

PAPER • OPEN ACCESS

Multidimensional analysis of excitonic spectra of monolayers of tungsten disulphide: toward computer-aided identification of structural and environmental perturbations of 2D materials

To cite this article: Pavel V Kolesnichenko *et al* 2021 *Mach. Learn.: Sci. Technol.* **2** 025021

View the [article online](#) for updates and enhancements.

You may also like

- [Erratum: "The GRIFFIN Project—Formation of Star Clusters with Individual Massive Stars in a Simulated Dwarf Galaxy Starburst" \(2020, ApJ, 891, 2\)](#)
Natalia Lahén, Thorsten Naab, Peter H. Johansson et al.
- [Erratum: "Optical Photometry of the Quasar 3C 454.3 during the Period 2006–2018 and the Long-term Periodicity Analysis" \(2021, ApJS, 253, 10\)](#)
J. H. Fan, S. O. Kurtanidze, Y. Liu et al.
- [Corrigendum: The credibility challenge for global fluvial flood risk analysis \(2016 *Environ. Res. Lett.* 11 094014\)](#)
M A Trigg, C E Birch, J C Neal et al.



PAPER

OPEN ACCESS

RECEIVED
24 August 2020REVISED
18 November 2020ACCEPTED FOR PUBLICATION
5 January 2021PUBLISHED
2 March 2021

Original Content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Multidimensional analysis of excitonic spectra of monolayers of tungsten disulphide: toward computer-aided identification of structural and environmental perturbations of 2D materials

Pavel V Kolesnichenko^{1,4} , Qianhui Zhang², Changxi Zheng^{3,5,6,7}, Michael S Fuhrer^{3,5}
and Jeffrey A Davis^{1,4}

¹ Optical Sciences Centre, Swinburne University of Technology, Melbourne, Victoria 3122, Australia

² Department of Civil Engineering, Monash University, Melbourne, Victoria 3800, Australia

³ School of Physics and Astronomy, Monash University, Melbourne, Victoria 3800, Australia

⁴ ARC Centre of Excellence in Future Low-Energy Electronics Technologies, Swinburne University of Technology, Melbourne, Victoria 3122, Australia

⁵ ARC Centre of Excellence in Future Low-Energy Electronics Technologies, Monash University, Melbourne, Victoria 3800, Australia

⁶ School of Science, Westlake University, Hangzhou 310024, Zhejiang Province, People's Republic of China

⁷ Institute of Natural Sciences, Westlake Institute for Advanced Study, Hangzhou 310024, Zhejiang Province, People's Republic of China

E-mail: jdavis@swin.edu.au

Keywords: two-dimensional materials, principal component analysis, hyperspectral imaging

Supplementary material for this article is available [online](#)

Abstract

Despite 2D materials holding great promise for a broad range of applications, the proliferation of devices and their fulfillment of real-life demands are still far from being realized. Experimentally obtainable samples commonly experience a wide range of perturbations (ripples and wrinkles, point and line defects, grain boundaries, strain field, doping, water intercalation, oxidation, edge reconstructions) significantly deviating the properties from idealistic models. These perturbations, in general, can be entangled or occur in groups with each group forming a complex perturbation making the interpretations of observable physical properties and the disentanglement of simultaneously acting effects a highly non-trivial task even for an experienced researcher. Here we generalise statistical correlation analysis of excitonic spectra of monolayer WS₂, acquired by hyperspectral absorption and photoluminescence imaging, to a multidimensional case, and examine multidimensional correlations via unsupervised machine learning algorithms. Using principal component analysis we are able to identify four dominant components that are correlated with tensile strain, disorder induced by adsorption or intercalation of environmental molecules, multi-layer regions and charge doping, respectively. This approach has the potential to determine the local environment of WS₂ monolayers or other 2D materials from simple optical measurements, and paves the way toward advanced, machine-aided, characterization of monolayer matter.

1. Introduction

Since the realization of exfoliation of a single layer of graphite (graphene) and confirmation of its extraordinary physical properties [1], a wave of efforts aiming at synthesizing other two-dimensional (2D) materials has naturally emerged. A broad spectrum of experimentally obtained ultra-thin materials covering metals [2, 3], semimetals [4], semiconductors [5], insulators [6], topological insulators [7], superconductors [8, 9] and ferromagnets [10] has been already reported with many others having been theoretically predicted [11–14]. This has opened an avenue to material engineering in the form of van der Waals heterostructures giving rise to novel potential devices such as single-molecule and DNA sensors [15, 16], photodiodes [17, 18], transistors [19], memory cells [20, 21], batteries [22, 23], magnetic field sensors [24], and spintronic logic gates [25, 26].

Despite monolayers holding great promise for a broad range of applications, the research around 2D materials suggests that proliferation of the potential devices and their fulfillment of real-life demands are still far from realization. In contrast to theoretical descriptions of the physical properties of various 2D materials, experimentally obtainable samples commonly experience a wide range of perturbations significantly deviating the properties from idealistic models, and thereby affecting the performance of the devices. Amongst these perturbations are the presence of ripples and wrinkles [27–29], point and line defects [30, 31], grain boundaries [32], strain fields [33], doping [34–37], water intercalation [38], oxidation [39–41], and edge reconstructions [42]. These perturbations, in general, can be entangled or occur in groups, forming complex perturbations. This, in turn, makes the interpretations of observable physical properties and the disentanglement of simultaneously acting effects a highly non-trivial task even for an experienced researcher, and advanced characterization methods are often desirable.

Due to the monolayer nature of 2D materials, their optical signatures are highly sensitive to fluctuations in structure and the local environment. This sensitivity results in non-trivial spatial variations, which are hard to analyze manually, and indicates that unsupervised machine learning algorithms applied to multimodal optical imaging data may aid attempts to identify the fluctuations in structure and the local environment distributed across monolayers.

Here we consider a semiconducting monolayer of tungsten disulphide (WS_2) grown via chemical vapour deposition (CVD) on a sapphire substrate. Optical properties of WS_2 monolayers are dominated by excitonic effects manifested as intense signatures in their absorption and emission spectra [43]. We apply absorption and photoluminescence (PL) hyperspectral imaging to gather data on the spatial variations of the excitonic properties, which arise from the various perturbations. The spectra are fully parameterised, leading to a multidimensional parametric phase-space (hypercube) where a single data point represents the set of values corresponding to all parameters at a given spatial location on the monolayer sample. This then allows us to apply principal component analysis (PCA) [44–46] to identify the parameters that vary together and ideally combine to quantify specific perturbations and how they vary across the monolayer flake. A projection of the multidimensional data-cloud onto a 2D plane with axes given by the two most significant principal components preserves the maximum variance in the data. By using unsupervised K-means clustering [47–49] of the data-points in this PCA-plane, regions of the sample with similar properties can be identified and provide further insight into how the perturbations combine and vary across the monolayer sample.

