Supporting Information for Strain Mapping of Two-Dimensional Heterostructures with Sub-Picometer Precision

Yimo Han[1], Kayla Nguyen[3], Michael Cao[1], Paul Cueva1[1], Saien Xie[1,2], Mark W. Tate[4], Prafull Purohit[4], Sol M. Gruner[4,5,6,7], Jiwoong Park[3], David A. Muller[1,6]*

[1.] School of Applied and Engineering Physics, Cornell University, Ithaca, NY 14853, USA

[2.] Department of Chemistry, Institute for Molecular Engineering, and James Franck Institute, University of Chicago, Chicago, IL 60637, USA

[3.] Chemistry and Chemical Biology Department, Cornell University, Ithaca, NY 14853, USA

[4.] Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY 14853, USA

[5.] Physics Department, Cornell University, Ithaca, NY 14853, USA

[6.] Kavli Institute at Cornell for Nanoscale Science, Ithaca, NY 14853, USA

[7.] Cornell High Energy Synchrotron Source, Cornell University, Ithaca, NY 14853, USA

*Correspondence to: david.a.muller@cornell.edu

Table of Content

1: Experimental conditions

All experiments were performed on an aberration-corrected FEI Titan Themis at 80 kV with ~20 pA beam current. A 1.6-mrad convergence angle was used, leading to a ~3 nm probe size, which is the spatial resolution we used for our experiments in this paper. More details are discussed in the fourth part of SI #2, optimizing convergence angle and spatial resolution.

In addition, an exposure time of 1.86 ms (1 ms acquisition time along with 0.86 ms readout time) was employed when acquiring the EMPAD 4D datasets. The scan size in real space (the number of pixels the beam scan across) can be set from 64×64 to 512×512. The scan size of the data used in this paper was 512×512, followed by binning into 128×128 during the post-experiment data processing for speed and higher signal to noise ratio.

Our data was collected under different magnifications. The full field-of-view ranges from 1.5 μm (such as Fig. 3d and Fig. 4d and e in the main manuscript) to 4 μm (such as Fig. 2a and b, Fig. 3a and c, Fig. 4c), which is indicated in the map's scale bars. We used a 1.5 m camera length on our microscope, giving a ~50 mrad collecting angle on the EMPAD.

2: Center of mass (CoM) accuracy

CoM calculation

The CoM is calculated from the diffraction pattern, $I(\vec{p})$, using the following equation:

$$\langle \vec{p} \rangle = \int \vec{p}\, I(\vec{p}) d\vec{p} \quad (1)$$

where $\vec{p}$ is the momentum in the diffraction space.

Fig. S1 shows the definition of our masks (green circles) that were applied to the diffraction disks and the center disk when we calculated their CoMs. The green dots label the calculated CoMs of all disks. Fig. S1c shows that the diffraction disk span across ~6 pixels in diameter. The mask with a diameter of 12 pixels is aligned to the disk with 1/3-pixel resolution by eye. We attempted aligning the mask as well as possible and, in fact, we achieved that the center of the mask is close to the measured CoM in Fig. S1d. In addition, measuring the centers for the EMPAD 4D data, generally a few gigabytes in size, requires a fast-computational algorithm. The CoM calculation is an *O(n)* algorithm (see code below), where n can usually be only tens of pixels for each diffraction disk. As a conclusion, CoM is a high-efficiency approach for measuring centers for the EMPAD 4D datasets.

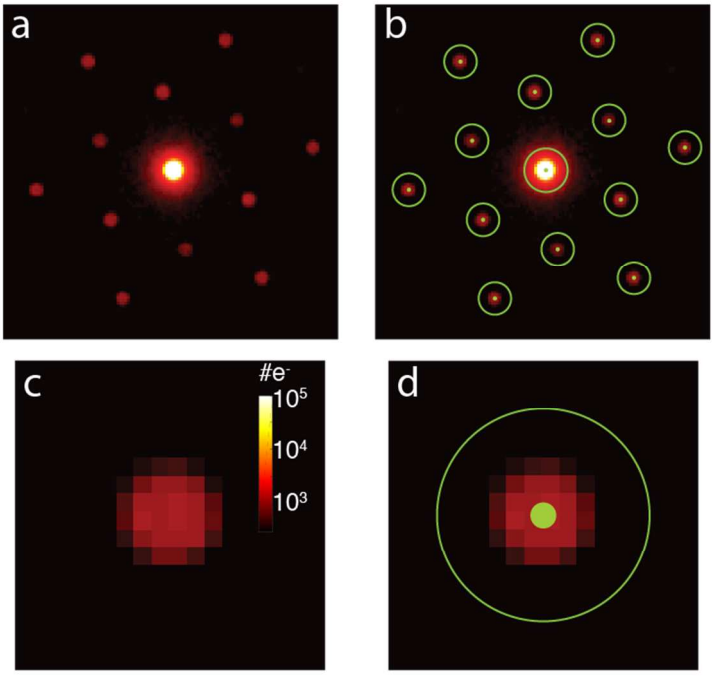

Figure S1 | CoM measurements. a, A diffraction pattern of WS_2 on a 5 nm SiN_x window taken by EMPAD. **b,** The diffraction pattern overlaid by the masks (green circles) and their CoMs (green dots). We manually placed the masks and aligned them to the diffraction spots by eye. **c,d,** A magnified diffraction disk (c) and its overlay with the mask and the CoM (d).

