Supplementary material for

2DMat.ChemDX.org: Experimental data platform for 2D materials from synthesis to physical properties

Jin-Hoon Yang[1], Habin Kang[2,3], Taeho Kim[2,3], Heonsu Ahn[2,3], Hyuk Jin Kim[4], Tae Gyu Rhee[4,5], Yeong Gwang Khim[4,5], Byoung Ki Choi[4,6], Moon-Ho Jo[2,3], Hyunju Chang[1], Jonghwan Kim[2,3,*], Young Jun Chang[4,5,7,*], and Yea-Lee Lee[1,*]

Affiliations:

1. Korea Research Institute of Chemical Technology, Daejeon 34114, Republic of Korea

2. Department of Materials Science and Engineering, Pohang University of Science and Technology, Pohang 37673, Republic of Korea

3. Center for van der Waals Quantum Solids, Institute for Basic Science, Pohang 37673, Republic of Korea

4. Department of Physics, University of Seoul, Seoul 02504, Republic of Korea

5. Department of Smart Cities, University of Seoul, Seoul 02504, Republic of Korea

6. Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

7. Department of Intelligent Semiconductor Engineering, University of Seoul, Seoul 02504, Republic of Korea

Corresponding Authors:

Jonghwan Kim (jonghwankim@postech.ac.kr)

Young Jun Chang (yjchang@uos.ac.kr)

Yea-Lee Lee (yealee@krict.re.kr)

In-situ RHEED acquisition during MBE growth and characterization

2D thin films were grown on various substrates, such as graphene-covered SiC (001), Al2O3 (0001), and SrTiO3, using a home-built MBE system in an ultrahigh vacuum (base pressure: 1×10–10 torr). We synthesized the 2D thin films by co-evaporating several sources using an electron-beam evaporator and a Knudsen cell while monitoring the film surface by in-situ RHEED. We captured still RHEED images before, during, and after the growth process and recorded RHEED videos in the ".mp4" format at a frame rate of 1 - 30 frames/sec.

The Raman spectroscopic measurements were performed using a 532 nm excitation laser source at room temperature. Scattered light from the samples was analyzed using a single-grating monochromator with a focal length of 50 cm and was detected by a liquid-nitrogen-cooled charge-coupled-device detector (LabRAM HR Evolution, HORIBA). AFM was performed to investigate the surface morphology under atmospheric conditions after the deposition (XE-100, Park system), and the samples were scanned in the non-contact mode. XPS measurements were carried out to examine the stoichiometry of the films (NEXSA, Thermo Fisher Scientific)

Photoluminescence and Raman mapping database

Raman/PL measurements were performed using a confocal microscope system (WITec alpha 300RS). A 532 nm diode-pumped solid-state laser was directed through a ×100 objective lens with a numerical aperture of 0.9 for focusing. To protect the sample from damage, a laser power of 50 μW and an integration time of 0.1 seconds were applied during the PL and Raman mapping. To acquire the PL mapping data with different resolutions in the same area, several data were generated by varying the number of steps.

Super-resolution of the spectral mapping data

Data structure

The spectral mapping data is stored in a plain text file, wherein each record is structured as a two-dimensional matrix with dimensions represented by ($\boldsymbol{M}$, $\boldsymbol{N} * \boldsymbol{N} + 1$). This format indicates that there are $\boldsymbol{M}$ rows, with each row corresponding to a specific wavelength data point. Within each of these rows, the first column contains the actual value of the wavelength at that point, followed by $\boldsymbol{N} * \boldsymbol{N}$ columns that document the intensity values across a grid of $\boldsymbol{N}$ by $\boldsymbol{N}$ pixels for that particular wavelength.

Data processing

Before utilizing the spectral mapping data in the image super-resolution model, it is crucial to first normalize the data. While the model is designed to process data within a fixed range, such as 0 to 1, the data can exhibit an infinite range of value. This normalization is especially important in adapting to varying intensity levels that may arise from different samples or laser powers. However, this process can unintentionally amplify noise present in the data, which typically appear as exceedingly high intensity values compared to the true spectrum signals. If not addressed, this noise can be falsely identified as peak intensity during normalization, significantly diminishing the spectrum signals from sample.

To mitigate this problem, we use a histogram-based method to remove bright noise from the images, enabling us to find the real highest and lowest intensities from the data. This method involves grouping at least 100k values, structured as a grid, such as (1000, 10, 10), into a finite number of bins (e.g., 2000) to determine the frequency of each value. The noise usually appears as isolated spikes in these bins because they are primarily unique points that stand out. By choosing a threshold (like 3), we can ignore bins with counts below this value, effectively identifying the true minimum and maximum values in the spectrum data. Subsequently, the data values are truncated to the minimum and maximum and normalized to the range from 0 to 1. This process eliminates noise and readies the dataset for processing by the image super-resolution model.