2. Results and discussion

Typical absorption and emission spectra of WS_2 monolayers are shown in figure 1(a). Absorption spectra are approximated here by differential reflectance [36, 50, 51] and feature two distinct peaks corresponding to spin-orbit split A- and B-exciton transitions occurring at K symmetry points in the first Brillouin zone [52]. Red-shifted PL emission is evident as an asymmetric peak formed as a result of recombination of excitons and trions [53]. Figures 1(b) and (c) show the spatially-resolved peak absorption amplitude and wavelength corresponding to the A-exciton transition, revealing trigonally-symmetric variations. Similar trends are observed in the spatial maps of PL emission (figures 1(d) and (e)): the absorption and emission are blue-shifted in the regions spanning from the center of the flake toward its apexes. This type of behaviour has been attributed previously to elevated n -doping levels in those areas [37, 54], and conversely, red-shifted absorption and emission peaks in the adjacent regions have been shown to result from greater tensile strain [37]. While the absorption and emission wavelength maps are somewhat similar, differences reveal variations in Stokes shift, which is closely related to charge doping. More obvious differences are observed between the patterns formed by absorption peak amplitudes (figure 1(b)) and emission peak intensities (figure 1(d)). The most significant difference is that the edges of the triangular monolayer flake can be clearly distinguished in the PL emission intensity map. Along the edges the PL intensity is enhanced as a result of combined effects of water intercalation progressing toward the interior over time [38] and oxidation [40, 41]. Additionally, three bright spots near the center of the flake can be clearly distinguished in the absorption amplitude map. These bright features are believed to represent multilayer WS_2 material formed at the nucleation centers of the monolayer since larger reflectance contrasts have been observed for TMdC multilayers [55]. All these observed differences point to the complementarity of absorption and PL measurements allowing for observations of nonzero correlations between various parameters.

Statistical correlation analysis has been already proven to be a powerful tool in the studies of optoelectronic properties of 2D materials [33, 36, 56–60]. For example, by correlating spectral shifts of prominent Raman peaks in graphene and graphene/TMdC heterostructures it can be possible to disentangle the effects of doping and strain [56, 59]. Another route to solve a similar problem for TMdC monolayers used correlations involving the PL Stokes shift [37]. Correlation analyses also facilitated the recognition of physically distinct edges of triangular TMdC flakes as domains hosting large number of point defects [57, 60]

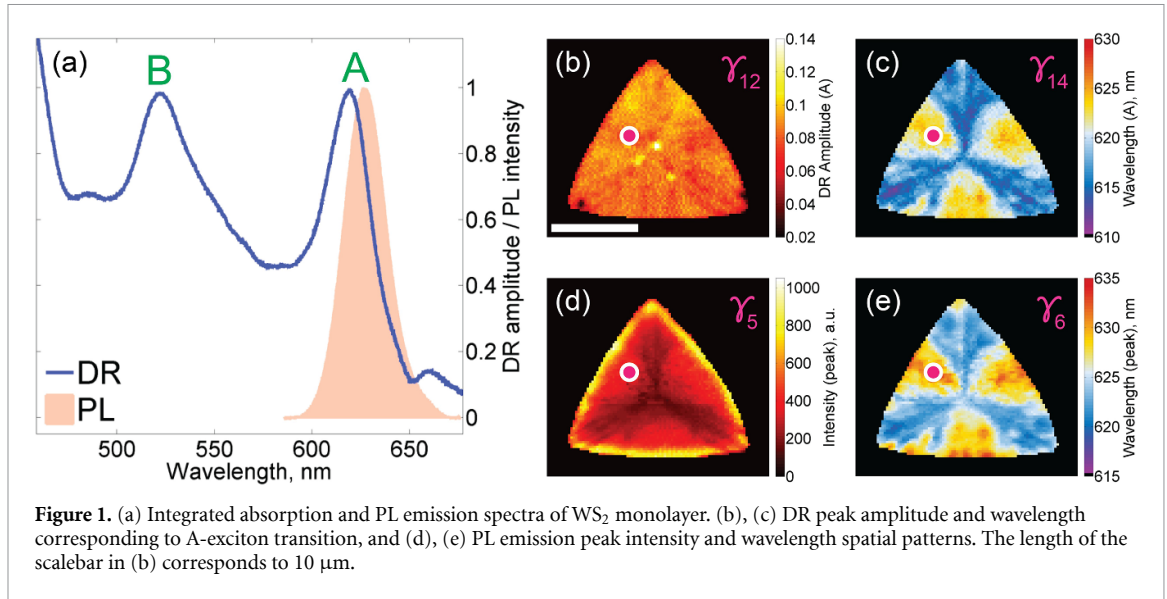


Table 1. List of parameters $\gamma_i, i = \{1, \dots, 17\}$, used as dimensions for a hypercube. ‘X’ stands for exciton; ‘T’ stands for trion; ‘A’ stands for A-exciton; ‘B’ stands for B-exciton; ‘SM’ stands for spectral median.