The accuracy of CoM measurements

CoM has the advantage of speed and simplicity compared to more elaborate curve fitting procedures. The CoM provides sufficiently accurate centers of diffraction patterns of 2D materials mainly because of the following two reasons: 1) the rod-like nature of diffraction patterns of 2D materials and 2) the high dynamic range of our EMPAD, which counts all transmitted electrons. The errors for CoM measurements come from the unavoidable Poisson noise from the detector.

Here, we discuss how Poisson noise affects the CoM. For each pixel in the EMPAD, the Poisson noise is proportional to the square root of the number of electrons hitting that pixel. By calculating the error propagation in Equation (1), we achieved that

$$\delta CoM = \sqrt{\frac{\langle p^2 \rangle}{I}} \tag{2}$$

where δCoM is the absolute error of CoM and I is the beam current. This result shows that reducing beam current will increase the error caused by Poisson noise, indicating that high beam current is preferred. However, there is a trade-off between high beam current for more accurate CoM and low beam current to avoid electron beam damage in 2D materials. Experimentally, we were working with a beam current of 20 pA, and used a binning factor of 4 in real space, resulting a total of $4x10^6$ electrons per frame and $\sim 10^4$ electrons in a second-order diffraction disk.

$\langle p^2 \rangle$ is the second moment which will be defined in Equation (10). The second moment is a measure of the beam broadness and has a unit of length square in the diffraction space. As a result, Equ. (2) implies a linear growth of the error corresponding to the disk diameter. Fig. S2a shows the diffraction disks with different diameters, which can be measured in the unit of pixels of the detector. We simulated δCoM for different disk diameters by averaging the errors from 1,000 diffraction patterns with Poisson noise (Fig. S2b). The results indicate the absolute error (δCoM) is proportional to the disk diameter for a given dose.

There are two ways to reduce the disk diameters: decreasing the camera length (equivalent to building less pixels in the detector) or spread out the beam on the sample to focus the diffraction patterns in momentum space. We will discuss these two cases separately in the following sections.

Designing number of pixels in the detector

The parameter determines the angular resolution is the percentage error, σ, which is the ratio between the absolute error (δCoM) and the vector length, k:

$$\sigma = \delta CoM / k \tag{3}$$

We note that k, as well as the disk diameter, changes correspondingly as we change the pixel size in the detector (or change the camera length to magnify or demagnify the diffraction patterns), as shown in Fig. S2c. In addition, since k is proportional to the disk diameter here, we ignored a constant scaling prefactor to assume k equals to the disk diameter. The percentage errors are plotted in Fig. S2d. The percentage errors are close to a constant with small increment at the smaller disk diameters, especially for the low-dose case. For example, if each diffraction disk only contains 10 electrons (blue curve in Fig.

S2d), we would choose 10-15 pixels for a disk diameter for reasonable angular resolution. For doses larger than 1000 electrons per disk, where we worked at, the optimized disk diameters will be any one larger than 5 pixels. Above that, the errors stay constant, indicating that we do not benefit from designing more pixels in the detector or magnifying the diffraction patterns.

Optimizing convergence angle and spatial resolution

For diffraction disks at a fixed camera length and pixel size (k is fixed), the disk diameter is proportional to the convergence angle θ. Thus, the percentage error becomes:

$$\sigma \propto \frac{\theta}{\sqrt{I}} \tag{4}$$

Smaller convergence angles lead to less percentage errors (as shown in Fig. S2e and f). This means more parallel beam is preferred. The lower bound is a 2x2-pixel-sized diffraction disk, which is similar to a differential phase contrast (DPC) detector.

However, in STEM mode, the spatial resolution is determined by the spot size. For a diffraction-limited probe, there is a trade-off between the spatial resolution and the angular resolution in k-space. Reducing the convergence angle (improving the angular resolution) will decrease the spot size:

$$d_0 = 1.22\frac{\lambda}{\theta} \tag{5}$$

d_0 is the spot size and λ is the wavelength of the electrons. Combining Eq. (4) and (5) gives the relationship:

$$\sigma \propto 1/(d_0 \times \sqrt{I}) \tag{6}$$

where improving the spatial resolution (reducing d_0) will cause a reduction of the mapping precision for the same dose or it requires a square-dependence increment of the current to compensate.

