Abstract
Three-dimensional (3D) imaging of thin, extended specimens at nanometer resolution is critical for a range of applications in biology, materials science, advanced synthesis, and manufacturing. Existing 3D imaging techniques are limited to surface features, or available only for selective cross-sections, or require a tilt series of a local region, hence making them unsuitable for rapid, non-sacrificial screening of extended objects, or investigating fast dynamics. Here we demonstrate a coherent imaging technique that recovers the 3D volume of a thin specimen with only a single, non-tomographic, energy-filtered, bright-field transmission electron microscopy (TEM) image. This technique does not require physically fracturing or sectioning the specimen, is fast since only modest electron doses of ~100 e Å$^{-2}$ are required, and can be readily calibrated for many existing TEMs; thus it can be widely deployed for rapid 3D metrology that complements existing forms of metrology.

1. Introduction
Nanometrology is an indispensable part of nanoscience and nanotechnology. Since the dimensions of a nano-object define its properties and functionalities, it is inevitable to get rapid feedback on the fabrication and synthesis strategies, failure analyses, reverse-engineering, and counterfeit verification$^{1-3}$. From investigating the self-assembly of natural photonic crystals (nanometer-scale) found in butterfly wing scales (micrometer-scale) to inspecting transistors (nanometer-scale) in an integrated circuit chip (millimeter-scale), three-dimensional (3D) nanometrology for extended flat volumes has a variety of applications$^{4-6}$. Especially, there is a pressing need for fast, high-throughput, high-resolution, in-situ nanometrology in the semiconductor industry, as it moves toward 3D power scaling, the new scaling paradigm to keep up with the increasing demands for denser and more energy-efficient compute and storage components$^{7,8}$. Existing metrology tools either cannot scale to large areas or fail to meet the complexity and resolution demands of 3D power scaling$^{7,8}$.

Optical scatterometry, scanning electron microscopy (SEM), and atomic force microscopy (AFM) are the most commonly-used nanometrology tools for process control$^{9,10}$. Scatterometry is the fastest method compared to other techniques available for inline process control. However, it uses heavy modeling to determine the average deviations of many nanostructure from an idealized model, which renders it unsuitable for measuring the features of unknown geometries/defects.

SEM metrology is predominantly used for estimating the 3D structures from a single 2D perspective. This approach lacks complete 3D measurement and is often destructive, i.e. we have to break the specimen across a feature for measuring the cross-section. SEM-based techniques, which provide primarily surface contrast, cannot probe deeply into the sample and are subject to feature occlusion$^{11}$.

Alternatively, AFM can image surfaces to atomic resolution, and is non-destructive. However, the raster scans of AFM are relatively slow, and higher scan rates cause high noise levels, distortion in the images, and damages the tip$^{12,13}$. Furthermore, different types of structures require different AFM probes. SEM and AFM techniques can only provide topographical information about one of the specimen surfaces; a separate scan is necessary.
for the underside, and a careful registration of these two scans is needed to compose a 3D relief map. These limitations make AFM and SEM impractical for dynamic 3D imaging or measuring fields of view in the micrometre range.

X-ray computed tomography and electron tomography can provide complete 3D measurements; hence, they are widely used for 3D structural analysis at the nanometer scale. However, the time required to acquire a full-tilt series typically limits the resolvable dynamics to a few seconds which is not sufficient for imaging deformation dynamics that require millisecond resolution. More important for extended samples common in the fabrication of electronic components, where there is usually a high aspect ratio between at least one pair of dimensions, collecting a full tilt-series is challenging either because of occlusion by the sample holder at high tilt angles or X-ray/electron absorption when viewing down the extended dimension. Ptychographic X-ray laminography, a coherent imaging technique that still requires a series of images for the reconstruction, is limited to large bright X-ray facilities such as synchrotrons and X-ray free-electron lasers.

Here, we present a coherent imaging technique that allows rapid sub-10 nm resolution 3D metrology of thin, extended materials from single TEM images that complements existing metrology modalities. We coin this technique single-shot pop-out 3D metrology, which simultaneously measures both the thickness and the imaging depth (i.e. z-position) of the material by utilising both the absorption and phase contrast information from a partially coherent TEM image. Because this method extracts 3D information from a single 2D image, the same field of view can be re-interrogated rapidly, which allows us to study fast structural dynamics of materials with nanometer resolution. The data efficiency of pop-out metrology also supports fast in-situ inline process control and defect detection in nanofabrication of 3D devices over extended surfaces, without physically sectioning or fracturing the sample as one would in critical dimension metrology. Finally, this technique can be readily calibrated for the many TEMs that are commonly found in many manufacturing, fabrication, and research facilities.

2. Principle of pop-out 3D metrology

While the local thickness of a specimen is routinely measured from its amplitude contrast in BF-TEM (Bright Field-TEM), its local depth along the optical axis is far more non-trivial. The goal of pop-out 3D metrology is to serially infer both the depth and thickness of overlapping regions of an extended sample. When a specimen is illuminated by a largely coherent, monochromatic electron beam in BF-TEM, its resultant image is in fact a near-field interference pattern. This interference pattern, which typically creates undesirable speckles that confound sample contrast, actually encodes the depth of local regions of a specimen akin to how the fringe spacings in the Fresnel diffraction pattern of an object changes as it is moved closer or farther away from a detector.

To sample this interference pattern sufficiently requires a large enough patch of pixels. To infer a sample’s local thickness, we measure the fraction of electrons lost to inelastic scattering (see Methods section). To estimate this fraction, the electrons from our BF-TEM images are energy-filtered via hardware, hence retaining only the elastically scattered electrons. SI Section 1 derives the mathematical foundations of this approach more rigorously, starting from the multislice scattering formalism.

Fig. 1a illustrates the pop-out principle with an amorphous specimen of homogeneous density. The specimen’s right region is thinner and closer to the back focal plane than the left region. Hence, the thinner right region elastically scatters fewer electrons, which makes the right half of the energy-filtered TEM image appear lighter. Quantitatively, the relative thickness of these two regions is determined from the log-ratio of the number of electrons received at each region.

The depth of the centre of mass of each region in Fig. 1a is apparent in their local power spectrum (i.e., squared-amplitude of their Fourier transforms). The Thon rings in the
power spectrum from the right region are spaced farther apart compared to those from the left region. This ring-spacing is related to the relative defocus of each region with respect to the TEM's plane of focus. Consequently, each region's centre of mass depth can be determined from its defocus parameter from a semi-empirical fit to these Thon rings \(^{24-28}\). The contrast transfer function (CTF) for amorphous, homogeneous-density specimens is modeled by Eq. (1), which accounts for spherical aberration, astigmatism, inelastic scattering, multiple scattering, and the incoherence of the electron beam (see Methods section). Note that the visibility of these Thon rings is strongly affected by the spatial coherence of the electron source in TEMs\(^{29}\).

Combining the inferred local sample thickness and its centre of mass depth in Fig. 1a, we can correctly "pop-out" a thicker block of material on the left half of the field of view that is farther away than the thinner, nearer block on the right. Repeating this recipe across small, overlapping patches in the 2D TEM image allows us to recover a 3D structure of the entire field of view: the depths of these overlapping patches are stitched together to reconstruct how the specimen's centre of mass depth (i.e., along z-axis) varies transversely (i.e., along the x-y plane).

A proof of concept for pop-out 3D metrology is shown in Fig. 1b using a simulated BF-TEM image. Starting from a ground truth 3D object made of amorphous silicon nitride (bottom left of Fig. 1b), we used a multislice approach that includes absorption effects\(^ {29}\) to simulate its energy-filtered BF-TEM image. From this BF-TEM image, we inferred the patch-wise defocus and thickness maps of the 3D object which were combined to "pop-out" the 3D reconstruction of the object (bottom right of Fig. 1b).

3. Factors affecting the resolution of pop-out metrology

Understanding which factors affect the resolution limits for a given specimen on a given microscope can help us calibrate its key parameters for optimal pop-out reconstructions. The in-plane (x-y) and out-of-plane (z) resolutions of a pop-out reconstruction depend on the sample material, sample thickness, range of defocus parameters, patch size, and total electron dose chosen in the experiment. Here we use realistic simulations to study the interplay between these factors and how they affect both in-plane and out-of-plane resolutions.

A critical factor in a pop-out reconstruction is the patch size. Since we determine the sample's average thickness and centre of mass depth over an image patch, the size of this patch limits our x-y resolution. Furthermore, the size of the patch is proportional to the total number of elastically scattered electrons within the patch, as well as how finely the Thon rings are sampled in the Fourier domain. Hence the patch size, in turn, impact the z resolution of a pop-out reconstruction: the precision of each patch's determined centre of mass depth.

3.1 Effect of sample thickness on the resolution

The thickness of a specimen along the optical axis, \(T_{max}\), determines both the first node of the envelope function (dotted vertical lines in Fig. 2a) and the number of electron counts received at the detector. The envelope node suppresses the CTF undulations, which further affects our depth (z) resolution. Similarly, when the received electron counts are too low for model fitting, we have to increase the patch size, which lowers our x-y resolution.

The effects on x-y and z resolutions do not change linearly with sample thickness. Given a fixed incident electron dose, a thicker sample elastically scatters more incident electrons, which imprints clearer Thon rings over the unscattered beam in the sample's TEM power spectrum. This is a positive effect because clear rings allow better fits for the sample's centre of mass depth. However, a thicker sample also inelastically scatters more incident
electrons, which reduces the number of electrons that reach the downstream image detector. This has a negative impact because the entire power spectrum is now noisier, which increases the uncertainty in our semi-empirical fits for centre of mass depths. Fig. 2c shows this positive effect dominates over the negative effect at modest sample thicknesses.