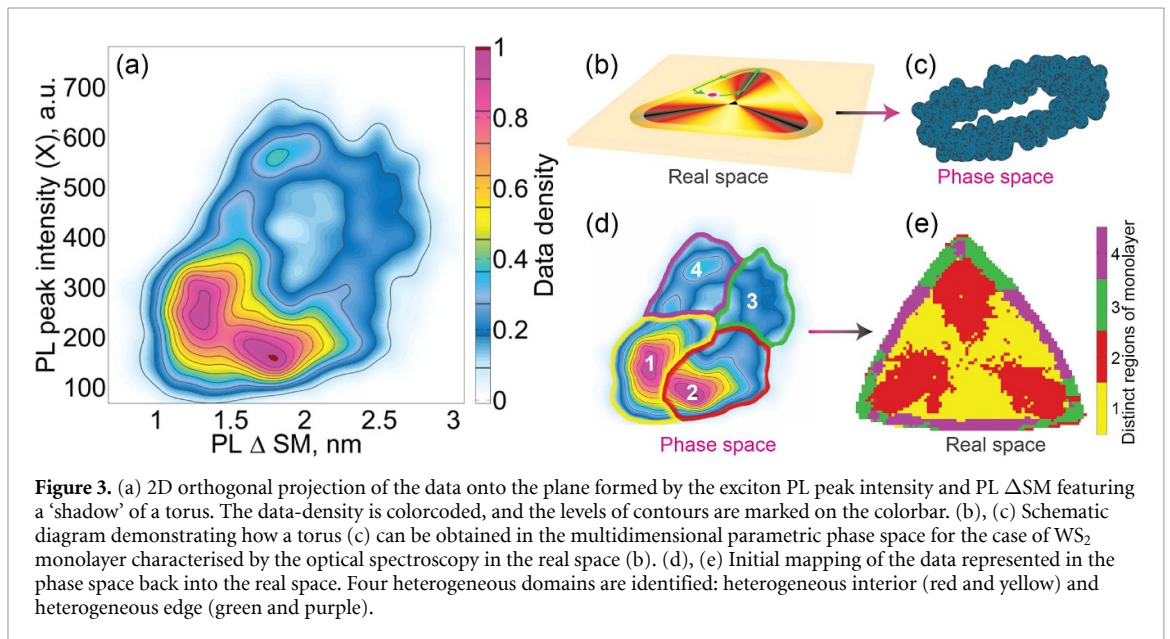
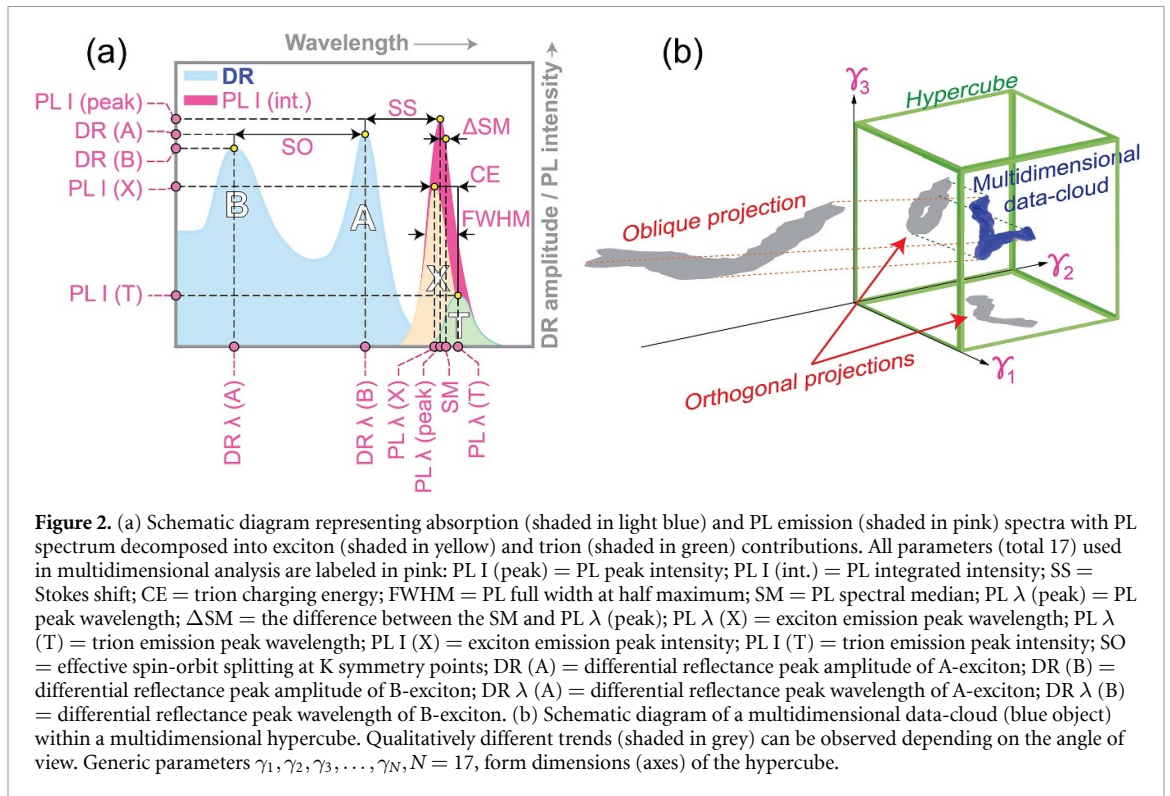
Parameter	Meaning	Parameter	Meaning
γ_1	PL peak intensity (X) ^a	γ_{10}	PL FWHM
γ_2	PL peak wavelength (X) ^a	γ_{11}	Trion charging energy ^a
γ_3	PL peak intensity (T) ^a	γ_{12}	DR peak intensity (A)
γ_4	PL peak wavelength (T) ^a	γ_{13}	DR peak intensity (B)
γ_5	PL peak intensity	γ_{14}	DR peak wavelength (A)
γ_6	PL peak wavelength	γ_{15}	DR peak wavelength (B)
γ_7	PL integrated intensity	γ_{16}	Effective spin-orbit splitting
γ_8	PL spectral median	γ_{17}	PL Stokes shift
γ_9	PL ΔSM		

^a These parameters were derived from fittings.

and the effects of strain on optoelectronic properties of various TMdC monolayers, including direct/indirect nature of the bandgap [33]. All these results, however, were based on scatter plots between specifically chosen pairs of parameters missing out other possible correlations, and were not able to recognise the presence of any subtle variations in the data. Here, we generalise statistical correlation analysis to an N -dimensional case to acquire more insights into the optoelectronic variations commonly found in 2D materials.

To make the N -dimensional correlation analysis possible we fully parameterise absorption and emission spectra (figure 2(a)) and use each of the parameters to represent a dimension of an N -dimensional parametric space (figure 2(b)), with $N = 17$ in our case. The parameters used are shown in table 1, and given a label γ_i , and their spatial distributions are shown in the supporting information (S1) (available online at stacks.iop.org/MLST/2/025021/mmedia). These parameters include all fit-free parameters we extracted and some basic fitting parameters: exciton and trion peak intensity and wavelength as well as the trion’s charging energy, even though they are strongly correlated with such fitting-free quantities as PL peak intensity and wavelength, SM and ΔSM [37]. Despite the fact that several parameters measure similar quantities we did not exclude any at this stage, in case there were small differences that could be correlated with other changes and specific perturbations. This parameter space is represented by an N -dimensional hypercube (N -cube) encapsulating an N -dimensional data-cloud where each data-point $\vec{\gamma}$ is described by a set of N values (coordinates), i.e. $\vec{\gamma} = \{\gamma_1, \gamma_2, \dots, \gamma_N\}$. The parameters whose spatial variations are mapped in figures 1(b)–(e) correspond to $\gamma_5, \gamma_6, \gamma_{12}, \gamma_{14}$. A specific location (pink point in figures 1(b)–(e)) on the monolayer island can, therefore, be assigned a set of $N = 17$ numbers corresponding to the values of the 17 parameters chosen to describe the optical properties of the material.

The natural approach to visualization of the geometry of a multidimensional object (data-cloud) is to look at its projections onto 2D planes (figure 2(b)). Amongst infinite number of possible planes and projecting angles, a particular case of orthogonal projections onto the sides of the N -cube is the simplest to realise. It is this particular case that was considered in the previously reported correlation analyses of optoelectronic properties of 2D materials where certain physical trends and clusters have been identified



[33, 36, 37, 56, 57, 59, 60]. In some instances, oblique projections can provide greater separation of the data, and in some cases correspond to meaningful parameters. For example, charging energy is defined as the difference between the exciton and trion energy, and would correspond to an oblique projection, as detailed in supporting information (S5).