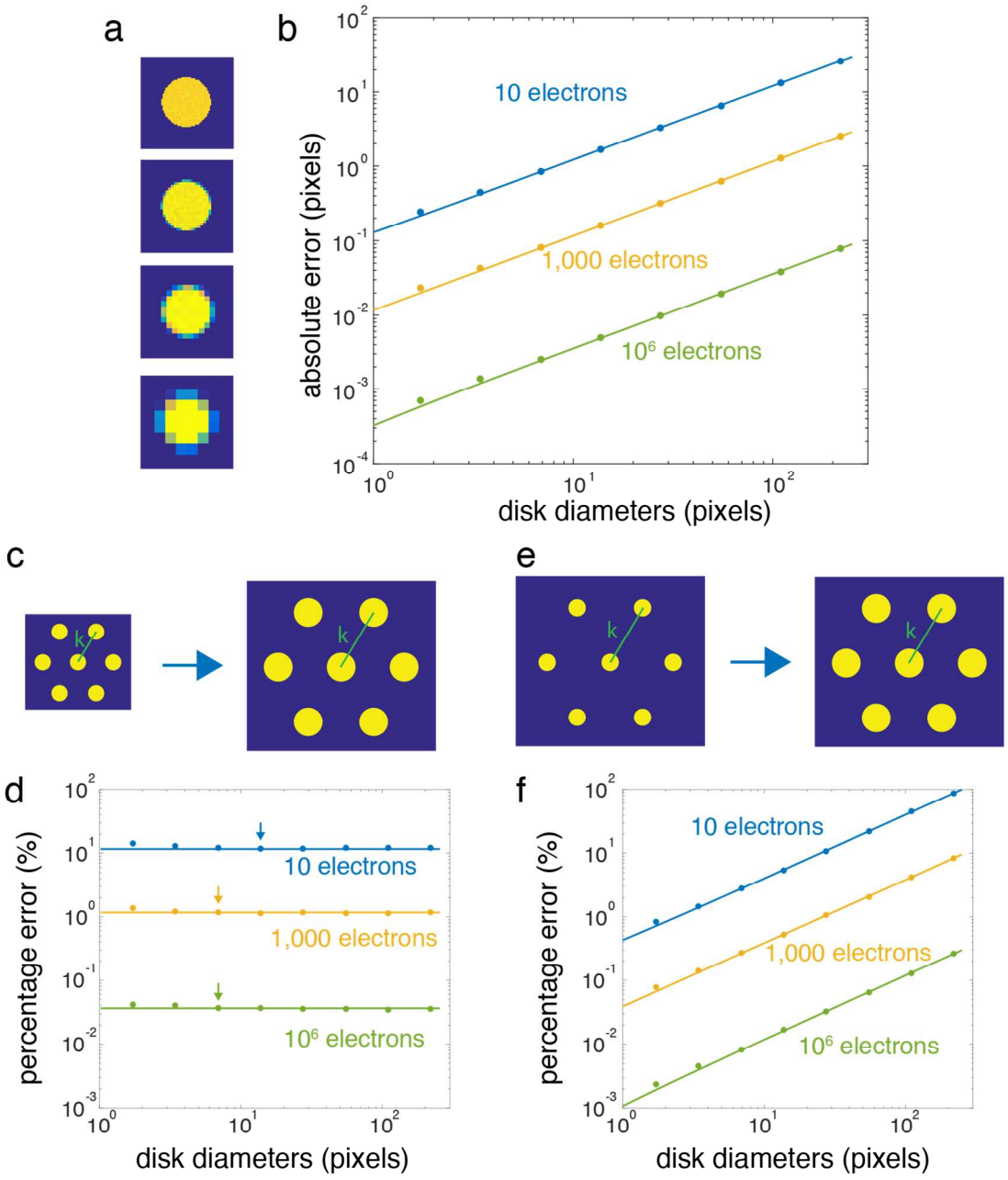

Figure S2 | Accuracy of CoM measurements. a, Simulated diffraction patterns of different diameters (in pixels) with Poisson noise. **b,** Absolute errors of CoM (δCoM) proportional to the disk diameters. **c,** Schematics of diffraction patterns when we increase the number of pixels in the detector (or magnify the diffraction patterns by changing the camera length). **d,** Percentage error plot ($\delta CoM/k$) under the situation described in (c), with arrows indicating the lower bounds of the optimized disk diameters. (We used k =

disk diameter.) **e,** Schematics of diffraction patterns when we change the convergence angle. **f,** Percentage error plot ($\delta CoM/k$) for cases in (e), showing that reducing the convergence angle will reduce the error and improve the angular resolution dramatically. (We used k = 30 pixels.)

3: Mapping lattice constants

To calculate the lattice constants from a single diffraction pattern, we averaged the distances between diffracting beams and the center beam, d_1 to d_6, as shown in Fig. S3a. The averaged lattice constant is:

$$a_{ave} = \frac{6a_0 d_0}{\sum_{i=1}^{6} d_i} \tag{7}$$

where a_0 and d_0 are the calibrated ones from a referenced region. We used flat WS_2 as the reference.

In STEM, EMPAD acquires diffraction patterns at each scan position with 1.86 ms/frame (1 ms exposure time and 0.86 ms readout time) – so a 4D data (x and y in real space and k_x and k_y in momentum space) at 256×256 scan points can be reached in about two minutes. Using the 4D dataset, we can map the lattice constant throughout the entire sample.

