

Supplemental Material for “ Increasing the resolution of electron microscopy by computational ghost imaging”

P.Rosi^{1,2}, L. Viani², E. Rotunno^{1*}, S. Frabboni^{1,2}, A. Tavabi³, R. E .Dunin-Borkowski³, A. Roncaglia⁴, V. Grillo¹

1) Institute of Nanosciences CNR-S3, via G.Campi 213, 41125 Modena, Italy

2) FIM Department, University of Modena and Reggio Emilia, via G. Campi 213/A, 41125 Modena, Italy

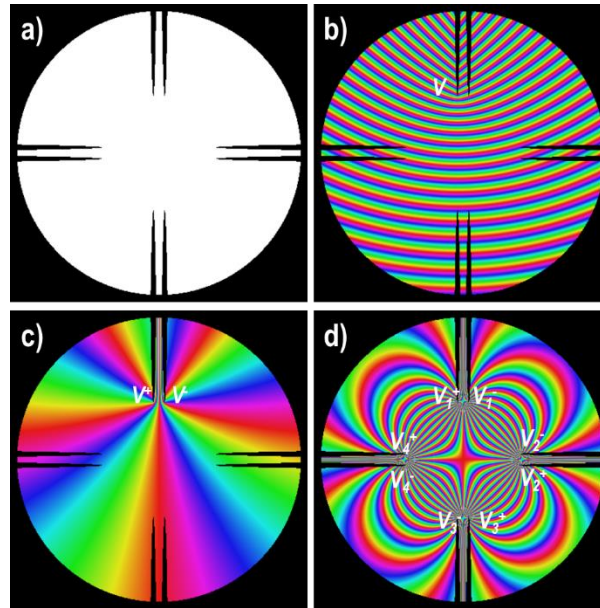
3) Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons, Forschungszentrum Jülich, 52425 Jülich, Germany

4) Institute for Microelectronics and Microsystems- CNR, 40129 Bologna, Italy

S1: Database creation and optimisation

We optimised the design of an electron modulator able to produce structured illumination featuring a high contrast intensity fringe, intended for mimicking the structure of an atomic lattice. This serves the purpose of increasing the similarity with crystalline samples in order to achieve the highest possible lateral resolution.

The design of our electron modulator is shown in figure Supp. 1a. It consists of 8 needle-like electrodes arranged in 4 pairs along two perpendicular axes. Each pair of electrodes has a separation of 1 μm between the needles. Opposing pairs have a distance, tip to tip, of 20 μm .



Supp. 1: a) image of the device, b) image of the phase shift induced by a single needle modeled by equation S1, c) phase shift image obtained when two adjacent needles are oppositely biased which generates an azimuthally varying phase profile typical of a spiral phase plate (SPP), d) phase shift image when the device is biased so that opposing SPPs generate the same topological charge while adjacent SPPs generate opposite topological charge

The phase landscape generated by each needle is modelled as [1]:

$$\varphi(x, y) = ak \left(y \tan^{-1} \frac{y}{x} - x \log \left(\frac{\sqrt{x^2 + y^2}}{L} \right) + x \right) \quad (S1)$$

where k is the electron's wavevector, L is the needle's length, and a is a scaling constant depending on the device thickness (Supp. 1b). The overall phase φ_T generated by the device is obtained by a linear combination of the phase shifts induced by the single needles φ_i weighted by their respective bias V_i :

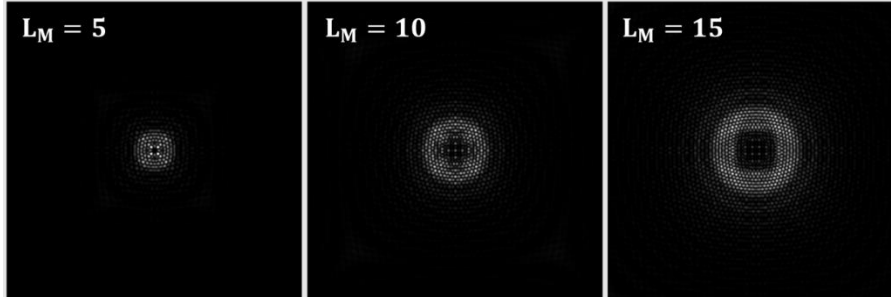
$$\varphi_T(x, y) = \sum_i V_i \varphi_i(x, y) \quad (\text{S2})$$

Even if each needle is modelled independently, it is more useful to consider them as pairs. In fact, if correctly biased, each pair behave as a spiral phase plate (SPP) generating a vortex with topological charge ℓ proportional to their potential difference (Supp. 1c):

$$\ell \propto V^+ - V^- \quad (\text{S3})$$

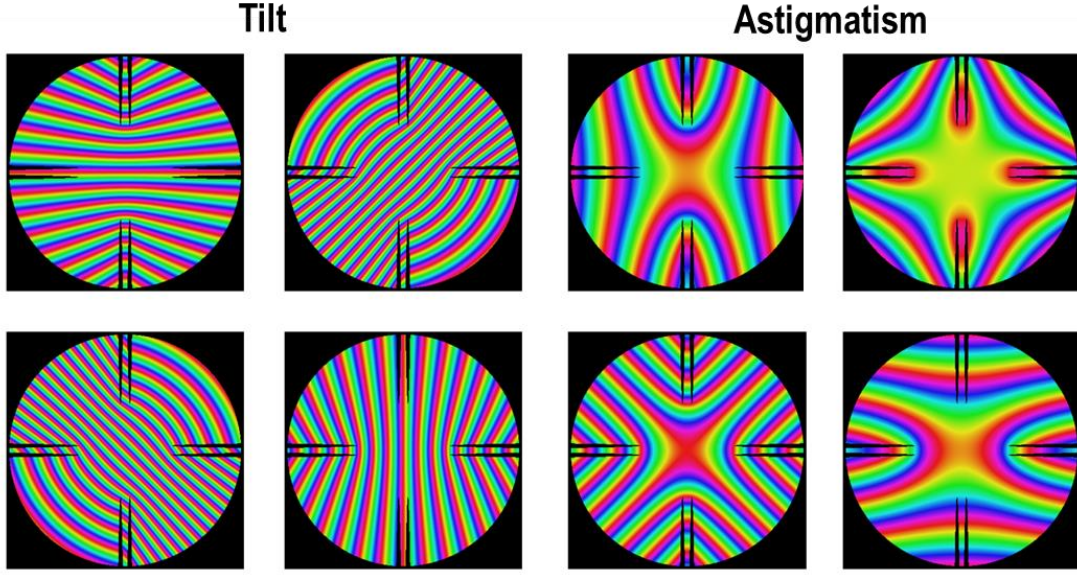
When the device is biased in such a way the opposing SPPs generate the same topological charge while adjacent SPPs generates opposite topological charge, $\ell_1 = \ell_3 = -\ell_2 = -\ell_4 = L_M$, it is possible to maximize the interference between the vortex beam, while imposing a centre-symmetry to the phase plate (Supp. 1d). Due to this symmetry, the electron wavefunction on the sample plane is real and that all the information encoded in the modulator's phase is translated into an intensity oscillation.

Increasing the value of L_M simultaneously increases both the extension of the illuminated area and the spatial frequency of the fringe patterns, for a given convergence angle, as shown in Supp. 2. If increasing the illuminated area is desirable (we can sample a larger area of the sample without moving the probe) this makes the pattern more susceptible to decoherence effects. A good compromise between pattern resolution and illuminated area is found for $L_M = 10$.



Supp. 2: example of illuminating probes at different L_M values, where $L_M = \ell_1 = \ell_3 = -\ell_2 = -\ell_4$ and ℓ_i is the topological charge of the i -th SPP.

Since the spiralling phase generated by the pair only depends on the bias difference between the two needles, it is possible to add a common offset without altering the topological charge. This offset can be used to induce a tilt in the beam along the main axis of the electrodes. Since we have four pairs of electrodes, we can use the four offsets to add four new degrees of freedom to the device: tilt x/y and astigmatism x/y (Supp. 3).



Supp. 3: Phase images of the electron wave-function modified by our electron modulator, it is possible to see that by properly adjusting the offset of each pair of electrodes we can induce arbitrary tilt or astigmatism.

To predict the structured illumination formed on the sample when the device is loaded in the conjugated condenser aperture plane, we developed a custom python3 algorithm.

The simulation space spans a field of view (FoV) of $25\text{nm} \times 25\text{nm}$ sampled on a $1024\text{px} \times 1024\text{px}$ grid, thus the pixel size is 24.41 pm . The probe forming aperture is 15.4 mrad at an electron energy of 300keV , thus the expected aperture-limited resolution is 77.98pm . On this grid the phase shift induced by the device is computed using equations S1 and S2 and then a Fast Fourier Transform (FFT) is performed to obtain the illumination pattern expected on the sample plain.