3.2 Effect of spatial sampling and patch size on the resolution

Increasing the image patch size increases the sampling frequency of the power spectrum, which improves their fit to the semi-empirical CTF functions (Fig. 2b). This, in turn, improves $z$ resolution but notably at the expense of $xy$ resolution. For a decent fit, we adopt a simple criterion for sampling frequency of at least three intensity maxima within the first node of the envelope (exemplified by Fig. 2a), and at least eight frequency samples between maxima ($\sigma_z = 8$) (see SI section 1.1). To accommodate for unforeseen uncertainties in actual experiments, we recommend having slightly higher sampling frequencies (hence, larger image patch sizes).

In Fig. 2b, we performed a numerical experiment to analyse the influence of the image patch size on pop-out reconstructions. Here, we simulated multislice TEM images of an amorphous silicon nitride pillar (50 nm tall) on top of two different silicon nitride substrates (thicknesses of 25 nm or 100 nm). The corresponding maximum sample thicknesses are $T_{\text{max}} = 75$ nm and 150 nm (pillar on the substrate), respectively, which lead to minimum image patch sizes of $w_{\text{min}} = 18$ nm and 27 nm (see SI S1.1). Fig. 2b illustrates the effect on pop-out reconstructions of the $T_{\text{max}} = 75$ nm sample with sufficiently large patch sizes (20 nm patch in case i), or with sizes that are too small (12.8 nm in case ii). When the maximum sample thickness increases to $T_{\text{max}} = 150$ nm (case iii), even the larger image patch of 20 nm produces a poorer reconstruction as it is less than the minimum patch size required (27 nm) for $T_{\text{max}} = 150$ nm.

3.3 Effect of electron dose on depth resolution

The precision and accuracy of depth and thickness determination are impacted by the total number of elastically scattered electrons measured at the detector for each image patch. This number, in turn, depends on the total incident electron dose and the sample’s thickness.

To study the effects of electron dose, we simulated a series of multislice TEM images with different specimen thicknesses and integrated electron dose. Fig. 2c shows the depth errors in our pop-out reconstructions of these images with an image patch size of 20 nm. This figure illustrates that 5 nm accuracy in $z$ depth is possible for silicon features that have a thickness less than half of the inelastic mean free path (in this case, 66.5 nm for SiNx) with exit doses measured at the detector that are $\sim 100$ e Å$^{-2}$. The error map plotted against exit dose instead of incident dose to account for electrons lost to inelastic scattering. The inelastic mean free path assumed for multislice simulations is 133 nm which is calibrated from our experiments$^{3,30}$ (Fig. 4). We expected a monotonous increase in the error values with further thickness increase. However, the vertical artefacts that show deviation from the expected trend highlight the limitation of our model. The precision in error maps improves as the exit dose increases dramatically up to 100 e Å$^{-2}$, with negligible improvement with further dose increases. Finally, when comparing the reconstructions in Fig. 2b, it is evident that increasing the patch size lowers depth errors. However, this lowered error, which improves the $z$ resolution because of the larger patch size, is at the expense of $xy$ resolution. Although the studies of $xy$ and $z$ resolutions vs. electron dose and patch size in Fig. 2 were from idealized simulations, similar studies can be calibrated for known samples at actual TEMs. Such calibrations were performed for the applications shown below.
4. Applications in nanometrology

Pop-out 3D metrology can be used for rapid nanometrology in inline process control or quality inspection over large fields of view. A series of overlapping TEM images of an extended thin sample can be stitched together using pop-out 3D metrology. Single exposure nanometrology can also be done for selected sample regions.

4.1 Simulations of a large-scale nanometrology.

The left panel of Fig. 3a shows a multislice TEM image of various nanometer-scale features commonly found in semiconductor devices. The smallest feature size in this simulated model in the lateral $x,y$ plane and longitudinal $z$ directions are 15 nm and 12.5 nm, respectively. The entire field of view spans $1.84 \times 1.23 \mu m^2$, which is stitched together using $6 \times 4$ smaller TEM images, each with $2048 \times 2048$ pixels that are 0.15 nm wide. Details are elaborated in SI section 4. The accompanying pop-out 3D metrology reconstruction can resolve all the nanometer features with only a single TEM exposure on each region of the sample. Pop-out metrology can even recover the features on the object's underside, which is much slower for scanning probe metrology techniques such as atomic force microscopy (AFM) because the sample has to be inverted (and registered) to expose the underside to the probe.

Assuming that each TEM sub-image takes 1 s to acquire at an electron flux of $100 \text{ e}^{-2} \text{ Å}^{-2} \text{s}^{-1}$, the acquisition time of the entire Fig. 3a only takes 24 s. The following pop-out reconstruction of the entire field of view can be completed in 75 s on a 48 CPU-core machine.

The process flow chart in Fig. 3b describes the steps and checks to implement pop-out metrology, showing that the entire process can be automated. The pertinent details are elaborated as a checklist in the Methods section.

4.2 Experimental demonstrations

To validate the pop-out principle, we demonstrated its feasibility in three proof-of-concept experiments. In each case, we sought to recover 3D features that are challenging to infer directly from their TEM images, other scanning probe measurements (i.e., AFM, SEM, etc), or tomography (because of the extended substrate). In all three cases, the TEM images were collected with either a brief 2–6 s exposure, or multiple brief exposures without having to rotate the sample. Taken together, these demonstrations justify why we coined our method TEM-based single-shot 3D pop-out metrology.

First, we started with a relatively simple 3D nanochannel that was etched onto one side of an amorphous silicon nitride (a- SiN$_x$) substrate. The leftmost panel of Fig. 4a shows a single-exposure TEM image of this nanochannel from which we reconstructed the 3D profile using the pop-out principle (the middle and rightmost panels of Fig. 4a show the top and bottom of the channel). The linear size of the floating cube (reconstructed voxel size) in this reconstruction corresponds to the size of the image patch used for pop-out reconstruction, hence limiting the transverse $x,y$ resolution to 30 nm. Remarkably, our reconstruction correctly shows that the specimen's bottom surface is flat, and that the nanochannel was etched from the top surface. Incidentally, our reconstruction also shows that the substrate on which the nanochannel was etched has a small tilt ($\sim 5^\circ$).

In our second proof-of-principle demonstration, we wanted to reconstruct more complex 3D features than the first demonstration, again from a single-exposure TEM image. Here, we imaged a $\sim 80$ nm pit that was etched on an amorphous silicon nitride substrate. A TEM image of this pit and the corresponding pop-out reconstruction are shown in Fig. 4b, (resolution 40 nm, limited by patch size used in pop-out). The dark “blob” in the TEM image represents the debris from the etching piled near the rim of the pit, and the irregular etching around the rim causes the lighter “petals”. The cross-section shows that both the debris and the petals are on the top surface where the ion beam used for etching was incident. Importantly, our reconstruction reveals that the hollow region of the pit forms a double-conical
structure: first narrowing as we etch deeper into the pit, then “blowing-out” to a larger width on
the bottom side. SI section 6 shows evidence that this double-conical structure is not an
artefact of pop-out reconstruction. Moreover, this double-conical shape has also been
reported in nano-pore etching on a silicon nitride membrane \(^{31}\). This hidden feature, which is
neither visible from the top surface nor easy to scan with a probe, is hard to measure using
an AFM or SEM.

In the final proof-of-principle demonstration, we show how to retrieve the 3D structure
of a wide field of view of a mostly planar object from only their overlapping TEM images. We
image a quantifoil R2/2 holey carbon grid, Fig. 5a, which is commonly used to mount samples
for cryo-electron microscopy. In particular, we wanted to recover the 3D warp profile of the
carbonaceous region around a hole in this grid. By measuring the fraction of elastically
scattered electrons, we obtain the thickness map in Fig. 5b, which revealed that the rim of the
hole is distinctly thicker than the rest \(^{32}\). The corresponding pop-out reconstructions in Fig. 5c
show that the sample was tilted farther away from the detector going from the left to the right:
\(\sim 110\) nm tilt across the 3.2 \(\mu\)m specimen. We fitted for and subtracted away this linear tilt in
the warp map, and fitted the residual depth map to a quadratic model in Fig. 5d. Remarkably,
we were able to determine that the carbonaceous region warped 30 nm, approximately 2 \(\mu\)m
from the centre of the hole, and the clear signature of two different radii of curvatures in the
sample. The depth maps of different TEM images had slight mismatches in their overlapping
regions owing to unaccounted higher-order aberrations. Hence each depth map was corrected
before stitching the depth maps (see Methods section).

Combining these proof-of-principle experiments, we expect TEM-based single-shot
pop-out metrology to be able to rapidly recover the 3D structure of low-dimensional amorphous
materials without rotating the sample and the usual sample preparation needed for
tomography (e.g., ion milling, microtoming)\(^{33}\). These rapid 3D reconstructions can resolve
nanometer longitudinal and transverse strain dynamics of micron-sized laminae that are
stressed \textit{in operando}, which complements atomic-resolution studies of the same but restricted
to a small region\(^{34}\).

