ADVANCED OPTICAL MATERIALS
www.advopticalmat.de

Excitonic Landscapes in Monolayer Lateral Heterostructures Revealed by Unsupervised Machine Learning

*Maninder Kaur, Nicolas T. Sandino, Jason P. Terry, Mahdi Ghafariasl, and Yohannes Abate**

Two-dimensional (2D) in-plane heterostructures including compositionally graded alloys and lateral heterostructures with defined interfaces display rich optoelectronic properties and offer versatile platforms to explore one-dimensional (1D) interface physics and many-body interaction effects. Graded $Mo_xW_{1-x}S_2$ alloys show smooth spatial variations in composition and strain that continuously tune excitonic emission, while MoS_2–WS_2 lateral heterostructures contain atomically sharp interfaces supporting 1D excitonic phenomena. These single-layer systems combine tunable optical and electronic properties with potential for stable, high-performance optoelectronic devices. Hyperspectral and nano-resolved photoluminescence (PL) imaging enable spatial mapping of optical features along with local variations in composition, strain, and defects, but manual interpretation of such large datasets is slow and subjective. Here, a fast and scalable unsupervised machine-learning (ML) framework is introduced to extract quantitative and interpretable information from hyperspectral PL datasets of graded $Mo_xW_{1-x}S_2$ alloys and MoS_2–WS_2 heterostructures. Combining principal-component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), and density-based spatial clustering of applications with noise (DBSCAN), spectrally distinct domains associated with composition, strain, and defect variations are uncovered. Decomposition of representative spectra reveals multiple emission species, including band-edge excitons and defect-related transitions, demonstrating that ML-driven analysis provides a robust and automated route to interpret rich optical properties of 2D materials.

1. Introduction

Two-dimensional (2D) materials have remarkable electronic, optical, and mechanical properties, including layer-dependent bandgaps, exciton physics, strain-induced band modifications, and defect-bound states. Among these, transition metal dichalcogenides (TMDCs) such as MoS_2 and WS_2 form a particularly important class, widely investigated for their excitonic and defect-related properties.[1–7]

Compositionally graded 2D alloys such as $Mo_xW_{1-x}S_2$ are of great importance because they enable precise control of electronic and optical properties within a single crystal.[8] By gradually varying the ratio of Mo and W across the layer, the alloy allows continuous tuning of the bandgap and excitonic transitions, bridging the properties of pure MoS_2 and WS_2. This built-in compositional gradient also produces smooth variations in strain and defect concentration, giving rise to localized changes in exciton energy and intensity. Such spatial control over the band structure and defect landscape provides an ideal platform for exploring how composition, strain, and vacancies together shape light–matter interactions, exciton dynamics, and charge transfer in 2D semiconductors.[5,6,8]

In-plane (lateral) TMDC heterostructures represent another key platform for exploring nanoscale optical and electronic behavior. Unlike graded alloys, these systems contain atomically sharp covalent interfaces between dissimilar monolayer domains that form through seamless lattice stitching during growth.[9,10] Such 1D junctions give rise to unique interface physics, including exciton diffusion, charge-carrier funneling, and strain relaxation across the boundary.[10,11] The presence of alloyed transition regions at the interface can also stabilize the structure against oxidation and photodegradation, providing long-term optical and chemical stability under illumination.[10,12] These lateral heterostructures thus serve as ideal building blocks for planar monolayer *p*–*n* junctions, excitonic circuits, and high-mobility field-effect transistors, while offering a platform to investigate 1D confinement, defect-assisted recombination, and interface-driven exciton dynamics in 2D semiconductors.[13] Advanced imaging and spectroscopy techniques, including PL mapping, scanning near-field optical microscopy (SNOM), and other scanning probe methods, enable spatially resolved visualization of such phenomena at nanometer scales.[10,13–18] However, these experiments generate enormous

M. Kaur, N. T. Sandino, J. P. Terry, M. Ghafariasl, Y. Abate
Department of Physics and Astronomy
University of Georgia
Athens, GA, USA
E-mail: yohannes.abate@uga.edu

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adom.202503674

DOI: 10.1002/adom.202503674

hyperspectral datasets, making manual analysis and interpretation challenging.

Machine learning (ML) has become a crucial part of modern scientific data analysis, providing tools for dimensionality reduction, clustering, and pattern recognition in complex datasets.[19–33] ML techniques have been successfully used to identify phase transitions,[34,35] discover order parameters,[36] and analyze diverse spectroscopic and imaging data, including scanning tunneling microscopy, near-field optical microscopy, and PL studies of 2D materials.[19,23–28,30–33,37–40] Recent developments demonstrate that ML can classify multiple 2D materials from Raman or PL spectra with high accuracy, underscoring its ability to handle complex experimental data.[29,41–47]

In this work, we apply PCA, t-SNE, DBSCAN clustering, and multi-peak Gaussian fitting to hyperspectral PL datasets of graded $Mo_xW_{1-x}S_2$ monolayers and MoS_2–WS_2 lateral heterostructures. Our analysis reveals spatially distinct optical domains corresponding to composition gradients, strain-induced spectral shifts, and defect-related emission. By quantitatively decomposing individual spectra, we establish a direct connection between unsupervised ML clusters and physical material properties, offering a comprehensive strategy to characterize nanoscale optical heterogeneity in 2D materials.

2. Experimental Section

Graded monolayer $Mo_xW_{1-x}S_2$ alloys were synthesized (as described in ref [8]) using an alkali-metal-halide-assisted chemical vapor deposition (CVD) process that produces a continuous lateral gradient in composition from Mo-rich centers to W-rich edges.[8] The growth utilized MoO_3 and WO_3 powders as precursors, elemental sulfur as the chalcogen source, and NaCl as a growth promoter. The reaction was carried out at 750–800 °C on SiO_2/Si, yielding large-area monolayers with smooth compositional transitions.

In-plane MoS_2–WS_2 heterostructures were grown (as described in Ref [10]) by a sequential CVD process in which the Mo and W oxide precursors were introduced at different stages of the reaction to promote lateral epitaxy.[10] MoS_2 domains nucleated on the SiO_2/Si substrate at ~750 °C under sulfur vapor. Once the MoS_2 growth was complete, the Mo precursor was removed, and WO_3 vapor was introduced to drive edge-selective substitutional growth of WS_2 at the periphery of existing MoS_2 crystals. This process yielded laterally stitched monolayer heterostructures with atomically sharp junctions and narrow alloy transition regions. All optical measurements were performed directly on the as-grown samples without any post-growth transfer to a different substrate. Hyperspectral PL data cubes were acquired under ambient conditions (T = 300K) using a scattering-type near-field optical microscope (s-SNOM) integrated with a broadband excitation source. Each hyperspectral cube contained 10^3 spectral channels per pixel, with a nanoscale spatial resolution. The x and y axes of the cube correspond to the lateral dimensions of the sample surface, while the z-axis represents the photon energy range (1.85–2.0˜eV). Each spectrum was acquired with a 1˜s integration time, resulting in a total acquisition time of approximately 2˜h per image. No additional denoising or normalization was applied beyond experimental preprocessing; all analyses were performed directly on the acquired hyperspectral cube $I(x, y, E)$.

