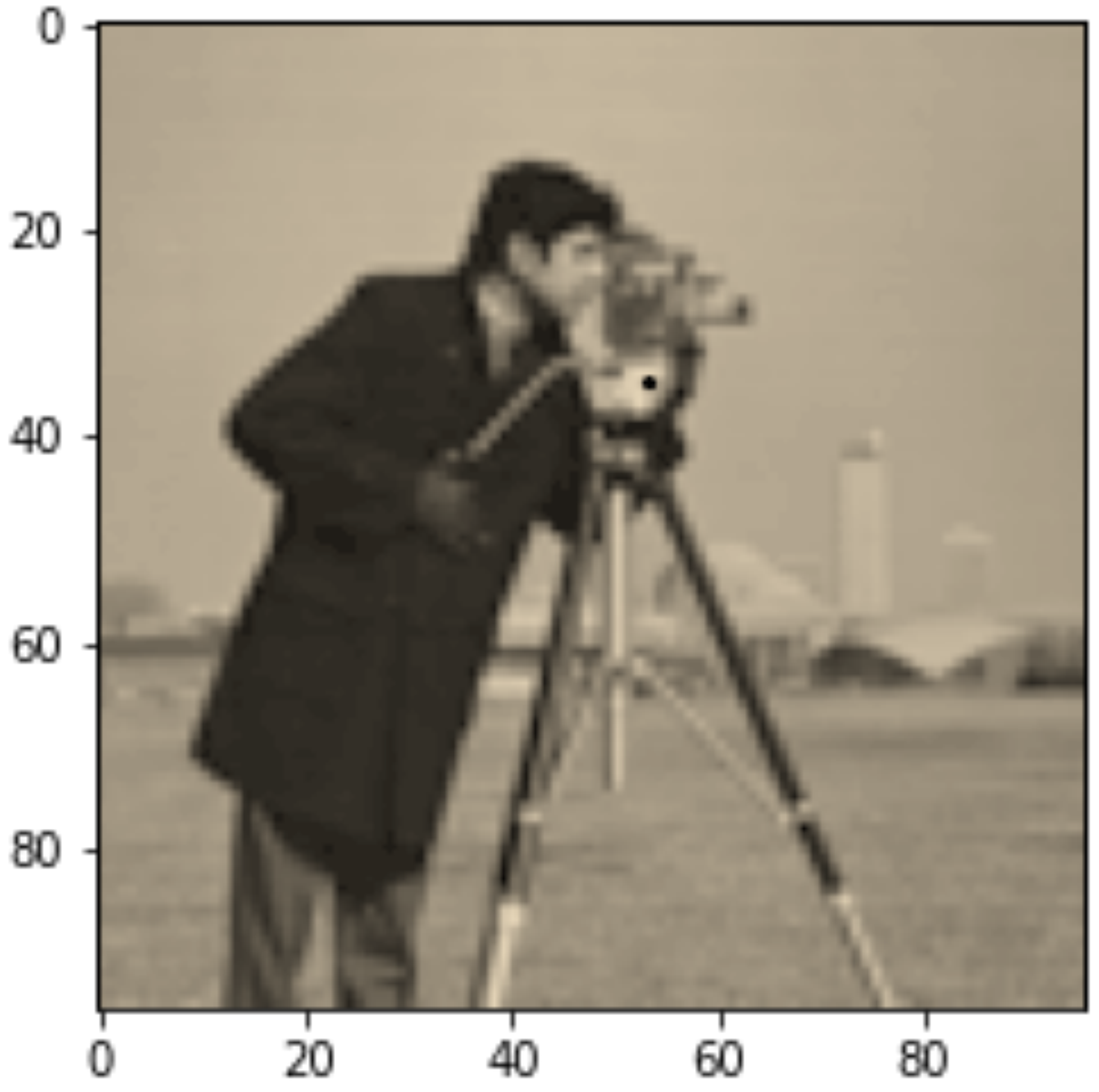
$o_j \in \mathbb{C}^{n \cdot n}$

Wave Model:

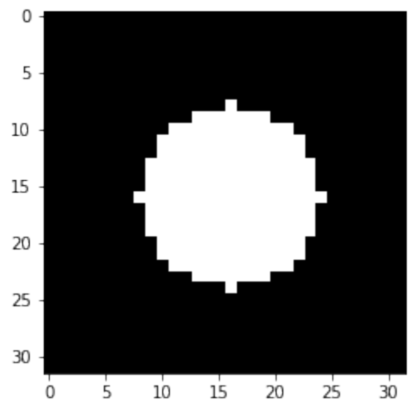
$$w_j = F(\, o_j\, p \,)$$

where F is the Discrete Fourier Transform (in 2D)

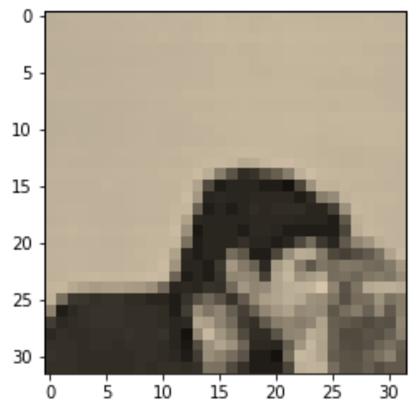
Forward Model



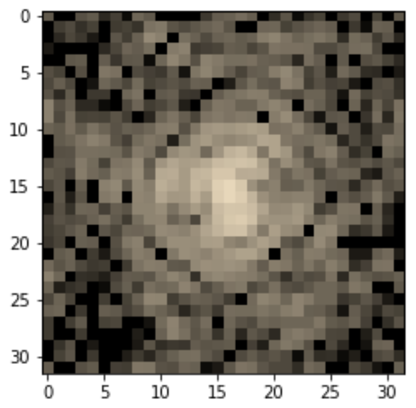
$o \in \mathbb{C}^{m \cdot m}$



$p \in \mathbb{C}^{n \cdot n}$



$o_j \in \mathbb{C}^{n \cdot n}$



$i_j \in \mathbb{R}^{n \cdot n}$

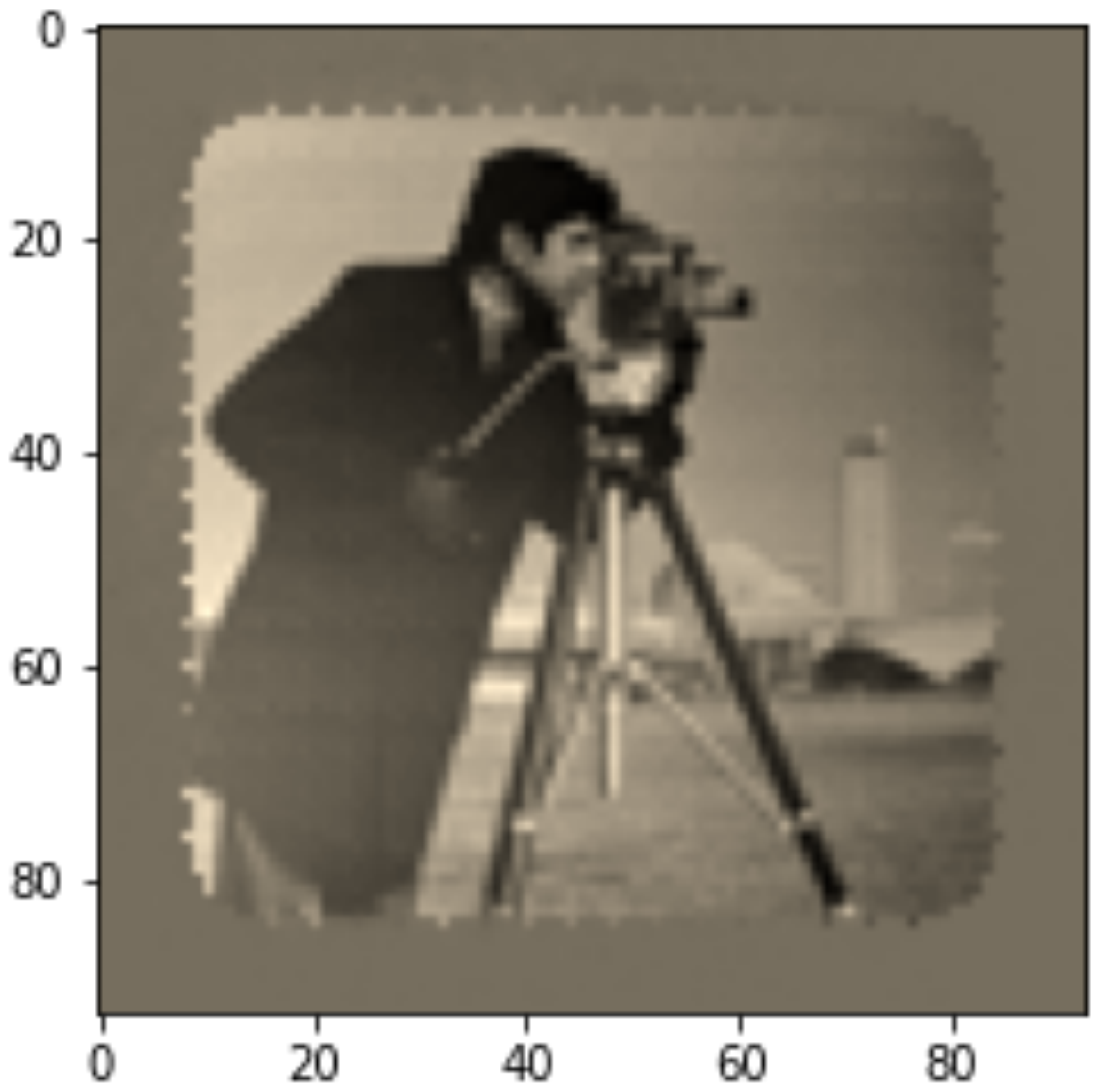
Forward Model:

$$|w_j|^2 = |F(\,o_j\,p\,)|^2 \approx i_j$$

where F is the Discrete Fourier Transform (in 2D)