4.3 Dual-layer pop-out metrology

The demonstrations of the pop-out principle have so far been limited to single-layered
specimens. Fig. 6a shows how this principle can be extended to a two-layered sample made
from a single type of amorphous material. We first consider a scheme for estimating the depth
of each layer. Notably, the power spectrum of image patches within this sample exhibits
intensity modulations that resembled those from the two layers separately (Fig. 6b). By
modeling this power spectrum as an incoherent addition of two single-layered models (see
methods section), we found that we can separately estimate the depth of each of the two
layers. A full pop-out reconstruction, however, also requires the thickness information from
both layers, despite measuring only the total fraction of electrons lost by both layers.
Nevertheless, should the thickness of one of the layers be known (or the relative thickness
between both layers), the thickness of the other layer can be deduced.

Fig. 6a shows a proof of concept for a dual-layer pop-out with such a thickness prior. This
sample comprises a 25-nm-thick top silicon nitride membrane (thickness known), and a
lower membrane whose (unknown) thickness linearly increases in one direction from 2–50
nm. Our dual-layer pop-out principle correctly reconstructs the entire structure. The radial
average of the fitted CTF from the additive model and the actual Thon rings from the wedge
and membrane are plotted in Fig. 6b, which validates the incoherent additive modeling
approach for multi-layered structures. Nevertheless, dual-layer pop-out is still considerably
more challenging compared to single-layer pop-out (Fig. S8).
5. Conclusions

We have demonstrated that the 3D density distribution of a sample that is a few hundred nanometers thick and several micrometers across can be recovered to sub-10 nm resolution using only a single energy-filtered TEM image. Using the “pop-out” principle, 3D sample reconstruction is possible without having to rotate the sample (e.g., tomography or laminography) or destroy it (e.g., critical dimension metrology). We detailed this “pop-out” principle, the key imaging parameters that control resolution, and described how to generalize to multi-layer structures.

Considering how TEMs are already routinely used to characterize nanostructures in biology, material science, and semiconductor fabrication, we speculate that with suitable automation, this “pop-out” principle can be useful for fast 3D characterisation of the structural dynamics within a large field of view. The rapid feedback afforded by this “pop-out” technique with little to no sample modification on many existing TEMs makes it suitable as a fast screening tool, which fills an important gap amongst existing nanometer-scale metrology modalities. Furthermore, we speculate this method to be relevant for imaging nanometer features of complex structures commonly found in physical and biological sciences.

6. Methods

The simulations are carried out using a TEM simulator that was developed in the programming language Python. All the simulations are generated for a microscope with energy 200 keV, spherical aberration of 1.2 mm, and a detector with pixel size of 6 µm. The experiments are carried out on a JEOL 2200 TEM equipped with a DE16 direct electron detector and an omega energy filter with a 20 eV window around the zero-loss peak.

6.1 Checklist

To optimize the resolution achieved from the pop-out metrology, the TEM parameters should be calibrated for the specimen. A checklist is provided here for this calibration, and a schematic of the process flow is shown in Fig. 3b.

1. Ensure that the spherical aberration parameter $C_s$ of the TEM is known (see Eq. (1-3)). Otherwise, conduct experiments to fit the spherical aberration value for the TEM.
2. Ensure that the pixel size of the detector is known.
3. Calculate the resolution limit for the specimen thickness from Eq. (S27) and determine the feasible range of magnification/resolution for the specimen given this resolution limit.
4. Calculate the theoretical reconstruction voxel size limit for the specimen from Eq. (S31).
5. Ensure the dose limit for the specimen is known, and fix the total dose exposure for the experiment. We can now calculate the window size for the defocus fit, which should be just large enough to capture the required signal.
6. Before imaging the specimen, capture the electron beam without any specimen at the chosen magnification values for different electron doses. The electron beam might vary for various reasons in a TEM and the electron dose cannot be measured accurately at every pixel due to different gain responses of the detector pixels. This series of images will help remove all the uncertainties if used instead of the total electron dose values.
7. Ensure that the energy filter is applied and choose the proper cutoff threshold for the specimen so that all the inelastically scattered electrons are filtered out.
8. While capturing the TEM images with the specimen, ensure that at least three prominent CTF rings are visible at every part of the TEM image by adjusting the defocus value.
9. Once we are set with the magnification, total dosage and the defocus value, start capturing TEM images. For large specimens, capture a series of TEM images by lateral scanning. These TEM images can be stitched together and used as a single image during the reconstruction.
10. If the thickness of any part of the specimen is known, we can use that to calculate the electron mean free path in the material for the current TEM setting. Otherwise, we have to image a calibration specimen made of the same material with the known thickness.

6.2 Estimating sample depth from defocus parameter in the contrast transfer function

Our method needs to determine the representative centre of the mass plane of scattering volume elements in each patch of pixels of the specimen in its TEM image. According to SI section 1, the depth of this centre of scattering mass plane along the optical z-axis is encoded in spatial frequency domain \( k = (k_x, k_y) \) as the relative defocus \( \Delta f \) parameter of the patch’s Thon rings. This prescription can be derived from the multislice formalism, assuming that higher-order scattering terms are relatively small (see SI section 1).

Assuming that a monochromatic electron plane wave of wavelength \( \lambda \) impinges the sample, the spatial frequency \( k \)-dependent contrast transfer function of its exit wave can be modeled by the semi-empirical model shown here in its azimuthally averaged form \( (k = |k|) \):

\[
|l_{det}(k)| \approx A \ enve(k) |CTF(k)| + \text{noise}(k),
\]

\[
CTF(k) = w_1 \sin(\chi) - w_2 \cos(\chi),
\]

\[
\chi = \pi \lambda k^2 (0.5 \lambda^2 k^2 C_s - \overline{\Delta f}) + \Delta \varphi,
\]

\[
\text{env}(k) = e^{-(b_1 + b_2 k^2 + b_3 k^4)},
\]

\[
\text{noise}(k) = n_4 e^{-n_1 \sqrt{|k|}} + n_5 e^{-n_2 k} + n_6 e^{-n_3 k^2}.
\]

Above \( A \) is an overall amplitude of elastically scattered electrons; the \( \text{env} \) term models the envelope caused by the relative thickness of the sample and the spatial, temporal coherence of the electron wave (see SI section 1). The \( \text{noise} \) term models the background noise unrelated to the specimen; \( \text{CTF} \) is the contrast transfer function, \( \chi \) is the aberration function that depends on spherical aberration \( C_s \), defocus \( \overline{\Delta f} \) (\( \overline{\Delta f} = \Delta f + T/2 \)), specimen thickness \( T \), electron wavelength \( \lambda \) and any known overall phase shifts \( \Delta \varphi \) (e.g., which might be caused by a phase plate). The values for the constants \( w_1 \) and \( w_2 \) are obtained from the amplitude contrast ratio \( Q \), which is related to the ratio between the real and imaginary parts of the scattering potential \( \epsilon \) (see SI section 1).

To account for astigmatism, the defocus \( \Delta f \) term in the aberration function can be modified as:

\[
\overline{\Delta f} \rightarrow [\overline{\Delta f}_0 + \overline{\Delta f}_{\text{ast}} \cos(2(\theta_k - \theta_{\text{ast}}))],
\]

where \( \overline{\Delta f}_0 \) and \( \overline{\Delta f}_{\text{ast}} \) are the average defocus value and effective astigmatism which is half of the difference between defocuses in major and minor axes, \( \theta_{\text{ast}} \) is the angle between the major axis and \( x \)-axis, and \( \theta_k \) is the angle between the scattering vector and \( x \)-axis \( 24-27 \). We compare power spectrums from a simulated and an experiment micrograph with their corresponding fits which accounted for astigmatism in Fig. S2.

Overall, from our derivation in SI section 1, we see that \( \overline{\Delta f}_0 \) in Eq. (6) is the relative defocus, or depth, of the sample’s centre of scattering mass from the focal plane \( (\overline{\Delta f}_0 = \Delta f_0 + T/2) \).
From Fig. 2a and Fig. S1, we can see that the attenuation of CTF by a “sinc-like” envelope suppresses the Thon rings, which in turn reduces the signal for depth estimation. In Fig. S3, we show that this “sinc-like” function can be approximated with a Gaussian envelope (Eq. (4)).

6.3 Thickness estimation

The thickness map of the specimen is determined from the absorption contrast i.e., from the ratio of elastically scattered electrons.

\[ T(x,y) = \ell_{\text{mfp}} \ln\left(\frac{I_0(x,y)}{I_t(x,y)}\right), \]  

(7)

where \( I_0(x,y) \) is the total electron dose, \( I_t(x,y) \) is the transmitted unabsorbed electrons detected in a BF-TEM image, and \( \ell_{\text{mfp}} \) is the inelastic electron mean free path of the specimen. In Eq. (7), it is assumed that all the inelastically scattered electrons are removed before reaching the detector. Inelastically scattered electrons form an overall background that does not directly contribute depth or thickness information about the sample. Hence, an energy filter should be applied to filter the inelastically scattered electrons; otherwise, an error will be introduced in the thickness map. We compare the experimental results for TEM without and with energy filtering in Figs. S5 and S6 vs. Fig. 5, respectively.