An example of an orthogonal projection is shown in figure 3(a). In this case, the data is distributed in a ring, which indicates that the full data-cloud is a torus-isomorphic object (other projections showing torus-shaped data-shadows are shown in supporting information, S7, for data density estimation). This topology can then also be related to the spatial variations of material properties assuming they vary smoothly, which is typically the case. Specifically, the shape of the data-cloud indicates that it is possible to define loops where each point on the loop has distinct spectral properties (figures 3(b) and (c)). These loops will be around a specific point or points on the material.

To help visualise this we note that around the data-cloud ring in figure 3(a), there appears to be four main clusters of data, as identified in figure 3(d). The boundaries are defined as the lines connecting those points of the data density contours that have high negative curvature [61, 62] (see supporting information, S4, for more details on how these boundaries were identified). The data-points within each cluster are mapped back into their real-space location in figure 3(e), showing that the clusters are indeed related to specific regions on the sample. The points about which the circular paths can be identified, are the points where the four colours (corresponding to the four clusters) meet.

It is apparent then that by selecting specific orthogonal projections and subsequent clustering analysis, we can gain some insight into how the optical properties (and corresponding structural/environmental properties) vary across the sample. However, it is likely that there is further fine structure in these clusters, indeed, it is expected that many of the sample perturbations vary smoothly and continuously. These perturbations will affect the 17 different parameters in different ways, and in most cases will affect more than one parameter. To identify the parameters that vary together and maintain the maximum variance of the data, we apply the PCA [44–46]. By identifying the orthogonal principal components (i.e. specific linear combinations of the 17 parameters) that maintain the maximal variance, the ability to resolve fine structure and small variations in the data cloud is enhanced. Furthermore, in identifying the parameters that vary together, this approach has the potential to separate each specific sample perturbation (e.g. strain, doping, molecular adsorption, etc) and the specific changes to the optical properties induced by each of them. It may then become possible to map the structural and environmental properties across the sample.

The method for PCA has been described previously [44–46] and the details of the approach used here are included in supporting information (S6). The result is a new set of axes (the principal components) for the data hypercube. These principal components are defined such that the variance Δ_i , $i = \{1, \dots, 17\}$ of the data along the principal components (PC i) is maximised for PC1 and decreases for each successive component (i.e. $\Delta_1 \geq \Delta_2 \geq \dots \geq \Delta_{17}$). The corollary of defining the principal components in this way is that it also identifies the measurement parameters that vary together, and distills the variance of the 17-dimensional data-cloud into a hypercube of a lower dimensionality. In the case of the hypercube defined by the parameters of absorption and emission spectra considered here, the PCA approach showed that it is possible to reduce the number of dimensions from 17 (defined by the parameters $\gamma_1, \dots, \gamma_{17}$) down to four (defined by the parameters PC1, ..., PC4) and still preserve as much as 87.9% of the total variance (see supporting information, S6, figure 6).

Each of the principal components are formed by a linear combination of the measurement parameters. The relative weights of the parameters for each PC are shown in the supporting information (S6), tables 2 and 3. Among other things, this shows that the parameters that measure similar quantities are grouped together, as would be expected. For example, the fitted PL peak intensity (γ_1), PL maximum intensity (γ_5), and integrated PL intensity make contributions to each of the first eight principal components that are within 10% of each other, as can be seen in figure S7 and table S2. For components PC9–PC17, it is not clear that there is any significance associated with them since they make up less than 1% of the total variance of the data, combined. Furthermore, the derived parameters, such as ΔSM also identify the expected relationship with their source parameters. For example, the weight of ΔSM (γ_9), which is calculated by the difference between PL peak wavelength (γ_6) and SM (γ_8), is close to zero for PC1 where the weightings for γ_6 and γ_8 are relatively large and essentially equal. In PC2, however, ΔSM has the largest weight, and the weights of γ_6 and γ_8 are significant, but of opposite sign to each other, meaning that when the contributions are added to form the PC2, the difference between PL peak wavelength and SM appears naturally. This gives further confidence that the PCA is performing reliably and giving meaningful results.

The amplitude of each principal component then varies across the sample and can be mapped spatially in the same way the measurement parameters were mapped in figures 1(b)–(e). Based on the make-up of each PC, their spatial variations across the WS₂ flake, and previously reported understanding of these materials, we are able to correlate each PC with a specific perturbation (or group of perturbations) of the sample structure and/or environment. Specifically, we link PC1 with variations in strain, PC2 with disorder induced by adsorption and/or intercalation of environmental molecules, PC3 with multilayers and PC4 with charge doping.

The attribution of PC1 to strain variation is based on the observation that the dominant contributions to this component are the PL intensity and wavelength parameters, as well as the absorption wavelength for the A-exciton, consistent with previous measurements reporting the effect of strain on optical properties [55, 63, 64]. In addition, the spatial variation of PC1 across the sample (figure 4(a)) is consistent with previous observations of how the strain varies from the apexes to the middle of the sides in CVD-grown WS₂ monolayers [37, 65, 66]. There are other perturbations that can also alter the PL intensities and wavelengths, however, these also affect other parameters that are absent from this principal component (e.g. doping also leads to substantial Stokes shift). PC2 (figure 4(b)) is dominated by variations in the PL FWHM, ΔSM , and