Science and
Technology
Facilities Council

Reconstruction



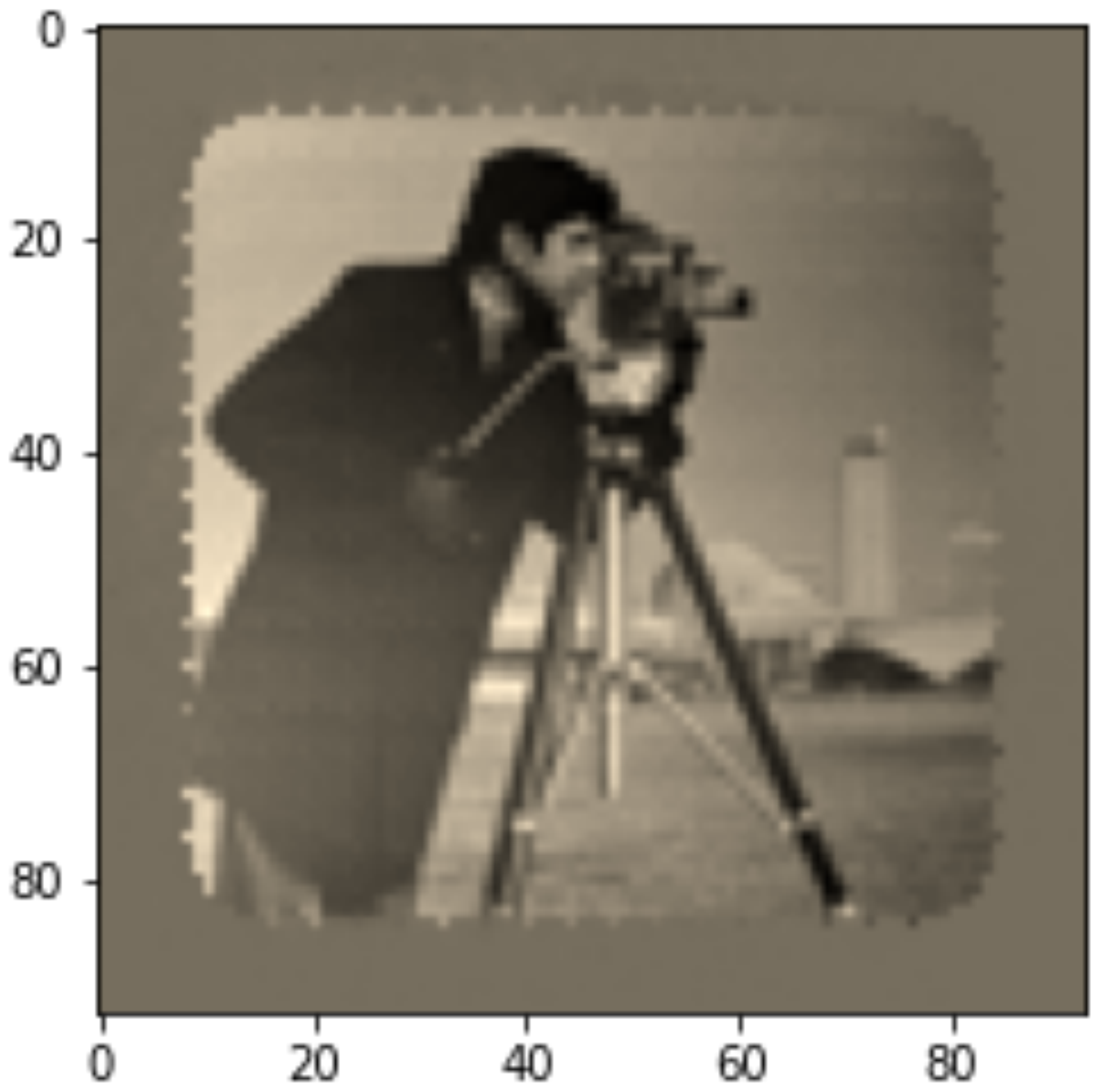
Reconstruct object (and probe) from measured intensities:

$$\min_{o_j\, p} \sum_j \left\| \, |F(o_j\, p)|^2 - i_j \, \right\|^2$$

The choice of error nom here is key (e.g. l^2, likelihood, etc).

Science and Technology Facilities Council

Inverse Problem



Reconstruct object (and probe) from measured intensities:

$$\min_{o_j\,p} \sum_j \left\| \, |F(o_j\,p)|^2 - i_j \, \right\|^2$$

a sum of nonlinear least-squares problems in $\mathbb{C}^{n \cdot n}$.

Science and
Technology
Facilities Council

Inverse Problem

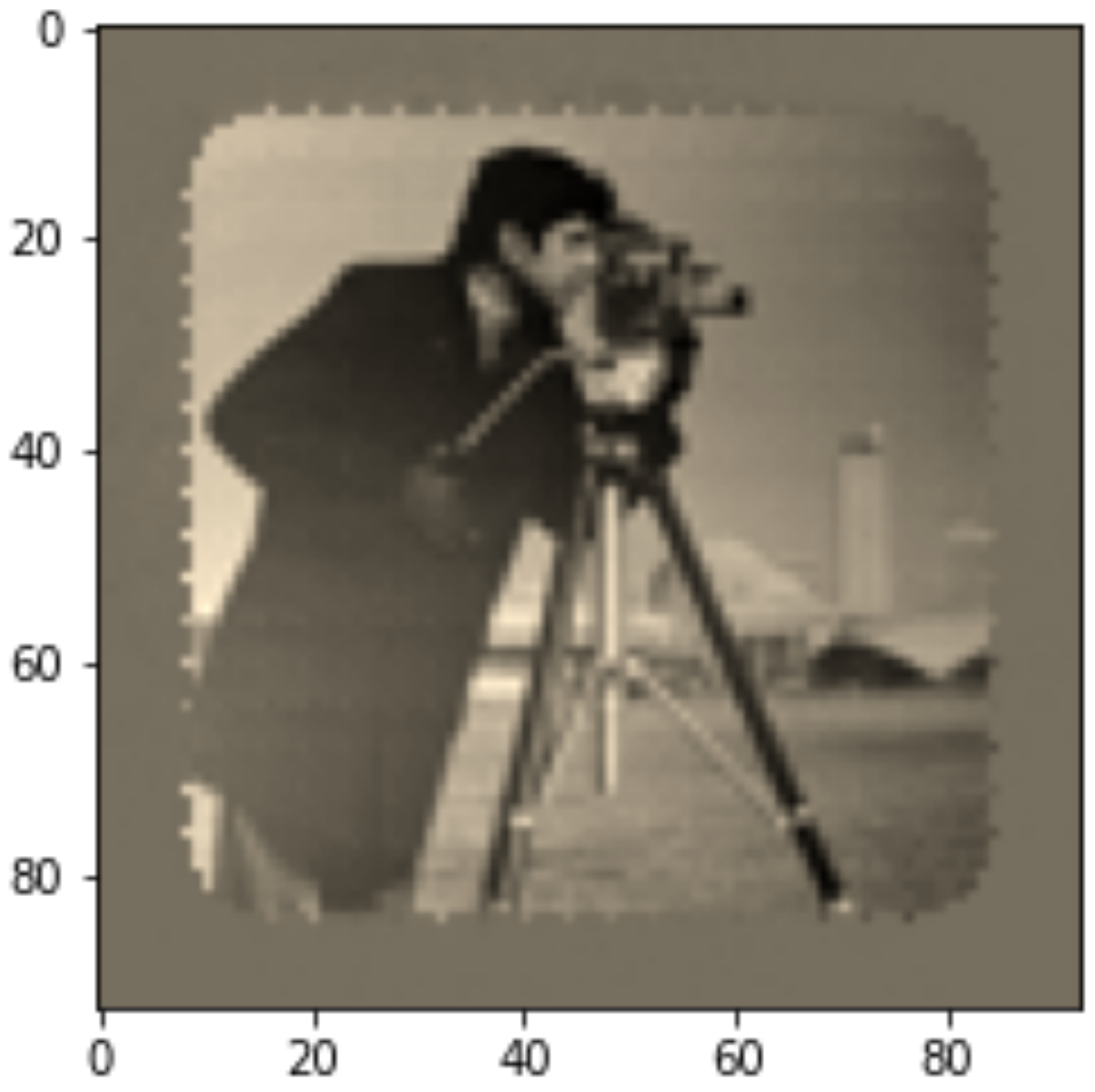
Unfortunately the inverse problem is not very well-posed, issues include:

- constant amplitude scaling: $o_j\, p = (A o_j)\,(A^{-1} p)$
- constant phase offset: $o_j\, p = (e^{i\phi} o_j)\,(e^{-i\phi} p)$
- global probe and object translation
- linear phase ramp (i.e. DFT shift theorem)

However most of these issues can be addressed with suitable constraints.

Science and Technology Facilities Council

Nonlinear least-squares

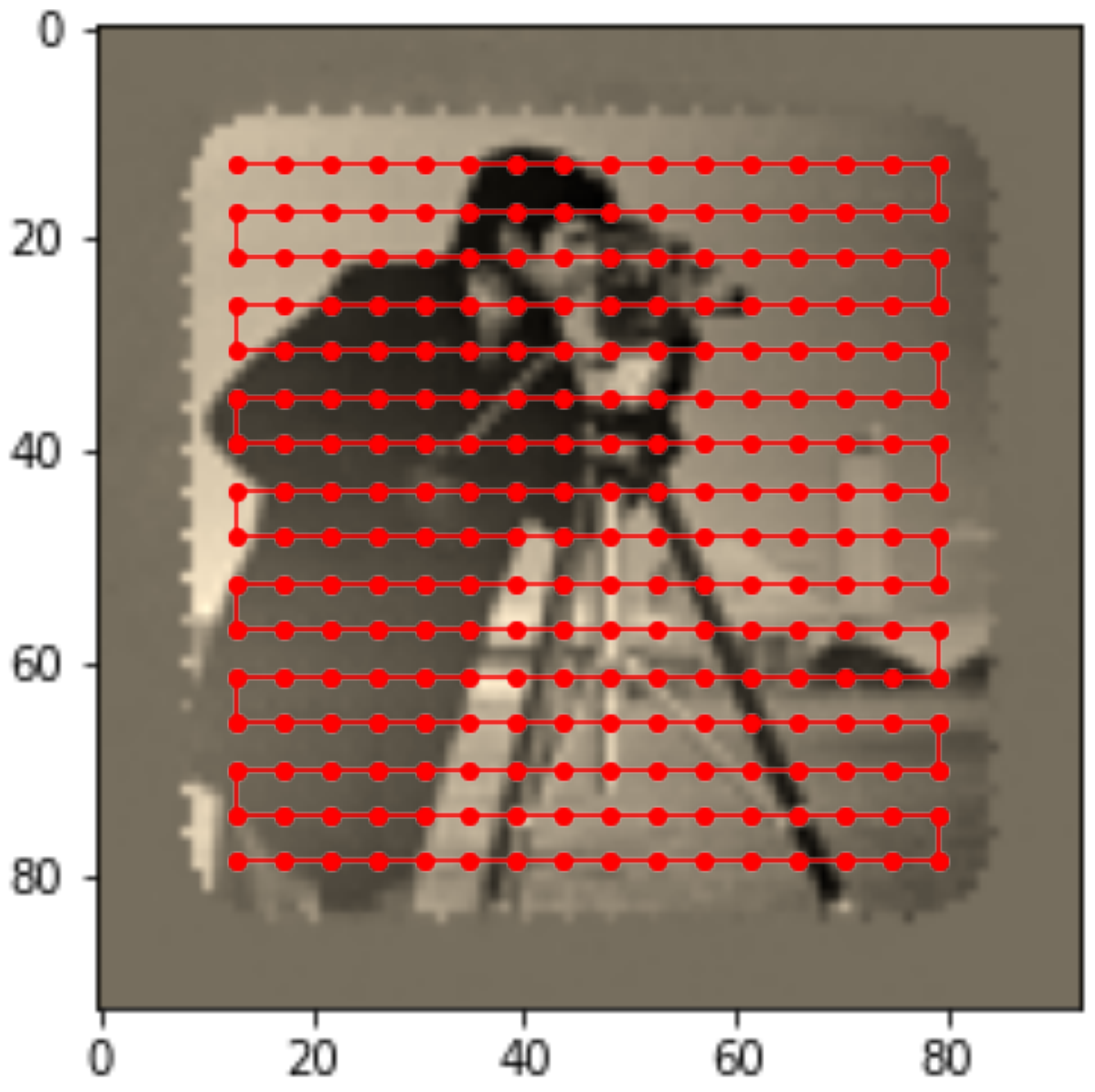


Reconstruct object (and probe) from measured intensities:

$$\min_{o_j\, p} \sum_j \left\| \, |F(o_j\, p)|^2 - i_j \, \right\|^2$$

a sum of nonlinear least-squares problems in $\mathbb{C}^{n \cdot n}$.

Nonlinear least-squares



Reconstruct object (and probe) from measured intensities:

$$\min_{o_j\, p} \sum_j \left\| \, |F(o_j\, p)|^2 - i_j \right\|^2$$

a sum of nonlinear least-squares problems in $\mathbb{C}^{n \cdot n}$.