Since the device can work even when strong aberrations are present, we incorporated their effect in our model. The phase of the aberration being:

$$\varphi_{ab}(x, y) = \frac{\pi C_s}{2\lambda} \left(\frac{r}{f}\right)^4 - \frac{\pi df}{\lambda} \left(\frac{r}{f}\right)^2 \quad (\text{S4})$$

where λ is the electron wavelength, r is the distance from the optical axis, f is the focal length of the probe forming lens and df is the defocus. C_s is the spherical aberration coefficient. In this work we assumed a C_s value of 2.7mm . This is a quite severe aberration that can be encountered in conventional uncorrected microscopes. All other aberrations are neglected.

Defocus df can be used as a further degree of freedom whose main effect is to enlarge the illuminated region, here we set it to a constant value equal to four times the Scherzer defocus:

$$df_{\text{Scherzer}} = 1.2\sqrt{C_s\lambda} \quad (\text{S5})$$

It is worth noticing that, in an un-corrected STEM one would use a convergence angle equal to $\left(\frac{4\lambda}{C_s}\right)^{\frac{1}{4}} = \alpha = 7.3\text{ mrad}$ to limit as much as possible the effects of the spherical aberration. This limit is less than half the convergence angle used in our simulations.

In the main text we discussed the necessity to generate a database of structured illumination patterns characterised by similarity to the target image and an identical independent distribution

(i.i.d.) of the pixels intensities among the different patterns. The similarity criterion translates in the case of crystalline samples to having pattern characterised by fringes with a periodicity on the same length-scale of the typical interatomic distance. The size of the fringes and their spacing depends on the convergence of the beam. This is a staple of electron microscopy and in particular of STEM techniques where the size of the probe is reduced to sub angstrom in order to fit in between atoms. This is achieved using large convergence angles and $L_M > 8$. In principle, the more similar the structured illumination is the more information is in the measured signal and faster it is to reach convergence in the reconstruction. A further step to maximize the similarity between the sample and the structured illumination is to define a metric for the similarity which can either be the single pixel signal (choosing the patterns that give the highest and lowest signals should improve the speed), or well-known statistical distribution divergences that measure similarity such as cross-entropy or Kullback–Leibler divergence. Given the few degrees of freedom that our electron modulator has (we are able only to control the applied bias on 8 electrodes) when we enforced these other similarity metrics we only had a very marginal improvement with respect to the big improvements of provided by having patterns with fringes with a periodicity on the same length-scale of the typical interatomic distance. Naturally, when more degrees of freedom in the generation of structured illumination can be used these criterions provide a great chance to further refine the database and speed up the reconstruction.

On the other hand, enforcing the i.i.d. means that no to minimal correlations between pixels should exists along the dataset. This condition guarantees that the total intensity integrated in the whole dataset is completely flat and that redundant information is limited.

A simple analytical way to measure the extension of the pattern is computing the L1 norm of the pattern after it is normalised to its maximum:

$$\mathbb{S}_n = \left\| \frac{p_n}{\max(p_n)} \right\|_1 \quad (S6)$$

\mathbb{S}_n can be understood as the average number of illuminated pixels in the pattern: a few, bright pixels will lead to a smaller value of \mathbb{S}_n , signalling the presence of a caustic.

In order to impose the identical distribution constraint, we want to minimise as much as possible overlap \mathbb{O}_n between the n -th pattern and the sum of the previous ones.

$$\mathbb{O}_n = \sum_{i,j} p_n \sum_{m=0}^{n-1} p_m \quad (S7)$$

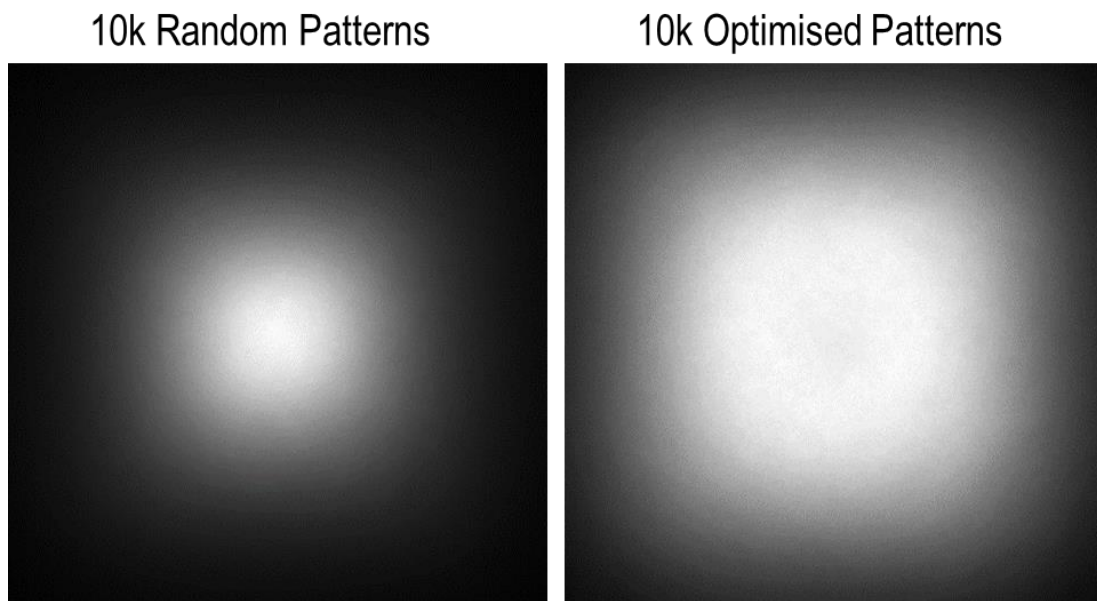
This metric allows us to filter away patterns which do not contribute to the smoothness of the overall intensity distribution.

As a final constraint we imposed that the pattern should be completely contained within the desired FoV. If a conspicuous amount of the beam intensity falls outside the image boundary it will hamper the reconstruction process. We define the lost intensity ratio \mathbb{L}_n as the ratio between the intensity within the FoV area and the total intensity of the beam:

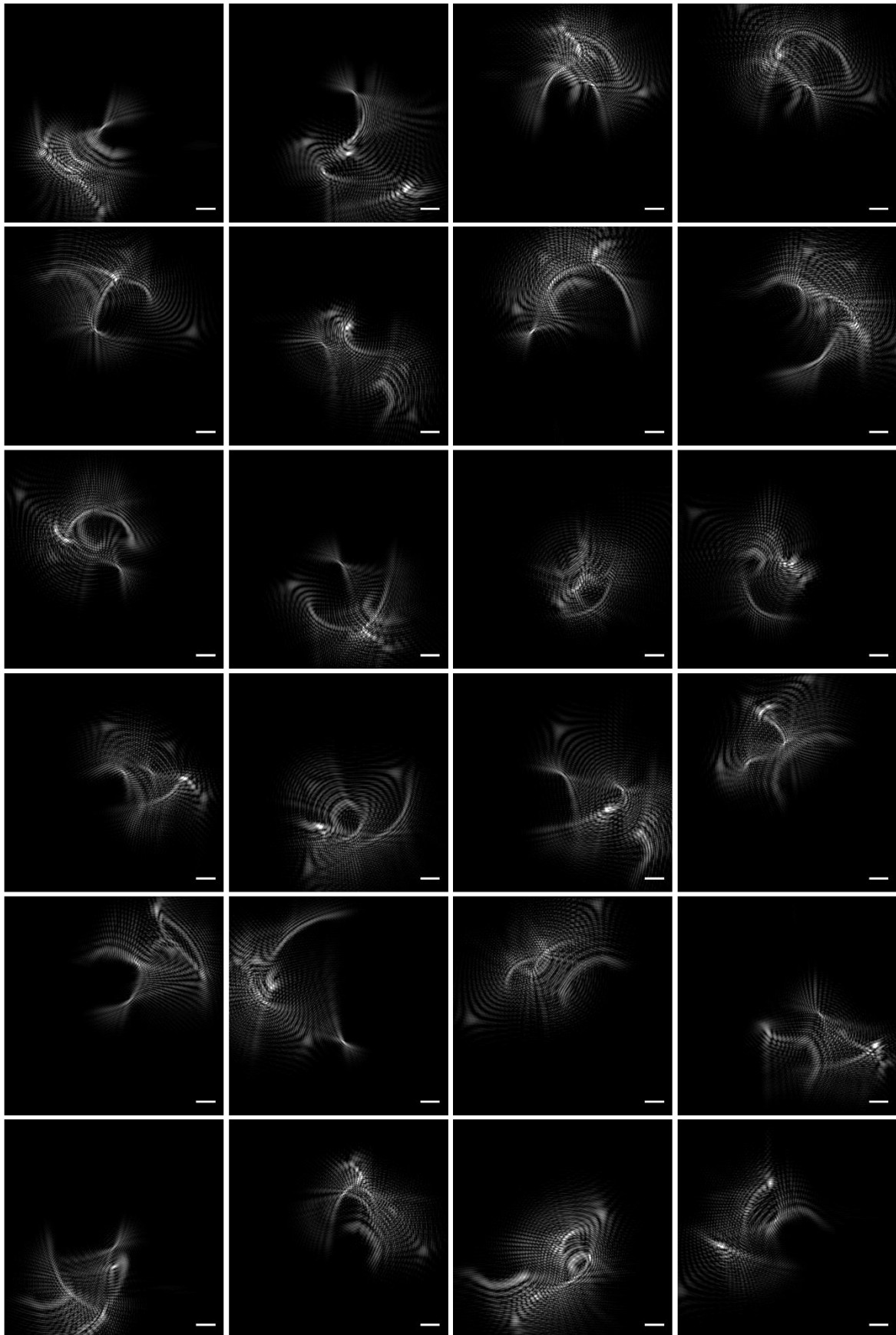
$$\mathbb{L}_n = 1 - \frac{\sum_{i,j}^{FoV} p_n(i,j)}{\sum_{i,j} p_n(i,j)} \quad (S8)$$