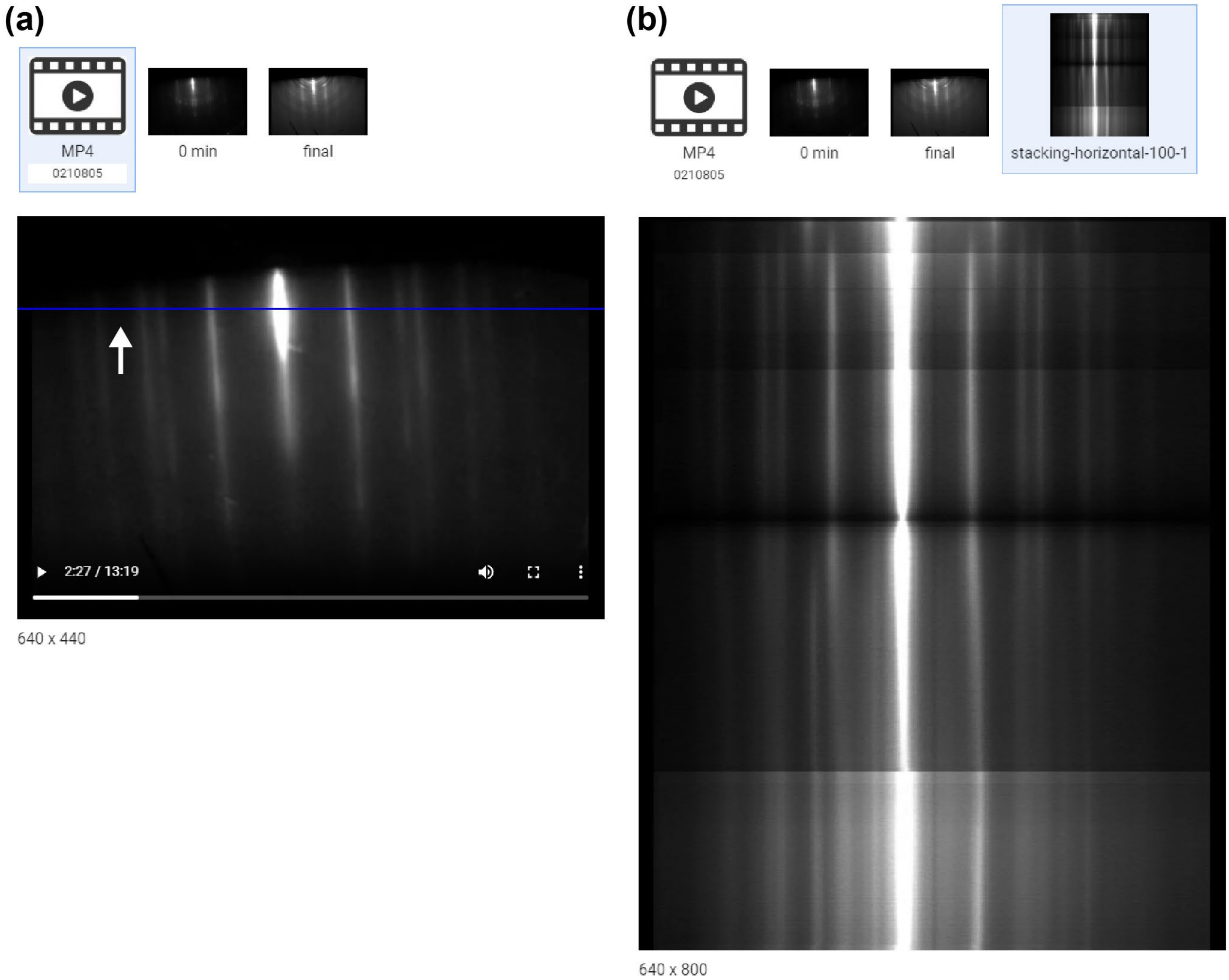

Figure S1. The visualization GUI for the RHEED data in the DATA-UOS group showcases the main section (a) before and (b) after the addition of a guideline to video data along the horizontal axis at position 100 with a 1-pixel width. The resulting stacked image is appended to the list, identified as “stacking-horizontal-100-1”. The dimensions of the image are 640 by 800, derived from the size of the horizontal axis of video and the product of frame count and guideline width. In this instance, spanning 800 frames (13min 20sec) and width of 1.