Science and
Technology
Facilities Council

Nonlinear least-squares

Given a nonlinear least-squares problem

$$f(x) = \|r(x)\|^2$$

with nonlinear residual $r: \mathbb{R}^n \to \mathbb{R}^m$, the gradient is given by

$$\nabla f(x) = 2J(x)^T r(x)$$

where $J(x) = [\partial r_i / \partial x_j]_{ij}$ is the Jacobian matrix, and the Hessian is given by

$$\nabla^2 f(x) = 2J(x)^T J(x) + 2\sum_i r_i(x)\,\nabla^2 r_i(x)$$

However, often in practice the Hessian is too expensive to compute.

Gauss-Newton approximation

The first order optimality conditions for unconstrained optimization are:

$$\nabla f(x) = 0$$

applying Newton's method to this equation gives the step s^k from x^k as

$$\nabla^2 f(x^k)\, s^k = -\nabla f(x^k)$$

The idea behind Gauss-Newton is to use the approximation

$$\nabla^2 f(x^k) \approx 2J^T(x^k)J(x^k) \cancel{+ 2\sum_i r_i(x^k)\nabla^2 r_i(x^k)}$$

In particular this is asymptotically exact for zero residual problems.

Science and Technology Facilities Council

L-BFGS Approximation

A BFGS step approximates the Hessian with the matrix B^k

$$s^{k+1} = -\left(B^{k+1}\right)^{-1} \nabla f(x^{k+1})$$

using past gradients $y^k = \nabla f(x^{k+1}) - \nabla f(x^k)$ as the rank-2 update

$$B^{k+1^{-1}} = \left(I - \frac{s^k y^{k^T}}{y^{k^T} s^k}\right) B^{k^{-1}} \left(I - \frac{y^k s^{k^T}}{y^{k^T} s^k}\right) + \frac{s^k s^{k^T}}{y^{k^T} s^k}$$

L-BFGS only uses a fixed number (limited memory) of past s^k, y^k .

Science and Technology Facilities Council

BFGS

Broyden, Fletcher, Goldfarb, Shanno

Nonlinear CG

Nonlinear CG is a generalisation of CG to general nonlinear objectives.

Starting from the steepest descent direction

$$s^0 = -\nabla f(x^0)$$

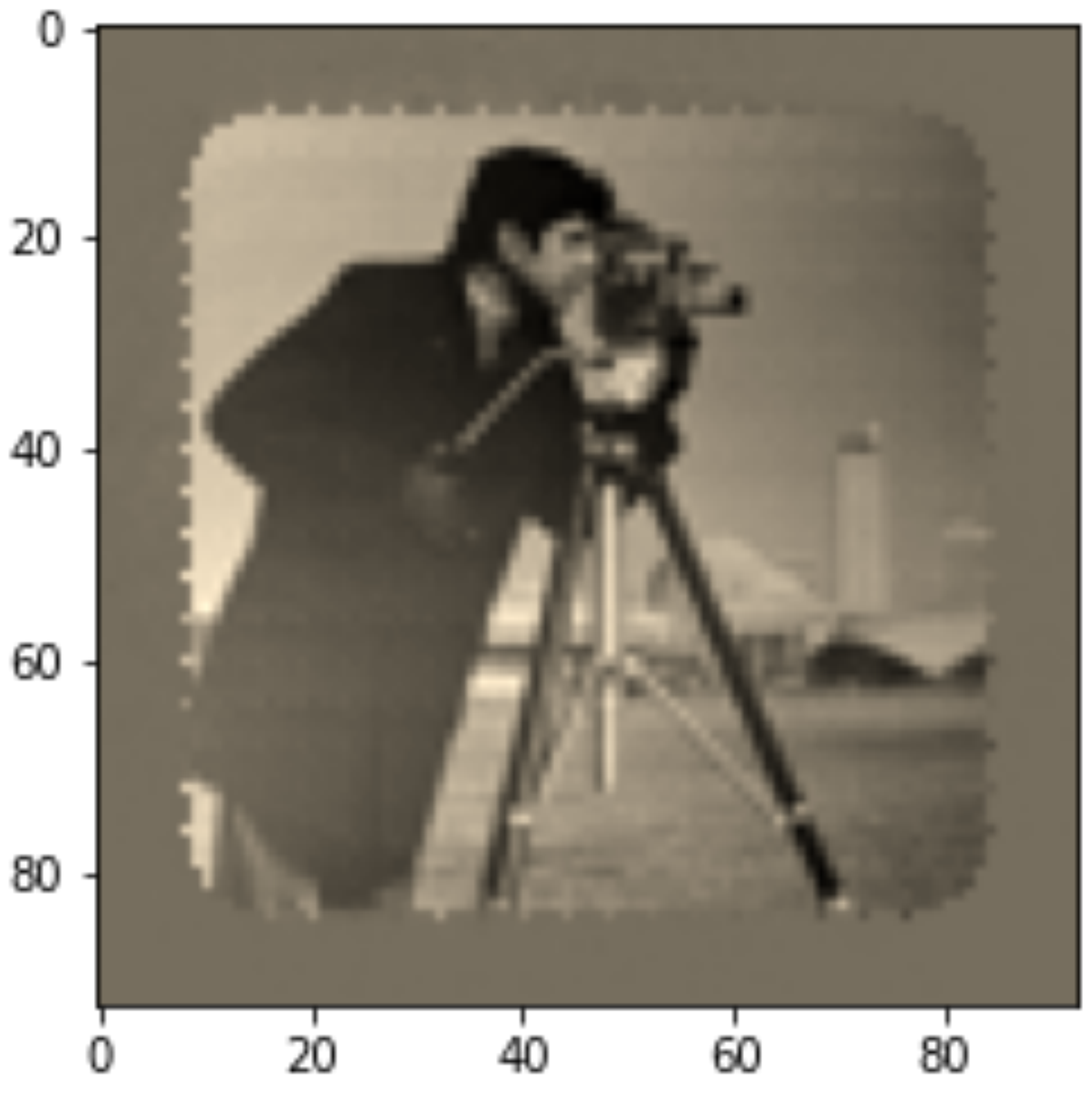
nonlinear CG proceeds along the approximately conjugate directions

$$s^{k+1} = -\nabla f(x^{k+1}) + \beta^k s^k$$
$$\beta^k = \frac{\|\nabla f(x^{k+1})\|^2}{\|\nabla f(x^k)\|^2}$$

Many choices of β^k are possible here (Fletcher-Reeves, Polak-Ribière, etc).

Science and Technology Facilities Council

Optimization: least-squares



Notice that the objective is not analytic (due to the modulus):

$$\min_{o_j\, p} \sum_j \left\| \, |F(o_j\, p)|^2 - i_j \, \right\|^2$$

hence can only optimize it by identifying $\mathbb{C} \cong \mathbb{R} \times \mathbb{R}$.

Wirtinger Derivatives

Wirtinger derivatives neatly extend complex derivatives to $\mathbb{C} \cong \mathbb{R} \times \mathbb{R}$, for $z = x + iy \in \mathbb{C}$ define:

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right)$$
$$\frac{\partial}{\partial z^*} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)$$

where z^* denotes complex conjugation. To convert to (x, y)-space derivatives we can then use:

$$\frac{\partial}{\partial x} + i\frac{\partial}{\partial y} = 2\left(\frac{\partial}{\partial z}\right)^*$$

For analytic functions the complex derivative agrees with the first Wirtinger derivative (the second is zero).

Science and Technology Facilities Council

Wirtinger Calculus

For example, for the modulus function:

$$s(z) = |z|^2 = x^2 + y^2$$

we have that

$$\frac{\partial s}{\partial x} = 2x \qquad \frac{\partial s}{\partial y} = 2y$$

and therefore that

$$\frac{\partial s}{\partial z} = x - iy = z^*$$

$$\frac{\partial s}{\partial z^*} = x + iy = z$$

And for the product function:

$$m(z_1, z_2) = z_1 z_2 = x_1 x_2 + i(x_1 y_2 + x_2 y_1) - y_1 y_2$$

we have that

$$\frac{\partial m}{\partial x_1} = x_2 + iy_2 \qquad \frac{\partial m}{\partial y_1} = ix_2 - y_2$$

and therefore that

$$\frac{\partial s}{\partial z_1} = x_2 + iy_2 = z_2$$

$$\frac{\partial s}{\partial z_1^*} = 0$$

Science and
Technology
Facilities Council

Nonlinear Least-squares

Consider the objective for jth scan position:

$$f_j(o_j, p) = \| r_j \|^2 = \left\| \, |w_j|^2 - i_j \, \right\|^2$$

where $w_j = F(o_j p)$ is the wave model. Then:

$$\nabla_{o_j} f_j = 2J_{o_j}^T r_j = 4F^{-1}\left(w_j r_j\right) p^*$$

$$\nabla_p f_j = 2J_p^T r_j = 4F^{-1}\left(w_j r_j\right) o_j^*$$

are the gradients when viewing $\mathbb{C} \cong \mathbb{R} \times \mathbb{R}$

Science and Technology Facilities Council

Gauss-Newton

Approximates Hessian matrix using only first-order terms:

$$\nabla^2 f_j \approx 2 \begin{pmatrix} J_{o_j}^T J_{o_j} & J_{o_j}^T J_p \\ J_p^T J_{o_j} & J_p^T J_p \end{pmatrix}$$

where J are Jacobians. Too big to form, consider Hessian-vector products:

$$\nabla^2 f_j \begin{pmatrix} v_{o_j} \\ v_p \end{pmatrix} \approx 8 \begin{pmatrix} p^* F^{-1}\left(w_j^2 F^{-1}(p^* v_{o_j})\right) + p^* F^{-1}\left(w_j^2 F^{-1}(o_j^* v_p)\right) \\ o_j^* F^{-1}\left(w_j^2 F^{-1}(p^* v_{o_j})\right) + o_j^* F^{-1}\left(w_j^2 F^{-1}(o_j^* v_p)\right) \end{pmatrix}$$

unfortunately these turn out to be extremely ill-conditioned.

L-BFGS

L-BFGS steps use approximate object and probe Hessians

$$s_o^{k+1} = -\sum_j (B_{o_j}^{k+1})^{-1} \nabla_{o_j} f_j^{k+1}$$
$$s_p^{k+1} = -\sum_j (B_p^{k+1})^{-1} \nabla_p f_j^{k+1}$$

calculated using a two-loop recursion for the BFGS formula:

$$B^{k+1^{-1}} = \left(I - \frac{s^k y^{k^T}}{y^{k^T} s^k}\right) B^{k^{-1}} \left(I - \frac{y^k s^{k^T}}{y^{k^T} s^k}\right) + \frac{s^k s^{k^T}}{y^{k^T} s^k}$$

which is computationally much more efficient.