Owing to the two conditions for an optimal database (i.i.d. and similarity to the target image), we build a database of 200.000 illumination patterns using the following values of the previously defined constraints: $\mathbb{L}_n < 0.1$, $\mathbb{S}_n > 2000$, $\mathbb{O}_n < 0.75$. In Supp.4 it is possible to see the difference between the illuminated region by a partial database of 10k random patterns versus the illuminated

region by a partial database of 10k optimised patterns. In here, even though the illumination is not completely flat and uniformly distributed, for the optimised case the illuminated region is almost three times bigger with respect to the unoptimized database.



Supp. 4: Comparison between 10000 patterns generated with random conditions and 10000 patterns generated by imposing the constraints: $\mathbb{L}_n < 0.1$, $S_n > 2000$, $\mathbb{O}_n < 0.75$.



Supp. 5: Examples of structured illumination patterns obtained by imposing the constraints $\mathbb{L}_n < 0.1$, $\mathbb{S}_n > 2000$, $\mathbb{O}_n < 0.75$. The white scale bar in the bottom right corner is 1nm long.

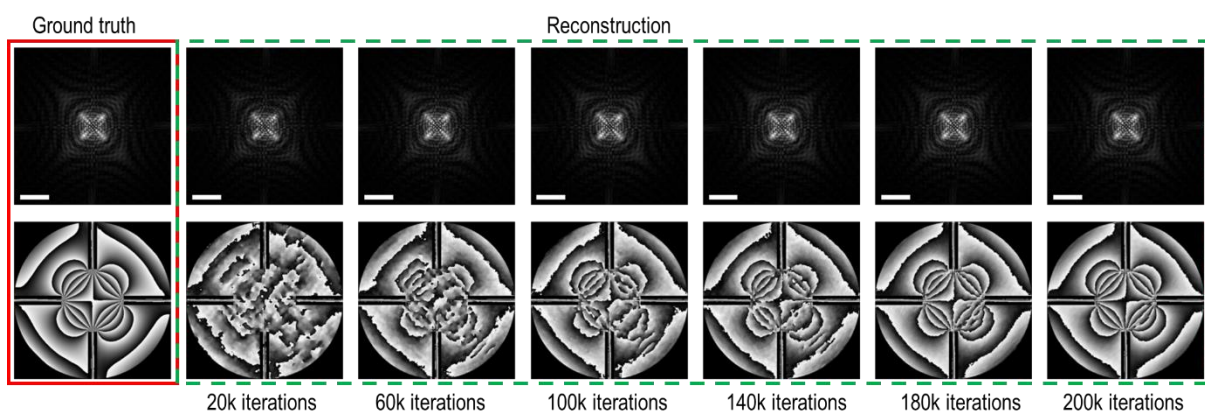
S2: Probe reconstruction

For a future experimental application of CGI, it is paramount to have an accurate simulation of the experimental structured illumination patterns. For this reason, we have devised a probe reconstruction scheme that does not rely on the use of an image aberration corrector. This scheme contains several key steps:

1. **Initial phase guess:** Prior to the experiments, an initial guess of the phase imparted by the electron beam modulating device is obtained through finite element simulations (e.g., by utilizing COMSOL) or experimental techniques such as the transport of intensity equation or electron holography.
2. **Aberration characterization:** By relying on established diagnostic methods, (e.g., Ronchigram analysis), high-precision characterization of the aberrations of the imaging system is achieved.
3. **Pattern acquisition:** The device is introduced into the appropriate condenser aperture plane and images of the generated patterns are captured under various bias conditions.
4. **Phase retrieval:** An iterative algorithm is used to reconstruct the phase for representative cases of different applied bias. The initial guess (as outlined in point 1) aids in convergence of the algorithm towards an accurate solution.
5. **Aberration subtraction:** The effect of the aberrations of the imaging system (point 2) is separated from that of the pattern generator phase.
6. **Pattern prediction:** Our goal is to predict probe shapes for various applied biases, thereby establishing a predictive model.

The critical step is number 4, which relies on the use of an iterative algorithm. While the Gerchberg-Saxton algorithm [2] is a natural choice, its convergence is not guaranteed, as it can become trapped in local minima. In contrast, the hybrid input-output algorithm (HIO) [3] offers a more dependable approach.

Through testing in simulated yet realistic scenarios, we are able to demonstrate its functionality. A specific instance of its successful application is described in Supp.6.



Supp. 6: Convergence behavior of the HIO algorithm. The scale bar in the upper row (structured patterns) is 2 nm, while for the phase images in the device plane the limiting aperture is 7 mrad.

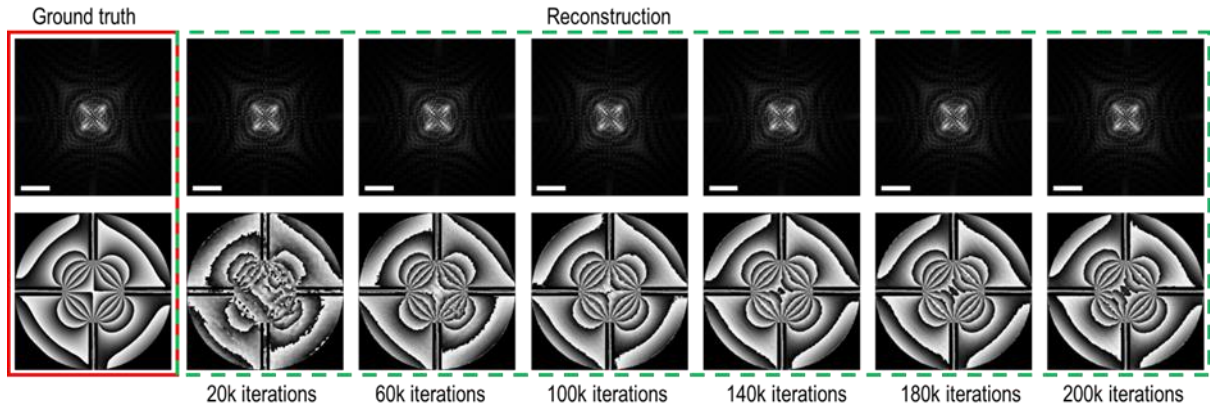
In our simulations, we used a spherical aberration of 1.35 mm in the probe and imaging systems, maintaining symmetrical biasing conditions. Specifically, we used alternating positive and negative OAMs on the four SPPs, as described in the main text.

Unfortunately, the HIO exhibits slow convergence, nearing an exact solution after 200 thousand iterations. In order to improve its speed, we focused on refining and customizing the algorithm to match the specific geometry of our computational ghost imaging setup.

One useful aspect that we considered is the constraint that, leaving aberrations to one side, the phase should follow the constraint that $\nabla^2\varphi = 0$ outside the electrodes. This harmonic constraint implies that the phase must change smoothly in these regions.

In our modified algorithm, the input interference pattern undergoes a convolution with a Gaussian cut-off centered on the barycentre of the interference pattern. Initially, the radius of the Gaussian in Fourier space is small ($\sim 10\%$ of the probe extension). With increasing iterations, it is extended gradually until it covers the entire image, thereby losing its effect.

This approach is similar to that proposed by Roger Vincent [4]. It ensures that an initial guess converges towards a bandwidth-limited solution. At the outset, only low frequency features are reconstructed. As the reconstruction progresses, the bandwidth is relaxed and higher frequency features are reintroduced and reconstructed. Although this approach does not guarantee smoothness of the final solution, it expedites convergence towards such a result, as shown in Figure 2.