Although there are strain and tilt at some regions in the sample, due to the averaging of the six spots in different directions, the strain and tilt effects are negligible. Fig. S3b and c show the schematics of how strain and tilt affect the diffraction pattern. The calculation below describes that the strain and tilt are higher order effects in the lattice constant calculation:

$$a'_{ave} \cong \frac{3a_0 d_0}{d_1(1+\varepsilon')+2d_2\left(1-\frac{\sqrt{3}}{2}\nu\varepsilon'\right)} \tag{8}$$

where ν is the Poisson's ratio (0.25 for WS_2) and ε' is a small uniaxial strain (we used compressive strain here).

$$a''_{ave} \cong \frac{3a_0 d_0}{d_1\left(1+\frac{\theta^2}{2}\right)+2d_2\left(1+\frac{\theta^2}{8}\right)} \tag{9}$$

where θ is the small tilt angle.

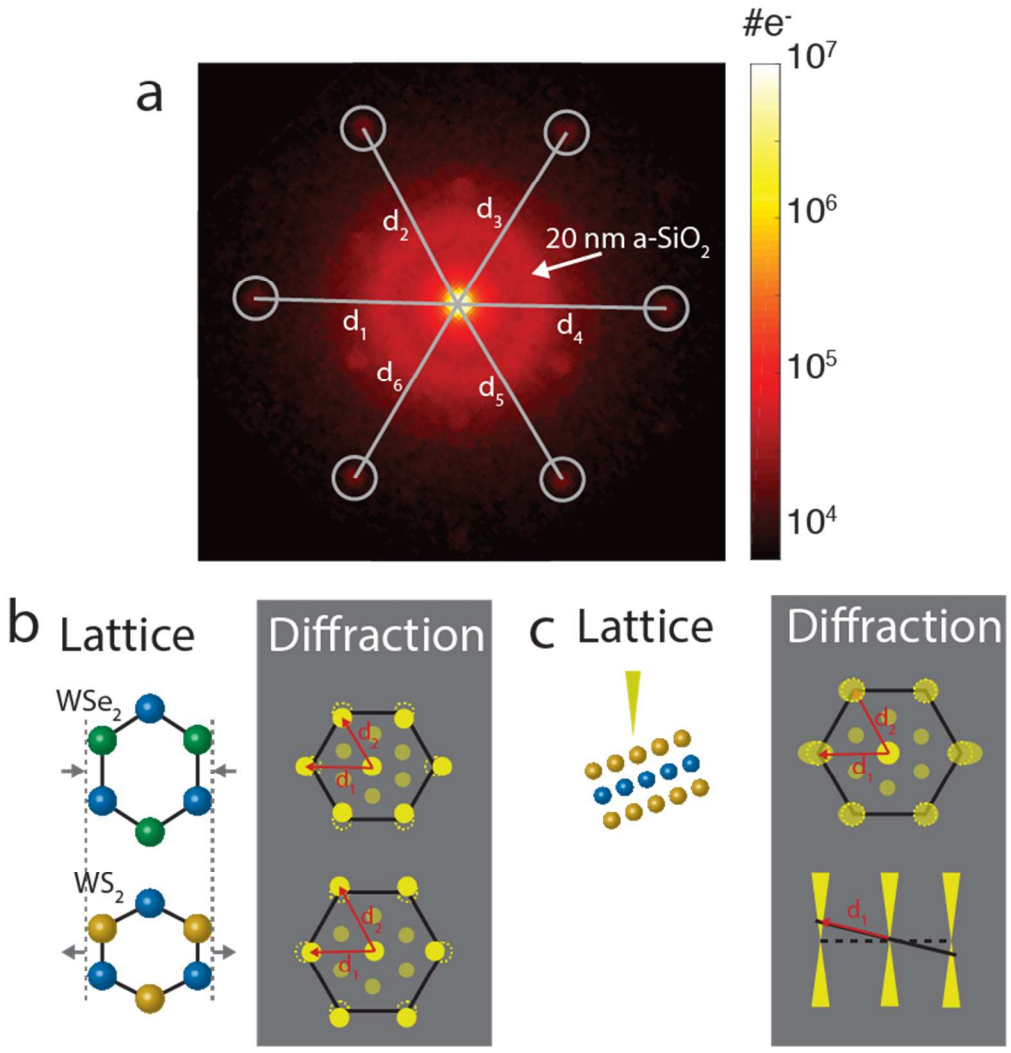

Figure S3 | Lattice constant map. a, Diffraction pattern of WS_2-WSe_2 on 20 nm SiN_x windows. The gray circles are the masks we used to calculate the CoMs. The reciprocal lattice constants were measured, shown as $d_1 - d_6$. **b,** Schematic showing how strain affects the lattice constant measurements. **c,** Schematic showing how small tilt affects the lattice constant measurements. The small strain and tilt are higher order effects for the lattice constant calculation.