Science and Technology Facilities Council

Nonlinear CG

Proceed along conjugate directions, starting from steepest descent:

$$s_o^0 = -\sum_j \nabla_{o_j} f_j = -\sum_j 4F^{-1}\left(w_j r_j\right) p^*$$

$$s_p^0 = -\sum_j \nabla_p f_j = -\sum_j 4F^{-1}\left(w_j r_j\right) o_j^*$$

and updating the directions in the standard way (with suitable β):

$$s_o^{k+1} = -\sum_j \nabla_{o_j} f_j^{k+1} + \beta^k s_o^k$$

$$s_p^{k+1} = -\sum_j \nabla_p f_j^{k+1} + \beta^k s_p^k$$

$$\beta^k = \frac{\sum_j \|\nabla_{o_j} f_j^{k+1}\|^2 + \|\nabla_p f_j^{k+1}\|^2}{\sum_j \|\nabla_{o_j} f_j^k\|^2 + \|\nabla_p f_j^k\|^2}$$

Science and Technology Facilities Council

Exact Linesearch

Taking object and probe steps with step-size α :

$$o_j(\alpha) = o_j + \alpha s_o$$
$$p(\alpha) = p + \alpha s_p$$

and substituting into objective for jth scan position gives

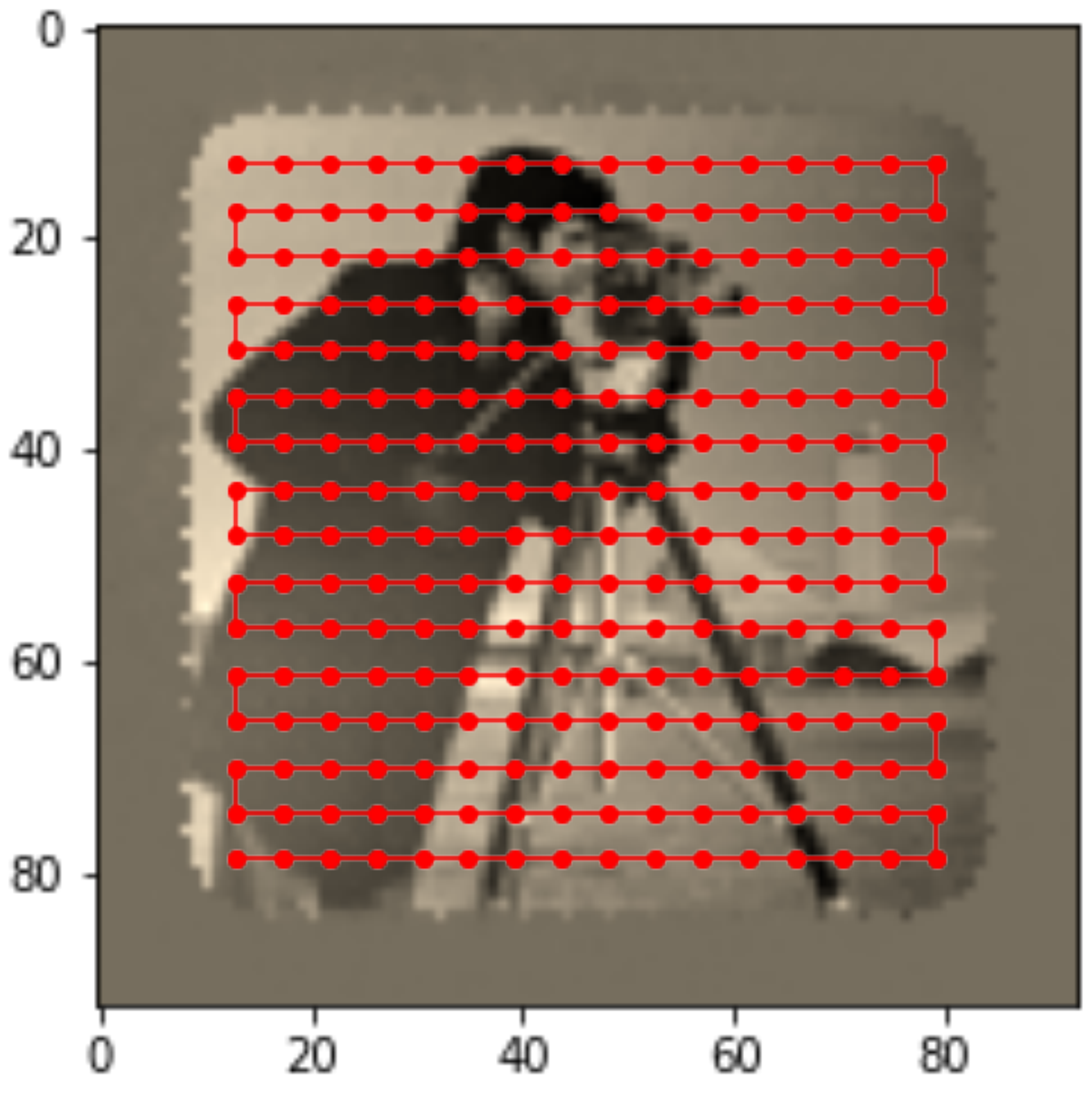
$$f_j(o_j(\alpha), p(\alpha)) = \left\| \, |F(o_j(\alpha)p(\alpha))|^2 - i_j \right\|^2 = \left(\sum_{n=0}^{4} c_n \alpha^n \right)^2$$

Thus to find the optimal step length we can solve

$$0 = \sum_j \nabla_\alpha f_j(o_j(\alpha), p(\alpha)) = \sum_j \left(\sum_{n=1}^{4} n c_n \alpha^{n-1} \right)^2$$

by finding the smallest real root of this real polynomial.

ePIE Algorithm



ePIE algorithm considers more natural amplitude objective:

$$\min_{o_j\, p} \sum_j \left\| \, |F(o_j\, p)| - \sqrt{i_j} \, \right\|^2$$

and performs *for each j at random* a two-stage optimization.

Science and
Technology
Facilities Council

Fourier and Real Spaces

All ptychography algorithms have to satisfy two constraints

Firstly, the modulus constraint in Fourier space:

$$|F(e_j)| = \sqrt{i_j}$$

Secondly, the overlap constraint in real space:

$$e_j = o_j\, p$$

which is enforced for all j scan positions.

Science and Technology Facilities Council

ePIE Algorithm

First, consider amplitude objective for jth scan position:

$$f_j(w_j) = \left\| \, |w_j| - \sqrt{i_j} \, \right\|^2$$

where $w_j = F(o_j p)$ and perform a wave model update:

$$w_j^{k+1} = w_j^k - \alpha_j \nabla_{w_j} f_j$$

Second, consider the difference between exit waves:

$$g_j(o_j, p) = \left\| \, o_j \, p - F^{-1}(w_j^{k+1}) \, \right\|^2$$

and perform a steepest descent update wrt o_j, p .

Science and Technology Facilities Council

ePIE Algorithm

First update is equivalent to the modulus constraint:

$$\hat{e}_j = F^{-1}\left(\sqrt{i_j}\exp(iArg[w_j])\right)$$

where $\hat{e}_j$ is the exit wave $o_j\,p$ with replaced modulus.

Second update takes a step along the gradients of g_j:

$$\nabla_{o_j} g_j = 2\left(o_j\,p - \hat{e}_j\right)p^*$$

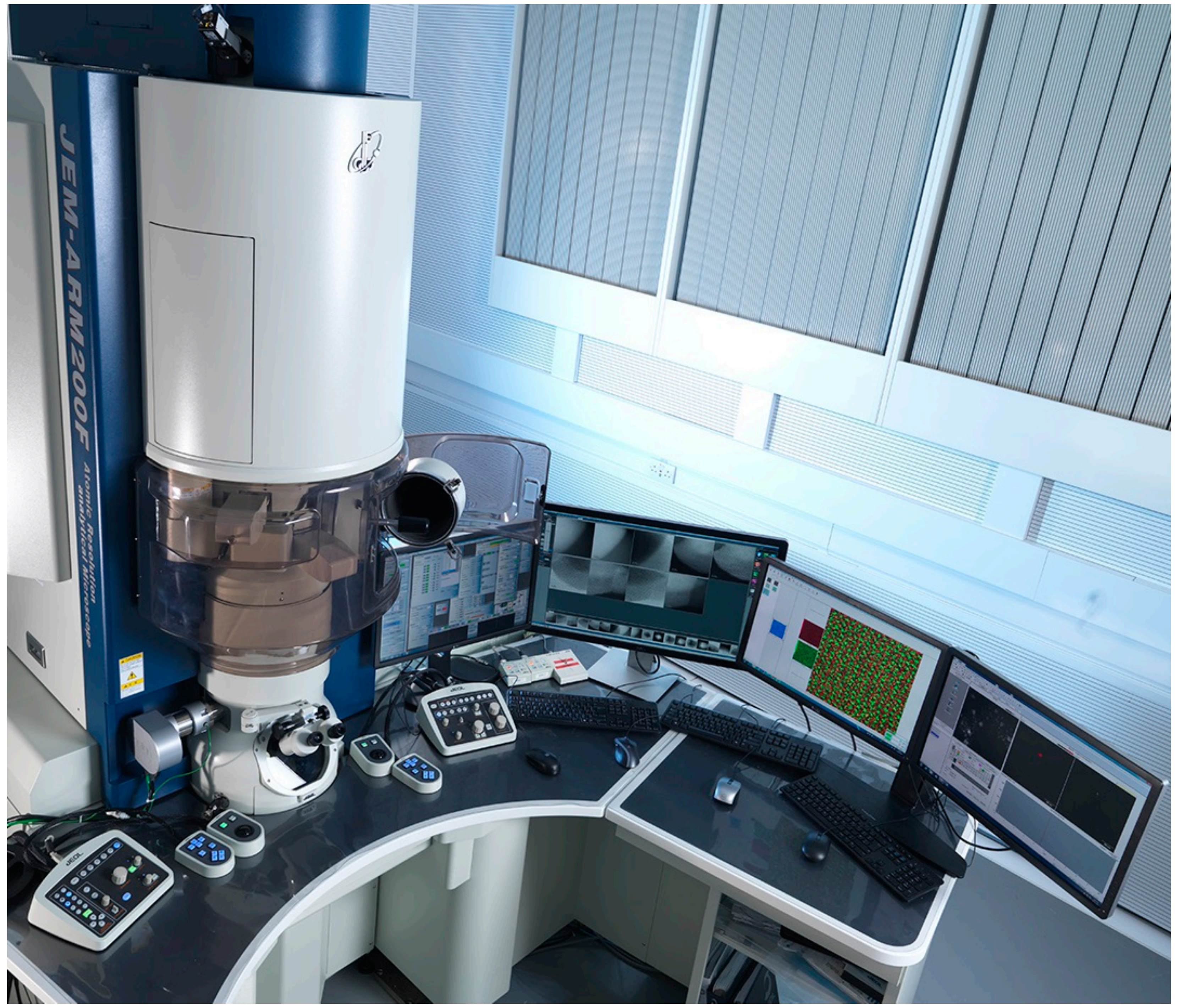
$$\nabla_p g_j = 2\left(o_j\,p - \hat{e}_j\right)o_j^*$$

with step-sizes $1/2|p|^2_{\max}$ and $1/2|o_j|^2_{\max}$ respectively.