Since \( \ell_{\text{mfp}} \) varies with both material and imaging conditions, it is challenging to calculate \( \ell_{\text{mfp}} \) theoretically for any specimen. Hence, we should determine \( \ell_{\text{mfp}} \) from a BF-TEM image of a calibrated specimen with the same material with the same imaging conditions. Since we knew the thickness of a particular region in the specimen (e.g., the substrate thickness), we were able to use Eq. (7) to determine the value of \( \ell_{\text{mfp}} \). To minimize spurious thickness changes due to spatial variations in beam intensity and detector response, a reference TEM image without any specimen, \( I_0(x,y) \), was captured at the same imaging conditions as \( I_t(x,y) \).

6.4 Depth estimation for multi-layer samples

For multi-layered specimens, the thickness information cannot be readily resolved for each layer. However, the additive CTF model in Eq. (8) shows that we can infer the depth information of each layer if both layers are not too thick.

\[ |I_{\text{det}}(k)| \approx A_1 \text{env}_1(k) |\text{CTF}_1(k)| + A_2 \text{env}_2(k) |\text{CTF}_2(k)| + \text{noise}(k), \]  

(8)

where \( A_n \), CTF \( \text{CTF}_n(k) \), and env \( \text{env}_n(k) \) are amplitude, CTF and envelope functions for the corresponding top \( n=1 \) and bottom \( n=2 \) layers. Here the CTF and envelope terms are related to the terms in Eqs. (2-4). Since this is an additive model, the sum of additive noise terms from each layer can be combined into one term as defined in Eq. (5). This additive model is used to generate the dual-layer structure in Fig. 6a and we compare the radial profile of Thon rings from a dual-layer specimen and the additive model (Eq. (8)) in Fig 6b.

6.5 Running window averages of sample thickness and depth

In principle, the thickness value at every pixel in the image \( xy \) plane can be used to pop out material symmetrically along the \( z \) axis on either side of the centre of the mass value of that pixel (defocus map). Recall that we can only compute the average defocus for each image patch; hence this creates a resolution gap between our estimates for thickness versus depth. In practice, this gap is smaller because we compute a more noise-robust average thickness over a relatively small multi-pixel window. This is done for a more noise-robust estimate of the sample thickness from Eq. (7). Nevertheless, the side length of the patches used for CTF fitting is still larger than those of the windows used to estimate the average
sample thickness. For example, in Fig. 4a, we used a CTF-fitting patch size of 30 nm (300 pixels), while average thickness windows of 7.5 nm (75 pixels). Hence, the former sets the conservative $x_y$ resolution of our pop-out reconstructions.

For similar noise-robustness in our estimates of sample depth, we also computed the average defocus map $\Delta f'(i,j)$ with overlapping windows with a stride length less than the patch size. In Fig. 4a, the stride length of 7.5 nm (75 pixels) for the 30 nm patch size provides three overlapping patches between any two non-overlapping patches. Choosing the stride size similar to the thickness window size resolves the resolution gap issue. Nevertheless, we can further determine the resultant running-window average defocus values for every $(x,y)$ pixel in the image from Eq. (9).

$$
\Delta f_{(x,y)} = \sum_{i,j} \frac{\Delta f'(i,j)V_{(i,j,x,y)}}{\sum_{i,j} V_{(i,j,x,y)}},
$$

where $\Delta f'(i,j)$ is the average defocus calculated from a particular window indexed by $(i,j)$. $V_{(i,j,x,y)}$, the visitation weights, is an array of ones and zeros: array element $V_{(i,j,x,y)}$ takes on the value of one only if pixel $(x,y)$ is visited by a particular window indexed by $(i,j)$.

Apart from the shot noise, the crosstalk between the phase contrast and amplitude contrast affects the thickness and defocus determination. The phase-contrast produces light-dark fringes, and these fringes are more prominent near the sharp edges of the specimen. Thus, the number of electrons in the image pixels near these edges does not correspond to the material thickness. Similarly, the defocus map would be affected near such edges in the specimen as the defocusses within the window varies sharply. As a workaround, a standard deviation error for the defocus parameter in the CTF fitting can be used to determine which defocus values are erroneous (standard deviation error above a certain threshold) and are discarded from the visitation weights $V_{(i,j,k,l)}$. We can also set a confidence parameter that registers only the defocus values for pixels with more than a certain number of visitations. If there are gaps in the final defocus map due to high error values or fewer visitations, the nearest neighbour interpolation is used to fill the gaps in the final defocus map $\Delta f_{(x,y)}$.

With defocus map $\Delta f_{(x,y)}$ and thickness map $T_{(x,y)}$ at reconstruction resolution, a 3D model is obtained using Eq. (10). To obtain an isometric reconstruction, either the defocus and thickness values should be scaled to match their $x_y$ reconstruction voxel size, or the reconstruction voxel size should be set to match the units of their values ($z$ axis). Otherwise, the reconstruction would be anisometric.

$$
P_{(x,y,z)} = \begin{cases} 
1, & \Delta f_{(x,y)} - \frac{T_{(x,y)}}{2} \leq z \leq \Delta f_{(x,y)} + \frac{T_{(x,y)}}{2} \\
0, & \text{else} \end{cases}
$$

6.6 Implementation of pop-out metrology
The linear regression for depth estimates in Eq. (1) is performed using the Levenberg-Marquardt algorithm implemented in the SciPy package. The total computation time required for the pop-out 3D metrology reconstruction is based on the search space of the CTF fitting, TEM image size, window size, and step size (for overlapping window). The parameter search space for the entire TEM image can sometimes be narrowed down after fitting is performed for a single patch from any part of this image. Should the determined astigmatism be negligible in this first patch, we have the option of ignoring this parameter to speed up parameter regressions for other patches. The knowledge of predetermined ranges of defocus values and astigmatism values also accelerates our parameter regressions by searching in a much smaller parameter space.

6.7 Corrections for electron beam shape and z-drift
We assume an electron beam is flat and devoid of higher-order aberrations in our algorithm. However, TEMs are not aberration-free, and the beam alignment process demands expertise. A deformed beam due to the higher-order aberrations causes distortion in the determined depth as a function of $x, y$ position on the detector. Moreover, when we scan an extended sample, beam misalignment might introduce a $z$-drift between the scans, which cannot be rectified by mere registration. We encountered the $z$-drift and deformed electron beam problem for our carbon grid reconstructions (Fig. 5). The defocus maps obtained from adjacent scans were not aligned perfectly as any continuous object should.

$$
\delta \Psi_0 = a_{30} x^3 + a_{21} x^2 y + a_{12} xy^2 + a_{03} y^3 + b_{20} x^2 + b_{11} xy + b_{02} y^2 + c_{10} x + c_{01} y + d. \tag{11}
$$

We have used a cubic model in Eq. (11) to fit the beam shape such that all the defocus maps we have from various scans align perfectly. The offset term $d$ from the model corresponds to the $z$-drift, and other coefficients correspond to various optical aberrations. As we have a smooth continuous specimen (carbon grid), we were able to use the scans directly to determine the beam shape. However, in the case of measuring any discontinuous specimen, we suggest calibrating the beam shape for the given microscope setting with any smooth continuous specimen. Hence the electron beam shape estimation is an additional step that is required when we encounter warped electron beam and $z$-drift between scans due to misalignment and higher-order aberrations.

**Author contributions.**
D.B. and N.D.L. conceived the project. D.B., N.D.L., M.B., and U.M. conceptualized the idea. N.D.L. and D.B. derived the math. D.B. wrote the code for the simulation and developed/implemented the algorithm, under N.D.L.’s advice. S.W.C., Z.B., and U.M. provided the samples. D.B., S.W.C., and Z.B. collected the experimental micrographs. D.B. and N.D.L. wrote the manuscript with contributions from all authors.

**Acknowledgements.**
This work was supported by the Singapore National Research Foundation’s Competitive Research Program funding (NRF-CRP16-2015-05), National University of Singapore Early Career Research Grant awarded to N.D.L., and the Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 1 grant (A-0004462-00-00). The authors would like to thank Jian Shi for pointing out the higher-order aberrations causing deviations in the depth determination, the IT administrators at the National University of Singapore Centre for Bio-imaging Sciences for their computing support, and Paul T. Matsudaira for insightful discussions.

**References**
1. Koumoulos, E. P. et al. Metrology and nano-mechanical tests for nano-manufacturing and nano-bio interface: Challenges & future perspectives. Mater. Des. 137, 446–462 (2018).
2. Witold, L., Rasit, T., Ana, P. & Agnieszka, D. Eighth NanoForum Report. (2006).
3. Miele, E., Raj, S., Baraissov, Z., Král, P. & Mirsaidov, U. Dynamics of Templated Assembly of Nanoparticle Filaments within Nanochannels. Adv. Mater. 29, (2017).
4. Holler, M. et al. Three-dimensional imaging of integrated circuits with macro-
nanoscale zoom. *Nature Electronics* **2**, 464–470 (2019).

5. Witte, K. *et al.* From 2D STXM to 3D Imaging: Soft X-ray Laminography of Thin Specimens. *Nano Lett.* **20**, 1305–1314 (2020).

6. Sala, S. *et al.* Ptychographic X-ray computed tomography at a high-brilliance X-ray source. *Optics Express* vol. 27 533 (2019).

7. Conte, T. M. & Gargini, P. A. On the foundation of the new computing industry beyond 2020. *Preliminary IEEE RC-ITRS Report* (2015).

8. Orji, N. G. *et al.* Metrology for the next generation of semiconductor devices. *Nat Electron* **1**, (2018).

9. Dai, G. *et al.* Development of a 3D-AFM for true 3D measurements of nanostructures. *Meas. Sci. Technol.* **22**, 094009 (2011).