4: Mapping strain

To map the strain from the 4D dataset, we calculated the diffraction vectors $\boldsymbol{g}_{i\,(i=1,2)}$ (i.e. the reciprocal lattice vectors) as shown in Fig. S4a. The reference diffraction

vectors $\boldsymbol{g}_i^{ref}$ were set by averaging 200 scan positions (or pixels) in real space where half of them are on WS_2 and the other half are on the WSe_2 region. The choice of these $\boldsymbol{g}_i^{ref}$ is for mapping convenience. Afterwards, we derived the transformation matrix T using $\boldsymbol{g}_i = T\,\boldsymbol{g}_i^{ref}$. T can be polar-decomposed into a rotation matrix R and a strain matrix U, from which the uniaxial strain $\varepsilon_{xx}=1-U_{11}$ and $\varepsilon_{yy}=1-U_{22}$, shear strain $\varepsilon_{xy}=U_{12}$, and rotation $\varepsilon_{rot}=\mathrm{asin}(R_{12})$ can be calculated.

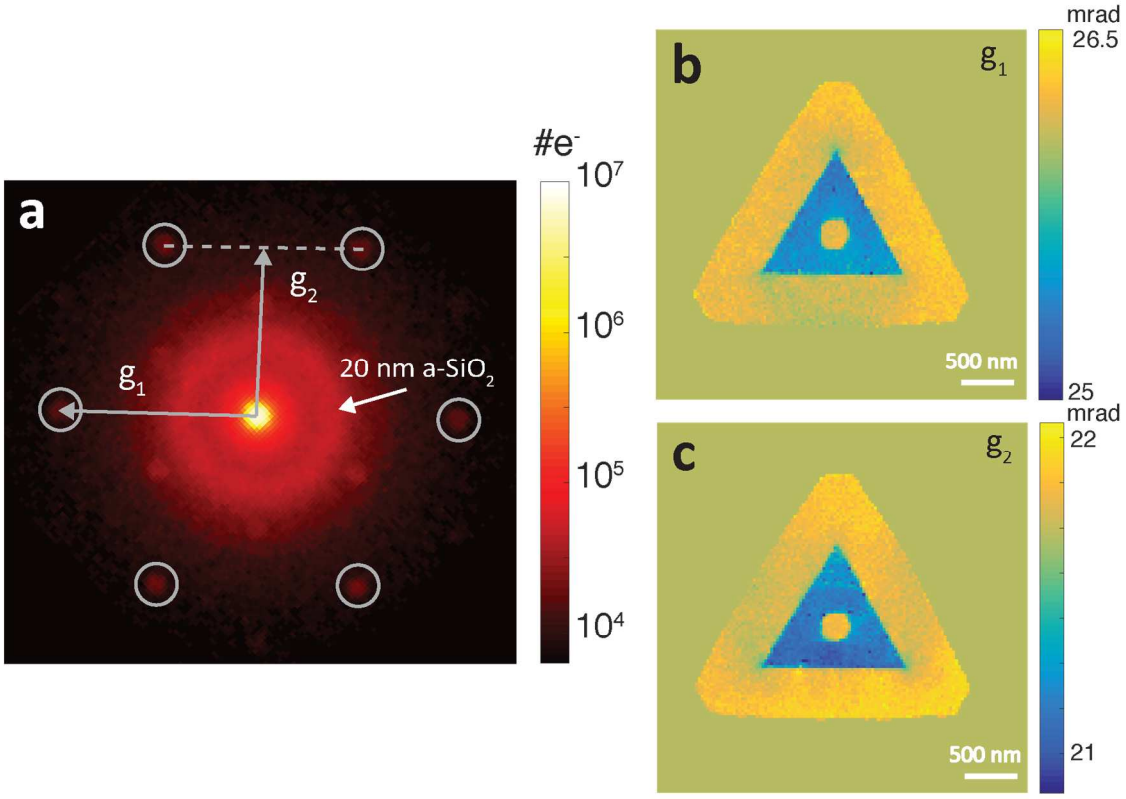

Figure S4 | Maps of diffraction vectors. a, A single diffraction pattern of WSe_2 taken by EMPAD with second order diffraction spots highlighted by the masks. For each spot, we calculated the CoM and achieved the diffraction vectors as labeled by $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$. **b,c,** $|\boldsymbol{g}_1|$ and $|\boldsymbol{g}_2|$ maps over the entire triangle.

