

Charting the low-loss region in Electron Energy Loss Spectroscopy with machine learning

Laurien Roest^{1,2}

¹*Kavli Institute of Nanoscience, Delft University of Technology, 2628CJ Delft, The Netherlands*

²*Nikhef Theory Group, Science Park 105, 1098 XG Amsterdam, The Netherlands*

³*Department of Physics and Astronomy, VU, 1081 HV Amsterdam, The Netherlands*

Abstract

Electron energy-loss spectroscopy (EELS) within the transmission electron microscope provides valuable information on the structural, chemical, and electronic properties of materials at the nanoscale. Exploiting the information contained in EEL spectra requires reliable access to the low-loss region ($\Delta E \lesssim 5$ eV), where the contribution from the zero-loss peak (ZLP) overwhelms that from the inelastic scatterings off the sample. Here we deploy machine learning techniques, inspired in particle physics applications, to realise a model-independent multidimensional determination of the ZLP with a faithful uncertainty estimate. Our method is then used to disentangle ZLP from the sample contributions in low-loss EEL spectra acquired in WS₂ nanostructures. This makes possible determining the value and type of the WS₂ bandgap as a function of the underlying crystalline morphology of the nanostructures.

Keywords: Transmission Electron Microscopy, Electron Energy Loss Spectroscopy, Neural Networks, Bandgap, Transition Metal Dichalcogenides.

Contents

1	Introduction	4
2	Fundamentals of Electron Energy Loss Spectroscopy	6
2.1	EEL spectra	7
2.2	Energy resolution	8
2.3	ZLP subtraction	9
3	Transition metal dichalcogenides and WS₂	11
4	Fundamentals of Neural Networks	13
5	Neural network determination of the ZLP	16
5.1	ZLP parametrisation	16
5.2	Uncertainty propagation	18
5.3	Training strategy	20
6	Results I. Vacuum spectra	26
6.1	Training initialization	26
6.2	Fit quality	28
6.3	Dependence on the electron energy loss	30
6.4	Dependence on beam energy and exposure time	32
6.5	Stability and reliability checks	33
7	Results II. Sample spectra	37
7.1	Training dataset	37
7.2	Subtraction procedure	39
7.3	Bandgap analysis of sample A	43
7.4	Bandgap analysis of sample B	45
8	Summary and outlook	47
A	Installation and usage of EELSfitter	49

1 Introduction

Electron energy-loss spectroscopy (EELS) within the transmission electron microscope (TEM) provides a wide range of valuable information on the structural, chemical, and electronic properties of nanoscale materials. Thanks to recent instrumentation breakthroughs such as electron monochromators [1, 2] and aberration correctors [3], modern EELS analyses can map these properties with unprecedented spatial and spectral resolution. A particularly important region of the EEL spectra is the low-loss region, defined by those electrons that have lost less than a few eV ($\Delta E \lesssim 5$ GeV) following their inelastic interactions with the sample. The analysis of this low-loss region makes possible charting the local electronic properties of nanomaterials [4], from the characterisation of bulk and surface plasmons [5], excitons [6], inter- and intra-band transitions [7], and phonons to the determination of their bandgap and band structure [8].

Provided the sample is electron-transparent, as required for TEM inspection, in EELS the bulk of the incident electron beam will traverse it either without interacting or restricted to elastic scatterings with the atoms of the sample's crystalline lattice. These electrons are recorded as a narrow, high intensity peak centered at energy losses of $\Delta E \simeq 0$ and known as the zero loss peak (ZLP). The energy resolution of EELS analyses is ultimately determined by the electron beam size of the system, often expressed in terms of the full width at half maximum (FWHM) of the ZLP [9]. In the low-loss region, the contribution from the ZLP often overwhelms that from the inelastic scatterings arising with the interactions of the beam electrons with the sample. Therefore, relevant signals of low-loss phenomena such as excitons, phonons, and intraband transitions risk being drowned in ZLP tail [10]. An accurate removal of the ZLP contribution is thus crucial in order to efficiently chart and identify the features of the low-loss region in EEL spectra.

The properties of the ZLP in monochromated EELS depend on the electron energy dispersion, the monochromator alignment, and the sample thickness [8, 11]. The first two factors arise already in the absence of a specimen (vacuum operation), while the third one is associated to elastic scatterings with the sample such as atomic scatterings, phonon excitation, and exciton losses. This implies that measurements of vacuum EEL spectra can be used for calibration purposes but not to subtract the ZLP from spectra taken on specimens, since their shapes will differ in general.