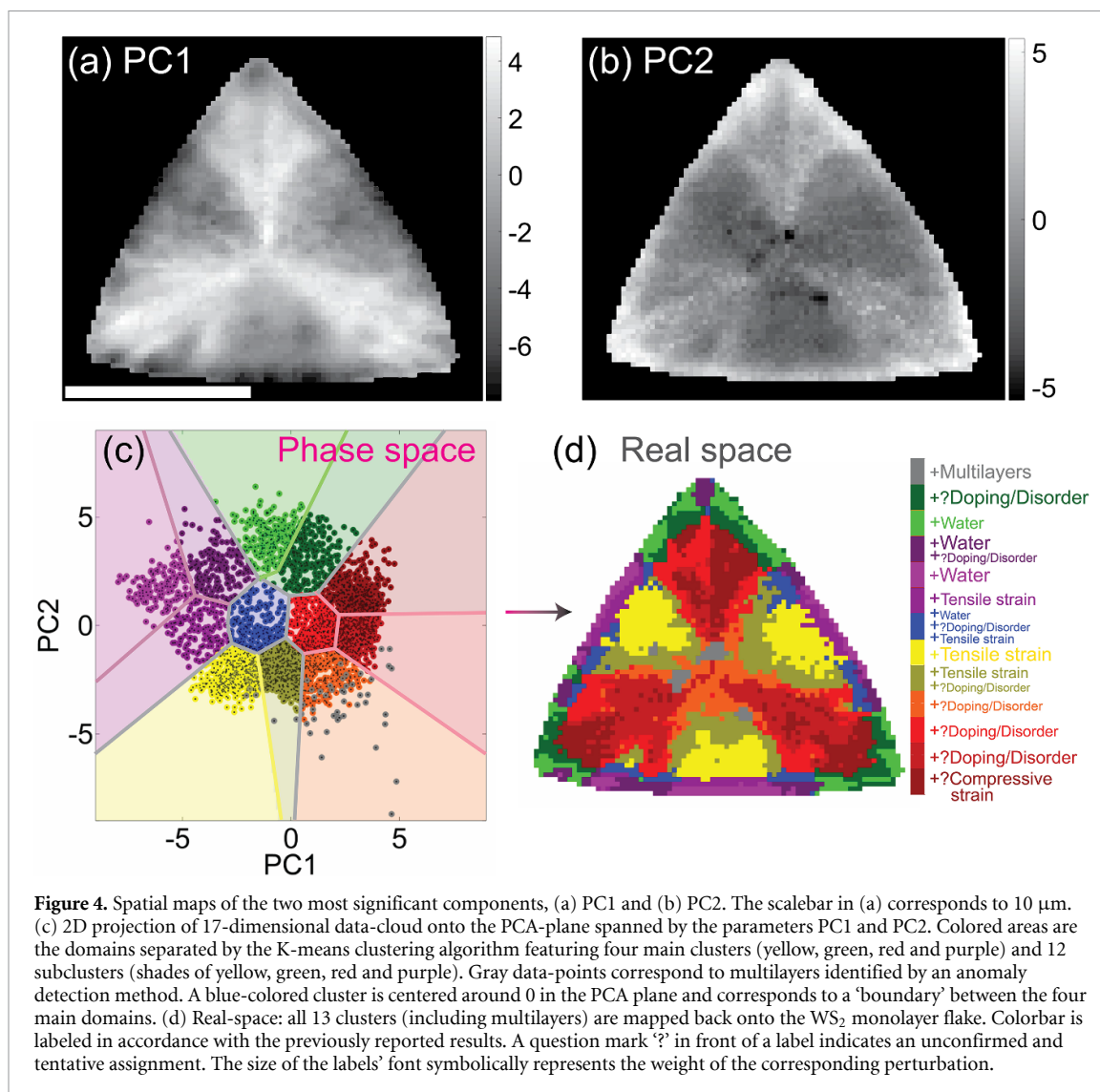


Figure 4. Spatial maps of the two most significant components, (a) PC1 and (b) PC2. The scalebar in (a) corresponds to 10 μm . (c) 2D projection of 17-dimensional data-cloud onto the PCA-plane spanned by the parameters PC1 and PC2. Colored areas are the domains separated by the K-means clustering algorithm featuring four main clusters (yellow, green, red and purple) and 12 subclusters (shades of yellow, green, red and purple). Gray data-points correspond to multilayers identified by an anomaly detection method. A blue-colored cluster is centered around 0 in the PCA plane and corresponds to a ‘boundary’ between the four main domains. (d) Real-space: all 13 clusters (including multilayers) are mapped back onto the WS_2 monolayer flake. Colorbar is labeled in accordance with the previously reported results. A question mark ‘?’ in front of a label indicates an unconfirmed and tentative assignment. The size of the labels’ font symbolically represents the weight of the corresponding perturbation.

charging energy. Alone, ΔSM and charging energy can be associated with doping density, however, a clearer signature of doping is the Stokes shift [36, 37, 54], which does not make a significant contribution to this PC. Furthermore, this is a flake that has been exposed to air for some time and previous work has shown that where freshly grown flakes have large amounts of n -doping in the region of the apexes, the aged flakes adsorb environmental molecules, which reduce the density of free charges, and increase the FWHM [36, 37, 67, 68]. The spatial variation of PC2 adds further weight to this assignment, as in addition to the expected variations in the bulk of the 2D flake, it also reveals the edges where water intercalation has occurred [37, 38].

PC3 is dominated by the DR peak intensity for both A and B excitons and the spatial map shows the most significant variance occurs in small regions near the centre of the flake. This is consistent with previous measurements that have attributed significant increases in DR intensity to increased scattering from multi-layer regions on the sample [69, 70]. The dominant contribution to PC4 is the PL Stokes shift, which is strongly linked with charge doping [36, 37, 54]. The relatively low level of variation in PC4 is consistent with previous observations of small variations of doping density on aged CVD-grown flakes. Interestingly, the other significant contribution to PC4 is the effective spin-orbit splitting, or in other words the energy difference between the A and B excitons in the DR measurements. It was previously speculated that these variations could be due to increased doping [37], but other possibilities could not be ruled out. This observation adds further weight to the case that it is due to variations in doping density.

The ability to self-consistently attribute specific sample perturbations to the four primary principal components that arise naturally from PCA of a single flake is potentially of great value. It points to the possibility of developing a set of well-defined principal components, consisting of linear combinations of spectroscopic parameters, that could be applied to determine the precise structural and environmental perturbations at a specific location in a 2D semiconductor. Further analysis of how these perturbations vary

across the flakes could then provide insight into growth mechanisms and causes of the variations from pristine materials. We note, however, that to have a greater level of confidence in the make-up of the significant principal components and their relationship to physical perturbations, analysis of a larger dataset and materials prepared under different conditions with different combinations of perturbations is required. This will be the subject of future work.

One of the challenges that may be resolved by this type of PCA is the ability to distinguish different contributions, where multiple perturbations occur at the same place. To demonstrate this we project the data onto the plane spanned by the first two principal components (PC1 and PC2) (figure 4(c)), associated with strain and the adsorption of environmental molecules/intercalation of water. This projection shows the maximum spread of the data and reveals some clustering of data points. In order to acquire more insights into the projected data and correlate the coordinates in this PC1-PC2 plane with the spatial location on the WS₂ flake, we applied the K-means-clustering unsupervised learning algorithm [47–49]. This algorithm, for a given input number K , tries to classify the data-set into K labeled clusters (see supporting information, S9, for details). The value of K , however, cannot be automatically identified by the algorithm, and, therefore, cluster identification methods are commonly used. Here we used the so called ‘elbow’ method [71] as one of the most popular methods for identification of the natural number of clusters, if there are any (see supporting information, S8). As expected, in the case of the WS₂ monolayer considered here, the ‘elbow’ method revealed the presence of four prominent clusters in the data-cloud (see supporting information, S8) corresponding to the two heterogeneous interior domains and two heterogeneous edge domains within the flake, as identified above in figure 3. However, the method also revealed that there are other natural cluster sets ($K = 2$, $K = 8$ and $K = 12$) present in the data, although they are not as prominent as the set of four clusters ($K = 4$). With increasing the number of clusters K , additional ‘shades’ are introduced to the $K = 4$ cluster set (figure 4(c)). We note that we excluded multilayer regions (gray data-points in figure 4(c), from the K-means analysis by treating them as ‘anomalies’ (or ‘outliers’) (see supporting information, S7) [72]. We do this because the multilayer regions have vastly different optical properties from the rest of the sample, and if we were to plot the data as a function of the first three principal components, they would form their own cluster due to being separated along the PC3 axis, again due to their vastly different properties.