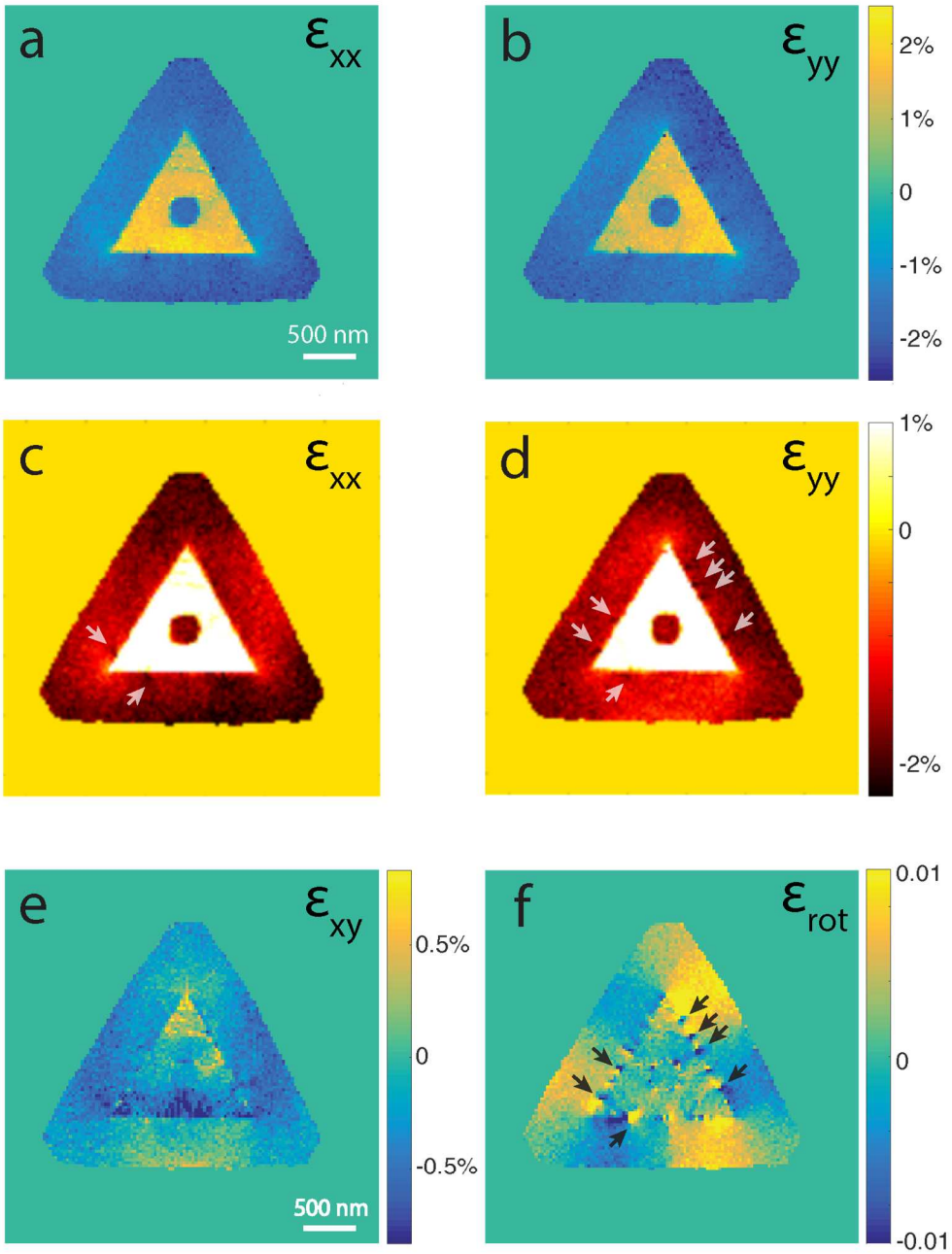

Figure S5 | Strain maps of broad WS_2-WSe_2 junctions. a, x-direction uniaxial strain map. **b,** y-direction uniaxial strain map. **c,d,** the same maps as (**a**) and (**b**) with a different color scale, chosen to highlight the misfit dislocation cores (indicated by arrows). **e,** shear map. **f,** rotation map with the dislocations shown in (**c**) and (**d**) indicated by arrows.