Several approaches to ZLP subtraction [8, 12, 13] have been put forward in literature. These are often based on specific model assumptions about the ZLP properties, specifically concerning its parametric functional dependence on the electron energy loss ΔE , from Lorentzian [14] and power laws [6] to more general multiple-parameter functions [15]. Another approach is based on the mirroring the $\Delta E < 0$ region of the spectra, assuming that the $\Delta E > 0$ region is fully symmetric [16]. More recent studies use integrated software applications for background subtraction methods [17–20]. These subtraction methods are however affected by three main limitations. Firstly, they rely on specific model assumptions *e.g.* with the choice of fit function, introducing a methodological bias whose size is difficult to quantify. Secondly, they lack an estimate of the associated uncertainties, which in turn affects the reliability of any physical interpretations of the low loss region such as band gap extraction. Thirdly, *ad hoc* choices of such as those of the fitting ranges introduce a significant degree of arbitrariness in the procedure.

Here we bypass these limitations by developing a model-independent strategy that makes possible a multidimensional determination of the ZLP with a faithful uncertainty estimate. Our approach is based on machine learning (ML) techniques developed in high-energy physics to study quark and gluon substructure of protons particle collisions [21–24]. It is based on the Monte Carlo replica method to construct a probability distribution in the space of experimental data and artificial neural networks as unbiased interpolators to parametrise the ZLP. The result is a faithful sampling of the probability distribution in the

space of ZLP which can be used to subtract its contribution to EEL spectra while propagating the associated uncertainties. Further, we can extrapolate this ZLP parametrisation to other TEM operation conditions beyond those used in the training dataset.

Our work is divided into two main parts. In the first one, we construct a ML model of ZLP spectra taken in vacuum able to accommodate an arbitrary number of input variables corresponding different operation conditions of the TEM. We demonstrate how the model describes successfully the input spectra and assess its extrapolation for other operation conditions. In the second part, we construct a one-dimensional model of the ZLP as a function of ΔE from spectra acquired on tungsten disulfide (WS_2) nanoflowers [25]. The resulting subtracted spectra are used to determine the value and type of the WS_2 bandgap and its dependence on the underlying crystalline morphology of these nanostructures.

The paper is organized as follows. First of all, in Sect. 2 we review the main features of the EELS technique and its application to nanostructres built upon transition metal dichalcogenide (TMD) materials, with emphasis on WS_2 . In Sect. 5 we describe our machine learning methodology for the ZLP parametrisation. Sects. 6 and 7 contain our results for the ZLP parametrisation for spectra acquired in vacuum and in sample respectively, which in the latter case allows us to probe the local band structure properties of the WS_2 nanoflowers. Finally in Sect. 8 we summarise and outline possible future developments. Our results have been obtained with an open-source PYTHON code, **EELfitter**, presented in App. A together with some installation and usage instructions.

2 Fundamentals of Electron Energy Loss Spectroscopy

Electron energy loss spectroscopy (EELS) is a method used in combination with transmission electron microscopy (TEM), which analyses the energy distribution of initially monoenergetic electrons after they have interacted with a specimen [?]. In this introductory chapter, we present an introduction to the basics of EELS, followed by an overview of energy loss spectra and the limitations of the technique.

Electron energy loss spectroscopy was developed by James Hillier and RF Baker in the mid-1940s [?], but it was only becoming more widespread in research in the 1990s due to improvements in microscope instrumentation. Since modern instrumentation became widely available in laboratories in the mid-1990s, the scientific developments regarding electron microscopes grew rapidly. Especially since the introduction of modern aberration correctors and monochromated electron sources, energy resolutions of 100 meV or even higher could be achieved [], which enabled measurements of single (columns of) atoms. Transmission electron microscopy can provide structural information with excellent spatial resolution down to atomic dimensions. These structural data can be supplemented by chemical information from the same specimen region, obtained using analytical techniques such as EELS [?]. EELS instrumentation is typically incorporated into a scanning transmission electron microscope (STEM) or in a conventional TEM (c-TEM). These microscope types use high energy electrons, typically 60 - 300 keV, to interrogate the sample. The transmitted electrons are deflected through a uniform magnetic field of the order of 0.01 T, generated by an electromagnet with carefully shaped polepieces. Electrons that scattered inelastically will stray from the central trajectory, giving rise to a greater or lesser deflection angle, and are sorted and detected according to their energies. The existence of different kinetic energies thus results in a fringing field at the EELS detector. A schematic illustration of a typical EELS setup is shown in the left panel of Fig. 2.1.