We then mapped the data-points in the PCA-plane within each of the clusters back onto the monolayer flake (figure 4(d)) revealing fine-structure of the four main regions of the sample mentioned above. These four clusters can be clearly grouped in pairs, with the purple and green regions mapping the edges affected by water intercalation, and the red and yellow regions mapping the interior of the flake. This roughly correlates with the higher values for PC2 around the edges, although it can also be seen that the value of PC2 increases when going from purple to green. A similar trend is seen when going from yellow to red in the interior. This indicates that while the main change in going from yellow to red and purple to green is along the PC1 axis, and due to reducing tensile strain, this trend is accompanied by an increase in disorder due to adsorbed molecules, and hence a shift along the PC2 axis.

We note also that water intercalation does not change appreciably the amount of strain along the edges, as evidenced by the consistent variation along PC1 for both the edges and interior. This suggests that on average the strain field vectors are aligned angularly around the center of the monolayer island so that the radially-propagating water intercalation does not release strain (see also supporting information, S10).

3. Conclusions

These results indicate that the PCA based on the spectral parameters from PL and DR hyperspectral imaging has the potential to disentangle and quantify different types of perturbations in monolayer materials. This approach is effectively an extension of specific 2D correlation plots that have been used to help understand the variations across a monolayer flake [33, 36, 37, 56, 57, 59, 60], and which were used here to reveal different regions of the monolayer flake with clearly different combinations of perturbations. The PCA, however, is a more systematic and quantitative approach.

The PCA applied to the data here produced four dominant, orthogonal principal components. By examining the combination of spectral parameters that makes up each of these, and the variation of these PCs across the flake, we were able to assign a specific sample perturbation to each: tensile strain, disorder induced by adsorption/intercalation of environmental molecules, multi-layers, and charge doping. These assignments, the spatial variations and spectral parameters contributing are fully consistent with previous measurements and understanding developed from similar flakes [37, 55, 63–66, 69, 70]. However, these assignments are not definitive and may not be able to reliably predict the specific perturbations on a different flake. It does, however, point to the possibility of using this approach for this purpose. To achieve this, a larger sample size including multiple flakes with different levels of the different perturbations is needed, and a refinement of the parameters may be necessary to remove the intensity/amplitude parameters, which

depend on the measurement system. Subsequent steps could involve a large labeled dataset, and combinations of PCA and cluster analysis to train neural networks and enable real-time identification of spatially-varying perturbation. Regardless, the demonstration here that a self-consistent attribution can be made using PCA on a single flake indicates that this approach is promising, and may allow creation of a tool capable of identifying the perturbations at a given location in a given 2D material with optical transitions simply from PL and DR spectra.

For other 2D materials a similar unsupervised learning approach could also be used, however, the data preparation may need to be different. For example, semiconducting monolayers grown on a SiO₂/Si substrate may lead to additional thin-film interference effects which have to be compensated [73]. Other 2D materials, such as graphene, may lack distinct absorption/emission peaks, in which case other parameters can be used for the described multidimensional analysis. For example, in the case of graphene the parameters derived from Raman spectra may be more useful [74]. In general, the reported analysis is not limited to PL/DR/Raman spectroscopy methods and can be applied to imaging data obtained from numerous other techniques such as nano-ARPES, CARS microscopy, SHG imaging, etc. The list of the possible imaging methods that have been already applied to 2D materials to yield new insights is given in the last section of supporting information (S11).

4. Experimental section

4.1. Sample preparation

The sample preparation was performed in a similar way as described in [75]. Briefly, monolayers of WS₂ were grown on sapphire substrate via CVD using WO₃ and sulphur precursors. WO₃ precursor was placed in the middle of the chamber (high-temperature zone, 860 °C) while the sulphur precursor was placed further upstream (low-temperature zone, 180 °C). The substrate was placed in a direct proximity to the WO₃ precursor. Heating the chamber and maintaining the hot environment lasted over ~45 min, after which a cooling process was initiated.

4.2. Experimental realization

The PL and DR imaging setups were implemented in the same way as described in [37]. In a nutshell, in PL experiments, linearly polarised cw radiation (~410 nm) was focused on the sample through a 100× objective lens (NA = 0.95). The detection scheme was implemented in an epi-fluorescence geometry in a confocal way with the detection spatial resolution reaching ~300 nm. DR measurements were performed using a broadband (400–800 nm) incoherent white light radiation with the detection spatial resolution reaching ~380 nm.

Data availability statement

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

Acknowledgment

This work was supported by the Australian Research Council Centre of Excellence in Future Low-Energy Electronics Technologies (CE170100039).

ORCID iDs

Pavel V Kolesnichenko  <https://orcid.org/0000-0003-0379-7695>

Michael S Fuhrer  <https://orcid.org/0000-0001-6183-2773>

Jeffrey A Davis  <https://orcid.org/0000-0003-4537-4084>

References

- [1] Novoselov K S, Geim A K, Morozov S V, Jiang D, Zhang Y, Dubonos S V, Grigorieva I V and Firsov A A 2004 Electric field effect in atomically thin carbon films *Science* **306** 666–9
- [2] Courty A, Henry A-I, Goubet N and Pileni M-P 2007 Large triangular single crystals formed by mild annealing of self-organized silver nanocrystals *Nat. Mater.* **6** 900–7
- [3] Ma Y, Li B and Yang S 2018 Ultrathin two-dimensional metallic nanomaterials *Mater. Chem. Frontiers* **2** 456–67
- [4] Kim J *et al* 2015 Observation of tunable band gap and anisotropic Dirac semimetal state in black phosphorus *Science* **349** 723–6
- [5] Splendiani A, Sun L, Zhang Y, Li T, Kim J, Chim C-Y, Galli G and Wang F 2010 Emerging photoluminescence in monolayer MoS₂ *Nano Lett.* **10** 1271–5