10. Askary, F. & Sullivan, N. T. Redefining critical in critical dimension metrology. in *Metrology, Inspection, and Process Control for Microlithography XV* vol. 4344 815–827 (International Society for Optics and Photonics, 2001).

11. Chon, Choi, Kim & Kim. Development of automated contact inspection system using in-line CD SEM. *Nouv. Com. Int. Cathol. Infirm. Assist. Med. Soc.*

12. Xue, B., Yan, Y., Hu, Z. & Zhao, X. Study on effects of scan parameters on the image quality and tip wear in AFM tapping mode. *Scanning* **36**, 263–269 (2014).

13. Zhang, Y. *et al.* Real-time scan speed control of the atomic force microscopy for reducing imaging time based on sample topography. *Micron* **106**, 1–6 (2018).

14. Arslan, I., Yates, T. J. V., Browning, N. D. & Midgley, P. A. Embedded Nanostructures Revealed in Three Dimensions. *Science* **309**, 2195–2198 (2005).

15. Midgley, P. A. & Dunin-Borkowski, R. E. Electron tomography and holography in materials science. *Nat. Mater.* **8**, 271–280 (2009).

16. Van Aert, S., Batenburg, K. J., Rossell, M. D., Erni, R. & Van Tendeloo, G. Three-dimensional atomic imaging of crystalline nanoparticles. *Nature* **470**, 374 (2011).

17. Zhu, C. *et al.* Towards three-dimensional structural determination of amorphous materials at atomic resolution. *Phys. Rev. B Condens. Matter* **88**, 100201 (2013).
18. Ercius, P., Alaidi, O., Rames, M. J. & Ren, G. Electron Tomography: A Three-Dimensional Analytic Tool for Hard and Soft Materials Research. *Adv. Mater.* **27**, 5638–5663 (2015).

19. Hata, S. *et al.* Electron tomography: An imaging method for materials deformation dynamics. *Curr. Opin. Solid State Mater. Sci.* **24**, 100850 (2020).

20. Kwon, O.-H. & Zewail, A. H. 4D electron tomography. *Science* **328**, 1668–1673 (2010).

21. Joseph Kline, R. Multiscale 3D X-ray imaging. *Nature Electronics* **2**, 435–436 (2019).

22. Malis, T., Cheng, S. C. & Egerton, R. F. EELS log-ratio technique for specimen-thickness measurement in the TEM. *J. Electron Microsc. Tech.* **8**, 193–200 (1988).

23. Thon, F. Zur Defokussierungsabhängigkeit des Phasenkontrastes bei der elektronenmikroskopischen Abbildung. *Zeitschrift Naturforschung Teil A* **21**, 476–478 (1966).

24. Zhu, J., Penczek, P. A., Schröder, R. & Frank, J. Three-Dimensional Reconstruction with Contrast Transfer Function Correction from Energy-Filtered Cryoelectron Micrographs: Procedure and Application to the 70SEscherichia coliRibosome. *J. Struct. Biol.* **118**, 197–219 (1997).

25. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).

26. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

27. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).

28. Bonhomme, P. & Beorchia, A. The specimen thickness effect upon the electron microscope image contrast transfer of amorphous objects. *Journal of Physics D: Applied Physics* **16**, 705 (1983).

29. Kirkland, E. J. *Advanced Computing in Electron Microscopy*. (Springer Science & Business Media, 2013).

30. Jungjohann, K. L., Evans, J. E., Aguiar, J. A., Arslan, I. & Browning, N. D. Atomic-scale
imaging and spectroscopy for in situ liquid scanning transmission electron microscopy. *Microsc. Microanal.* 18, 621–627 (2012).

31. Kim, M. J., Wanunu, M., Bell, D. C. & Meller, A. Rapid Fabrication of Uniformly Sized Nanopores and Nanopore Arrays for Parallel DNA Analysis. *Adv. Mater.* 18, 3149–3153 (2006).

32. Taylor, D. W., Kelly, D. F., Cheng, A. & Taylor, K. A. On the freezing and identification of lipid monolayer 2-D arrays for cryoelectron microscopy. *J. Struct. Biol.* 160, 305–312 (2007).

33. Wagner, F. R. *et al.* Preparing samples from whole cells using focused-ion-beam milling for cryo-electron tomography. *Nat. Protoc.* 15, 2041–2070 (2020).

34. Feist, A., Rubiano da Silva, N., Liang, W., Ropers, C. & Schäfer, S. Nanoscale diffractive probing of strain dynamics in ultrafast transmission electron microscopy. *Struct Dyn* 5, 014302 (2018).

35. Ludike, S. J. & Chiu, W. Image restoration in sets of noisy electron micrographs. in *Proceedings IEEE International Symposium on Biomedical Imaging* 745–748 (2002).

36. Jones, E., Oliphant, T., Peterson, P. & Others. SciPy: Open source scientific tools for Python. (2001-).
Fig 1. The reconstruction principle of pop-out 3D metrology. a) Schematic of bright-field transmission electron microscopy. A thinner feature (right half) scatters fewer electrons and forms the brighter right half of the TEM image; it is also placed nearer to the focal plane. Hence its CTF has fewer Thon rings than the thicker feature, as we can see from the Fourier transforms of the patches p1 and p2. b) Applying the pop-out metrology technique to a 2048×2048 pixel simulated BF-TEM image of a 3D model (ground truth). The recovered average longitudinal centre of mass (defocus map) and the sample thickness map shown in the image were used to reconstruct the 3D volume.
Fig 2. Factors that limit in-plane (xy) and out-of-plane (z) resolutions. a) Plots show the first zero-crossing of the envelope for thin and thick amorphous silicon nitride specimens. The thicker specimen has a steeper envelope which limits the resolution; hence it needs finer sampling of the spatial frequencies to fit CTF accurately. b) 3D reconstructions with the case i) optimal defocus-fit patch size (20 nm) for the given thickness (substrate 25 nm and pillar 50 nm), case ii) insufficient patch size (12.8 nm) for the same thickness, and case iii) 20 nm patch size for a thicker specimen (substrate 100 nm and pillar 50 nm). The histograms show the spread in defocus values in each case, i.e., the CTF-fitting precision. The same patch size, which was optimal in case i), is insufficient for a thicker specimen (case iii) as expected. c) Defocus fit error map to understand the influence of dose across various specimen thicknesses on defocus fitting for a given patch size (20 nm); Increasing the dose improves the defocus fit accuracy significantly at lower doses until 100 e Å⁻² exit dose. Increasing the patch size helps to sample the frequencies finer in the Fourier space; hence it improves the accuracy of depth fitting. As the patch size (xy resolution) is sufficiently large (20 nm), the accuracy value (z resolution) stays below the patch size.
Fig 3. A large-scale scanning implementation of pop-out metrology. a) A simulated large scan 1.84 μm × 1.23 μm TEM image and a pop-out 3D metrology reconstruction of the structure computed from the TEM image. All the features (Wel - well, SP - stepped pyramid, Pil - pillars, Ch - Channel, Tr - Transistors, FF - FinFET, Gr - grid, Cub - Cuboid, and Arc - Arch) have been reconstructed to 15 nm resolution. b) A process flow chart to explain the necessary steps and inputs for pop-out metrology.
Fig 4. Experimental validations of pop-out 3D metrology. a) An energy-filtered BF-TEM image (total dose 2000 e Å⁻²) of a specimen with features on one side, i.e., a nano-channel etched on an amorphous SiNₓ membrane. The top and bottom sides of a volumetric reconstruction show that the channel is etched only on the top surface, while the bottom surface remains relatively flat. b) An energy-filtered BF-TEM image (total dose 2500 e Å⁻²) of a specimen with features on either side, i.e., a nano-pit etched all the way on an amorphous SiNₓ membrane; the reconstruction shows that all the labelled rim, petals, and the blob of debris are present on the top surface. Although the substrate was etched from the top, the reconstruction shows that the opening of the nanopit was widened towards the bottom surface.
Fig 5. Experimental proof of concept for large area scanning pop-out 3D metrology. a) A series of 15 TEM images captured around a quantifoil carbon grid hole are stitched together (14500×14500 pixels). b) The thickness map of the carbon grid is shown in a log-scale. c) Pop-out reconstruction of the large area scan; CTF-fitting is carried out on each TEM image, and the defocus maps are manually stitched to compensate for the residual stage drift along all three axes. The heat map shows the overall tilt of the carbon grid, and the left view and side view show the amount of tilt in horizontal and vertical directions. d) When the defocus map is fitted to a quadratic profile, and the linear tilt component is subtracted, the curvature (warp) of the carbon grid is obtained. The carbon grid shows a 30 nm height difference along the rim of the 2 µm diameter hole.
Fig 6. Numerical validation for dual-layer pop-out metrology. a) A dual-layer 3D model is simulated, and the pop-out reconstruction is generated using the prior information, i.e., the thickness of the top layer. The side panels show the grid search fitness values from additive CTF model of corresponding color-coded regions marked in the reconstruction. b) The radial average of Thon rings from a dual-layer specimen (simulated) and the radial average of the sum of CTFs from the defocus values of both layers. Both layers are 25 nm in thickness, and they are 180 nm apart; the defocus applied on the exit wave is 500 nm. Hence the centre of mass of both layers from the image plane is 512.5 nm and 717.5 nm.
1. Derivation of semi-empirical multislice scattering.