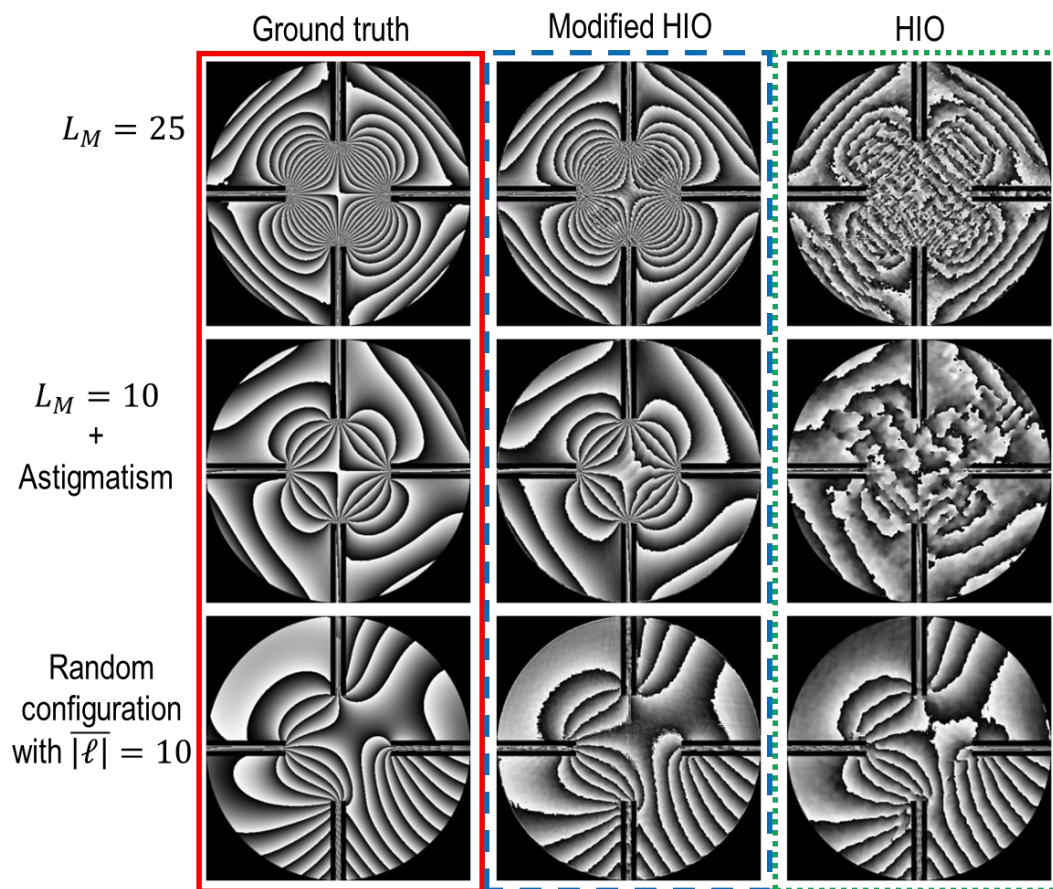
Supp. 7: Convergence behavior of the modified HIO algorithm. The scale bar in the upper row (structured patterns) is 2 nm, while for the phase images in the device plane the limiting aperture is 7 mrad.

By using the modified HIO algorithm, we achieved an almost exact solution in less than half the number of iterations required by the conventional algorithm.

It is important to note the significant advantage derived from the dominance of an even (i.e., center-symmetric) potential in the generative phase towards attaining a final solution. In instances where the condenser exhibits a perfectly even phase landscape, including aberrations, the probe becomes real-valued. Other than a sign ambiguity, this condition would enable an exact analytical solution for the phase to be obtained without the need for an iterative method.

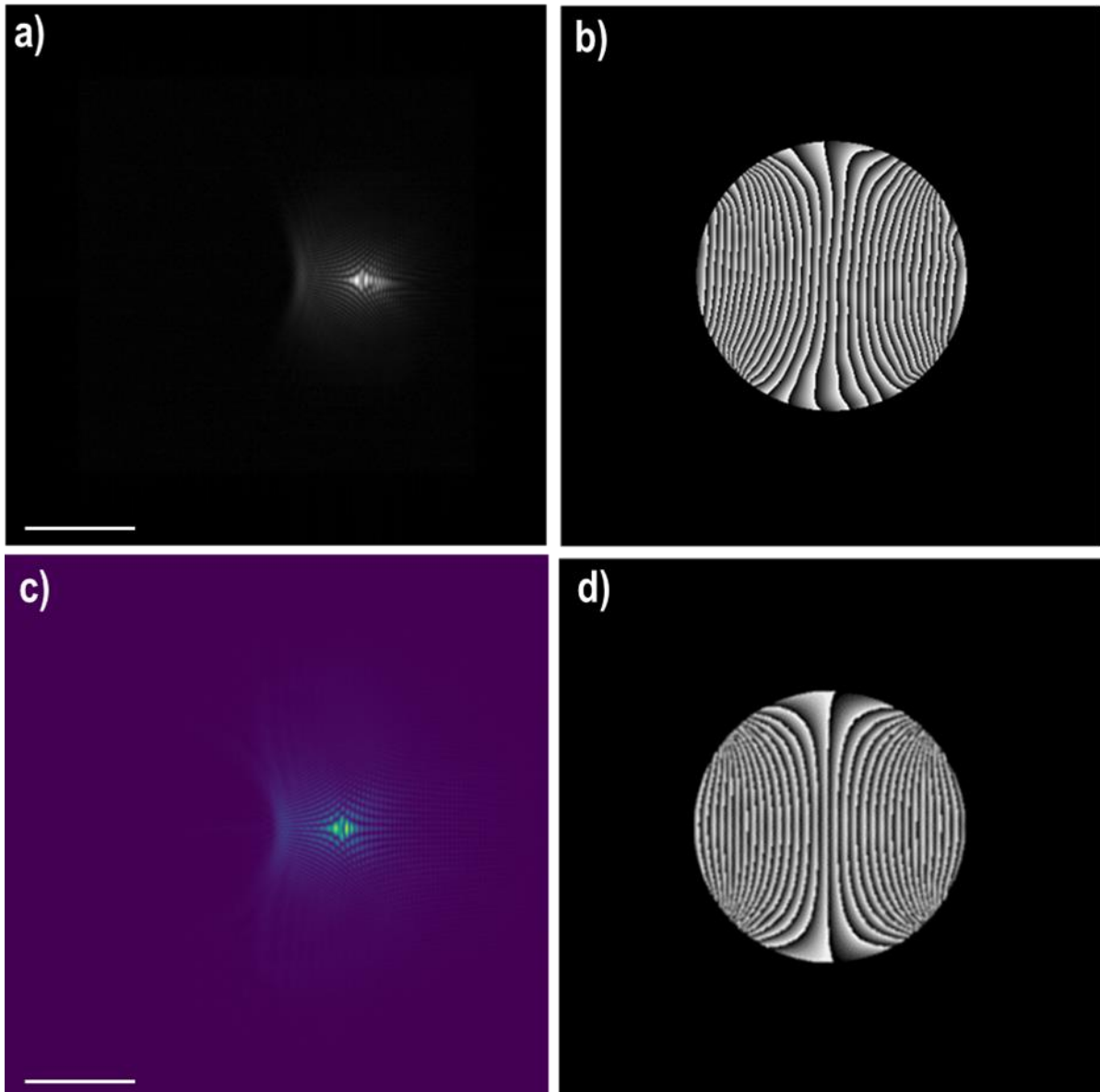
It is important to assess the reliability of both algorithms, in particular for the case of more conspicuous phase shifts or asymmetrical phase profiles. Upon increasing the absolute OAM values applied by each spiral phase plate (SPP) and performing reconstructions using both algorithms, a

significant disparity in performance becomes evident (Supp. 8, upper row). The conventional HIO algorithm struggles to achieve an accurate reconstruction under these stronger phase gradient conditions. However, our modified algorithm maintains its ability to effectively approximate the phase profile. The same behavior occurs for asymmetrical phase profiles (second and third rows of Supp. 8). Interestingly, in the most asymmetrical case (lowest row of Supp. 8), the conventional HIO algorithm still achieves a reasonable result in spite of the complicated target profiles but is slower in reaching an approximate solution. Taken together with other examples that are not shown here, these results show that the conventional HIO algorithm may provide a good approximate reconstruction in some cases, but often fails for asymmetrical phase profiles and has a poor reconstruction success rate.



Supp. 8: Phase reconstructions for more complex phase landscapes. For each algorithm, the reconstructed phase images were obtained after 200 000 iterations.