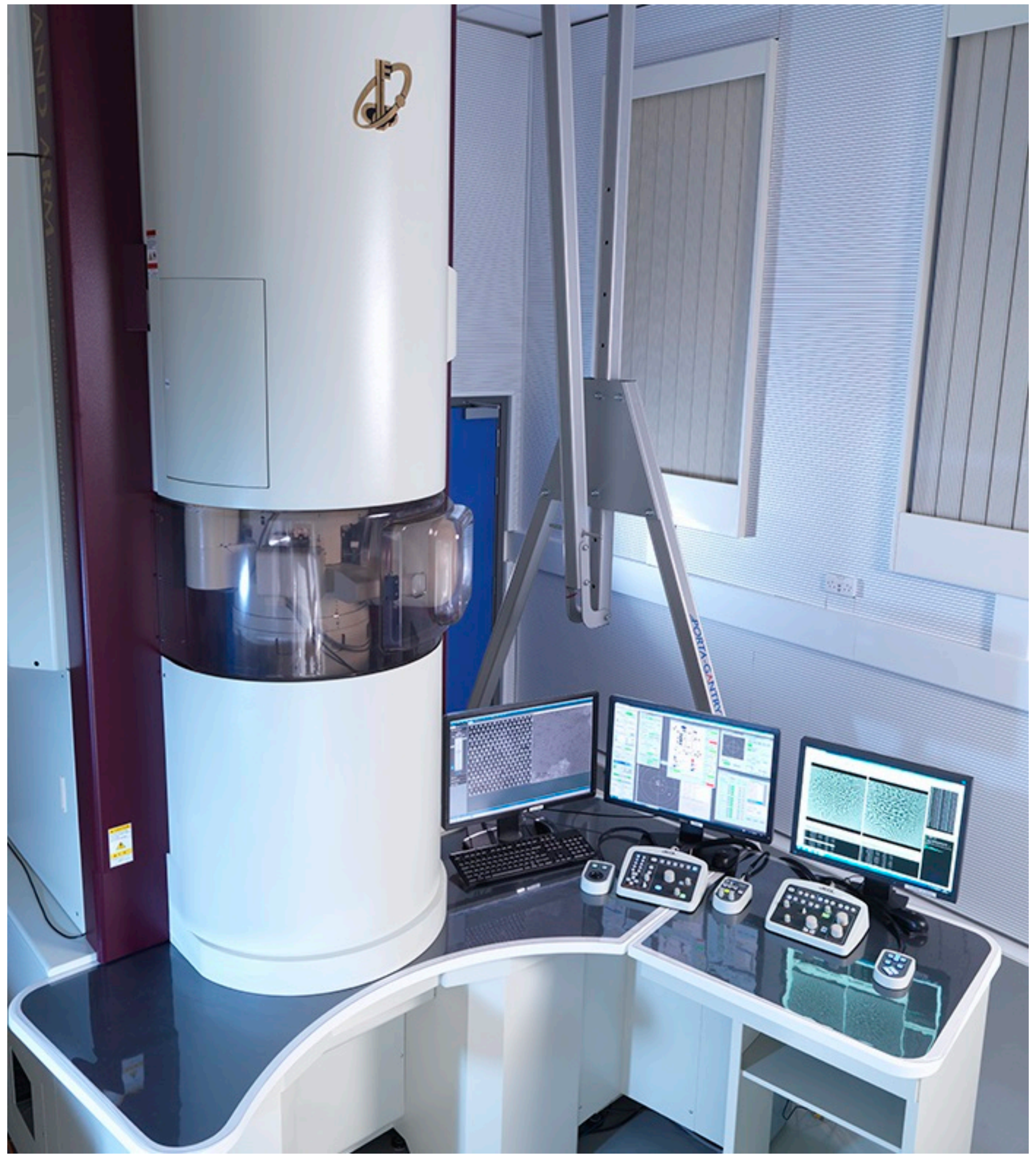
UKRI Science and Technology Facilities Council

Example: ePSIC

electron Physical Science Imaging Centre



JEOL ARM200F

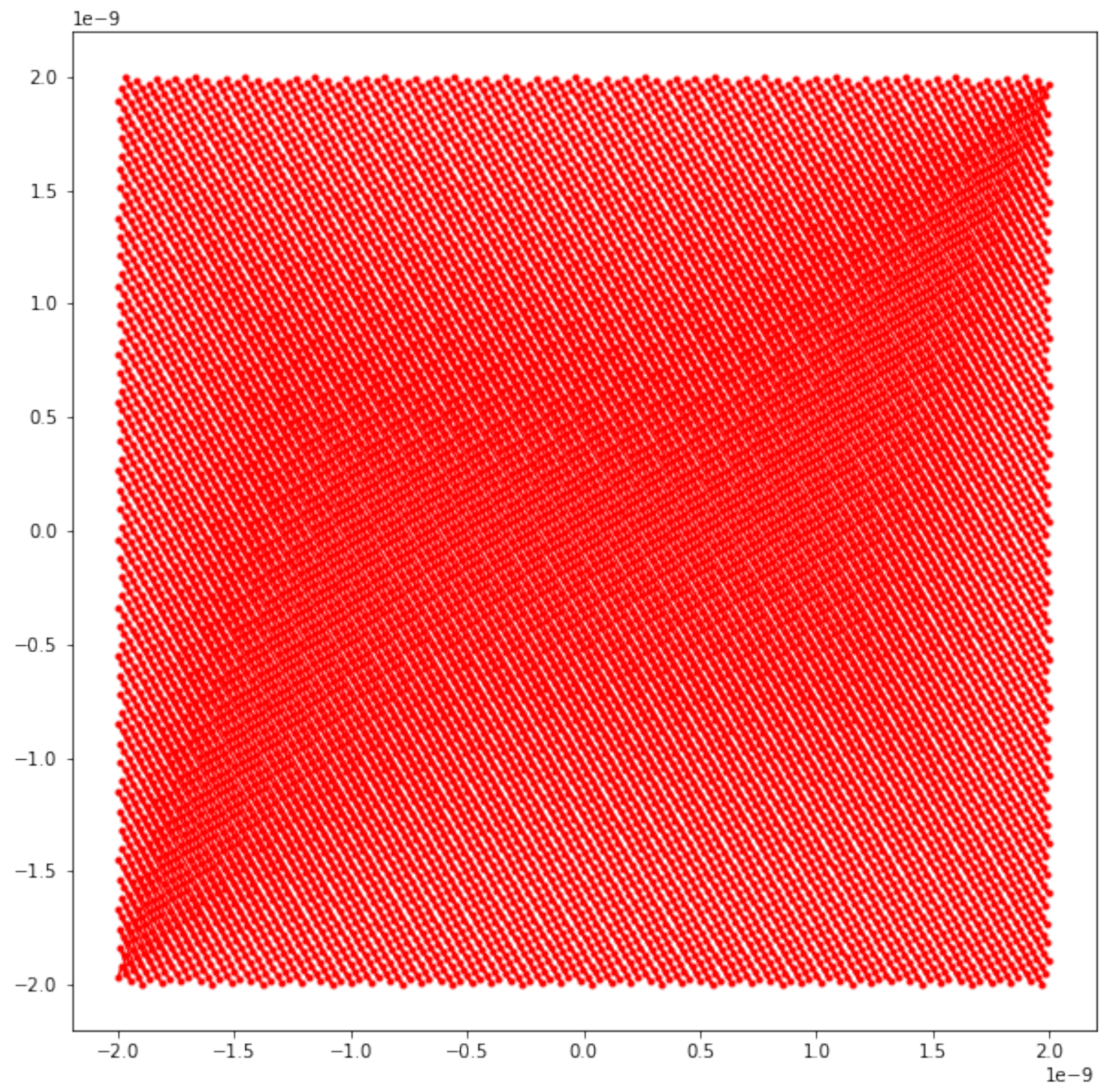


JEOL ARM300F

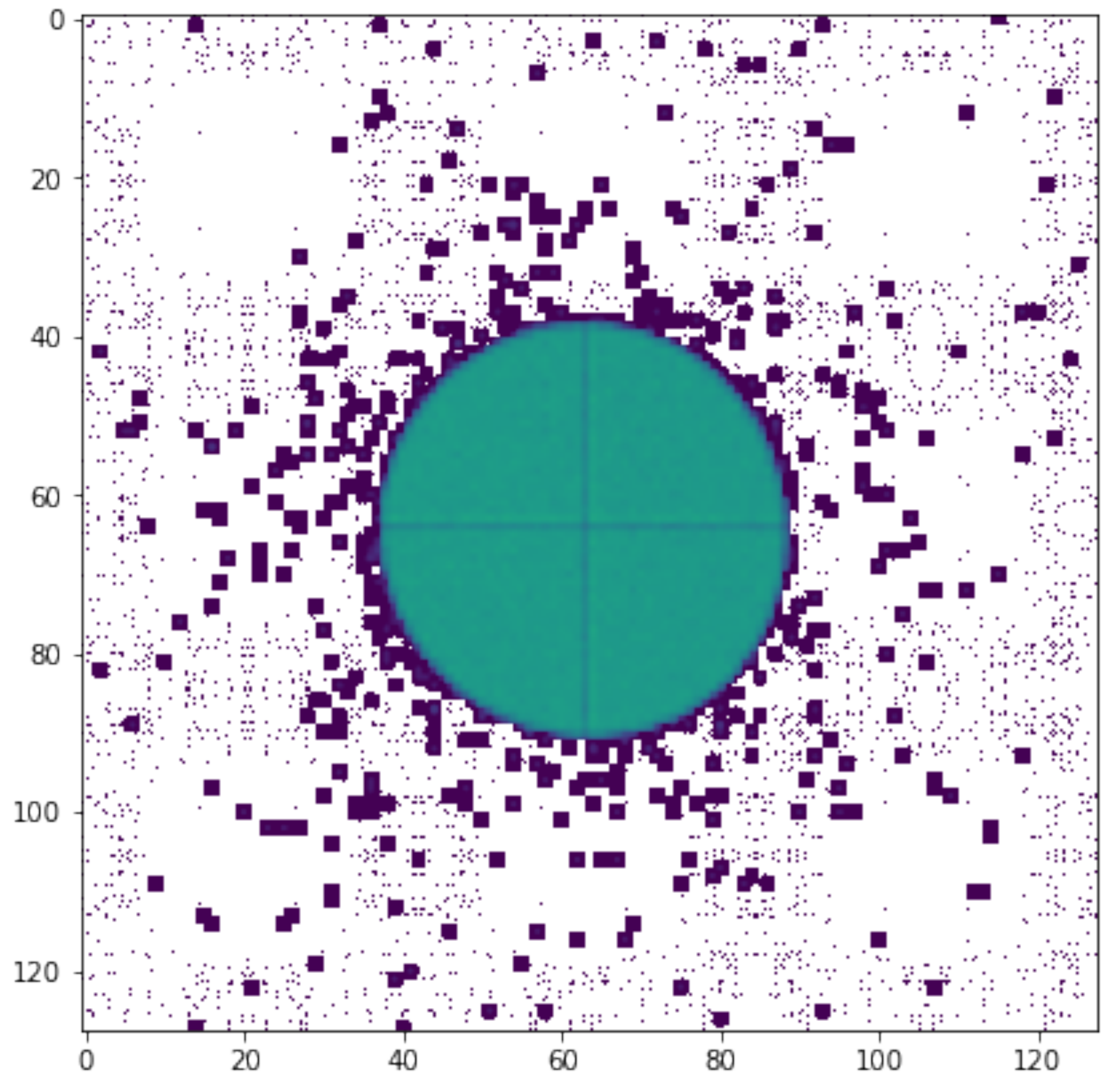
Science and Technology Facilities Council