For ML analysis, each hyperspectral cube was reshaped so that individual pixel spectra formed rows in a 2D array of size $(H \cdot W) \times C$, where H and W are the spatial dimensions and C is the number of spectral channels. For the alloy dataset ($75 \times 75 \times 1024$), this produced a matrix of $75^2 \times 1024 = 5625 \times 1024$. For the heterostructure dataset ($48 \times 48 \times 1024$), the resulting matrix was $48^2 \times 1024 = 2304 \times 1024$. This representation enabled PCA to reduce the high-dimensional spectral data and identify the main modes of variation, i.e., principal components (PCs). Prior to PCA, the spectra were mean-centered at each photon-energy channel (without variance scaling), and the component weights (eigen-spectra) shown in Figures 2c and Figure 5c were extracted using standard PCA via singular value decomposition (SVD) of the centered hyperspectral data matrix. The PCA-reduced data were then visualized using t-SNE to project the spectral relationships into two dimensions, allowing intuitive assessment of similarities and differences across the sample. For unsupervised clustering, both the alloy and heterostructure datasets were analyzed using DBSCAN applied to the PCA-reduced spectra. The complete workflow of the ML analysis is illustrated in **Figure 1**. In each case, the data were first reduced to three principal components and then embedded into a 2D t-SNE space using a fixed random seed for reproducibility. DBSCAN clustering was performed with *eps* = 2.5 and *min_samples* = 6 to group pixels according to their spectral similarity in the reduced feature space. For the alloy dataset, representative spectra from each cluster were modeled using one-, two-, or three-component Gaussian fits to decompose the PL emission into individual peaks. Parameters such as the amplitude and center wavelength of the dominant emissive state were extracted from the best-fitting models and spatially mapped to establish correlations between the emission properties and the ML –identified clusters. For the heterostructure dataset, representative pixels from each cluster were directly plotted to visualize raw spectral differences across the dataset.

3. Results and Discussion

3.1. Unsupervised ML Analysis of Graded Alloyed 2D $Mo_xW_{1-x}S_2$

The analysis in this section is based on graded $Mo_xW_{1-x}S_2$ alloyed monolayer experimental data acquired using tip-based PL mapping in an s-SNOM.[8] **Figure 2** provides a full overview of how the PL emission varies across a graded $Mo_xW_{1-x}S_2$ monolayer, both in terms of total intensity and in the underlying spectral features extracted using PCA. It combines a raw intensity map, a summary of spectral variation, and spatial distribution of the top three principal components. Figure 2a shows the total experimental PL intensity map which is acquired by integrating each pixel's spectrum over all photon energies, and is used as a reference for evaluating spatial patterns in PCA-derived components. This map reveals that the monolayer has a compositional gradient, with the Mo-rich region near the center exhibiting lower PL intensity and the alloy/W-rich transition zones at the periphery showing stronger emission, resulting in a tunable bandgap of 1.85 to 1.95˜eV when moving from the center to the edge of the flake.[8] We then perform PCA to reduce the dimensionality of

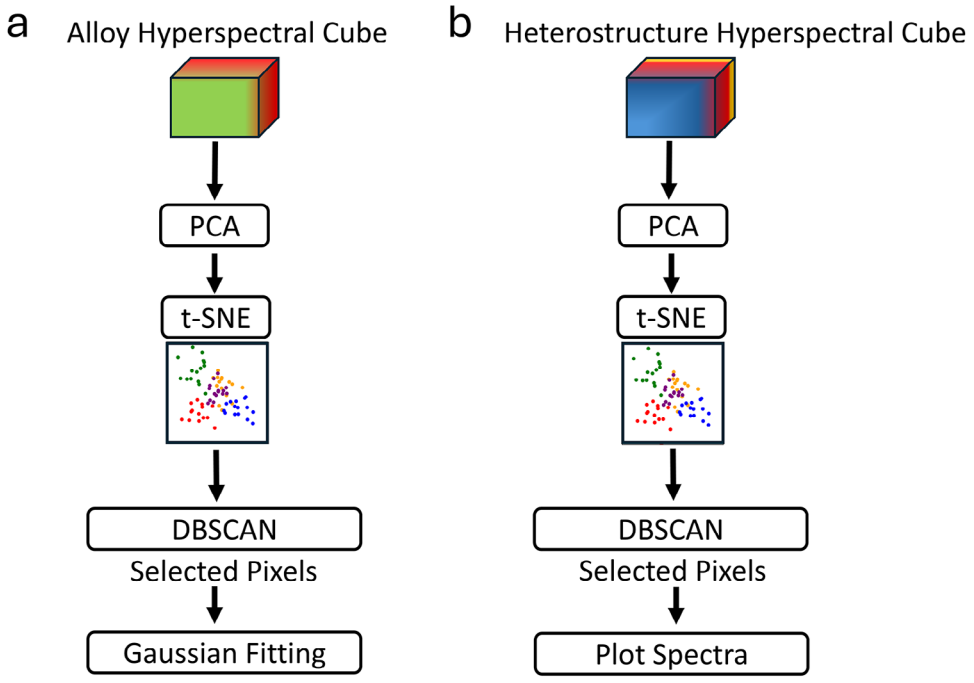

Figure 1. Workflow for unsupervised ML analysis of hyperspectral PL data from alloy and heterostructure monolayers. a) For alloy samples, the hyperspectral cube is reduced using PCA, followed by t-SNE visualization and DBSCAN clustering; selected pixels are then subjected to Gaussian fitting. b) For heterostructure samples, the same dimensionality-reduction and clustering workflow is applied.