Finally, we report here early experimental results that are intended for a separate publication, since these experiments are ongoing. First, we demonstrate in a simplified geometry and for low convergence that we are able to reconstruct the phase (Supp. 9b) in good agreement with COMSOL-based simulations (Supp. 9d). The experiment was carried out in a low convergence condition to initially avoid complications from the main high order aberrations. The results are shown in Supp. 9 below.

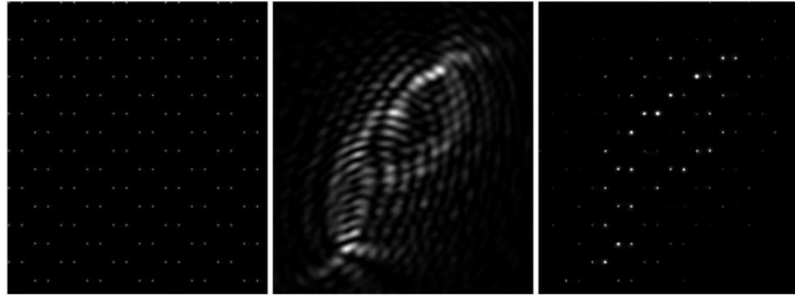


Supp. 9: Test of our modified reconstruction algorithm on an experimental probe/ interference pattern. The pattern was obtained for a low convergence case, with the electron beam passing only through the central part of the device without illuminating the eight electrodes. We also applied a different OAM configuration, since in this case two neighboring pairs of SPPs were kept at the same positive OAM, while the opposite two were kept at negative OAM ($\ell_1 = \ell_2 = -\ell_3 = -\ell_4 = L_M$). a) Recorded experimental probe/ interference pattern. b) Reconstructed phase shift obtained using our algorithm. c) Simulated probe/interference pattern obtained using the same experimental conditions in a COMSOL-based pattern simulator. d) Simulated phase shift in the device plane used to generate the pattern in c). The scale bar in the two images of the structured patterns in a) and c) is 20 nm. The limiting aperture in b) and d) corresponds to 1.8 mrad.

S3: Application limits

In order to produce an image, CGI requires the signal collected by the single pixel detector to be linear, i.e., it must be proportional to the overlap integral of the illuminating beam and transfer function of the sample. To verify that our imaging system satisfies this linearity requirements we

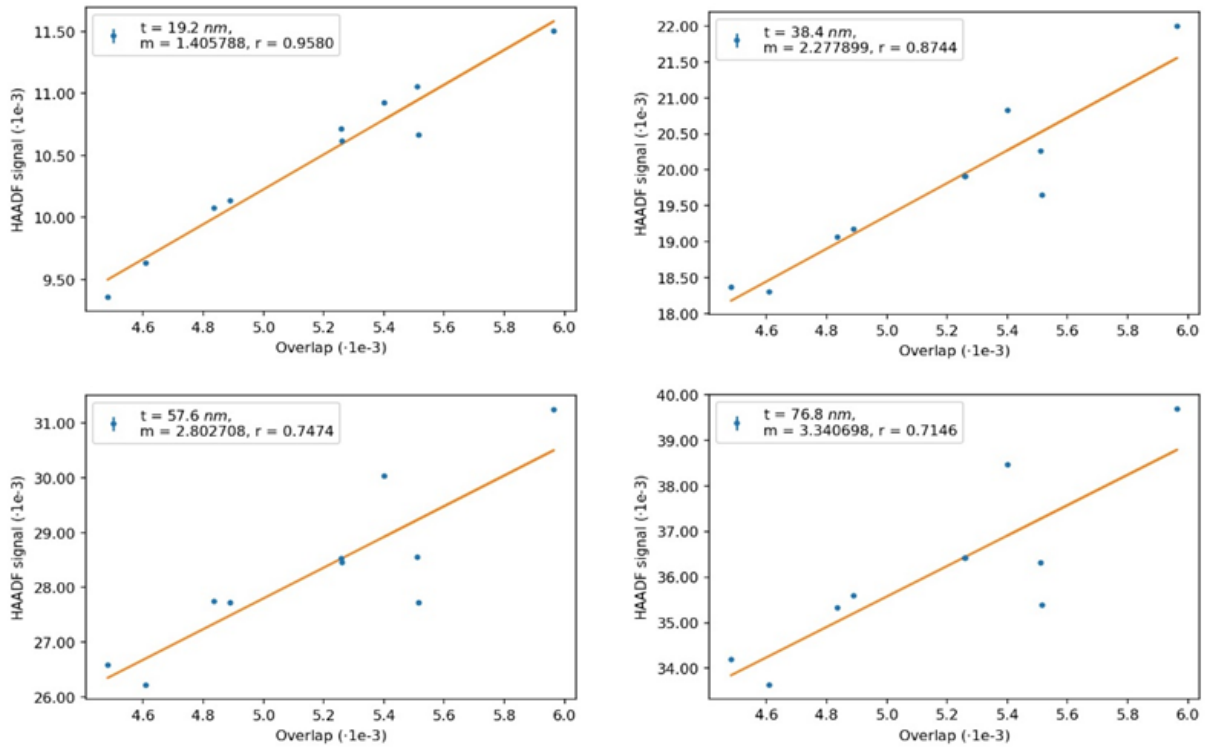
make use of multislice simulations to predict the intensity of the ADF signal. To carry out the multislice simulation we used STEM-cell [5], which is largely based on Kirkland [6] multislice algorithm that is an implementation of frozen lattice multislice. As a benchmark sample we used a 80nm thick Si crystal seen along the [110] direction. The predicted ADF signal is correlated with the overlap between sample and probe (Supp. 10).



Supp. 10: Example of overlap processing: (a) square of the crystal potential, (b) pattern, (c) overlap of (a) and (b)

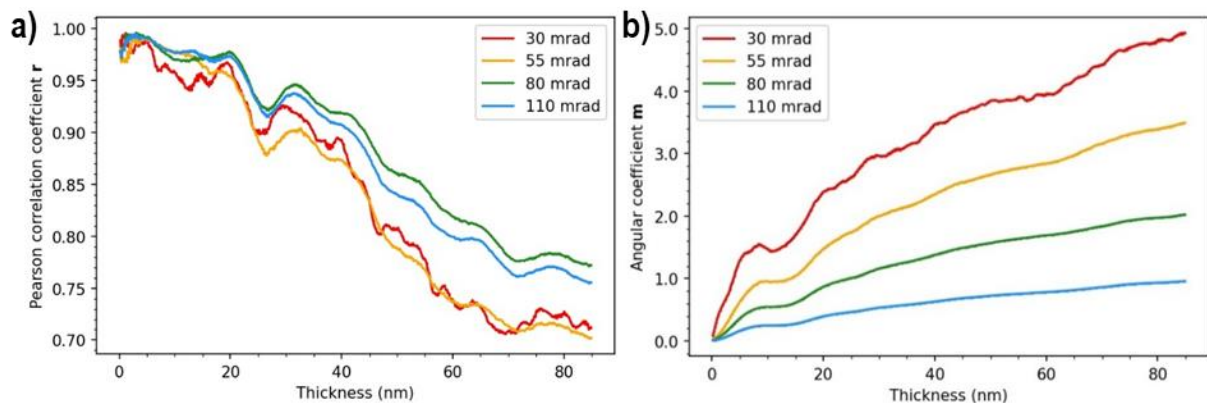
The degree of linearity between the overlap and the ADF signals was evaluated by means of linear regressions for every thickness. A selection of the fitting is reported in figure Supp. 11. As expected the linearity of the signal-overlap relation decreases as the thickness increases due to unavoidable dynamic effect due to the crystalline nature of the sample.

Plotting the Pearson correlation coefficient r and the angular coefficient m for regressions made at different thicknesses and four different value of the inner collection angle (figure Supp. 12) makes it clear that the linearity of the imaging scheme is valid only up to ~ 20 nm of sample thickness. r has an onset in the first few nanometers, stays almost constant up to ~ 20 nm and then starts decreasing for higher values of thickness. The angular coefficient grows progressively and shows that there is proportionality between ADF signal and the thickness.



Supp. 11: Linear regressions of ADF signal vs Overlap at 19.2nm, 38.4nm, 57.6nm, 76.8nm.

We also explored the effect of the collection angle ranges on the linearity. Inner detection angles higher than 55 mrad lead to a good linear correlation in line with what expected from conventional ADF imaging [7]. In any case, inner detection angles higher than 55 mrad do not provide any significant improvement for thin sample below 20 nm. This limit is of course referring to a Silicon sample and it is expected to change for different materials, however CGI is expected to work better on thin samples, as for the general rule of electron microscopy.