Example: Graphene Layers

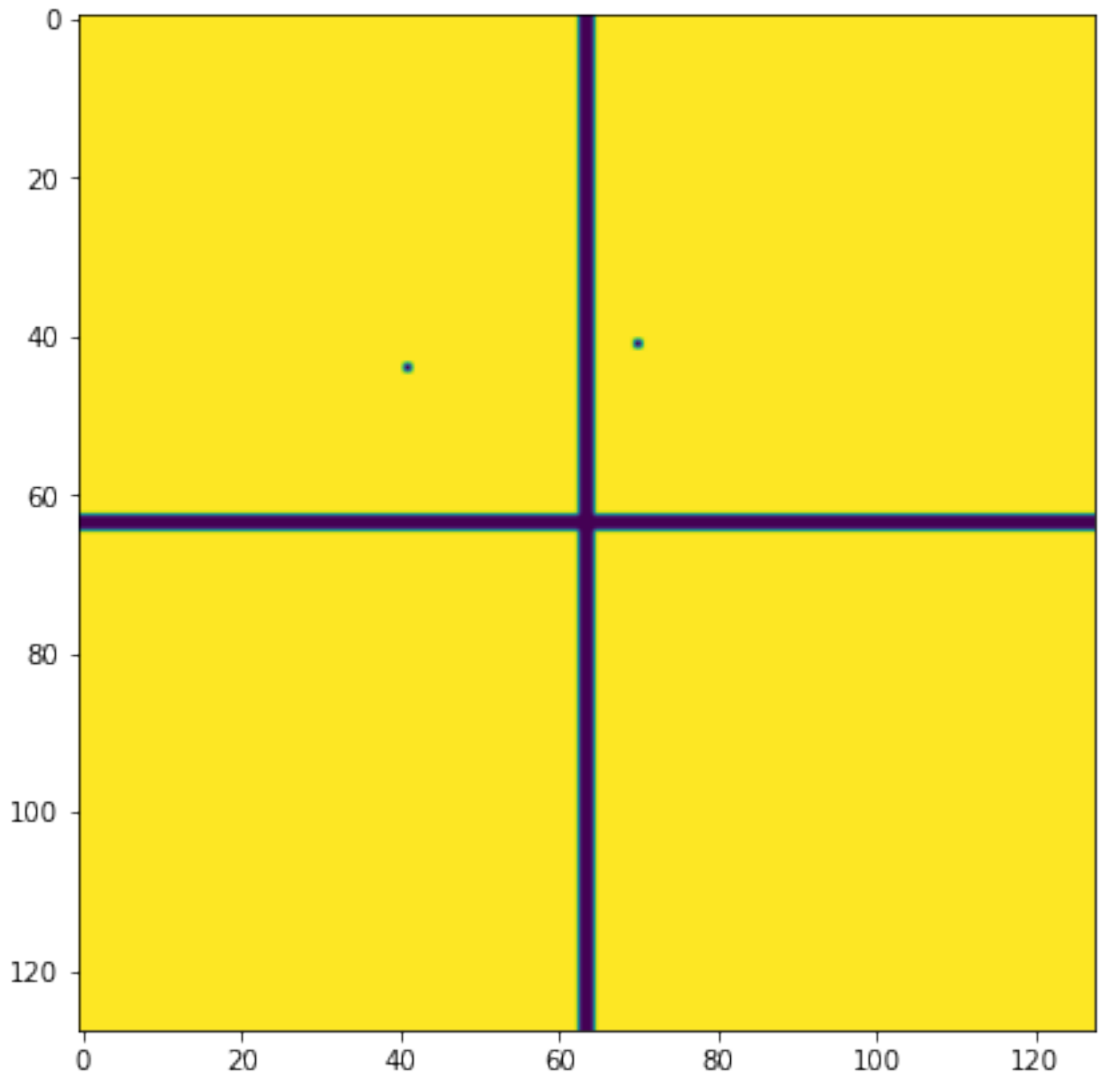
Graphene dataset courtesy of Chris Allen (ePSIC)



Scan Positions



Diffraction Intensity

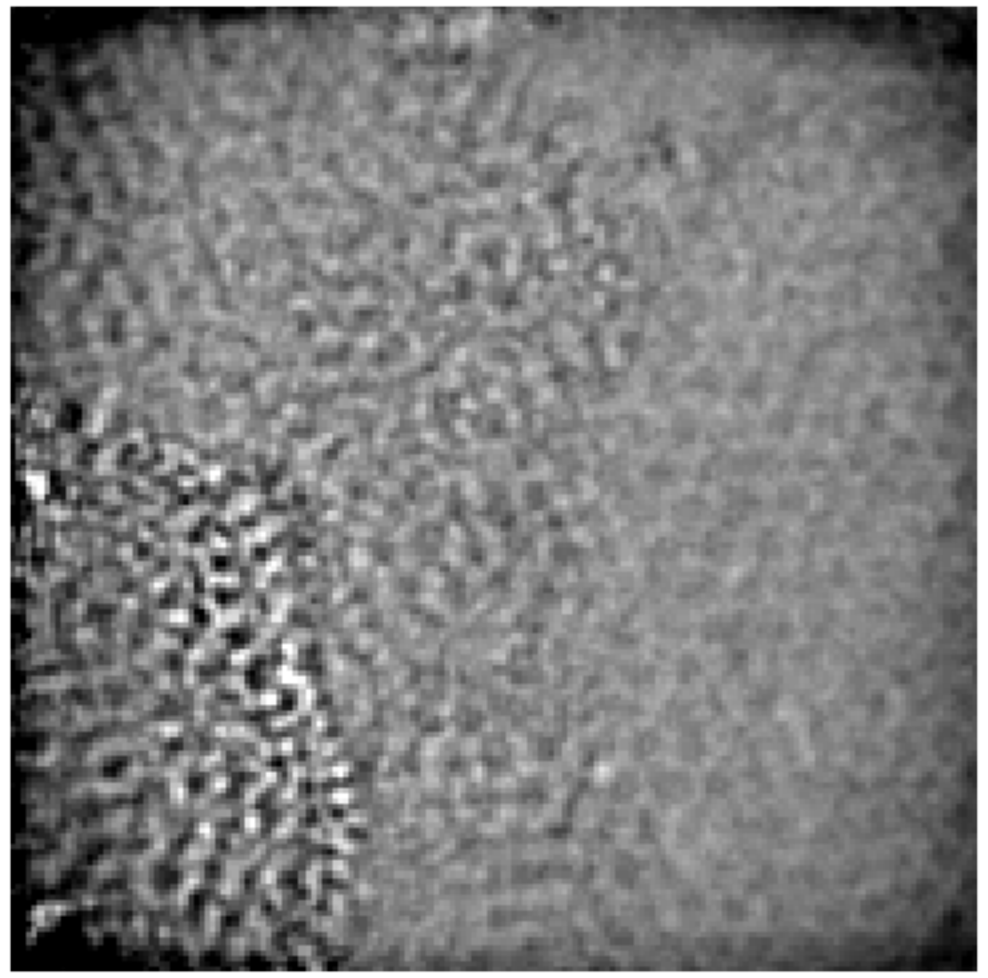


Detector Mask

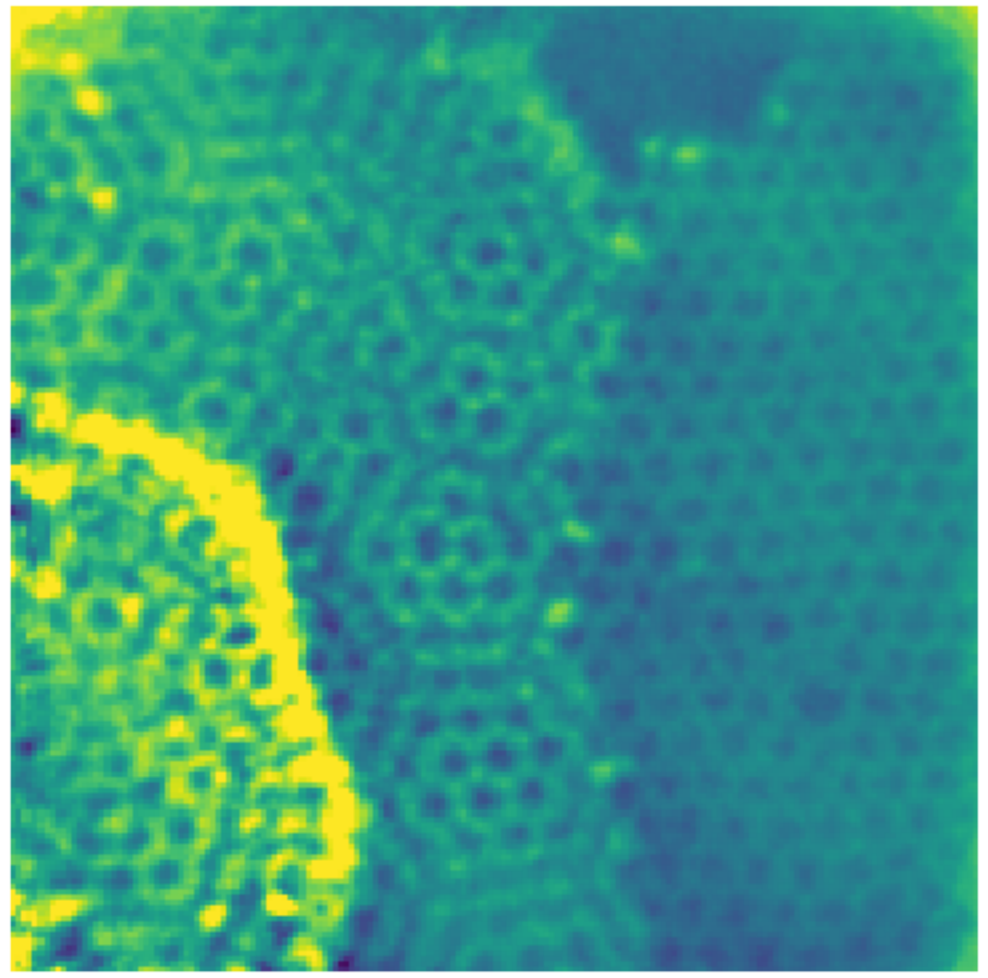
Science and Technology Facilities Council

Example: ePIE

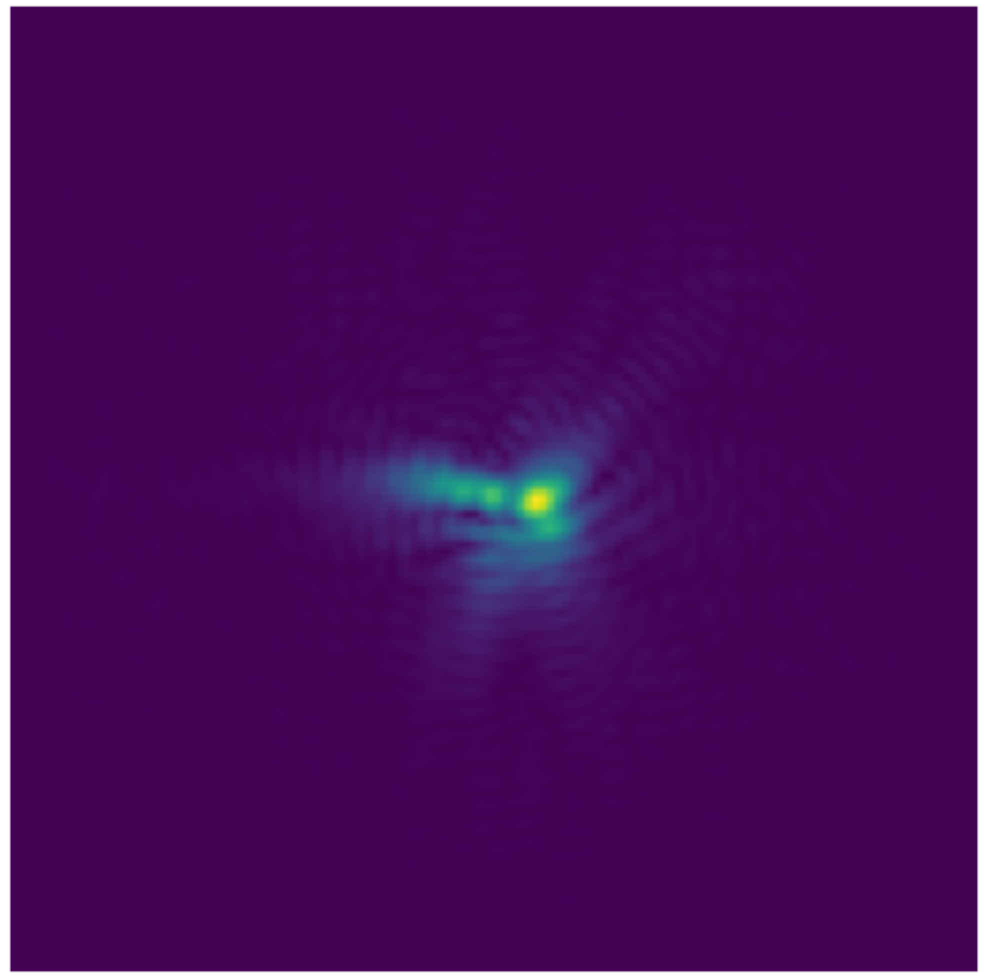
Graphene dataset courtesy of Chris Allen (ePSIC)