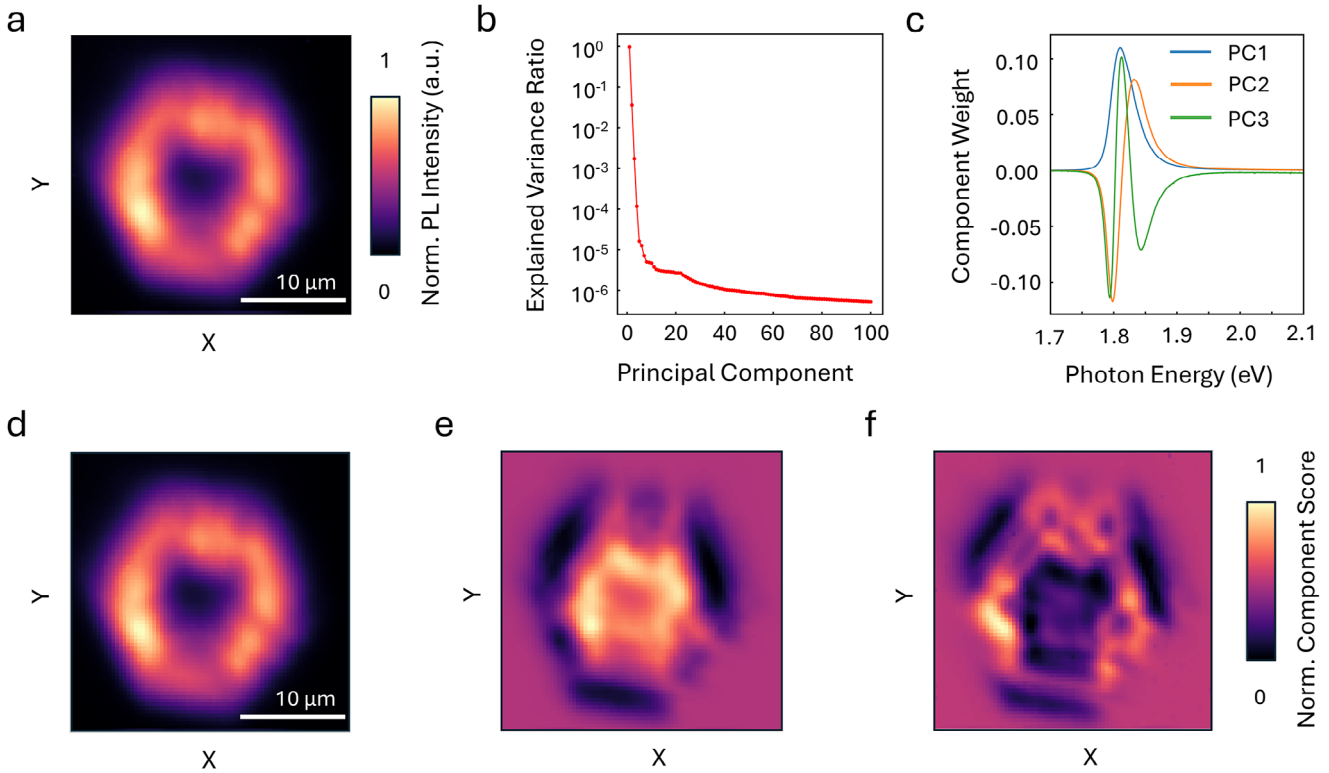

Figure 2. Hyperspectral analysis of a $Mo_xW_{1-x}S_2$ monolayer. a) Total PL intensity map obtained directly from experimental data (without machine learning). b) PCA explained variance plot showing the fraction of total variance captured by each component. c) Eigen-spectra (*PC1–PC3*) corresponding to the main modes of spectral variation. d–f) Spatial projections of (*PC1–PC3*) across the flake.

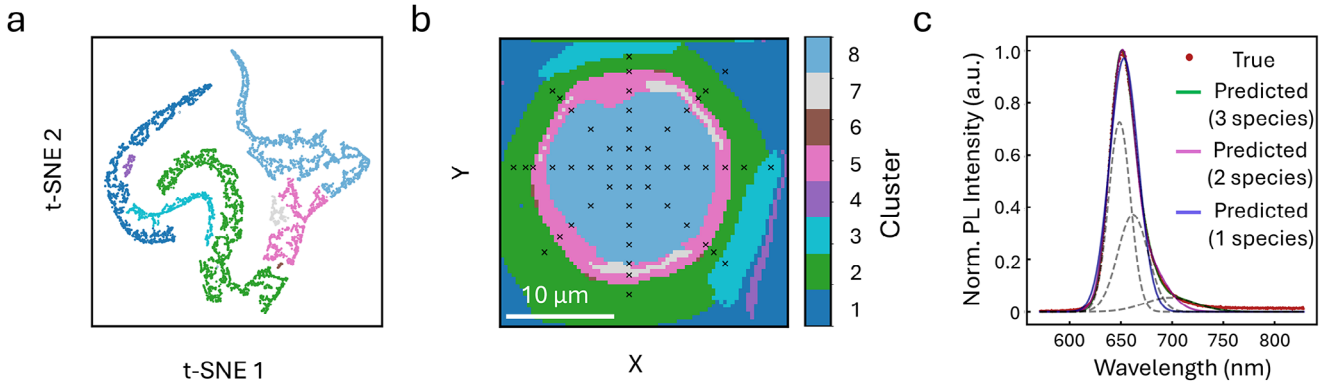

Figure 3. Unsupervised clustering and visualization of hyperspectral PL spectra from a graded $Mo_xW_{1-x}S_2$ monolayer. a) t-SNE projection of the PCA-reduced spectra colored by DBSCAN cluster labels. b) Real-space mapping of the same clusters across the flake. c) Representative Gaussian fits to a selected pixel spectrum (from the 51 marked in panel b), showing one-, two-, and three-component fitting models for comparison.

the spectral dataset and extract the main types of variation across the sample, shown in Figure 2b–f. This analysis helps us break down the complex dataset into just a few patterns that are easier to visualize and understand. Figure 2b shows how much variance each PC captures. The first three PCs together account for more than 99% of the total variance, indicating that they capture nearly all meaningful spectral variations in the dataset. This allows the analysis to be restricted to *PC1*, *PC2*, and *PC3* for identifying major trends in the PL spectra, rather than working with hundreds of individual spectral channels.

Figure 2c shows the spectral weight vectors (also called eigenspectra) for *PC1*, *PC2*, and *PC3*. These describe what each PC is looking for in the spectra. *PC1* has a strong positive peak near 1.82˜eV and resembles the average PL spectrum. *PC2* changes sign around 1.82˜eV, meaning it reacts to whether the peak shifts to lower or higher energy. *PC3* captures variations in the spectral line shape that are distinct from changes in overall intensity (*PC1*) and peak position (*PC2*). *PC3* exhibits a curvature-like profile, indicating sensitivity to linewidth and peak asymmetry rather than to a separate emission feature. Such line-shape variations are a well-established optical signature of inhomogeneous broadening arising from local disorder, including defects and microscopic strain gradients in 2D semiconductors.[8] Accordingly, in the alloy system, *PC3* is interpreted as a proxy for disorder-related spectral modifications. A derivative-based analysis supporting this interpretation is provided in the Supporting Information (Figure S1).