Supp. 12: Pearson correlation coefficient (r) and angular coefficient m as a function of thickness

S4: CGI reconstruction

The CGI imaging scheme assumes that the measurement is linear, i.e., the measured signal I corresponds to the overlap integral (scalar product) between the probe p and the sample transfer function T . The image formation process can then be expressed, in matrix notation, as:

$$\vec{I} = \hat{P}\vec{T} \quad (\text{S9})$$

where \vec{I} is the vector containing the N signals recorded by the single pixel detector corresponding to the N different illuminating probes that are gathered in the matrix \hat{P} . The \vec{T} is the transfer function describing the sample. In most cases the sample is assumed to be a pure amplitude object described by a 2D transfer function, i.e., the sample's transparency. The sample is discretised on a grid of points $\{i, j\}$. The illumination patterns p_n are defined on the same grid of points $\{i, j\}$. The signal I_n recorded when the pattern p_n passes through the sample is:

$$I_n = \sum_{i,j} T(i,j)p_n(i,j) \quad (\text{S10})$$

Many algorithms exist to solve for \vec{T} , here we compare the following three:

- the traditional CGI algorithm [8];
- the Alternated Projection algorithm (AP) [9];
- Conjugated Gradient Descent (CGD) algorithm [10,11];

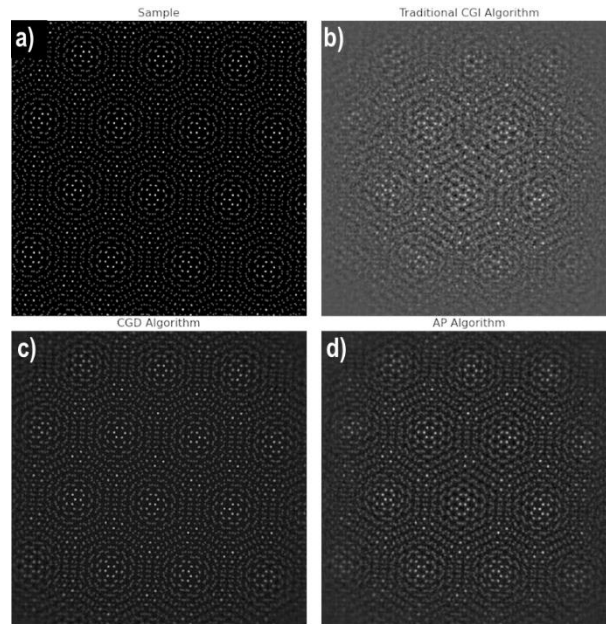
Traditional CGI (T-CGI) is the fastest algorithm, and it can provide a real time reconstruction during an experiment. Unfortunately, as reported many times in the literature, T-CGI has a non-optimal reconstruction quality. Among the plethora of iterative algorithms that may improve reconstruction quality, we chose CGD, for its reliability (it is proven to converge in n iterations, where n is the number of pixels) and AP which is more resilient to noise, even though its convergence is not always ensured. Even though both AP and CGD require long calculation times to reach convergence, as we will show, even with a limited number of iterations they provide better quality reconstruction than T-CGI with a resolution superior to the aberration limit.

Another feature common to all reconstruction algorithms is that they do not require a complete set of measurement (equal to the number of pixels) to converge. CGI reconstructions with a sampling ratio (SR), i.e., the ratio between the number of measurements and number of pixels, well below the Nyquist-Shannon limit are achievable. This can prove a desirable characteristic if high speed or dose reduction is sought for.

As reported in the main text of the paper, our sample of choice is a MoS₂ twisted bilayer composed of two MoS₂ monolayers stacked on top of each other at a 7° angle. We chose such sample since it offers an almost continuous range of interatomic distances and has proven very useful to estimate the resolution of imaging techniques [12]. The sample's squared atomic potential is shown in Supp. 13 a. In Supp. 13 b,c and d are shown the reconstructed CGI images with the three algorithms. For the traditional CGI reconstruction, we used a low-pass filter re-normalization where we filtered out frequencies below 1.83 nm⁻¹ (half the frequency corresponding to the Mo lattice parameter) since the traditional algorithm isn't able to reduce the low-frequency background. Moreover, it is possible to notice how all three of the reconstructed images are blurred near the image edges and this can be explained by looking at Supp. 4, where even in the optimised illumination case it is still not possible

to illuminate all the FOV. It is necessary to point out that T-CGI algorithm is non-iterative, and the reconstruction time was ~ 2 minutes, whereas both the CGD and AP algorithms are iterative.

To perform a fair comparison between the two iterative algorithms all reconstructions have been performed for the same amount of time (12h). In agreement with the literature [13], in absence of noise the CGD algorithm provides the best reconstruction quality.

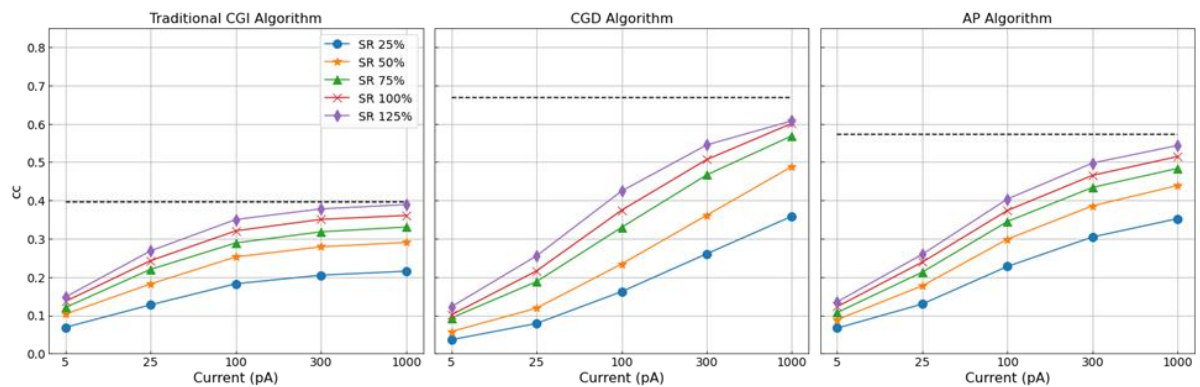
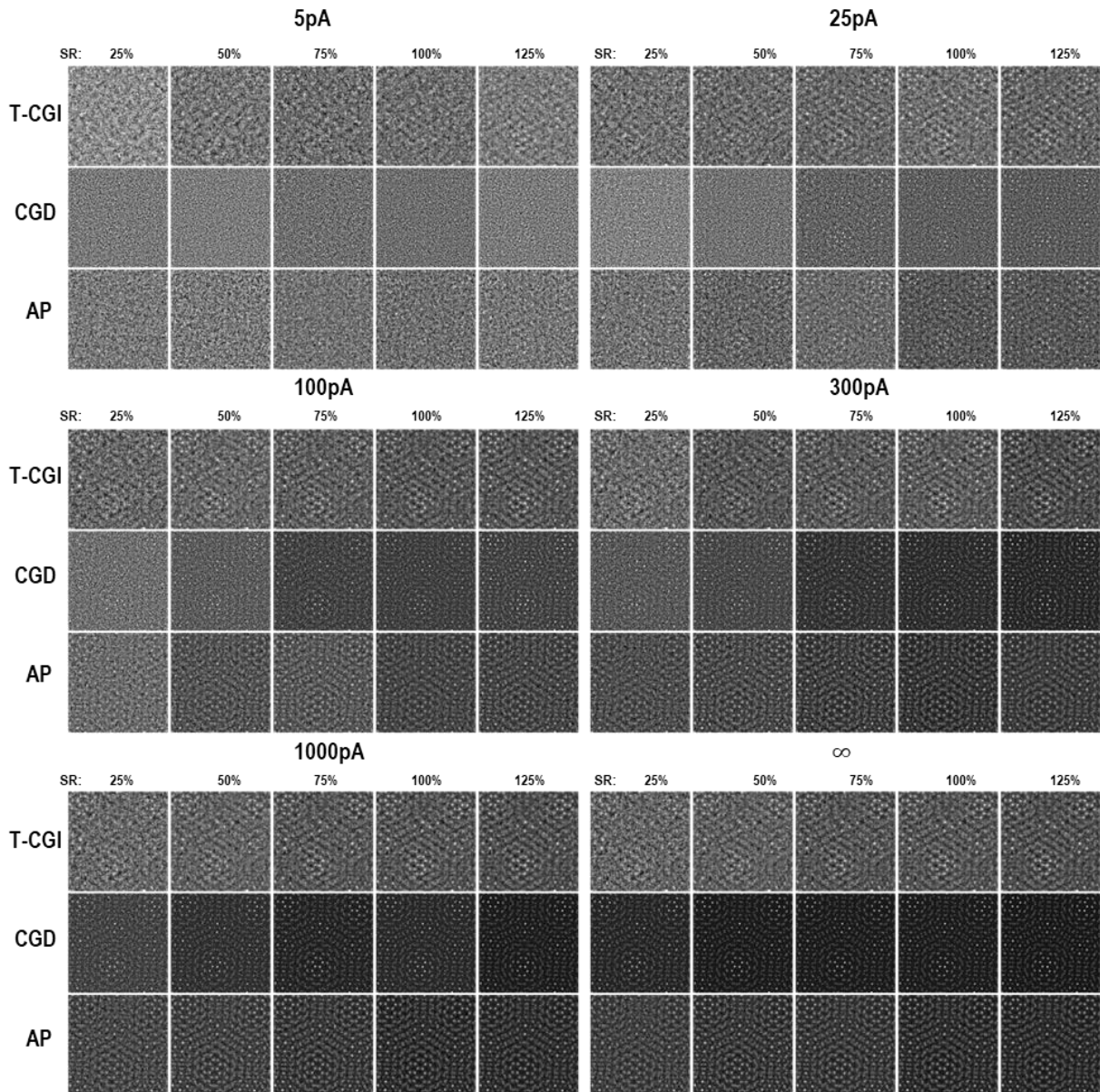


Supp. 13: a) MoS2 twisted bilayer squared atomic potential, b) reconstructed CGI image with the traditional CGI algorithm, c) reconstruction with the CGD algorithm and d) with the AP one. The images are $400px \times 400px$ corresponding to a $10nm \times 10nm$, and we used 160000 patterns for the reconstruction shown here.