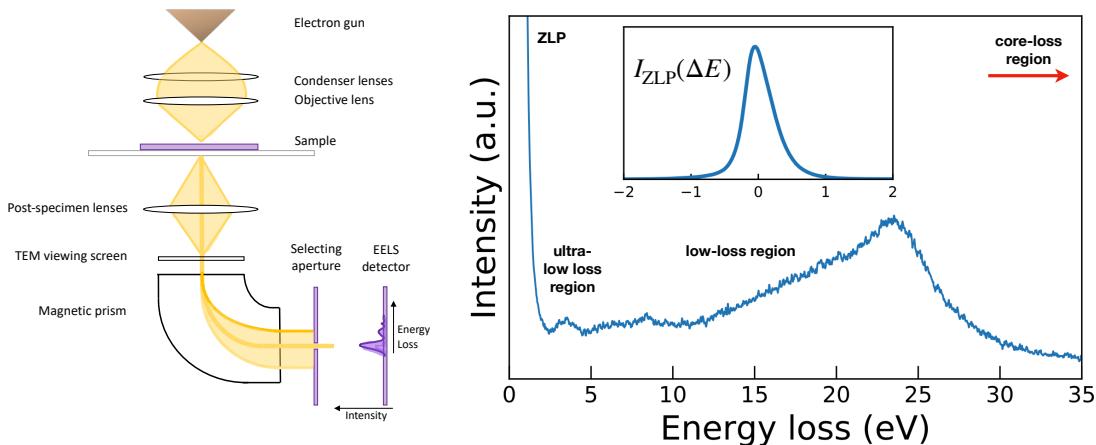


Figure 2.1. Left: in STEM-EELS, a magnetic prism is used to deflect the electron beam after crossing the sample so that the distribution of their energy losses ΔE can be recorded. Right: a representative spectrum for $\Delta E \leq 35$ eV acquired on a WS₂ nanoflower [25] with the inset displaying the corresponding ZLP.

Electron energy loss spectroscopy provides detailed information about the chemical components, bonding, and structure of materials down to the atomic scale. Thanks to recent progress in TEM instrumentation and data acquisition, state-of-the-art EELS analyses benefit from a highly competitive energy (spectral) resolution combined with an unparalleled spatial resolution. Especially scanning transmission electron microscopy (STEM) equipped with a monochromator is extremely useful for high resolution imaging [26]. The right panel of Fig. 2.1 displays a representative EELS spectrum in the region

$\Delta E \leq 35$ eV, recorded in one of the WS₂ nanostructures presented in [25] and which will be further discussed later onwards.

2.1 EEL spectra

If we are to understand how the features in electron energy loss spectra are produced, we must consider how the interaction of the incident electron with the sample contributes to the spectrum. Roughly speaking, EEL spectra can be divided into three main regions.

Zero-loss region. The first is the zero-loss region, which is centered around $\Delta E = 0$ and contains the contributions from electrons that are transmitted without suffering measurable energy loss. Provided the thickness of the sample is small, the greatest part of the incident electron beam transfers through the sample elastically, implicating the energy exchange is too less than the experimental energy resolution. A strong and narrow intensity peak around 0 eV loss can be observed called the zero loss peak (ZLP) or "elastic" peak. The width of the ZLP, typically 0.2-2 eV, reflects the energy distribution of the electron source [?].

The inset in Fig 2.1 displays the ZLP, illustrating how nearby $\Delta E \simeq 0$ its magnitude is larger than the contribution from the inelastic interactions with the sample by several orders of magnitude. This can be explained by the fact that a nucleus is thousands of times more massive than an electron, and therefore the energy transfer involved in elastic scattering is usually negligible. The probability of elastic scattering for a single incident electron (per unit solid angle Ω) can to a first approximation be described by the differential cross section [13],

$$d\sigma_e/d\Omega = \frac{4Z^2/k_0^2 T}{(\theta^2 + \theta_0^2)^2} \quad (2.1)$$

where Z is the atomic number, k_0 the electron wavenumber, and T is not the temperature but the incident electron energy. θ is the scattering angle of the electron of interest, and θ_0 represents the angular width of the scattered beam.