Figure 2d shows where *PC1* is strong or weak across the sample. It looks nearly identical to the total intensity map in Figure 2a. This similarity confirms that *PC1* mainly tracks overall PL intensity, so it can be used as a simpler version of the full intensity map. Figure 2e shows the spatial pattern of *PC2*. There is a clear contrast between center and edges. This contrast matches the spectral weight vector of *PC2*, which flips sign near the peak energy. This means that in the center of the sample, the PL peak is slightly red-shifted (lower energy), while toward the edge it is blue-shifted (higher energy). *PC2* therefore visualizes how the exciton emission spectra change from Mo-rich to W-rich regions.

Figure 2f shows the spatial distribution of *PC3*. Unlike *PC1* and *PC2*, this map highlights finer-scale spatial features, appearing as localized patches across the sample. Consistent with the curvature-like spectral profile of *PC3*, these regions correspond to areas where linewidth and spectral asymmetry vary locally. Such spatially inhomogeneous line-shape variations are commonly associated with disorder effects, including non-uniform strain and defect-related inhomogeneity in alloyed monolayers.[8] While the spectral weight vectors tell us what kind of spectral feature each PC responds to, the spatial maps tell us where those features appear on the sample. Both are important because they connect physical causes to their spatial effects.

Figure 3 aims to uncover discrete groups of pixels in the hyperspectral PL dataset that exhibit similar spectral characteristics. Instead of analyzing gradual variations as in PCA, this approach identifies clusters of similar spectra by applying unsupervised clustering (DBSCAN) on PCA-reduced data, with t-SNE used for nonlinear dimensionality reduction and visualization. Figure 3a shows the result of applying t-SNE to reduce the dimensionality of PCA-transformed data into two dimensions. Each point represents one pixel, and points are colored according to their assigned cluster label. The result of Figure 3a reveals that the spectra naturally group into distinct curved branches in t-SNE space, indicating differences in similarity-based groups of PL intensities or energies. These clusters correspond to physically distinct regions in the sample, identifying W-rich, Mo-rich, and transition zones. Figure 3b remarkably re-projects the same clusters identified in t-SNE space back onto the real-space coordinates of the sample. Each color corresponds to the same cluster as in Figure 3a. The spatial map in Figure 3b shows that these spectral clusters form concentric ring-like regions across the flake, consistent with a compositional gradient. This confirms that the spectral similarity strongly correlates with the physical location, reinforcing the connection between the emission properties and the alloy composition. The topographic spatial map of the graded $Mo_xW_{1-x}S_2$ monolayer sample analyzed here is shown in the atomic force microscopy (AFM) image (see Supporting Information, Figure S3a). This method provides a fast and objective way to visualize how different alloy compositions and emission features are distributed across the sample, pixel by pixel. Unlike manual inspection, which is slow, subjective, and prone to oversight, this automated approach immediately reveals spatial trends that would otherwise be difficult to detect.

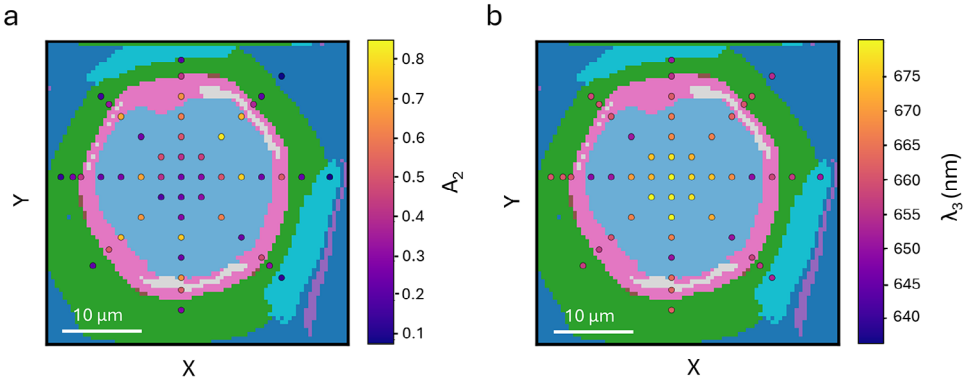

Figure 4. Spatial mapping of dominant emission parameters from three-component Gaussian fits in a graded $Mo_xW_{1-x}S_2$ monolayer. (a) Map of A_2 from the second Gaussian component. (b) Map of λ_3 from the third Gaussian component. Both maps are overlaid on the spectral cluster map from Figure 3b, where the 51 marked pixels (black crosses) are color-coded here according to the fitted amplitude and wavelength values. For completeness, the full model yields six fitted parameters (A_1, A_2, A_3, and λ_1, λ_2, λ_3); however, only the two dominant quantities (A_2 and λ_3) are shown here for brevity.

Figure 3c shows a representative Gaussian model fit to the PL spectrum from one of the 51 selected pixels marked with black crosses in Figure 3b. These spectra were fit using models with one, two, or three Gaussian components, and the fits are overlaid on the experimentally measured spectrum. The results of Figure 3c demonstrate that a single-Gaussian model fails to fit the emission spectrum, whereas a three-Gaussian fit closely matches the measured data. This is attributed to the fact that multiple emitting quasiparticles, such as band-edge excitons and defect-bound excitons, all contribute to the observed PL signal. The distinct spectral energy and linewidth of these three components can be used as a basis for the tentative spectral assignment, which is further validated by their spatial distributions correlation across the monolayer as discussed below.

Figure 4 displays spatial maps of emission characteristics extracted from three-component Gaussian fits to individual PL spectra across the graded $Mo_xW_{1-x}S_2$ flake. Each spectrum from 51 selected pixels was decomposed into three components: species 1 (band-edge exciton), species 2 (alloy disorder), and species 3 (sulfur-vacancy or defect-bound emission).[8] By amplitude, species 1 dominates in 42% of pixels, reflecting strong W-rich band-edge emission, while species 2 dominates in 50%, identifying alloy-disorder states as the most widespread contributors to PL intensity. Species 3 is amplitude-dominant in only 8% of pixels, indicating that defect-related states rarely provide the brightest peak. By wavelength, however, species 3 dominates, appearing as the reddest feature in 86% of pixels, compared to 12% for species 2 and only 2% for species 1. These results show that the optical response is governed primarily by species 2 in terms of intensity and by species 3 in terms of wavelength, while species 1 remains important in W-rich domains but is generally confined to higher-energy emission.