Next, we compared the three algorithms against noise and sampling ratio. We assume the single pixel detector to a perfect detector with quantum efficiency equal to 1, leaving counting noise, i.e. the beam current at which the experiment is conducted, as the sole source of noise. The results are shown in Supp. 14. As reported in the main paper text, we defined a metric to compare the reconstructed image to the reference image:

$$CC = \frac{\sum_{i,j} \frac{(I_R(i,j) - \bar{I}_R)(I_{GT}(i,j) - \bar{I}_{GT})}{|(I_R(i,j) - \bar{I}_R)| |(I_{GT}(i,j) - \bar{I}_{GT})|}}{N_{pix}} \quad (S11)$$

where $I_R(i,j)$ is the pixel intensity of the pixel with indices (i,j) of the reconstructed image, \bar{I}_R is the average pixel intensity of the reconstructed image, $I_{GT}(i,j)$ is the pixel intensity of the pixel with indices (i,j) of the ground truth image, \bar{I}_{GT} is the average pixel intensity of the ground truth image and N_{pix} is the total number of pixels. The correlation coefficient that we defined is a modified version of the cross-correlation coefficient defined by Thust and Urban [14].

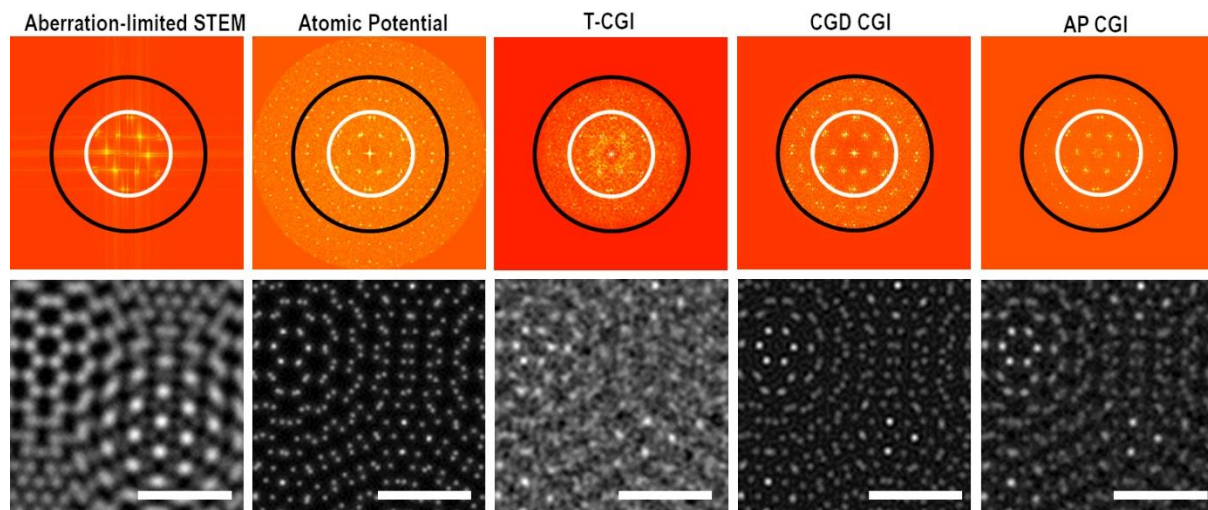


Supp. 14: Image reconstruction against beam current and SR for the three algorithms. For ease of view only a reduced portion of $3.6\text{nm} \times 3.6\text{nm}$ of the whole reconstructed image is shown. On the bottom row are reported the three graphs of cc vs current are shown, one for each algorithms.

As a confirmation of what said in the main text, the reconstruction quality can be improved by increasing the beam current or SR. At 5pA beam current, all three algorithms fail in reconstructing the image, but already at 25pA the reconstruction quality improves with the three algorithms

performing almost identically. A difference between the three algorithms appears at 100pA beam current, here the T-CGI starts to be outperformed by CGD and AP, and it is also possible to see that for low SR (25 and 50%) the AP algorithm performs better than CGD, but from 75% the CGD performs better. This tendency is also true at 300pA, but at 1000pA (very low noise) CGD becomes the best performing one. It is worth to emphasize that the CGD algorithm even at 1000pA is far from reaching its limit (defined from reconstructing images with maximum SR and without noise, so infinite current), while T-CGI and AP are almost at the limit. In addition, both AP and CGD have been run only for 12h and were far from convergence, so the results would have been better in case of longer reconstruction times

It also possible to compare the resolution of the reconstructed images for the different algorithms. In the bottom row of Supp. 14 we show images from a small portion of the whole image. For T-CGI, CGD and AP are shown the reconstructed images with SR=100% and noiseless reconstruction. In the upper row are shown the FFTs of the images. The inner white circle corresponds to frequencies of $14.89nm^{-1}$ and it represents the maximum resolution achievable in the aperture-limited non aberration corrected STEM with aperture convergence equal to α . The black circle on the other hand correspond to a frequency of $29.78nm^{-1}$ which represents the maximum resolution achievable when using an aperture defined convergence angle equal to 2α , that is the one that we used to generate the illumination patterns. As it is possible to see the T-CGI has peaks in its FFT that correspond to frequencies higher than $14.89nm^{-1}$, however visually the reconstruction quality isn't as good as the simulated STEM image, making it more similar to a noisy experimental aberration-limited STEM image. Beside this, CGD and AP both have peaks near the black circle, but the peak intensity is better in CGD confirming that even though both algorithms have a resolution limited by the aperture used to generate the patterns, the CGD has a better image reconstruction fidelity and, in fact, the image appears sharper and less blurred than the AP one.



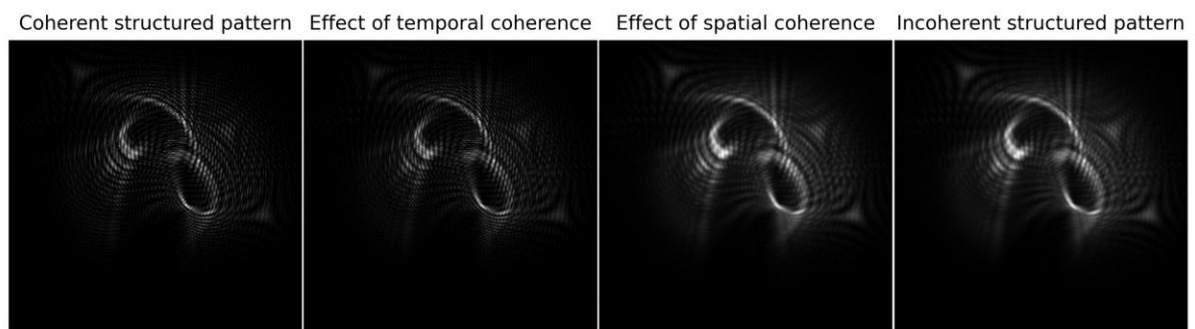
Supp. 15: Comparison between the aberration-limited STEM image, atomic potential and CGI reconstructed image of a region of the sample and the corresponding FFT (scalebar is 1nm). In the non-aberration corrected STEM image, it is possible to see that the resolution is limited by the 2α cutoff (marked with a white circle in the FFT images) used to in the simulation to filter out the aberrations, while the CGI image resolution is limited by the 4α cutoff (black circle in the FFT images). It is important to stress out that the illumination patterns have been generated considering an accelerating voltage of 300keV, a Cs of 2.7mm and defocus equal to four times Scherzer defocus.