In the multislice scheme, a slab of homogeneous scattering material of thickness $T$ is partitioned into $N$ thin slabs along the optical $z$-axis, each of thickness $\Delta z = T/(N - 1)$. The thinness of each slab allows the scattering potential of its constituent atoms to be projected to a single infinitesimally thin two-dimensional (2D) slice. Such a projection approximation effectively turns $N$ slabs into $N$ 2D slices.

The multislice scheme alternately applies two operations: (1) the scattering potential of each slice modifies the electron wavefunction that is incident upon it; (2) then a free-space propagator then propagates this modified wavefunction to the next slice, which in turn becomes the incident wavefunction for this next slice. This alternating operation takes the incident electron wavefunction from the first scattering slab through the final occupied slab. The exiting wavefunction from the final occupied slab is then propagated to the imaging plane, which includes the optical aberrations of the microscope's image-forming lenses.

Below, we will recast the scattering from multiple slices in the previous paragraph into that of a single effective slice. We start by considering the scattering contributions from each of these $N$ slices. In the weak phase approximation, the exit waves of electron plane waves of wavelength $\lambda$ after the first few of these slices (indexed $n = 1, 2, 3, ...$) are respectively

$$
\psi_1(r) = \exp[i \, s_1(r)] \approx 1 + i \, s_1(r), \\
\psi_2(r) \approx [\psi_1(r) \otimes p_1(r)] (1 + i \, s_2(r)),
$$

(S1)
\[
\psi_3(r) \approx [\psi_2(r) \otimes p_2(r)] (1 + i s_3(r)),
\] (S2)

where \(p_n(r)\) is the two-dimensional (2D) kernel function that propagates the wavefront from slice \(n\) to slice \(n + 1\), and \(\otimes\) is the 2D convolution operator; the scattering potential distribution of the \(n\)th slice is defined by

\[
s_n(r) \equiv v_n(r) - i \mu_n(r) \approx v_n(r)(1 + i \epsilon),
\] (S3)

where \(v_n\) and \(\mu_n\) are the real and imaginary parts of the \(n\)th-slice’s \(z\)-projected scattering potential. Here, we make the approximation that the real and imaginary parts of the scattering potential are related via a multiplicative constant \(\epsilon\).

For sufficiently thin slices, terms of order \(|s_n|^2\) can be ignored, the exit wave after \(N\) slices can be generalized as

\[
\psi_N(r) \approx 1 + \sum_{n=1}^{N} i s_n(r) \otimes p_{N-n}(r),
\] (S4)

where \(p_{N-n}(r)\) is the propagator to advance the exit wave through \(N - n\) slices. The result in Eq. (S4) essentially ignores multiple scattering and only accounts for the fact that the exit wave from farther slices must be propagated over longer distances to “match up” with the exit wave at the final \(N\)-th slice.

We denote \(\psi_{\text{det}}(r)\) as the wavefunction that is incident on the image forming detector, which includes aberrations due to post-sample optical elements (i.e., objective lens, etc.). The 2D Fourier transform of this wavefunction is

\[
\Psi_{\text{det}}(k) \approx \delta(k) + \sum_{n=1}^{N} i \exp[-i \chi(k)] S_n(k) P_{N-n}(k),
\] (S5)

with \(S_n(k)\) as the Fourier transform of \(s_n(x)\); the Fourier transform of the propagator \(p_{N-n}(x)\) is

\[
P_{N-n}(k) \equiv \exp[i (N - n)\theta] \equiv \exp[i(N - n)\pi \lambda k^2 \Delta z], \text{ where } \theta \equiv \pi \lambda k^2 \Delta z
\] (S6)

and the aberration function in the post-sample image-forming lenses as

\[
\chi(k) \equiv \frac{2\pi}{\lambda} \left[ \frac{C_g}{4} \lambda^4 k^4 - \frac{1}{2} (\Delta f) \lambda^2 k^2 \right], \text{ with } k \equiv |k|,
\] (S7)

\(\Delta f\) as the relative defocus of the final \(N\)-th slice from the plane of focus, and \(C_g\) as the spherical aberration parameter of the microscope’s image-forming lenses.

Now, if we defined \(\phi_n(k) = S_n(k) \exp(-i n \theta)\), and \(\tilde{\chi}(k) = \chi(k) - N\theta\), then we can rewrite Eq. (S5) as

\[
\Psi_{\text{det}}(k) \approx \delta(k) + i \exp(-i \tilde{\chi}(k)) \sum_{n=1}^{N} \phi_n(k).
\] (S8)

Hence, the probability of detecting electrons on the detector is (when dropping terms of order \(\phi^2\) or higher because the scattering from each thin slice is small), which is measured as intensities on the detector

\[
I_{\text{det}}(r) \equiv |\Psi_{\text{det}}(r)|^2 \approx 1 + h(r) \otimes \sum_{n=1}^{N} \phi_n(r) - i h^*(r) \otimes \sum_{n=1}^{N} \phi_n(r). \] (S9)
Fourier transforming this intensity gives

\[ I_\text{det}(k) = \delta(k) + i \sum_{n=1}^{N} \left[ \phi_n(k) \exp(-i \chi(k)) - \phi_n^*(-k) \exp(i \chi(k)) \right], \quad (S10) \]

where \( \delta(k) \) is the Dirac delta function. Using the approximation from Eq (S3) into \( \phi_n(k) \) and \( \phi_n^*(-k) \) in Eq. (S10), we obtain,

\[ I_\text{det}(k) = \delta(k) + 2 \sqrt{1 + \epsilon^2} \sum_{n=1}^{N} v_n(k) \sin(\chi + n\theta - \alpha), \quad \text{where } \alpha = \arctan(\epsilon). \quad (S11) \]

As an instructive curiosity, we can make the rather unphysical assumption that all slices are identical (i.e., \( v_n(k) = v(k) \)), although \( v(k) \) itself is random. In this case, we can pull out the \( v(k) \) term from the sum in Eq. (S11), which can be rewritten as

\[ I_\text{det}(k) \approx \delta(k) + 2 \sqrt{1 + \epsilon^2} v(k) \sum_{n=1}^{N} \sin(\chi - \alpha + \left( n - \frac{N}{2} \right) \theta), \quad (S12) \]

where \( \chi = \chi - N\theta/2 \) is the average aberration function as measured from the middle slice (i.e., \( n = N/2 \)). The summation in Eq. (S12) can be approximated as an integral (assuming sufficiently thin slices, \( \Delta z \to 0 \), see schematic at the beginning of this section) to give

\[ I_\text{det}(k) \approx \delta(k) + 2 \sqrt{1 + \epsilon^2} v(k) \int_{-T/2}^{T/2} \sin(\chi - \alpha + z\pi\lambda k^2) \, dz. \quad (S13) \]

Resolving the integral in Eq. (S13) gives

\[ I_\text{det}(k) \approx \delta(k) + 2 \sqrt{1 + \epsilon^2} v(k) \sin(\chi - \alpha) \frac{\sin(\xi/2)}{\xi/2}, \quad \text{where } \xi = \pi\lambda k^2 T. \quad (S14) \]

The resultant power spectra from Eq. (S14), then becomes

\[ |I_\text{det}(k)|^2 = \delta(k) + 2 \left( 1 + \epsilon^2 \right) |Tv(k)|^2 \left( 1 - \cos(2(\chi - \alpha)) \right) \left( \frac{\sin(\xi/2)}{\xi/2} \right)^2. \quad (S15) \]

Critically, the \( \cos(2(\chi - \alpha)) \) term in Eq. (S15) clearly shows how the effective defocus of the entire sample is now centered at the center of scattering mass of the sample (i.e., \( z = T/2 \)) as shown in the schematic at the beginning of this section. This conclusion was first observed by Bonhomme et al.\(^1\)

This unphysical “identical, random slice” assumption leads to the nodes of the squared sinc function in Eq. (S15) to occur at

\[ \frac{\xi}{2} = j \pi : j \in \mathbb{Z}^+, \quad \text{or} \quad \frac{\lambda k^2 T}{2} = 1, 2, 3, ..., \quad (S16) \]

which is the result first obtained by Bonhomme et al.\(^1\) These node positions, however, have been later shown by Tichelaar et al. to be incorrect using tomography.\(^2\)

If instead we assume that the more realistic scenario where the scattering potential of different slices \( v_n(k) \) are random and different, the power spectrum in Eq. (S11) now becomes

\[ |I_\text{det}(k)|^2 = 4 \left( 1 + \epsilon^2 \right) \left( \beta(k) + \gamma(k) \right), \quad (S17) \]
\[ \beta(k) = \sum_{n=1}^{N} |v_n(k)|^2 \sin^2(\bar{\chi} + n\theta - \alpha), \quad (S18) \]

\[ \gamma(k) = \sum_{n=1}^{N} \sum_{m>n}^N 2 \text{Re}[v_n(k)v_m(-k)] \sin(\bar{\chi} + n\theta - \alpha) \sin(\bar{\chi} + m\theta - \alpha). \quad (S19) \]

Fig. S1 shows that the cross multiplication term \( \gamma(k) \) is much smaller compared to the term \( \beta(k) \). To make progress, we approximated \( v_n(k) \) by creating thin slices of random SiN. Fig. S1 shows the one-dimensional angular average of the power spectrum \( \langle |I_{det}(k)|^2 \rangle \). Using such random slices in Eqs. (S17-S19), we see that the “sinc-like” nodes of the angularly averaged power spectrum occur when

\[ \lambda k^2 T = 1, 2, \ldots \in \mathbb{Z}. \quad (S20) \]

The node positions in Eq. (S20) are consistent with those shown in Tichelaar et al., which those authors experimentally validated using tomography. These same node positions were also proposed by McMullan et al., but with less rigor than the mathematical exposition presented in this section.