Low-loss region. The second region is the low-loss region or valence region, defined for energy losses $\Delta E \lesssim 50$ eV, which contains information about several important features such as plasmons, excitons, phonons, and inter- and intra-band transitions. Of particular relevance in this context is the ultra-low loss region, characterised by $\Delta E \simeq$ few eV. Here, the contributions of the ZLP and those from the inelastic interactions with the sample are of the same order of magnitude. Ranging between electron energy losses of 1 and 5 eV, difficultly distinctive features are the presence of relatively narrow exciton peaks, located close to the ZLP. The intensity of these valence features is over two orders of magnitude lower than the ZLP [?]. These peaks can be interpreted as arising from indirect and direct exciton transitions, the latter of which allows for bandgap determination [?].

The band gap refers to the energy difference between the top of the valence band and the bottom of the conduction band and the corresponding peak is expected to appear at energy losses where the joint density of states (JDOS) exhibits maxima. It has been shown by Bruley and Brown [?] that for parabolic bands, the JDOS probed by the electrons can be described by

$$\rho(E) = \frac{V}{4\pi^2} \left(\frac{2m^*}{\hbar} \right)^{(3/2)} \alpha \sqrt{E - E_{bg}} \quad (2.2)$$

for a direct band gap, where m^* is the mass of the electrons and holes in the valence and conduction band, E_{bg} is the band gap energy and α is the convergence angle of the electron beam. The nature (direct or indirect) and the band gap energy can be deduced

from the first few eV of the energy-loss function. From a fit of the band gap onset to equation 2.2, the value of $(3/2)$ for an indirect bandgap switches to $(1/2)$ for a direct bandgap, as demonstrated in Fig. 2.2.

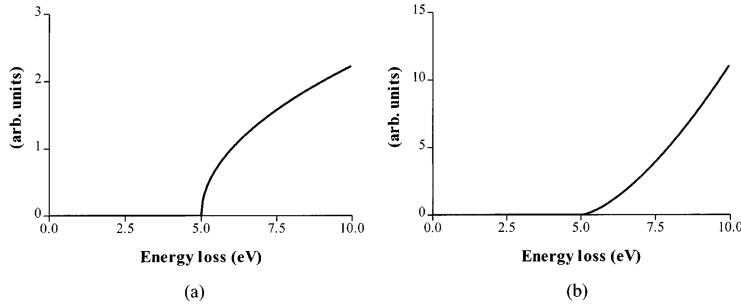


Figure 2.2. Schematic diagrams showing the contributions from the JDOS and matrix elements for (a) direct and (b) indirect transitions [?].

At even lower energy losses, $\Delta E \lesssim 100$ meV, vibrational modes can be revealed. These are the result of transmitted electrons that generate (and absorb) phonons while passing through the crystal. Phonon energies are of the order $k_b T$ and corresponding energy losses are below 0.1 eV, which requires very high resolution spectroscopy to record it. Limited vibrational spectroscopy becomes possible in an electron microscope at around 30 meV energy resolution. Vibrational spectroscopy, a field that didn't exist 5 years ago, includes vibrational mapping, analyzing the momentum dependence of vibrational states and determining local temperature from the ratio of energy gains to losses [26].

Core-loss region. The regime for $\Delta E \gtrsim 50$ eV is the core-loss region, which provides compositional information on the elements that constitute the sample. In this regime, the spectrum shows characteristic features called ionisation edges, formed when an inner-shell electron absorbs enough energy from a beam electron to be excited to a state above the Fermi level. The thicker the sample, the more prominent the ionisation edges are since multiple scattering events are more common. In this work however, the focus will be on the low-loss region of EEL spectra, therefore we will not go into more detail regarding the core ionisation peaks.