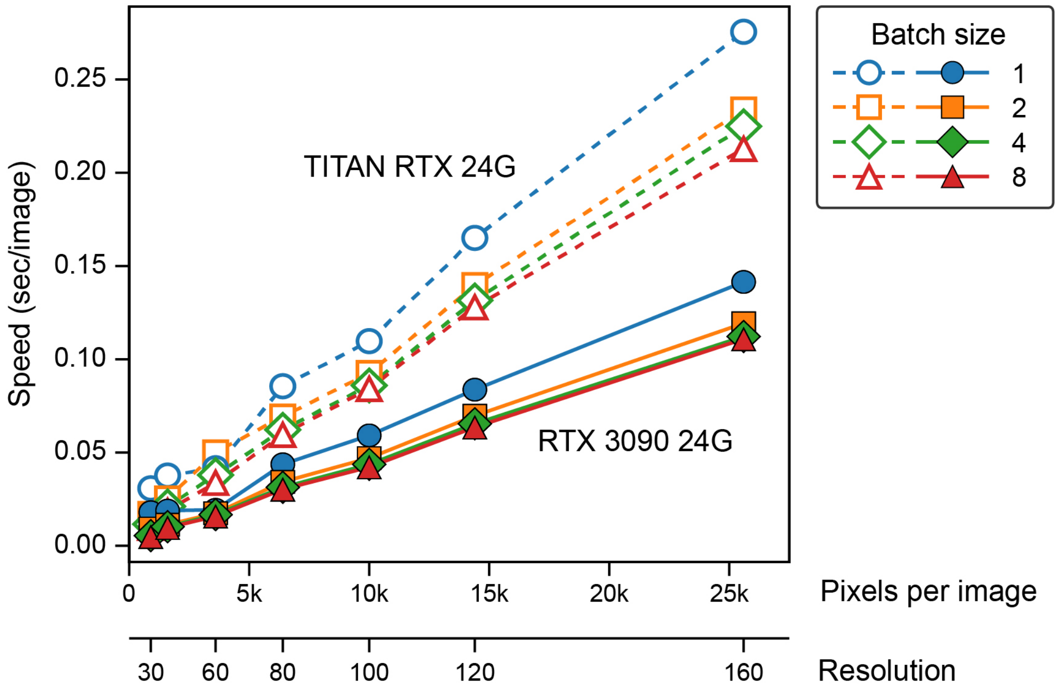

Figure S2. Results of the bench test for the SwinIR super-resolution model using TITAN RTX 24 GB (dashed lines) and RTX 3090 24 GB (solid lines) with different batch sizes of 1, 2, 4, and 8, represented by blue circle, orange square, green diamond, and red triangle, respectively. The required time to enhance the resolution of the image increases linearly with respect to the total pixels of the input image.

Table S1. List of input parameters for DATA-UOS group.

UOS sample name	effusion cell source1
DB name	effusion cell temperature1 (degC)
Date	effusion cell rate1 (A/sec)
Sample	effusion cell source2
Material	effusion cell temperature2 (degC)
Operator name	effusion cell rate2 (A/sec)
Growth rate (1ML duration, min)	Working pressure (torr)
Base pressure (torr)	growth temperature (degC)
substrate	growth duration (min)
Substrate resistance (Ohm)	Estimated thickness (monolayer)
Face of substrate	Post annealing temperature (degC)
Pre-annealing temperature	Post annealing duration (min)
Pre-annealing duration (min)	sample RHEED quality
substrate RHEED (Sub or Graphene)	RHEED video frame per sec (fps)
ebeam-source1	Capping material
e-beam flux1 (nA)	Capping source temperature (degC)
e-beam rate1 (A/sec)	Capping duration (min)
ebeam-source2	Capping thickness (nm)
e-beam flux2 (nA)	sample-z (mm)
e-beam rate2 (A/sec)	Note

Table S2. Input file form for DATA-POSTECH group.

Classification	**Item**		**Details (Value)**	**Unit**	**Remark**
Sample	Producer				
	Date				
	Sample ID				
	TMD composition				
	Stacked structure				
	Polymer Stamp				
	Method				
	Substrate				
	E-beam process				
	Annealing	Temperature		°C	
		Atmosphere			
		Time		hour	
	Laser	Wavelength		nm	
		Power		μW	
		Atmosphere			
		Time		sec	
Measure	Measurer				
	Date				
	Setup				
	Excitation Laser	Wavelength		nm	
		Type			
		Power		μW	
	Mapping condition	Size		μm	Square
		Data point			Square
		Integration time		sec	Per point
	Condition	Voltage		V	
		Temperature		°C	
Fitting	PL	Center wavelength		nm	
		Width		nm	
		Cutoff count			
	Raman	Center frequency		cm^{-1}	
		Width		cm^{-1}	
		Cutoff count			