Importantly, even for the random \( v_n(k) \) case, the effective defocus of the entire sample in \( \bar{\chi} \) is still centered at the center of scattering mass of the sample (i.e., \( z = T/2 \)). This has been verified in the multislice simulations in Fig. S1.

---

**Fig. S1** Comparison of angularly averaged power spectra from the multislice TEM micrograph (black) and from the equation (S17). The red curve from the equation (S17) provides a good estimation of the power spectrum. Even when we ignore the cross-multiplication term \( \gamma(k) \) and plot the \( \beta(k) \) term (blue), it already provides the expected nodes that occur at \( \lambda k^2 T \). We can notice that if we have identical slices (green), the nodes occur at \( \lambda k^2 T/2 \) (dashed lines) instead of \( \lambda k^2 T \). The green curve (identical slices case) is scaled down to match the amplitudes of the other curves.

Fig. S1 shows that \( \gamma(k) \) term is negligible, and CTF undulations follow \( \sin^2(\bar{\chi} - \alpha) \) with a “sinc-like” envelope caused by the specimen thickness. Apart from the specimen thickness, there are a plethora of effects such as the spatial and temporal incoherence, specimen motion, charging effects, beam-induced movement, and stage-drift contribute to the envelope function. Hence, a Gaussian envelope can be used as a cumulative envelope function in the model. The validation for choosing a Gaussian over a sin function for the envelope is provided in section S2.

Since \( \gamma(k) \) term is small and does not modify the undulations and the node positions, we can rewrite Eq. (S17) as
$$|I_{\text{det}}(k)| \approx A \text{env}(k) \sin(\bar{\chi} - \alpha) + \text{noise}(k), \quad (\text{S21})$$

where $A$ is a multiplicative constant proportional to $\sqrt{1 + \epsilon^2}$ and combined multiplicative and additive terms dependent on $k$ are modelled with env$(k)$ and noise$(k)$ (Eqs. (4-5) in the manuscript). The depth information is encoded in the term $\sin(\bar{\chi} - \alpha)$, which corresponds to the CTF function.

$$\sin(\bar{\chi} - \alpha) = \sin(\bar{\chi}) \cos(\alpha) - \sin(\alpha) \cos(\bar{\chi}). \quad (\text{S22})$$

Since $\alpha = \arctan(\epsilon)$, $\cos(\alpha) = \frac{1}{\sqrt{1+\epsilon^2}}$ and $\sin(\alpha) = \frac{\epsilon}{\sqrt{1+\epsilon^2}}$. To be consistent with the literature, the coefficient of cosine term is written as the amplitude contrast ratio $Q$, i.e., $Q = \frac{\epsilon}{\sqrt{1+\epsilon^2}}$ and $\sqrt{1-Q^2} = \frac{1}{\sqrt{1+\epsilon^2}}$, then we can rewrite Eq. (S22) as

$$\sin(\bar{\chi} - \alpha) = \sqrt{1-Q^2} \sin(\bar{\chi}) - Q \cos(\bar{\chi}). \quad (\text{S23})$$

Substituting this in Eq. (S21) gives

$$|I_{\text{det}}(k)| \approx A \text{env}(k) |\text{CTF}(k)| + \text{noise}(k), \quad (\text{S24})$$

$$\text{CTF}(k) = w_1 \sin(\bar{\chi}) - w_2 \cos(\bar{\chi}), \quad (\text{S25})$$

where $w_1 = \sqrt{1-Q^2}$ and $w_2 = Q$.

### 1.1. Resolution limit for a thick sample

The patch size for the CTF fitting (see Eq. S25) defines the $xy$ resolution of the pop-out 3D metrology. Many parameters determine the patch size, including TEM image resolution. Increasing the TEM image resolution increases the reconstruction resolution. However, for thicker samples, “sinc-like” nodes (Fig. S1), as mentioned earlier in section 1, limit our reconstruction resolution. Due to incoherence and noise, it is hard to obtain clear undulations beyond the first node (which occurs at $\lambda k_B^2 T = 1$). Since the resolution and the thickness are inversely proportional in determining the node position, the lower bound of our transverse resolution limit $k_{\text{lim}}$ for our reconstruction is set by the thickest part of the sample ($T_{\text{max}}$), regardless of the defocus parameter $\Delta f$:

$$\lambda k_{\text{lim}}^2 T_{\text{max}} = 1, \quad (\text{S26})$$

$$k_{\text{lim}} = \sqrt{1/\lambda T_{\text{max}}}, \quad (\text{S27})$$

where $T_{\text{max}}$ is the thickness of the thickest region of the specimen along the optical axis. The TEM’s magnification should be chosen correspondingly to achieve an image’s resolution greater than $k_{\text{lim}}$. This CTF fitting should be able to determine the defocus from both the nearest and the farthest points of the specimen. The point nearest to the focal plane, whose defocus parameter we denote as $\Delta f$, will have the fewest number of CTF rings in its image’s Fourier transform. In contrast, the farthest point, which we denote here to have depth $\Delta f + \Delta f_r$ from the focal plane, would have the most number of CTF rings. The aberration parameter $\chi(k_{\text{lim}}, \Delta f)$ (Eq. S7) of the nearest point at $k_{\text{lim}}$ will be:
\[ \Delta f \geq T_{\text{max}} \left( \eta + \frac{c_s \lambda}{2 T_{\text{max}}} \right), \quad (S28) \]

where \( \eta = \tilde{\chi}(k_{\text{lim}}, \Delta f)/\pi \) is the number of CTF rings up to \( k_{\text{lim}} \). To paraphrase, this last equation gives us the minimum defocus \( \Delta f \) needed to guarantee at least \( \eta \) CTF rings up to \( k_{\text{lim}} \).

However, these CTF rings of a patch have to be finely sampled enough to determine the patch’s average depth. This is equivalent to requiring that we satisfy a separate sampling criterion for the farthest point, which comprises \( \eta_r \) CTF rings at \( k_{\text{lim}} \) (i.e., \( \tilde{\chi}(k_{\text{lim}}, \Delta f + \Delta f_r) = \eta_r \)). In other others,

\[ -\eta_r = \frac{c_s \lambda^3 k_{\text{lim}}^2}{2} - \lambda k_{\text{lim}}^2 (\Delta f + \Delta f_r). \quad (S29) \]

The sampling criterion is that we have at least \( \sigma_s \) frequency samples (spaced apart by \( w^{-1} \) for patch sizes of side length \( w_{\text{min}} \) pixels) spanning between the \( \eta^\text{th} \) and \( (\eta_r - 1)^\text{th} \) CTF ring. Because the latter ring occurs at spatial frequency

\[ k_j = \sqrt{\frac{\lambda (\Delta f + \Delta f_r) - \sqrt{\left(\frac{\lambda (\Delta f + \Delta f_r)}{\lambda^3 c_s (\eta_r - 1)} \right)^2 - 2 \lambda^3 c_s (\eta - 1)}}{\lambda^3 c_s}}, \quad (S30) \]

this sampling criterion translates into having at least having patches whose side lengths are

\[ w_{\text{min}} \geq \frac{\sigma_s}{\sqrt{\lambda^4 k_{\text{max}}}} k_j. \quad (S31) \]
2. 2D CTF fitting with astigmatism

Fig. S2 Comparison of CTFs from TEM image and their corresponding fit. a) The radial profiles of the model fit and of the observed CTF rings from a simulated TEM image without any astigmatism. b) A hybrid image comparing the model fitted and the observed CTF rings from a simulated TEM image with astigmatism. c) A hybrid image comparing the model fitted and the observed CTF rings from an experimental TEM image with astigmatism. The radial profile shows that the zero-crossings of the observed CTFs matches those in the theoretical model, which provides validation for the curve-fitting and the model. In experiments, it is difficult to remove astigmatism completely. Hence we introduced astigmatism in a simulated TEM image; the simulation and experimental CTF show heavy astigmatism; nevertheless, the method is robust enough to fit a theoretical CTF in both cases.

3. Gauss envelope model vs. sinc envelope model

Fig. S3 Comparison of models with a Gauss envelope and a sinc envelope functions.
Eq. (S17) provides a theoretical sinc envelope model for fitting the CTF with the influence of specimen thickness. As we calculate the thickness information from the electron loss, we expected a sinc model with known thickness instead of a Gaussian envelope would provide a better fit. However, many things would cause an envelope in the actual spectrum, such as spatial and temporal incoherency, a beam fall-off, a detector point spread function, sample damage, and motion blur. Hence a generic Gaussian-only model in Eq. (4) is used to model these effects as shown in Fig. S3. Additionally, we also compared against a combination model comprising both sinc and Gauss envelopes, as well as the Gaussian-only model. Ultimately, we chose the Gaussian model for simplicity and fit efficiency.