Guided by this analysis, Figure 4 maps the amplitude A_2 of species 2 and the wavelength λ_3 of species 3, which together capture the dominant physical signatures of the flake. In Figure 4a, high A_2 values appear in intermediate, ring-like domains corresponding to Mo–W alloy regions (such as clusters 5 and 7), while both the Mo-rich center (cluster 8) and W-rich rim (cluster 2) show suppressed A_2. This spatial confinement validates the assignment of Species 2 to alloy-induced disorder states. Theoretical models and experimental studies of graded alloys indicate that disorder is maximized in transition regions where the Mo:W ratio approaches 1:1, exactly where the A_2 amplitude is observed to peak, whereas the purer Mo-rich and W-rich regions exhibit reduced disorder scattering.[8,48]

In Figure 4b, λ_3 is strongly red-shifted (~670 nm) in the central, defect-rich Mo-dominated region (cluster 8) and gradually blue-shifts toward the outer rim (cluster 2), characteristic of cleaner W-rich domains. The physical origin of Species 3 as defect-related emission is validated by this spatial correlation with the nucleation center. CVD growth typically yields a higher density of sulfur vacancies and structural defects at the nucleation site, consistent with the strong, low-energy (red-shifted) emission observed here.[8,49–51] Together, these spatial trends confirm that the spectral clusters identified in Figure 3 correspond directly to physically distinct material regions and that the Gaussian-decomposed species represent unique optical transitions tied to composition and defect density across the alloy flake. Unlike traditional PL analysis, where only a few representative spectra are manually fitted, this automated pixel-by-pixel decomposition enables comprehensive, unbiased mapping of emission species over the entire sample, capturing subtle spatial variations that would otherwise remain undetected.

3.2. Unsupervised ML Analysis of In-Plane 2D MoS_2–WS_2 Heterostructure

Following the analysis of graded $Mo_xW_{1-x}S_2$ alloys, where unsupervised clustering and Gaussian fitting were used to assign spectral meaning to spatial domains, we now apply a similar approach to probe the spatial and spectral heterogeneity in the 2D in-plane MoS_2–WS_2 sharp-interface heterostructure sample.[10]

Figure 5 provides a full overview of how the PL emission varies across the 2D in-plane MoS_2–WS_2 heterostructure. Figure 5a Integrated PL intensity map, normalized to highlight emission contrast across the sample. Figure 5b scree plot of explained variance ratio showing that 99% of the total variance is captured by the first three PCs. The corresponding PC weight vectors are plotted in Figure 5c. *PC1* represents the dominant excitonic

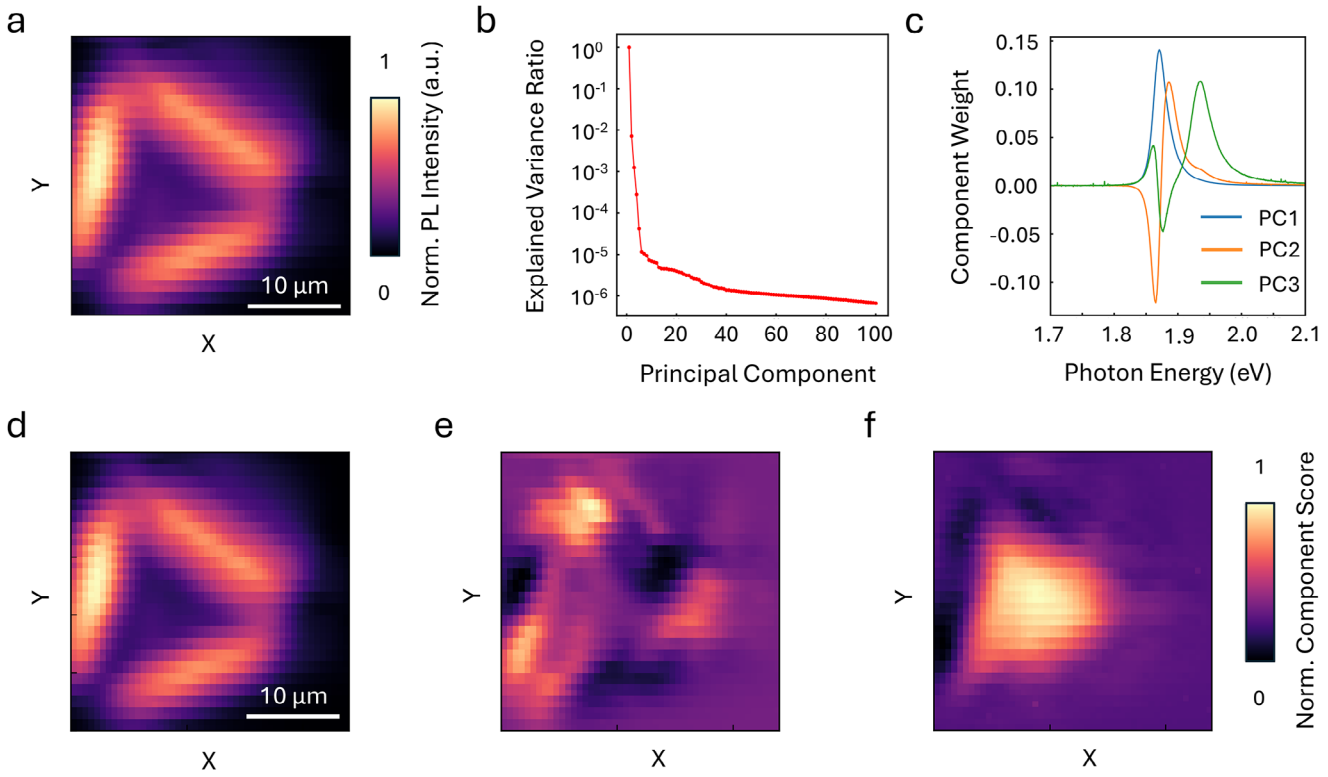

Figure 5. Hyperspectral analysis of the MoS_2–WS_2 heterostructure PL data. a) Total PL intensity map obtained directly from experimental data (without machine learning). b) PCA explained variance plot showing the fraction of total variance captured by each component. c) Eigen-spectra (*PC1–PC3*) corresponding to the main modes of spectral variation. d–f) Spatial projections of *PC1–PC3* across the flake.

emission profile centered near ~1.88–1.90 eV, reflecting the overall PL response of the system. *PC2* shows a derivative-like shape around the same energy, which indicates spectral shifts and intensity modulations arising from local strain fields, compositional gradients, or defects. *PC3* shows a distinct contribution at higher energy (~1.95 eV), consistent with additional spectral emission associated with junction-related excitonic effects at the heterojunction.[10] A derivative-based comparison highlighting the spectrally distinct character of PC3 in the heterostructure is provided in the Supporting Information (Figure S2).