- [6] Zhang K, Feng Y, Wang F, Yang Z and Wang J 2017 Two dimensional hexagonal boron nitride (2D-hBN): synthesis, properties and applications *J. Mater. Chem. C* **5** 11992–2022
- [7] Kou L, Ma Y, Sun Z, Heine T and Chen C 2017 Two-dimensional topological insulators: progress and prospects *J. Phys. Chem. Lett.* **8** 1905–19
- [8] Ménard G C et al 2017 Two-dimensional topological superconductivity in Pb/Co/Si(111) *Nat. Commun.* **8** 2040
- [9] Ménard G C, Mesaros A, Brun C, Debontridder F, Roditchev D, Simon P and Cren T 2019 Isolated pairs of Majorana zero modes in a disordered superconducting lead monolayer *Nat. Commun.* **10** 2587
- [10] Huang B et al 2017 Layer-dependent ferromagnetism in a van der Waals crystal down to the monolayer limit *Nature* **546** 270–3
- [11] Garcia J C, de Lima D B, Assali L V C and Justo J F 2011 Group IV graphene- and graphane-like nanosheets *J. Phys. Chem. C* **115** 13242–6
- [12] Andriotis A N, Richter E and Menon M 2016 Prediction of a new graphene like Si₂BN solid *Phys. Rev. B* **93** 081413(R)
- [13] Ding W, Zhu J, Wang Z, Gao Y, Xiao D, Gu Y, Zhang Z and Zhu W 2017 Prediction of intrinsic two-dimensional ferroelectrics in In₂Se₃ and other III₂-VI₃ van der Waals materials *Nat. Commun.* **8** 14956
- [14] Olsen T, Andersen E, Okugawa T, Torelli D, Deilmann T and Thygesen K S 2019 Discovering two-dimensional topological insulators from high-throughput computations *Phys. Rev. Mater.* **3** 024005
- [15] Arjmandi-Tash H, Belyaeva L A and Schneider G F 2016 Single molecule detection with graphene and other two-dimensional materials: nanopores and beyond *Chem. Soc. Rev.* **45** 476–93
- [16] Qiu H, Sarathy A, Schulten K and Leburton J-P 2017 Detection and mapping of DNA methylation with 2D material nanopores *NPJ 2D Mater. Appl.* **1** 3
- [17] Furchi M M, Pospischil A, Libisch F, Burgdörfer J and Mueller T 2014 Photovoltaic effect in an electrically tunable van der Waals heterojunction *Nano Lett.* **14** 4785–91
- [18] Lee C-H et al 2014 Atomically thin p–n junctions with van der Waals heterointerfaces *Nat. Nanotechnol.* **9** 676–81
- [19] Georgiou T et al 2012 Vertical field-effect transistor based on graphene–WS₂ heterostructures for flexible and transparent electronics *Nat. Nanotechnol.* **8** 100–3
- [20] Bertolazzi S, Krasnozhan D and Kis A 2013 Nonvolatile memory cells based on MoS₂/graphene heterostructures *ACS Nano* **7** 3246–52
- [21] Choi M S, Lee G-H, Yu Y-J, Lee D-Y, Lee S H, Kim P, Hone J and Yoo W J 2013 Controlled charge trapping by molybdenum disulphide and graphene in ultrathin heterostructured memory devices *Nat. Commun.* **4** 1624
- [22] Pomerantseva E and Gogotsi Y 2017 Two-dimensional heterostructures for energy storage *Nat. Energy* **2** 17089
- [23] Zhang P, Wang F, Yu M, Zhuang X and Feng X 2018 Two-dimensional materials for miniaturized energy storage devices: from individual devices to smart integrated systems *Chem. Soc. Rev.* **47** 7426–51
- [24] Jimenez V O, Kalappattil V, Eggers T, Bonilla M, Kolekar S, Huy P T, Batzill M and Phan M-H 2019 A highly sensitive magnetic sensor using a 2D van der Waals ferromagnetic material (arXiv: 1902.08365)
- [25] Kawakami R K 2015 Spin amplification by controlled symmetry breaking for spin-based logic *2D Mater.* **2** 034001
- [26] Han W 2016 Perspectives for spintronics in 2D materials *APL Mater.* **4** 032401
- [27] Brivio J, Alexander D T L and Kis A 2011 Ripples and layers in ultrathin MoS₂ membranes *Nano Lett.* **11** 5148–53
- [28] Tapasztó L, Dumitrică T, Kim S J, Nemes-Incze P, Hwang C and Biró L P 2012 Breakdown of continuum mechanics for nanometre-wavelength rippling of graphene *Nat. Phys.* **8** 739–42
- [29] Zhu W, Low T, Perebeinos V, Bol A A, Zhu Y, Yan H, Tersoff J and Avouris P 2012 Structure and electronic transport in graphene wrinkles *Nano Lett.* **12** 3431–6
- [30] Meyer J C, Kisielowski C, Erni R, Rossell M D, Crommie M F and Zettl A 2008 Direct imaging of lattice atoms and topological defects in graphene membranes *Nano Lett.* **8** 3582–6
- [31] Komsa H-P, Kurasch S, Lehtinen O, Kaiser U and Krasheninnikov A V 2013 From point to extended defects in two-dimensional MoS₂: evolution of atomic structure under electron irradiation *Phys. Rev. B* **88** 035301
- [32] van der Zande A M et al 2013 Grains and grain boundaries in highly crystalline monolayer molybdenum disulphide *Nat. Mater.* **12** 554–61
- [33] Hsu W-T et al 2017 Evidence of indirect gap in monolayer WSe₂ *Nat. Commun.* **8** 929
- [34] Liu B, Zhao W, Ding Z, Verzhbitskiy I, Li L, Lu J, Chen J, Eda G and Loh K P 2016 Engineering bandgaps of monolayer MoS₂ and WS₂ on fluoropolymer substrates by electrostatically tuned many-body effects *Adv. Mater.* **28** 6457–64
- [35] Kang Y and Han S 2017 An origin of unintentional doping in transition metal dichalcogenides: the role of hydrogen impurities *Nanoscale* **9** 4265–71
- [36] Borys N J et al 2017 Anomalous above-gap photoexcitations and optical signatures of localized charge puddles in monolayer molybdenum disulfide *ACS Nano* **11** 2115–23
- [37] Kolesnichenko P V, Zhang Q, Yun T, Zheng C, Fuhrer M S and Davis J A 2019 Disentangling the effects of doping, strain and disorder in monolayer WS₂ by optical spectroscopy *2D Mater.* **7** 025008
- [38] Zheng C, Xu Z-Q, Zhang Q, Edmonds M T, Watanabe K, Taniguchi T, Bao Q and Fuhrer M S 2015 Profound effect of substrate hydroxylation and hydration on electronic and optical properties of monolayer MoS₂ *Nano Lett.* **15** 3096–102
- [39] Zhang Y et al 2013 Controlled growth of high-quality monolayer WS₂ layers on sapphire and imaging its grain boundary *ACS Nano* **7** 8963–71
- [40] Hu Z et al 2019 The role of oxygen atoms on excitons at the edges of monolayer WS₂ *Nano Lett.* **19** 4641–50
- [41] Kotsakidis J C, Zhang Q, de Parga A L V, Currie M, Helmerson K, Gaskill D K and Fuhrer M S 2019 Oxidation of monolayer WS₂ in ambient is a photoinduced process *Nano Lett.* **19** 5205–15
- [42] Zhou W et al 2013 Intrinsic structural defects in monolayer molybdenum disulfide *Nano Lett.* **13** 2615–22
- [43] Gutiérrez H R et al 2012 Extraordinary room-temperature photoluminescence in triangular WS₂ monolayers *Nano Lett.* **13** 3447–54
- [44] Pearson K 1901 On lines and planes of closest fit to systems of points in space *London, Edinburgh Dublin Philos. Mag. J. Sci.* **2** 559–72
- [45] Hotelling H 1933 Analysis of a complex of statistical variables into principal components *J. Educ. Psychol.* **24** 417–41
- [46] Lever J, Krzywinski M and Altman N 2017 Principal component analysis *Nat. Methods* **14** 641–2
- [47] Steinhaus H 1956 Sur la division des corps matériels en parties *Bull. Acad. Pol. Sci. Cl. III* **4** 801–4 (in French)
- [48] MacQueen J 1967 Some methods for classification and analysis of multivariate observations *Proc. 5th Berkeley Symp. Mathematical Statistics and Probability* vol 1 (Berkeley, CA: University of California Press) pp 281–97 (<https://projecteuclid.org/euclid.bsm/ps/1200512992>)