4. Large field of view pop-out metrology for high-speed measurements

In Fig. 3a, a TEM image and its reconstruction are shown to explore the possibility of using pop-out for a large field of view metrology. In Fig. S4a, the ground truth of the specimen with nanostructures is shown. Several 2048 × 2048-pixel TEM images are separately simulated with the multislice formalism and stitched together to obtain a final TEM image of 12288 × 8192 pixels. The reconstruction shows that the method can retrieve every nanoscale feature. Various nanostructures in the model are labelled in the reconstructed model (pop-out 3D metrology). The well structure (Wel) has the smallest longitudinal feature, i.e., the substrate’s thickness inside the rim is 12.5 nm. The reconstruction can determine the thickness and the position accurately, and as we can see the feature is in line with the substrate’s base. The curvature of the rim of the well is smooth and not pixelated. The step pyramid structure (SP) reconstruction shows the method can resolve consecutive steps of size 25 nm. Even the 50 nm radius pillars (Pil) are resolved without pixelated curvatures. The smallest lateral feature is the valley between the channel (Ch) and the cuboids (Cub), which has been resolved accurately. The cuboids are affected by the “chamfered edge” artefacts caused by the sharp edges because the fringes are more prominent due to the drastic difference between the thickness of the cuboid (200 nm) and the substrate (25 nm). The slopes and the flat structures in the transistors (Tr) are resolved accurately; however, the sharp edges and the significant change in thickness caused artefacts around the edges. The fluctuating fins (FF) illustrate the pop-out 3D metrology’s strength better than all other features in the specimen. The thickness map does not show any distinction within the fins as all the fins have the same thickness; however, the defocus map brings out the hidden depth information to pop-out the 3D structure of the fins. AFM scanning technique used for nanometrology could not retrieve the topography of the hidden surface, but our fins reconstruction illustrates that the pop-out metrology can retrieve the complete 3D structure. The grid (Gr) and the arch (Arc) structures show that the method works well with perforated specimens.
5. Experimental results of TEM images without energy filter

The thickness of the specimen is calculated from the electrons absorbed by the specimen. However, imaging without an energy filter would let all the electrons, including the inelastically scattered electrons, reach the detector. These inelastically scattered electrons form a background, which does not contribute to the image contrast. Hence both the calibrated electron inelastic mean free path $\ell_{\text{mfp}}$ and the intensity at the detector $I_{t(x,y)}$ would be erroneous, and we cannot rely on the thickness map determined as shown in Eq. (S32).

$$T'(x,y) = \ell_{\text{mfp}} \ln(I_0/I_{t(x,y)})$$  \hspace{1cm} (S32)

An energy-filtered TEM image would resolve this issue as the inelastically scattered electrons are removed.
5.1. Amorphous silicon nitride nanochannel and nanopit

The TEM images of an amorphous SiN$_x$ nanochannel and a nanopit are images without energy filter and reconstructed with pop-out 3D metrology (Fig. S5). The reconstructions show that the defocus determination is robust enough to pop-out the shape of the channel and the pit. Though the reconstruction is not quantitatively valid as the pit and channel’s depth are inaccurate, the channel and the pit are popped-out only on the substrates’ etched side and not visible on the bottom side of the substrates. The edge artefacts visible on the bottom side of the substrates are within the reconstruction voxel size. However, due to the limitation that we cannot resolve whether there is any material in the centre of the pit or not, using pop-out without an energy filter produces erroneous reconstructions in such cases.

![Fig. S5](image)

**Fig. S5** Experimental demonstration of the pop-out 3D metrology on TEM images without an energy filter. a) TEM image of an amorphous SiN$_x$ (a-SiN$_x$) nanochannel, (total dose of 283 e Å$^{-2}$), and the volume rendering of the 3D structure. b) TEM image of an amorphous SiN$_x$ nanopit, (total dose of 850 e Å$^{-2}$), and the volume rendering of the 3D structure. The flat bottom of the rendered 3D structure nanochannel validates the depth estimation so the proposed method can reconstruct the 3D structures qualitatively even from a single BF-TEM image without energy filtering. However, due to the presence of inelastically scattered electrons in the centre of the nano-pit TEM image, the pop-out could not resolve whether there is any material in the pit or not.

5.2. Polycrystalline pillars on an amorphous silicon nitride substrate

The polycrystalline material comparatively produces either a few Bragg peaks or diffraction rings in the Fourier space. Hence the CTF rings should be fitted for the resolution range lesser than the diffraction ring’s resolution, and the SNR of the CTF rings is also very low. Because of this reduced fit range, polycrystalline materials require a higher electron dose than amorphous material, and the diffraction rings hinder the CTF undulation, which limits the resolution. The TEM image of a couple of nanopillars of height and diameter of 50 nm, the thickness map, the defocus map, and the volume rendering are shown in Fig. S6. The power spectrum from the amorphous substrate shows prominent Thon rings. However, the polycrystalline pillar’s power spectrum shows a halo
diffraction ring with more prominent Bragg peaks. The Thon rings are visible only at the lower frequencies as most of the electrons are sampled near the Bragg peaks. In the CTF fitting process, only the lower frequencies within the diffraction ring are used.

**Fig. S6 Experimental demonstration of pop-out 3D metrology on polycrystalline material.** a) TEM image of a polycrystalline Si nanopillar on an a-SiNₓ membrane (total dose of 3200 e Å⁻²). b) Observed CTFs from the amorphous substrate and polycrystalline pillars. c) The defocus map, which is determined with astigmatism. d) The volume rendering of the reconstructed 3D structure.

The amorphous SiNₓ provided a smooth, noiseless defocus map, but the polycrystalline pillars’ defocus values are noisy due to the aforementioned reasons. We can notice a small tilt in the substrate from the defocus values, which cannot be picked from the TEM image of the thickness map. The reconstruction is not smooth, and the curvature of pillars is pixelated due to poor lateral resolution as the polycrystalline Si requires a larger patch to sample frequencies within the Bragg peak resolution. Although pop-out metrology can be applicable to polycrystalline materials, the current implementation’s resolution is limited by the diffraction contrast.

6. Nano-pit simulation to validate the experiment reconstruction

The nanopit reconstruction shown in the manuscript (Fig. 4b) reveals that the nanopit has a double-cone structure. We have simulated a TEM image of a perfect cylindrical nanopit with a sharp edge and reconstructed it with the pop-out 3D metrology. Both microscope parameters and
specimen properties are matched with the experiment data, and the reconstruction is carried out with the same patch size (40 nm). Fig. S7a shows that the chamfered edge artefact (35 nm) produced by the reconstruction is within the reconstructed voxel size. However, the experiment reconstruction in Fig. S7b shows the slope size (90 nm) is much larger than the reconstructed voxel size. Hence we can validate that the double-cone shape is not an artefact of our algorithm, and is an actual feature of the nanopit.

![Fig. S7 Validation for the double-cone structure resolved by pop-out in a nanopit reconstruction.](image)

**Fig. S7** Validation for the double-cone structure resolved by pop-out in a nanopit reconstruction. a) The reconstruction of the simulated nanopit shows that the size of the chamfered edge artefact (35 nm) from the filtering process is less than the reconstruction voxel size (40 nm). b) The reconstruction of the experimental nanopit in Fig. 4b shows that the size of the slope (90 nm) is larger than the reconstruction voxel size (40 nm).

7. Challenges with complex 3D structures

![Fig. S8 A connected dual-layer specimen reconstruction to demonstrate the challenges in dual-layer pop-out metrology.](image)

**Fig. S8** A connected dual-layer specimen reconstruction to demonstrate the challenges in dual-layer pop-out metrology. a) The ground truth of the specimen. b) A dual-layer reconstruction with a 20 nm patch size and c) a dual-layer reconstruction with 12.8 nm patch size.

As mentioned in the main text, a CTF-fitting patch with more than one defocus (as in a step structure where the patch partly covers two steps) gives an erroneous fit value. This issue can be resolved for a single-layer pop-out by identifying the erroneous fit from the error map and
interpolating the defocus values from the neighbourhood region. However, for a complex specimen, as shown above, the pillars are a single layer that should be fitted with a single CTF model, and the regions around the pillars are dual-layered, which requires the additive CTF model to fit defocus from both the layers. We fitted the entire TEM image with both the single CTF model and additive CTF model, then the error values in the fit are used to pick the single-layered and dual-layered regions. This approach worked for the membrane regions except for the region around the pillars (Fig. S8b). The error values in either model are so high in these regions because the fitting patch partly covers the single-layer thick pillar and partly covers the dual-layer membranes. Reducing the patch size would confine this problem to a small region around the pillar. However, the smaller patch size could not sample frequencies fine enough to fit two CTFs accurately. The thin top membrane has poor reconstruction in Fig. S8c due to the inaccurate dual-layer fit caused by a smaller patch size (12.8 nm).

References

1. Bonhomme, P. & Beorchia, A. The specimen thickness effect upon the electron microscope image contrast transfer of amorphous objects. *Journal of Physics D: Applied Physics* **16**, 705 (1983).
2. Tichelaar, W., Hagen, W. J. H., Gorelik, T. E., Xue, L. & Mahamid, J. TEM bright field imaging of thick specimens: nodes in Thon ring patterns. *Ultramicroscopy*, **216** 113023 (2020).
3. McMullan, G., Vinothkumar, K. R. & Henderson, R. Thon rings from amorphous ice and implications of beam-induced Brownian motion in single particle electron cryo-microscopy. *Ultramicroscopy* **158**, 26–32 (2015).
4. Ludike, S. J. & Chiu, W. Image restoration in sets of noisy electron micrographs. in *Proceedings IEEE International Symposium on Biomedical Imaging*, 745–748 (2002).