Figure 5d reveals that the spatial distribution of *PC1* resembles the integrated PL intensity map as shown in Figure 5a, indicating that *PC1* captures the global brightness variations across the sample. The spatial map of *PC2*, as shown in Figure 5e, emphasizes more fragmented spatial features, particularly along boundaries and interface-like regions, suggesting *PC2* represents fine-scale spectral modulations arising from strain or interfacial degradation. Figure 5f displays the spatial contribution of *PC3*, which is localized in a triangular central domain of the sample. This localization hints at the presence of a distinct spectral component potentially higher-energy emission originating from the well-defined heterointerface formed during synthesis. The triangular pattern also aligns with expected geometric boundaries of the heterojunction,[10] supporting the hypothesis that *PC3* captures alloying or interfacial PL signatures unique to this structurally confined region. Such an accumulation of PC3 within the internal triangular region is also consistent with previously reported exciton lensing effects driven by strain gradients, as well as with charge-transfer–related excitonic states forming near lateral heterojunctions,[52,53] although no unique microscopic assignment is made here.

Figure 6 presents an unsupervised clustering analysis of the heterostructure hyperspectral dataset to identify spatially distinct emission domains based on spectral similarity. The result of Figure 6a reveals that the spectra naturally organize into six well-separated branches in the t-SNE embedding space, indicating meaningful differences in PL shape, peak position, or intensity. Each cluster reflects a set of spectra that share similar emission characteristics, suggesting that the heterostructure contains multiple optically distinct regions. Figure 6b reprojects the same clusters identified onto the real-space coordinates of the sample. Each color corresponds to the same cluster label shown in the t-SNE map. The resulting spatial segmentation reveals that these clusters are not randomly distributed; instead, they form physically coherent spatial patterns, with some clusters spanning extended interior regions while others trace narrow interfacial or boundary-like features, demonstrating that spectral similarity is linked to the structural and junction-related features of the heterostructure. The topographic spatial map of the MoS_2–WS_2 monolayer sample analyzed here is shown in the atomic force microscopy (AFM) image (see Supporting Information, Figure S3b).

Figure 6c shows representative PL spectra from individual pixels selected from each of the six clusters identified in the MoS_2–WS_2 heterostructure. These examples illustrate the spectral diversity across the sample. Cluster 5 exhibits a strong, broad emission centered near 640~nm, characteristic of emission covered

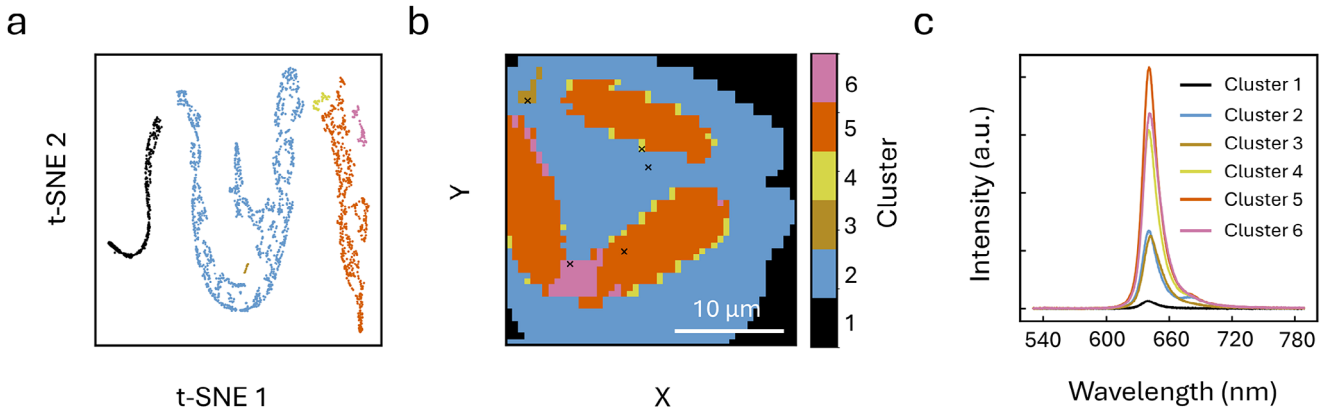

Figure 6. Unsupervised clustering and spectral variation in the MoS_2–WS_2 lateral heterostructure. a) t-SNE projection of spectra colored by DBSCAN cluster labels. b) Real-space mapping of the same clusters onto the sample coordinates using consistent colors; black crosses mark the pixels whose spectra are shown in panel (c). c) Representative PL spectra from the pixels marked in panel (b), corresponding to each cluster.

by pristine and alloy-stabilized WS_2 domains. In contrast, Cluster 1 is nearly quenched, consistent with sulfur-deficient or degraded regions near the MoS_2 side or exposed interfaces. Cluster 6 also exhibits strong WS_2-like emission but with reduced intensity compared to Cluster 5. Spatially, this cluster is localized to the vertices of the WS_2 domains. We note that in lateral heterostructures under near-field excitation, exciton transport and funneling toward the energetically favorable side can give rise to non-local PL responses near the junction.[52,54–56] In this context, Cluster 6 is best interpreted as a transport-affected WS_2-like recombination class, reflecting regions where proximity to the heterointerface modifies the effective recombination landscape. Cluster 4 corresponds to interfacial areas, giving intermediate intensity with broadened line shapes. Because the clustering results are mapped onto real-space coordinates, the spatial width of the interface-specific cluster (Cluster 4) provides a quantitative optical measure of the effective interfacial broadening surrounding the heterojunction. Cluster 3 appears at transitional pockets with weaker PL emission. Cluster 2 is associated with the MoS_2 core but shows an asymmetric profile with a clear shoulder, indicating spectral mixing or disorder-related states. The variation among these representative spectra reflects the underlying optical heterogeneity of the heterostructure and supports the interpretation that the clusters correspond to physically distinct regimes, including MoS_2-rich core regions, WS_2-dominated interior and edge regions, transport-affected transition zones near the junction, and interface-specific recombination regions, rather than assuming strictly local emission behavior.