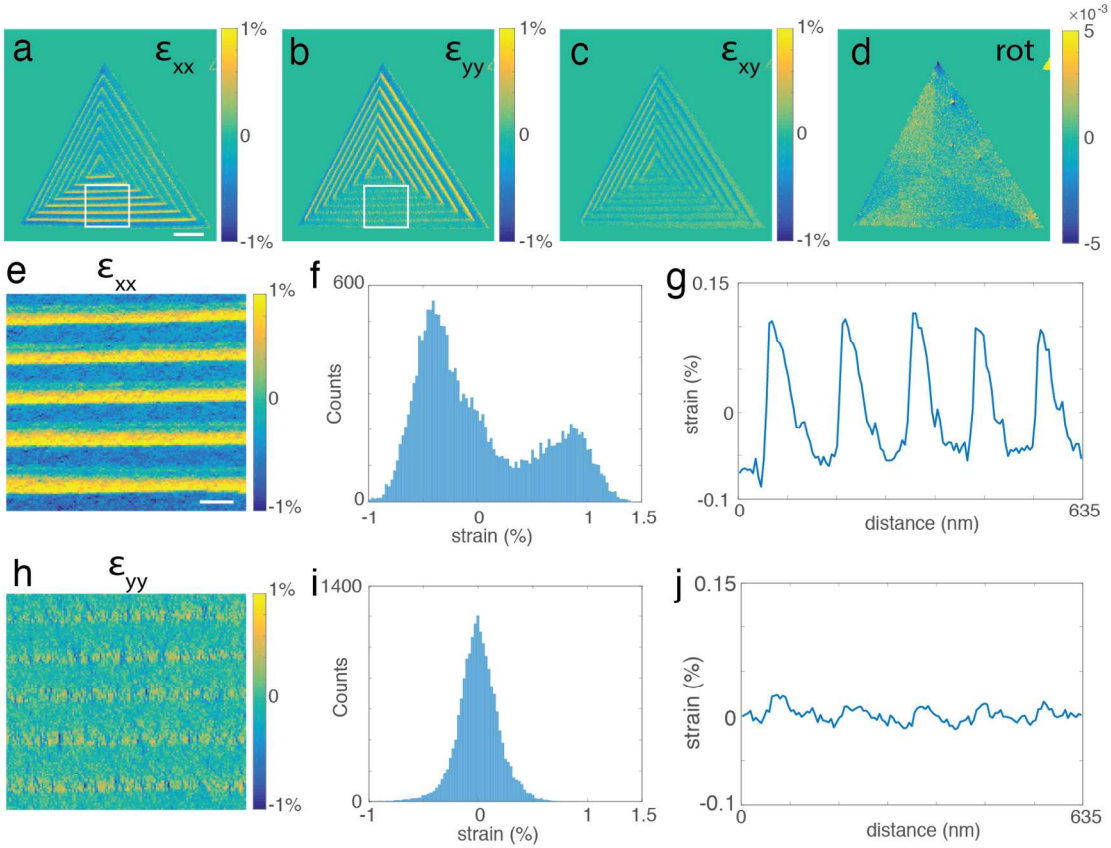

Figure S6 | Strain maps of WS_2-WSe_2 superlattices. a-d, x-direction uniaxial strain (a), y-direction uniaxial strain (b), shear strain (c), and rotation (d) maps. **e,h,** Magnified x-direction uniaxial strain and y-direction uniaxial strain maps from the white boxes in (a) and (b). Scale bar: 500 nm. **f,g,** Histogram and the line profiles of (e). **i,j,** Histogram and line profiles of (h). Scale bar: 100 nm.

5: Mapping ripples using second moments

Second moments are the variance, which describes the broadness of the diffraction spots, for the x-component of momentum transfer:

$$\langle p_x^2 \rangle = \int p_x^2 I(\vec{p}) d\vec{p} \tag{10}$$

We computed the second moments using the same masks we used for CoM calculation. From the calculated second moments for all diffraction spots, the broadness of the diffraction spots is

$$A\ (or\ B, C) = \sqrt{\langle {p_x}^2 \rangle + \langle {p_y}^2 \rangle} \tag{11}$$

where the A, B, and C are the corresponding second moments of the diffraction spots in Fig. 4a. To plot rotational ripple maps, we defined the complex ripple measure:

$$R = \ A + Be^{i2\pi/3} + Ce^{i4\pi/3} \tag{12}$$

The phase of R represents the direction (or orientation) of the ripple line, while the amplitude shows the tilt angle of the ripples.

To quantify the tilt angle, we can plot the projection of R along the tilt direction

$$Re[R] = \ A - 0.5B - 0.5C \tag{13}$$

whose intensity is proportional to the tilt angle, as shown in Fig. 5d (negative numbers represent the tilt perpendicular to the projection angle.

In addition, the amplitude of *R* is proportional to the tilt angle of the film. For a flat film, the amplitude of *R* can be calibrated by tilting the sample through a known set of angles and plotting *R* vs angle. For small curved surfaces, this is more challenging as the sample moves when the sample holder rotates. Here, we calibrated the maximum tilt using an atomic force microscope (AFM) to measure the peak to peak height difference of 3 nm (Fig. S7). Although the lateral resolution of the AFM is lower, the height difference is still accurate, and from this the maximum tilt angle is 8 degrees.