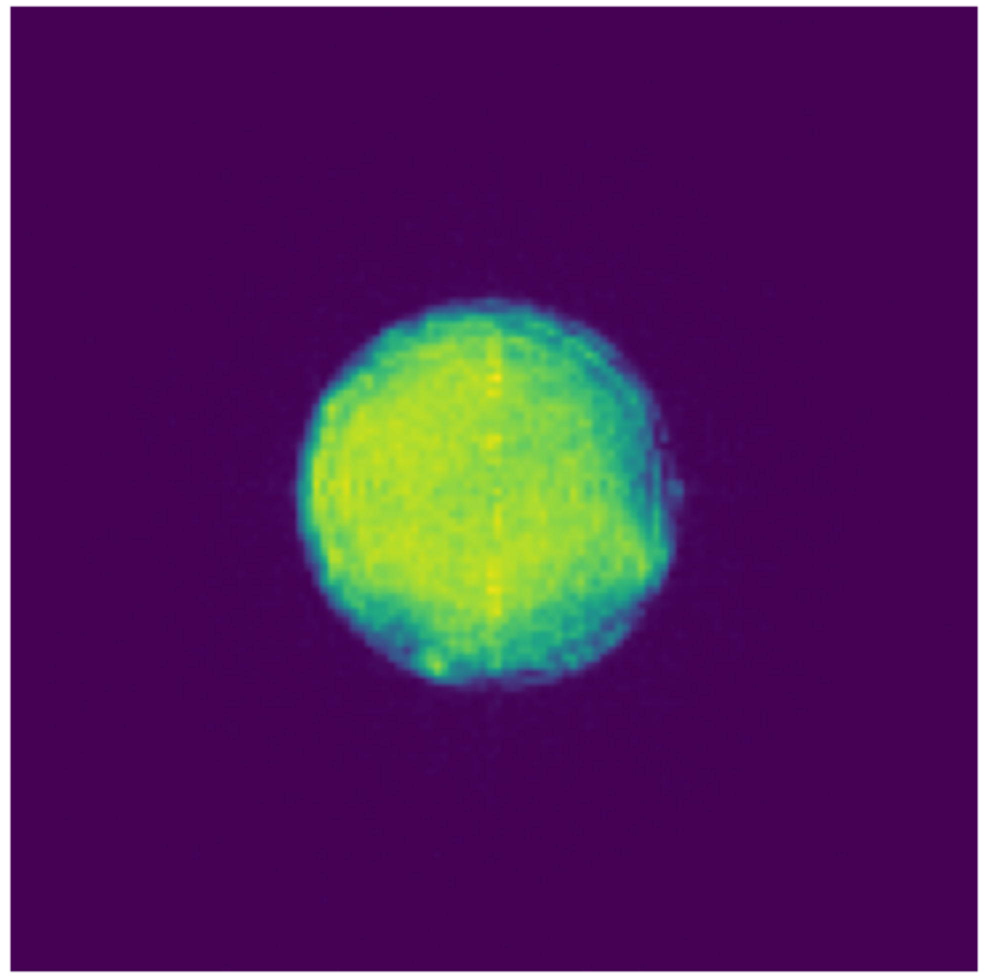
Object Amplitude



Object Phase



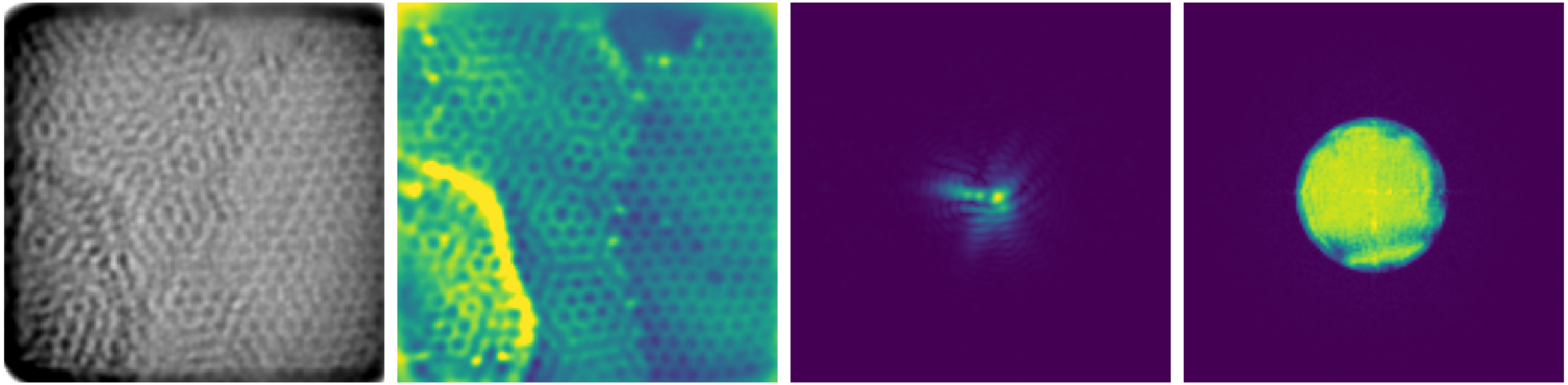
Probe Amplitude



FFT(Probe) Amplitude

Example: Nonlinear CG

Graphene dataset courtesy of Chris Allen (ePSIC)

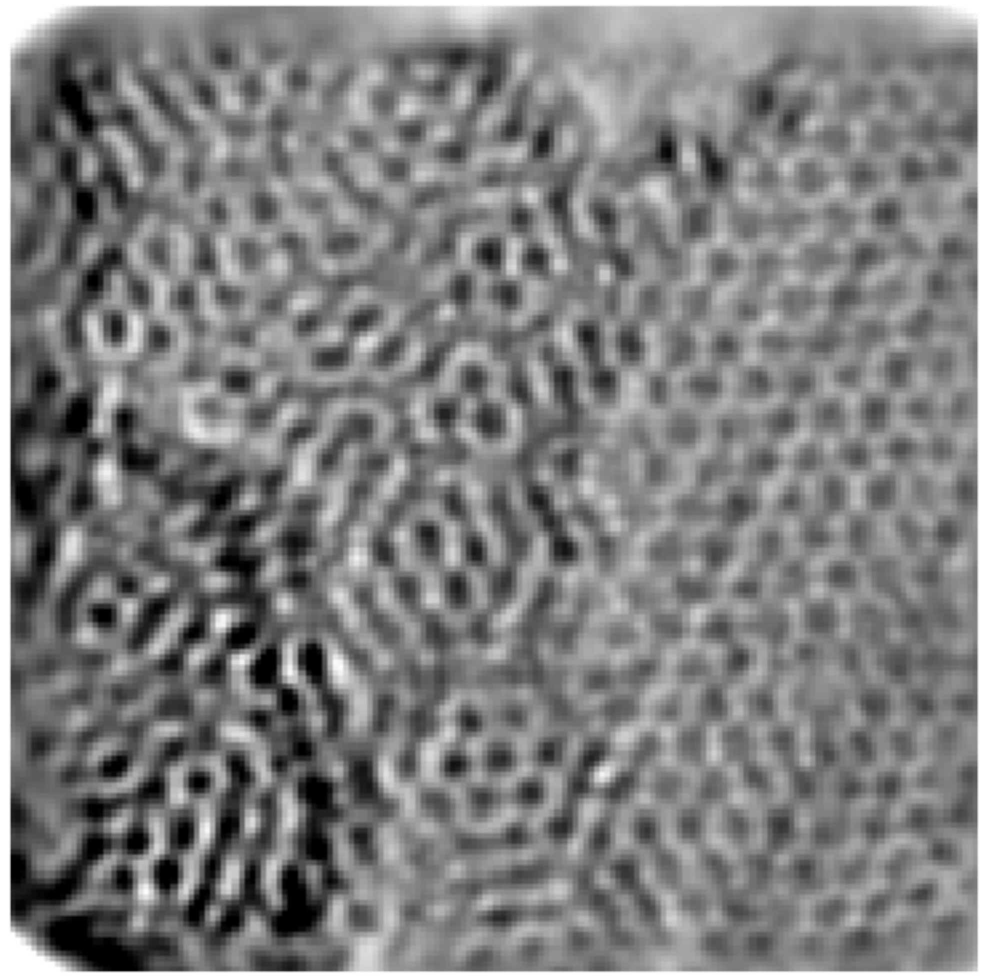


Object Amplitude | Object Phase | Probe Amplitude | FFT(Probe) Amplitude

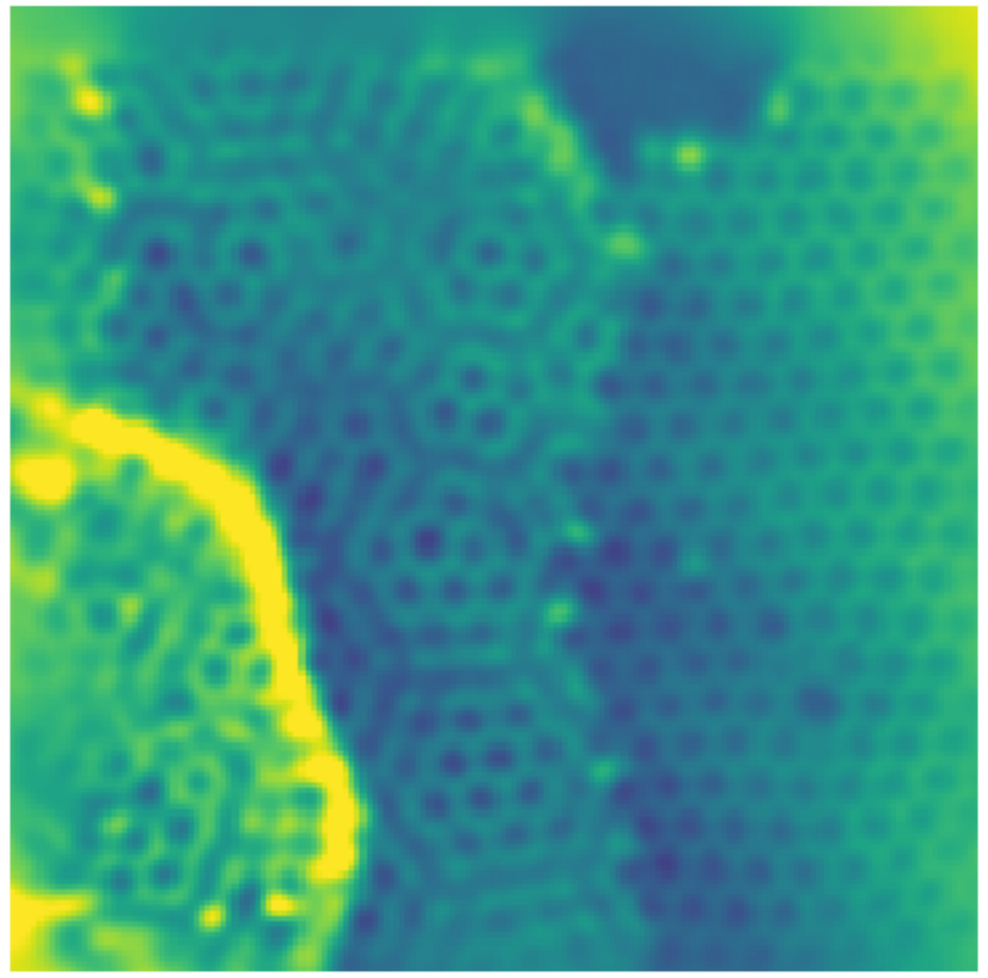
Science and Technology Facilities Council