4. Conclusion

The combination of high-resolution hyperspectral mapping and scalable data analysis applied to alloyed and interface-engineered 2D materials provides a powerful platform to uncover fundamental physics and guide the design of next-generation optoelectronic devices. We present an unsupervised machine learning framework that unifies PCA for dimensionality reduction, t-SNE for visualization, DBSCAN for clustering, and spectral fitting, enabling efficient decomposition of high-dimensional hyperspectral optical datasets into quantitatively interpretable features. Applying our method to complex TMDC monolayers–$Mo_xW_{1-x}S_2$ alloys and MoS_2–WS_2 heterostructures, we uncovered spatially distinct optical domains, including hidden features, and linked them to compositional gradients, interfacial regions, and defect-related emission. The key strength of this approach is its ability to reveal meaningful patterns in complex datasets without requiring prior labeling or manual analysis. Beyond the specific systems studied here, the same workflow can be applied broadly to connect hyperspectral imaging data with underlying material properties, providing a versatile tool for exploring nanoscale heterogeneity in 2D materials and beyond.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work was supported by the Gordon and Betty Moore Foundation, grant DOI 10.37807/GBMF12246 and the Air Force Office of Scientific Research (AFOSR) grant number FA9550-23-1-0375.Partial support for N.T.S. comes from the National Science Foundation (NSF) Grant No. 2152159 (NRT-QuaNTRASE).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

alloys, dimensionality reduction, excitonic emission, heterostructures, hyperspectral photoluminescence, unsupervised machine learning

Received: November 13, 2025
Revised: December 29, 2025
Published online:

[1] K. F. Mak, C. Lee, J. Hone, J. Shan, T. F. Heinz, *Phys. Rev. Lett.* **2010**, *105*, 136805.
[2] J. Hong, Z. Hu, M. Probert, K. Li, D. Lv, X. Yang, L. Gu, N. Mao, Q. Feng, L. Xie, J. Zhang, D. Wu, Z. Zhang, C. Jin, W. Ji, X. Zhang, J. Yuan, Z. Zhang, *Nat. Commun.* **2015**, *6*, 6293.
[3] K. F. Mak, J. Shan, *Nat. Photonics* **2016**, *10*, 216.
[4] G. Wang, A. Chernikov, M. M. Glazov, T. F. Heinz, X. Marie, T. Amand, B. Urbaszek, *Rev. Mod. Phys.* **2018**, *90*, 021001.
[5] K. He, C. Poole, K. F. Mak, J. Shan, *Nano Lett.* **2013**, *13*, 2931.
[6] S. Tongay, J. Suh, C. Ataca, W. Fan, A. Luce, J. S. Kang, J. Liu, C. Ko, R. Raghunathanan, J. Zhou, F. Ogletree, J. Li, J. C. Grossman, J. Wu, *Sci. Rep.* **2013**, *3*, 2657.
[7] X. Yin, C. S. Tang, Y. Zheng, J. Gao, J. Wu, H. Zhang, M. Chhowalla, W. Chen, A. T. S. Wee, *Chem. Soc. Rev.* **2021**, *50*, 10087.
[8] M. Ghafariasl, T. Zhang, Z. D. Ward, D. Zhou, D. Sanchez, V. Swaminathan, H. Terrones, M. Terrones, Y. Abate, *Adv. Opt. Mater.* **2024**, *12*, 2302326.
[9] M.-Y. Li, Y. Shi, C.-C. Cheng, L.-S. Lu, Y.-C. Lin, H.-L. Tang, M.-L. Tsai, C.-W. Chu, K.-H. Wei, J.-H. He, W.-H. Chang, K. Suenaga, L.-J. Li, *Science* **2015**, *349*, 524.
[10] A. Fali, T. Zhang, J. P. Terry, E. Kahn, K. Fujisawa, B. Kabius, S. Koirala, Y. Ghafouri, D. Zhou, W. Song, L. Yang, M. Terrones, Y. Abate, *ACS Nano* **2021**, *15*, 2447.
[11] C. Huang, S. Wu, A. M. Sanchez, J. J. P. Peters, R. Beanland, J. S. Ross, P. Rivera, W. Yao, D. H. Cobden, X. Xu, *Nat. Mater.* **2014**, *13*, 1096.
[12] H.-P. Komsa, A. V. Krasheninnikov, *J. Phys. Chem. Lett.* **2012**, *3*, 3652.
[13] S. Najmaei, Z. Liu, W. Zhou, X. Zou, G. Shi, S. Lei, B. I. Yakobson, J.-C. Idrobo, P. M. Ajayan, J. Lou, *Nat. Mater.* **2013**, *12*, 754.
[14] Y.-H. Lee, X.-Q. Zhang, W. Zhang, M.-T. Chang, C.-T. Lin, K.-D. Chang, Y.-C. Yu, J. T.-W. Wang, C.-S. Chang, L.-J. Li, T.-W. Lin, *Adv. Mater.* **2012**, *24*, 2320.
[15] Z. Fei, A. S. Rodin, G. O. Andreev, W. Bao, A. S. McLeod, M. Wagner, L. M. Zhang, Z. Zhao, M. Thiemens, G. Dominguez, M. M. Fogler, A. H. C. Neto, C. N. Lau, F. Keilmann, D. N. Basov, *Nature* **2012**, *487*, 82.
[16] Y. Abate, D. Seidlitz, A. Fali, S. Gamage, V. Babicheva, V. S. Yakovlev, M. I. Stockman, R. Collazo, D. Alden, N. Dietz, *ACS Appl. Mater. Interfaces* **2016**, *8*, 23160.
[17] R. Hillenbrand, Y. Abate, M. Liu, X. Chen, D. N. Basov, *Nat. Rev. Mater.* **2025**, *10*, 285.
[18] F. J. G. de Abajo, D. N. Basov, F. H. L. Koppens, L. Orsini, M. Ceccanti, S. Castilla, L. Cavicchi, M. Polini, P. A. D. Goncalves, A. T. Costa, N. M. R. Peres, N. A. Mortensen, S. Bharadwaj, Z. Jacob, P. J. Schuck, A. N. Pasupathy, M. Delor, M. K. Liu, A. Mugarza, P. Merino, M. G. Cuxart, E. C. Angel, M. Svec, L. H. G. Tizei, F. Dirnberger, H. Deng, C. Schneider, V. Menon, T. Deilmann, A. Chernikov, et al., *ACS Photonics* **2025**, *12*, 3961.
[19] B. Ryu, L. Wang, H. Pu, M. K. Y. Chan, J. Chen, *Chem. Soc. Rev.* **2022**, *51*, 1899.