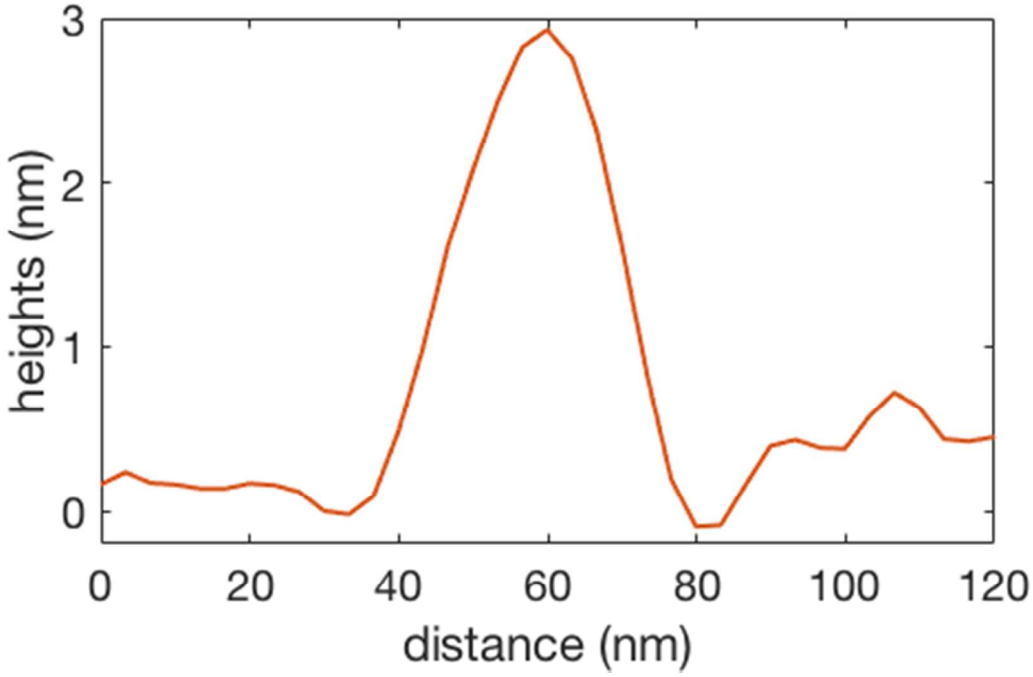

Figure S7 | Atomic force microscopy (AFM) line profile of a ripple. The plot shows a typical ripple topology in the specimen shown in Fig. 5. The ripple has a 3 nm height and roughly ~45 nm width. The maximum tilt angle of the ripple is roughly 8^{o}. Comparing it to Fig. 5d, we can estimate that the maximum intensity in the ripple map corresponds to an 8^{o} tilt.

6: Principle component analysis (PCA)

We performed PCA on the 4D dataset extract the most dominating features in the sample. To look at the diffraction spots only, we blocked the center beams for all diffraction patterns. The details of the PCA approach for multi-dimensional data can be found in reference 25 and 26 in the main manuscript. Fig. S8 shows additional principal components that are not included in Fig. 5 in the main manuscript.

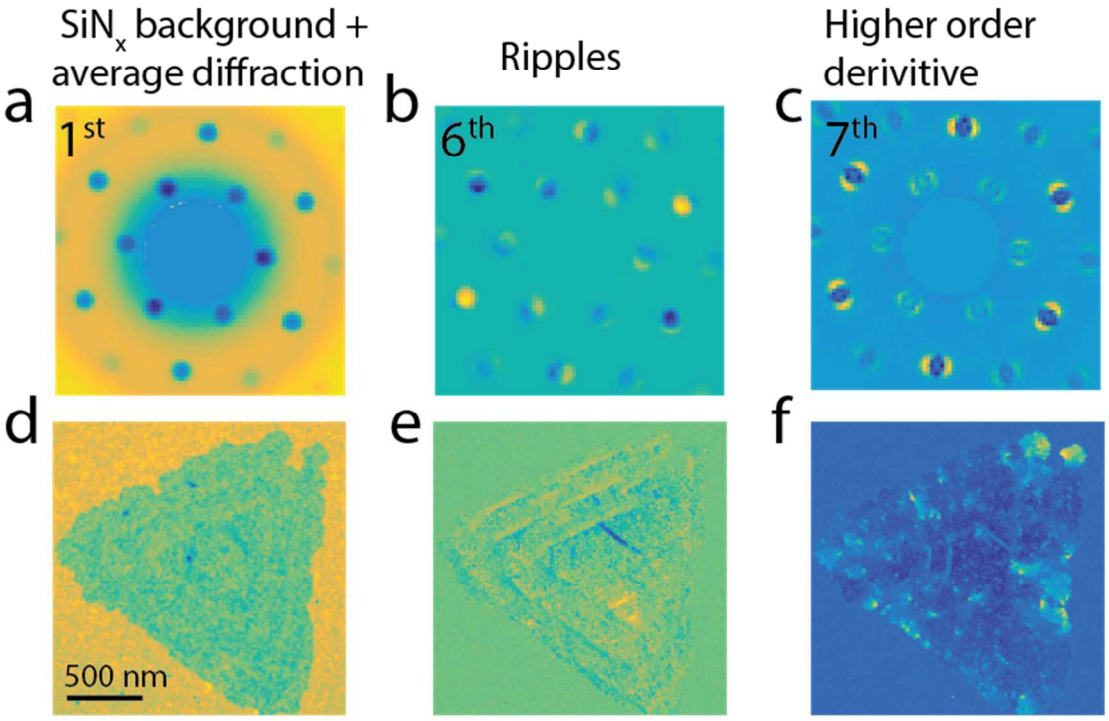

Figure S8 | PCA of 4D dataset. a,d, The first principle component which represents the SiN_x substrate and the averaged diffraction patterns from the lattice. **b,e,** The sixth principle component representing the ripple with a different orientation from the one in Fig. 5i and j. **c,f,** The seventh principle component showing higher order derivatives of the diffraction patterns. For principle components more than seventh order, the real space images start to show less feature as the higher order terms mainly contain noise.