In summary, the comparison of the three algorithms has led us to similar conclusions already present in literature, confirming that in absence of noise or in low noise regime the CGD algorithm is the best performing one, both from a point of view of image reconstruction quality and in our

specific case spatial resolution (which can be seen as another metric to test the reconstruction quality).

S5: Effect of coherence on reconstruction quality

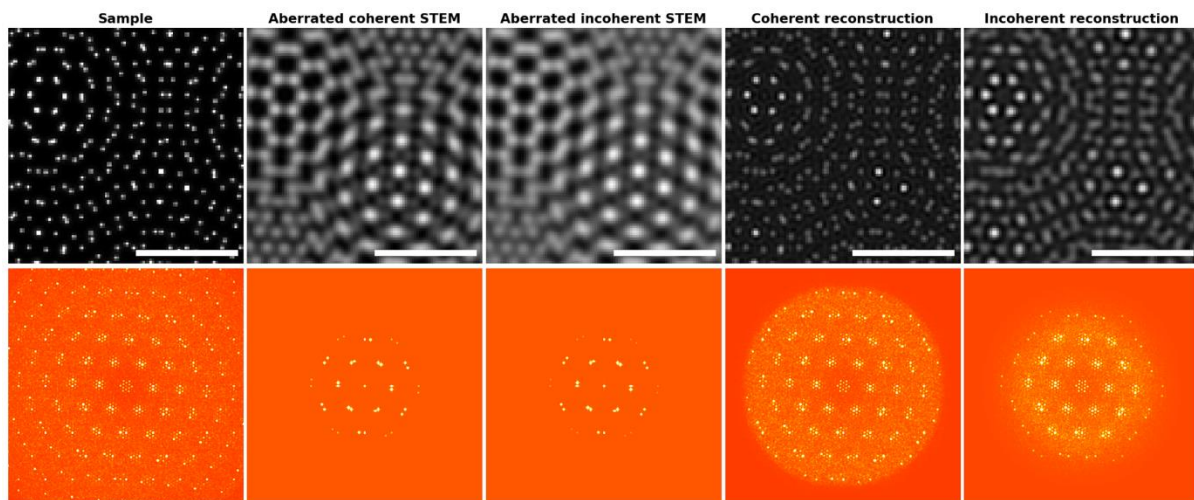
The coherence of the beam is the main factor that limits the reconstruction quality and thus the final spatial resolution of the image. As it is well known, both spatial and temporal coherence of the electron beam are related to the electron source. Spatial coherence is related to the effective source size of the electron beam, while temporal coherence is related to its energy spread. In fact, since the source has a finite size, at the sample the incident beam contains a distribution of incident directions limiting the possibility to achieve a point-like probe. This results in the blurring of the recorded images. The effect of the source size can be properly simulated using a Monte Carlo simulation, but studies [15] have shown that it can be well approximated by convolving a bidimensional gaussian function with the sample atomic potential distribution. In particular, in reference **Error! Bookmark not defined.** the authors estimate that the gaussian kernel for a 300kV TEM equipped in high resolution condition has a standard deviation of 34pm. On the other hand, since the electron source has an energy spread (ΔE) this translates due to the chromatic aberration of the objective lens in a spread of focus proportional to the energy spread ($\Delta f \propto \Delta E$). Even in this case a Monte Carlo simulation can be used to understand the effect of temporal coherence. In figure Supp.16 we see the effect of the two types of coherence on the structured probes that illuminate the sample, first separately and lastly on the rightmost frame together.



Supp. 16: Effect of coherence on the structured probe that illuminates the sample. From left to right we have a fully coherent probe, the same probe simulated taking into account the spread of focus that arises from the energy spread of the beam (temporal coherence), the same probe simulated taking into account the finite source size of the electron beam (spatial coherence) and lastly the same probe simulated taking both temporal and spatial coherence.

For both types of coherence, we run Monte Carlo simulations where the numbers were randomly generated using a normal distribution with standard deviation equal to 4.2nm of defocus for the temporal coherence and 34pm for spatial coherence. As it is possible to notice in Supp 16, the main source of decoherence is the spatial coherence (or lack of it). Thus, to generate the incoherent database we only considered the effect of spatial coherence. Furthermore, to speed up the database generation, instead of running a Monte Carlo simulation for every pattern, we simply convolve the coherent pattern with a properly designed a bidimensional gaussian function with a standard deviation equal to 34pm. As it can be expected the gaussian blurred pattern isn't able to fully reproduce the effect of both types of coherence, but it is still a good approximation. The incoherent database of 200000 pattern has been generate by simply convolving the previous 200000 coherent patterns with the 2D gaussian function.

With the incoherent database we re-performed the reconstruction using only the best performing algorithm (CGD) at 100% SR and without considering noise. The resulting incoherent reconstruction is shown in the rightmost column of Supp 17.



Supp. 17: Here we show the effect of spatial coherence (since we determined that it is the most prominent one) on the loss of resolution in the CGI reconstructed image. From left to right we first show the sample's atomic potential, the aberration limited simulated STEM image obtained considering a point-like source, the aberration limited simulated STEM image obtained considering a finite size source, the CGD reconstruction with the coherent probe database and lastly the CGD reconstruction with the incoherent probe database (in all the scale bar is 1nm). On the bottom row are reported the respective FFTs.

As a matter of fact, a Gaussian convolution in real space is equivalent to a multiplication in reciprocal space (Convolution Theorem) with a Gaussian whose FWHM is inversely proportional to the real space one. Owing to the linearity of CGI, the same Gaussian is expected to affect the Fourier transform of the reconstructed images, dampening high frequency.

Aberration limited STEM images appears to be only marginally affected by decoherence. In fact, there is no loss in resolution, but only a marginal reduction in the contrast. This is expected because the limiting aperture is much smaller than the multiplicative Gaussian that describes the coherence.

The dampening of high frequencies is instead clearly visible in the CGI image reconstruction where no limiting apertures are used and the decoherence itself is acting as a soft aperture cutting the highest frequency. This verify that coherence is the main limiting factor to the resolution of CGI that can only be overcome using a stable microscope with the smallest effective source size possible.

References

- [1] G. Pozzi, P.-H. Lu, A. H. Tavabi, M. Duchamp, and R. E. Dunin-Borkowski, *Generation of Electron Vortex Beams Using Line Charges via the Electrostatic Aharonov-Bohm Effect*, *Ultramicroscopy* **181**, 191 (2017).
- [2] R. W. Gerchberg and W. O. Saxton, *A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures*, *Optik (Stuttg)*. **35**, 237 (1971).
- [3] J. R. Fienup, *Phase Retrieval Algorithms: A Comparison*, *Appl. Opt.* **21**, 2758 (1982).
- [4] R. Vincent, *Phase Retrieval in TEM Using Fresnel Images*, *Ultramicroscopy* **90**, 135 (2002).

- [5] V. Grillo and E. Rotunno, *STEM_CELL: A Software Tool for Electron Microscopy: Part I—Simulations*, *Ultramicroscopy* **125**, 97 (2013).
- [6] E. J. Kirkland, *Advanced Computing in Electron Microscopy* (Springer US, Boston, MA, 2010).
- [7] S. J. Pennycook, *Z-Contrast Stem for Materials Science*, *Ultramicroscopy* **30**, 58 (1989).
- [8] J. H. Shapiro, *Computational Ghost Imaging*, *Phys. Rev. A* **78**, 061802 (2008)
- [9] K. Guo, S. Jiang, and G. Zheng, *Multilayer Fluorescence Imaging on a Single-Pixel Detector*, *Biomed. Opt. Express* **7**, 2425 (2016).
- [10] M. R. Hestenes and S. Eduard, *Methods of Conjugate Gradients for Solving Linear Systems*, *J. Res. Natl. Bur. Stand. (1934)*. **49**, 409 (1952).
- [11] D. G. Luenberger, *Introduction to Linear & Nonlinear Programming* (Addison-Wesley, Boston, MA, 1973).
- [12] Y. Jiang, Z. Chen, Y. Han, P. Deb, H. Gao, S. Xie, P. Purohit, M. W. Tate, J. Park, S. M. Gruner, V. Elser, and D. A. Muller, *Electron Ptychography of 2D Materials to Deep Sub-Ångström Resolution*, *Nature* **559**, 343 (2018).
- [13] L. Bian, J. Suo, Q. Dai, and F. Chen, *Experimental Comparison of Single-Pixel Imaging Algorithms*, *J. Opt. Soc. Am. A* **35**, 78 (2018).
- [14] A. Thust and K. Urban, *Quantitative High-Speed Matching of High-Resolution Electron Microscopy Images*, *Ultramicroscopy* **45**, 23 (1992).
- [15] C. Dwyer, R. Erni, and J. Etheridge, *Measurement of Effective Source Distribution and Its Importance for Quantitative Interpretation of STEM Images*, *Ultramicroscopy* **110**, 952 (2010).