[20] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, *Nature* **2018**, *559*, 547.
[21] O. A. von Lilienfeld, K. R. Müller, A. Tkatchenko, *Nat. Rev. Chem.* **2020**, *4*, 347.
[22] J. Schmidt, M. R. G. Marques, S. Botti, M. A. L. Marques, *npj Comput. Mater.* **2019**, *5*, 83.
[23] Z. Zhu, S. Yuan, Q. Yang, H. Jiang, F. Zheng, J. Lu, Q. Sun, *J. Am. Chem. Soc.* **2024**, *146*, 29199.
[24] X. Chen, Z. Yao, S. Xu, A. S. McLeod, S. N. G. Corder, Y. Zhao, M. Tsuneto, H. A. Bechtel, M. C. Martin, G. L. Carr, M. M. Fogler, S. G. Stanciu, D. N. Basov, M. Liu, *ACS Photonics* **2021**, *8*, 2987.
[25] M. Imamura, K. Takahashi, *Sci. Rep.* **2024**, *14*, 24200.
[26] M. Fu, S. Xu, S. Zhang, F. L. Ruta, J. Pack, R. A. Mayer, X. Chen, S. L. Moore, D. J. Rizzo, B. S. Jessen, M. Cothrine, D. G. Mandrus, K. Watanabe, T. Taniguchi, C. R. Dean, A. N. Pasupathy, V. Bisogni, P. J. Schuck, A. J. Millis, M. Liu, D. N. Basov, *Nano Lett.* **2024**, *24*, 2149.
[27] J. W. P. Khor, T. T. Tran, A. S. Sharbirin, S. X. B. Yap, H. Choo, J. Kim, *Adv. Opt. Mater.* **2024**, *12*, 2302195.
[28] H. Jia, C. Wang, C. Wang, P. Clancy, *ACS Nano* **2023**, *17*, 453.
[29] B. Lu, Y. Xia, Y. Ren, M. Xie, L. Zhou, G. Vinai, S. A. Morton, A. T. S. Wee, W. G. van der Wiel, W. Zhang, P. K. J. Wong, *Adv. Sci.* **2024**, *11*, 2305277.
[30] M. E. Abbassi, J. Overbeck, O. Braun, M. Calame, H. S. J. van der Zant, M. L. Perrin, *Commun. Phys.* **2021**, *4*, 50.
[31] P. V. Kolesnichenko, Q. Zhang, C. Zheng, M. S. Fuhrer, J. A. Davis, *Mach. Learn.: Sci. Technol.* **2021**, *2*, 025021.
[32] J. He, Y. Zhou, P. Zhang, H. Zhai, *Appl. Surf. Sci.* **2025**, *686*, 162110.
[33] B. Roter, N. Ninkovic, S. V. Dordevic, *Physica C:Superconductivity and its applications* **2022**, *598*, 1354078.
[34] J. Carrasquilla, R. G. Melko, *Nat. Phys.* **2017**, *13*, 431.
[35] S. J. Wetzel, *Phys. Rev. E* **2017**, *96*, 022140.
[36] L. Wang, *Phys. Rev. B* **2016**, *94*, 195105.
[37] N. Đorđević, J. S. Beckwith, M. Yarema, O. Yarema, A. Rosspeintner, N. Yazdani, J. Leuthold, E. Vauthey, V. Wood, *ACS Photonics* **2018**, *5*, 4888.
[38] K. Tanaka, K. Hachiya, W. Zhang, K. Matsuda, Y. Miyauchi, *ACS Nano* **2019**, *13*, 12687.
[39] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, *npj Comput. Mater.* **2017**, *3*, 54.
[40] Y. Mao, N. Dong, L. Wang, X. Chen, H. Wang, Z. Wang, I. M. Kislyakov, J. Wang, *Nanomaterials* **2020**, *10*, 2223.
[41] Y. Qi, D. Hu, M. Zheng, Y. Jiang, Y. P. Chen, *Appl. Mater. Today* **2024**, *41*, 102499.
[42] A.-Y. Lu, L. G. P. Martins, P.-C. Shen, Z. Chen, J.-H. Park, M. Xue, J. Han, N. Mao, M.-H. Chiu, T. Palacios, V. Tung, J. Kong, *Adv. Mater.* **2022**, *34*, 2202911.
[43] J. F. Joung, M. Han, J. Hwang, M. Jeong, D. H. Choi, S. Park, *JACS Au* **2021**, *1*, 427.
[44] K. Choo, G. Carleo, N. Regnault, T. Neupert, *Phys. Rev. Lett.* **2018**, *121*, 167204.
[45] S. S. Kalantre, J. P. Zwolak, S. Ragole, X. Wu, N. M. Zimmerman, M. D. S. Jr, J. M. Taylor, *npj Quantum Inf.* **2019**, *5*, 6.
[46] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, L. Zdeborova, *Rev. Mod. Phys.* **2019**, *91*, 045002, https://link.aps.org/doi/10.1103/RevModPhys.91.045002.
[47] N. C. Frey, D. Akinwande, D. Jariwala, V. B. Shenoy, *ACS Nano* **2020**, *14*, 13406.
[48] Y. Chen, J. Xi, D. O. Dumcenco, Z. Liu, K. Suenaga, D. Wang, Z. Shuai, Y.-S. Huang, L. Xie, *ACS Nano* **2013**, *7*, 4610.
[49] A. M. van der Zande, P. Y. Huang, D. A. Chenet, T. C. Berkelbach, Y. You, G.-H. Lee, T. F. Heinz, D. R. Reichman, D. A. Muller, J. C. Hone, *Nat. Mater.* **2013**, *12*, 554.
[50] Y. Zhu, J. Lim, Z. Zhang, Y. Wang, S. Sarkar, H. Ramsden, Y. Li, H. Yan, D. Phuyal, N. Gauriot, A. Rao, R. L. Z. Hoye, G. Eda, M. Chhowalla, *ACS Nano* **2023**, *17*, 13545.
[51] R. Tomar, B. Hsu, A. Perez, M. A. Stroscio, M. Dutta, *J. Electron. Mater.* **2023**, *52*, 5513.
[52] H. Lamsaadi, A. Cuche, G. Agez, I. Paradisanos, D. Beret, L. Lombez, P. Renucci, D. Lagarde, X. Marie, Z. Gan, A. George, K. Watanabe, T. Taniguchi, A. Turchanin, N. Combe, B. Urbaszek, V. Paillard, J.-M. Poumirol, *Adv. Opt. Mater.* **2025**, *13*, 2403009.

[53] K. W. Lau, Calvin, Z. Gong, H. Yu, W. Yao, *Phys. Rev. B* **2018**, *98*, 115427.

[54] D. Beret, I. Paradisanos, H. Lamsaadi, Z. Gan, E. Najafidehaghani, A. George, T. Lehnert, J. Biskupek, U. Kaiser, S. Shree, A. E. Real, D. Lagarde, X. Marie, P. Renucci, K. Watanabe, T. Taniguchi, S. Weber, V. Paillard, L. Lombez, J. Marie Poumirol, A. Turchanin, B. Urbaszek, *npj 2D Mater. Appl.* **2022**, *6*, 84.

[55] H. Lamsaadi, D. Beret, I. Paradisanos, P. Renucci, D. Lagarde, X. Marie, B. Urbaszek, Z. Gan, A. George, K. Watanabe, T. Taniguchi, A. Turcanin, L. Lombez, N. Combe, V. Paillard, J.-M. Poumirol, *Nat. Commun.* **2023**, *14*, 5881.

[56] R. Rosati, S. Shradha, J. Picker, A. Turchanin, B. Urbaszek, E. Malic, *Nano Lett.* **2025**, *25*, 11319.