2.2 Energy resolution

The energy resolution of an EELS spectrum is determined by several factors. Firstly, aberrations of the electron spectrometer cause blur of the incoming electron beam [?]. In general, the spectrometer dispersion becomes worse for higher electron losses, and therefore the resolution at ionization edges can be worse than close to the ZLP. Imaging quality can be improved by the implementation of an aberration corrector, to cancel the spherical aberration of the objective and condenser lenses. The energy resolution is then mainly determined by the angular distribution provided by the electron beam, usually expressed as the full width at half maximum (FWHM) of the ZLP. The peak width depends strongly on the electron source. An electron microscope equipped with a thermally assisted Schottky field-emission source typically has an energy resolution of about 800 meV under normal operation conditions [27]. While this width is small compared to the operating voltage of the STEM (usually between 60-300 keV), it sets a limit for the energy resolution of EELS and thereby hinders the ability to distinguish between peaks separated by less than this value. Furthermore, for low-loss phenomena such as excitons and intraband transitions, excitation probabilities can be quite low and these signals can be lost in the tails of the ZLP. The resolution can be drastically improved by implementation of a monochromator after the electric source. In monochromators, a

small magnetic prism and energy-selecting slit are installed directly after the electron source. This setup essentially works as an energy filter: the incoming electron beam is first dispersed before going through a narrow slit, restricting the energy distribution of the incoming electrons. After compressing it back into the electron probe, the width of the electron beam and the tail intensity are greatly reduced.

In a recent studies [26], a monochromated zero loss peak was obtained with a FWHM as small as 4.2 meV and, maybe even more importantly, the tail intensity at 20 meV loss has dropped below 10^{-3} of its maximum, allowing features in the very low-loss region to be resolved. The improving energy resolution opens new possibilities for accurate measurements on the bandgap and the dielectric function.

Apart from the increase in resolution, another advantage of a monochromated electron beam is its symmetric energy distribution. Asymmetries of absorption peaks in the spectrum can thus unambiguously be attributed to the response of the material [?]. What is more, the reduction of the energy spread of the incident electron beam often comes at the expense of the beam intensity.

In Fig 2.3, one can observe the effect of a monochromator on the zero-loss peaks of a Schottky field emission microscope. Due to thermal broadening, the unfiltered Schottky field-emission source shows a broad tail at the high-energy side, i.e. at negative energy values. The tails of the monochromated beam are highly symmetrical and the energy resolution (FWHM) is greatly increased.

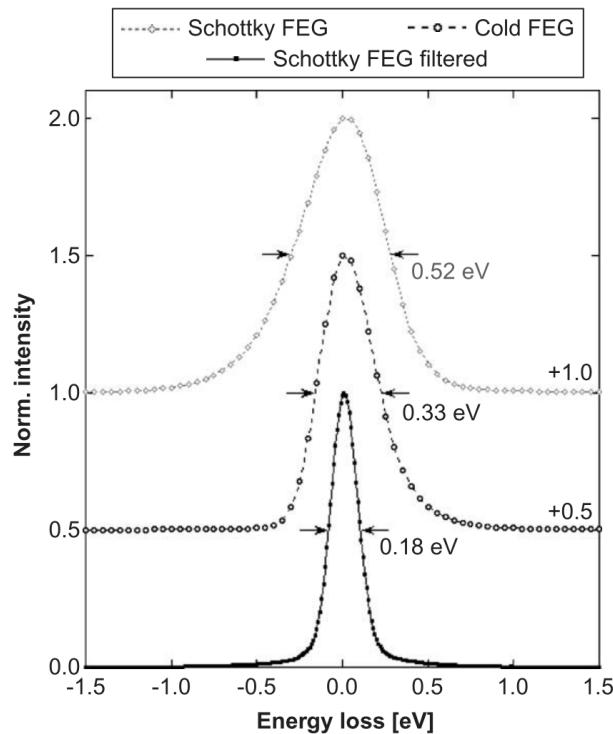


Figure 2.3. Comparison of the zero-loss peaks of an unfiltered Schottky field emission microscope (200 kV), a cold field-emission microscope (100 kV) and a monochromated electron beam with Schottky field-emission source (200 kV). The second ZLP shows a wider tail at the low-energy side and the tails of the monochromated beam are highly symmetrical. The FWHM indicated in each case provides a measure for the energy resolution.