Example: L-BFGS

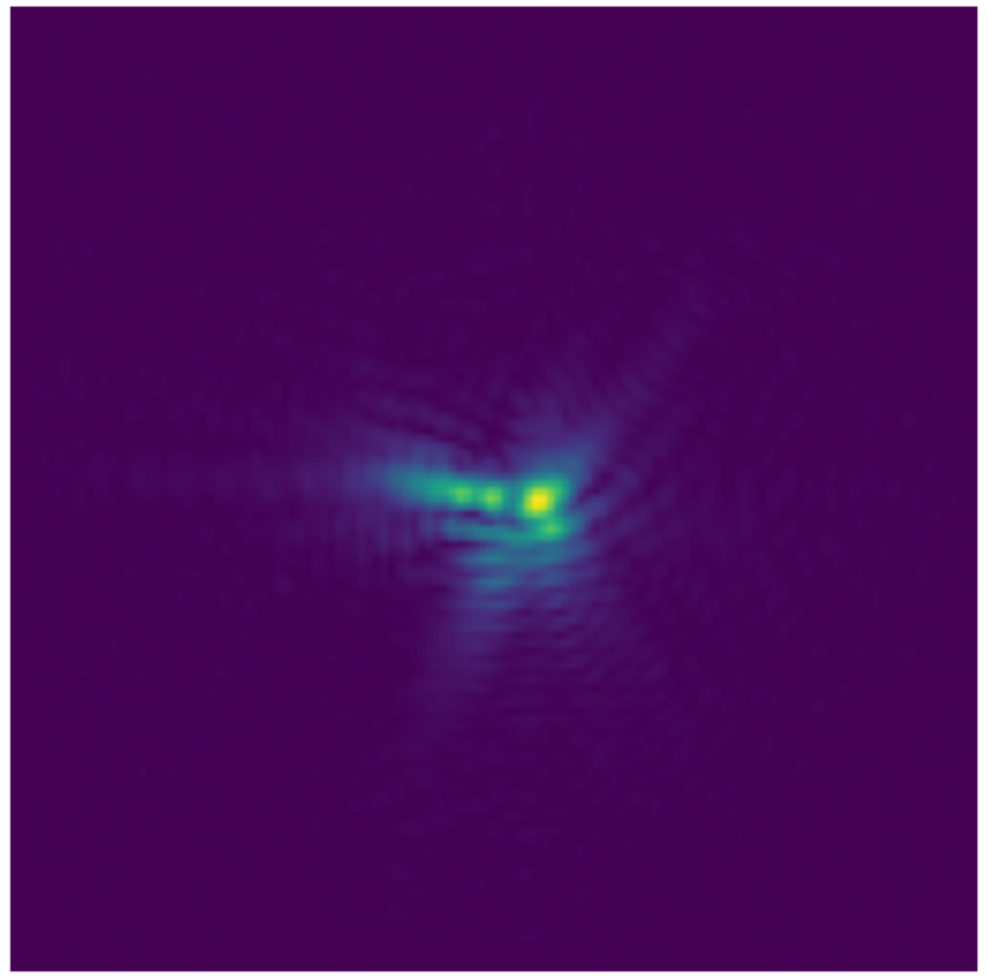
Graphene dataset courtesy of Chris Allen (ePSIC)



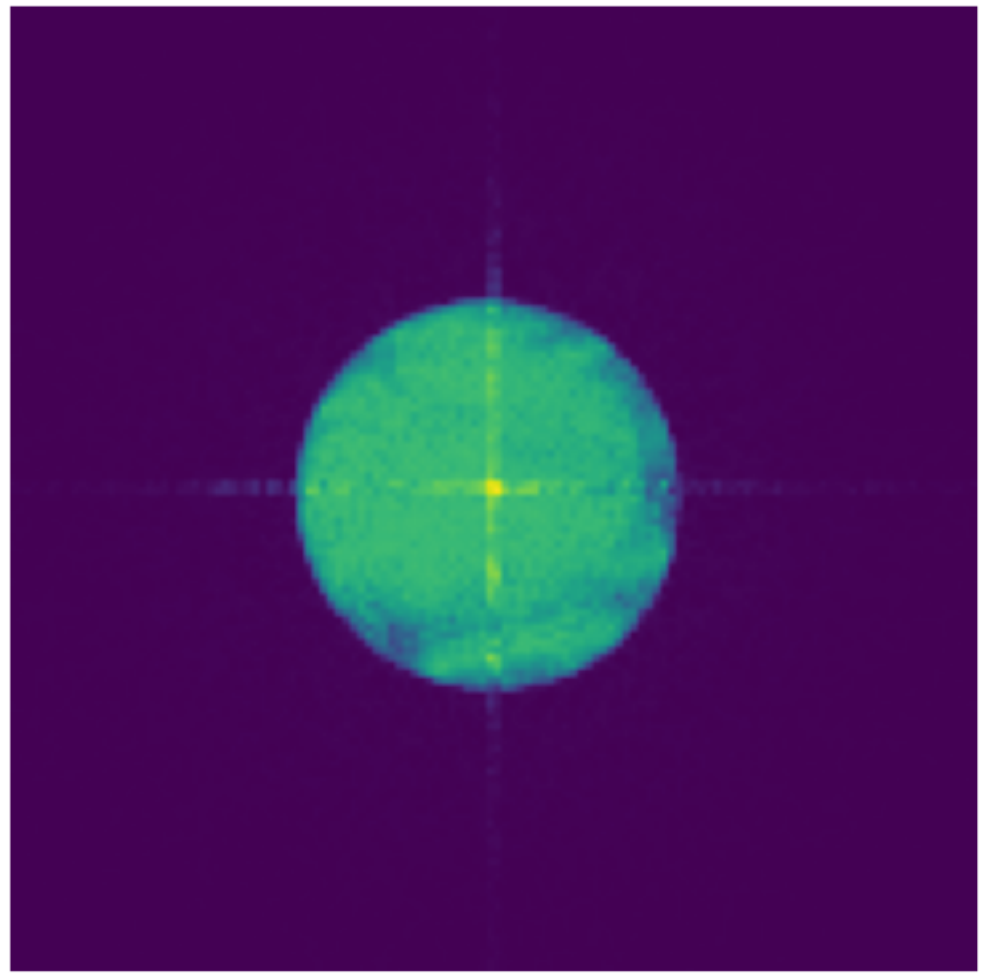
Object Amplitude



Object Phase



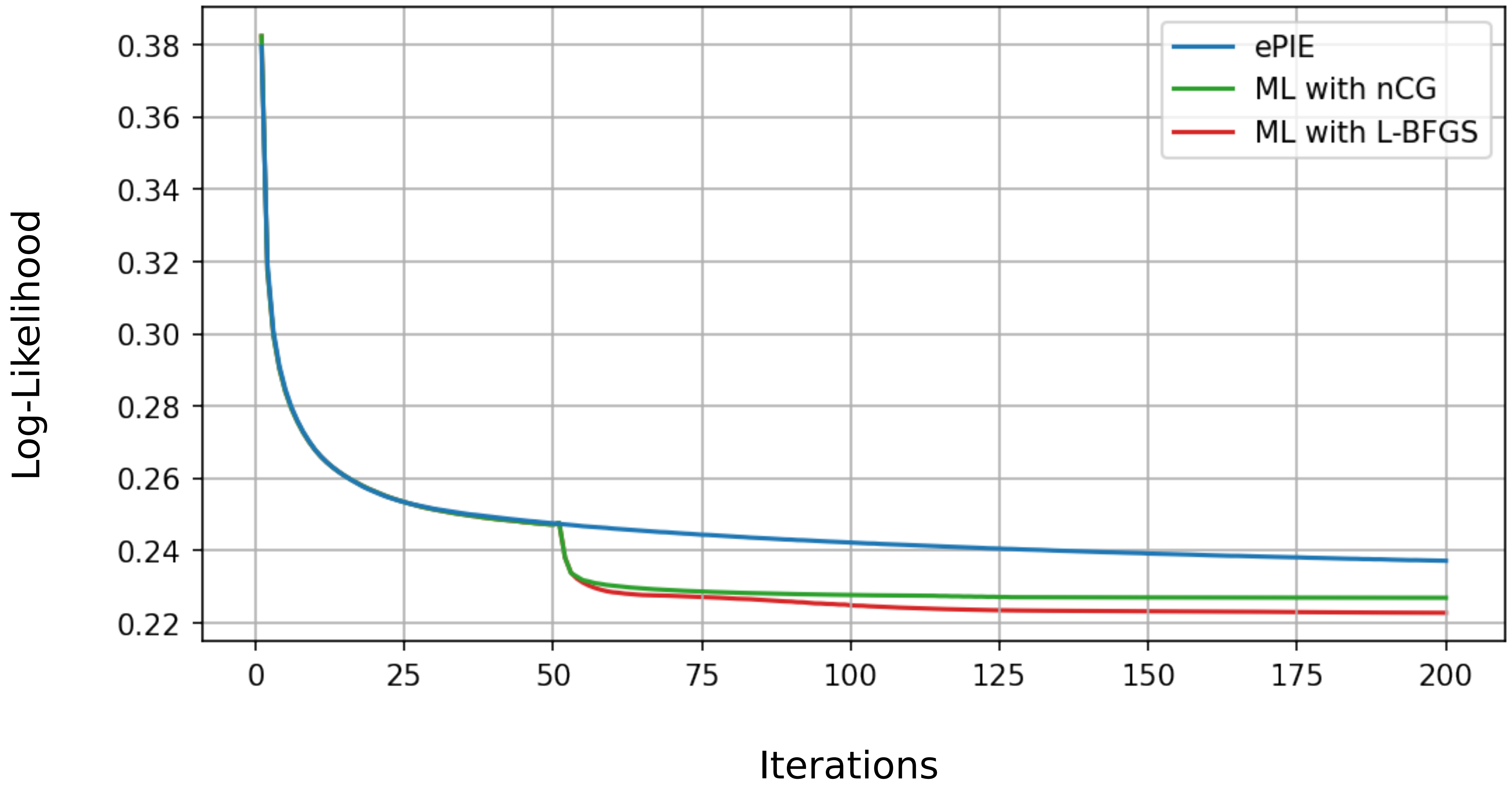
Probe Amplitude



FFT(Probe) Amplitude

Science and
Technology
Facilities Council

Example: Log-Likelihood Error



Science and
Technology
Facilities Council

Thank you! Questions?

References

- Fannjiang, A., & Strohmer, T. (2020). The numerics of phase retrieval. *Acta Numerica*, *29*, 125-228.
- Thibault, P., & Guizar-Sicairos, M. (2012). Maximum-likelihood refinement for coherent diffractive imaging. New Journal of Physics, 14(6), 063004.
- Maiden, A., Johnson, D., & Li, P. (2017). Further improvements to the ptychographical iterative engine. *Optica*, *4*(7), 736-745.
- Odstrčil, M., Menzel, A., & Guizar-Sicairos, M. (2018). Iterative least-squares solver for generalized maximum-likelihood ptychography. *Optics express, 26(3)*, 3108-3123.
- Fowkes, J., Poon, T., Thibault, P., & Daurer, B. Reviewing gradient-based methods for ptychograpic image reconstruction. *In preparation.*