- [49] Lloyd S 1982 Least squares quantization in PCM *IEEE Trans. Inf. Theory* **28** 129–37
- [50] McIntyre J and Aspnes D 1971 Differential reflection spectroscopy of very thin surface films *Surf. Sci.* **24** 417–34
- [51] Chernikov A, Ruppert C, Hill H M, Rigosi A F and Heinz T F 2015 Population inversion and giant bandgap renormalization in atomically thin WS₂ layers *Nat. Photon.* **9** 466–70
- [52] Liu G-B, Xiao D, Yao Y, Xu X and Yao W 2015 Electronic structures and theoretical modelling of two-dimensional group-VIB transition metal dichalcogenides *Chem. Soc. Rev.* **44** 2643–63
- [53] Christopher J W, Goldberg B B and Swan A K 2017 Long tailed trions in monolayer MoS₂: temperature dependent asymmetry and resulting red-shift of trion photoluminescence spectra *Sci. Rep.* **7** 14062
- [54] Mak K F, He K, Lee C, Lee G H, Hone J, Heinz T F and Shan J 2012 Tightly bound trions in monolayer MoS₂ *Nat. Mater.* **12** 207–11
- [55] Frisenda R *et al* 2017 Micro-reflectance and transmittance spectroscopy: a versatile and powerful tool to characterize 2D materials *J. Phys. D: Appl. Phys.* **50** 074002
- [56] Lee J E, Ahn G, Shim J, Lee Y S and Ryu S 2012 Optical separation of mechanical strain from charge doping in graphene *Nat. Commun.* **3** 1024
- [57] Bao W *et al* 2015 Visualizing nanoscale excitonic relaxation properties of disordered edges and grain boundaries in monolayer molybdenum disulfide *Nat. Commun.* **6** 7993
- [58] Wang L, Xu C, Li M-Y, Li L-J and Loh Z-H 2018 Unraveling spatially heterogeneous ultrafast carrier dynamics of single-layer WSe₂ by femtosecond time-resolved photoemission electron microscopy *Nano Lett.* **18** 5172–8
- [59] Rao R, Islam A E, Singh S, Berry R, Kawakami R K, Maruyama B and Katoch J 2019 Spectroscopic evaluation of charge-transfer doping and strain in graphene/MoS₂ heterostructures *Phys. Rev. B* **99** 195401
- [60] Kastl C *et al* 2019 Effects of defects on band structure and excitons in WS₂ revealed by nanoscale photoemission spectroscopy *ACS Nano* **13** 1284–91
- [61] Morisawa M 1957 Accuracy of determination of stream lengths from topographic maps *Trans. Am. Geophys. Union* **38** 86
- [62] Pelletier J D 2013 A robust, two-parameter method for the extraction of drainage networks from high-resolution digital elevation models (DEMs): evaluation using synthetic and real-world DEMs *Water Resour. Res.* **49** 75–89
- [63] Zhang Q *et al* 2016 Strain relaxation of monolayer WS₂ on plastic substrate *Adv. Funct. Mater.* **26** 8707–14
- [64] Niehues I *et al* 2018 Strain control of exciton–phonon coupling in atomically thin semiconductors *Nano Lett.* **18** 1751–7
- [65] McCreary K M *et al* 2016 The effect of preparation conditions on raman and photoluminescence of monolayer WS₂ *Sci. Rep.* **6** 35154
- [66] McCreary K M, Currie M, Hanbicki A T, Chuang H-J and Jonker B T 2017 Understanding variations in circularly polarized photoluminescence in monolayer transition metal dichalcogenides *ACS Nano* **11** 7988–94
- [67] Kim M S, Yun S J, Lee Y, Seo C, Han G H, Kim K K, Lee Y H and Kim J 2016 Biexciton emission from edges and grain boundaries of triangular WS₂ monolayers *ACS Nano* **10** 2399–405
- [68] Sun L *et al* 2017 Vacuum level dependent photoluminescence in chemical vapor deposition-grown monolayer MoS₂ *Sci. Rep.* **7** 16714
- [69] Dhakal K P, Duong D L, Lee J, Nam H, Kim M, Kan M, Lee Y H and Kim J 2014 Confocal absorption spectral imaging of MoS₂: optical transitions depending on the atomic thickness of intrinsic and chemically doped MoS₂ *Nanoscale* **6** 13028–35
- [70] Castellanos-Gomez A, Quereda J, van der Meulen H P, Agrait N and Rubio-Bollinger G 2016 Spatially resolved optical absorption spectroscopy of single- and Few-Layer MoS₂ by hyperspectral imaging *Nanotechnology* **27** 115705
- [71] Aldenderfer M S and Blashfield R K 1984 *Cluster Analysis (Quantitative Applications in the Social Sciences)* (Newbury Park, CA: SAGE Publications, Inc)
- [72] Chandola V, Banerjee A and Kumar V 2009 Anomaly detection *ACM Comput. Surv.* **41** 1–58
- [73] Hsu C, Frisenda R, Schmidt R, Arora A, Vasconcellos S M, Bratschitsch R, Zant H S J and Castellanos-Gomez A 2019 Thickness-dependent refractive index of 1L, 2L and 3L MoS₂, MoSe₂, WS₂ and WSe₂ *Adv. Opt. Mater.* **7** 1900239
- [74] Pakdehi D M *et al* 2019 Homogeneous large-area quasi-free-standing monolayer and bilayer graphene on SiC *ACS Appl. Nano Mater.* **2** 844–52
- [75] Zhang Q *et al* 2018 Reliable synthesis of large-area monolayer WS₂ single crystals, films and heterostructures with extraordinary photoluminescence induced by water intercalation *Adv. Opt. Mater.* **6** 1701347