2.3 ZLP subtraction

The most important aspect complicating the study of the low-loss regime of EEL spectra is the observation of the zero-loss peak, a very intense and ubiquitous feature which

right-hand tail suppresses the low loss features, which results in the loss of important information within the spectrum. Before analysing the low-loss region of EEL spectra, accurate removal of the ZLP is crucial. In the last several years different removal routines for the ZLP were introduced, as the increasing energy resolution of the instrumentation would allow for bandgap determination by means of VEELS. The most general suggestion is the subtraction of a fitted ZLP rather than using the vacuum recorded ZLP, because a separately recorded peak is always different in shape than the one recorded on a sample, due to phonon excitations, exciton losses, and broadening at the surface of the sample. Due to the obvious difference in shape between vacuum and on sample recorded ZLPs, direct subtraction of the first on the latter would introduce an extra source of error. For this reason, using a fitted ZLP and subtracting it from the EEL spectrum is the preferred way to go. Representative examples include the subtraction of a fitted Lorentzian distribution [14], directly subtracting the mirrored left-hand side of the ZLP [16], the subtraction of a power-law fit [6], and the use of a more general multi-parameter function [15]. These and several other attempts to model the ZLP distribution have had some success at describing the main intensity of the peak, but in the tails discrepancies can be as large as several tens of percent [28]. The standard method for background subtraction of the tails is to fit a power law to the tails, however this approach is not suitable in many circumstances [29–32]. Especially in the very-low-loss region, a simple functional fit completely removes all intensities belonging to losses within the bandgap energy. More recent studies use integrated software applications for background subtraction methods [17–20].

One common flaw of these subtraction methods is the fact that they are often based on specific model assumptions about the ZLP properties and thereby introduce a methodological bias which size is difficult to quantify. This bias arises from assumptions made *e.g.* on its functional form, symmetric properties or the fitting range that has been used, all introducing an arbitrariness to the procedure. Even more importantly, these subtraction methods lack an estimate of the associated uncertainties, which in turn affects the reliability of any physical interpretation of features that are observed in the ZLP-subtracted spectra.

Developing a model-independent strategy that allows for a determination of the ZLP with a faithful uncertainty estimate is very welcome. With the knowledge that the magnitude and shape of the ZLP intensity depends not only on the specific values of the electron energy loss ΔE , but also on other operation parameters of the TEM such as the electron beam energy E_b , the exposure time t_{exp} , the aperture width and the potential use of a monochromator, one cannot measure the ZLP for a given operating condition, for instance a high beam voltage of $E_b = 200$ keV, and expect to reproduce the ZLP distribution associated to different conditions, such as a lower beam voltage of $E_b = 60$ keV, without introducing model assumptions. Since it is not possible to compute the dependence of the ZLP on ΔE and the other operating conditions of the microscope from first principles, reliance on specific models appears to be unavoidable. Furthermore, even for identical operating conditions, the intensity of the ZLP will in general vary due to *e.g.* external perturbations such as electric or magnetic fields [12], the stability of the microscope and spectrometer electronics [33], the local environment (possibly exposed to mechanical, pressure and temperature fluctuations) and spectral aberrations [13]. Any model for the ZLP should thus account for this source of uncertainties.

3 Transition metal dichalcogenides and WS₂

In this work we will apply our machine-learning based method for describing the ZLP to a novel class of WS₂ nanostructures known as nanoflowers [25].

WS₂ is a transition metal dichalcogenide material, which belongs to a large family of materials known as 2D materials or van der Waals materials. They are named two-dimensional to emphasize their extraordinary thinness: TMDs are characterised by the remarkable property of being fully functional down to a single atomic layer. A MoS₂ monolayer is only 6.5 Å thick. Like graphite from graphene, TMD bulk crystals are formed of monolayers bound to each other by Van-der-Waals attraction. Over the past few years the exploration of these 2D layered materials has developed rapidly. In particular significant attention has been going to monolayers of transition metal dichalcogenides (TMDs), atomically thin semiconductor of the type MX_2 , here M is a transition metal atom (such as Mo or W) and X a chalcogen atom (such as S, Se, or Te). The characteristic crystalline structure of TMDs is one layer of M atoms that is sandwiched between two layers of X atoms. The electronic structure of TMDs strongly depends on the coordination of the transition metal atoms, giving rise to an array of electronic and magnetic properties [34]. In fact, most of the remarkable electronic and optical properties of TMDs can be traced back to the underlying periodic arrangements of their layers, the so-called stacking sequences [25]. Furthermore, the properties of this class of materials vary significantly with their thickness, for instance MoS₂ exhibits an indirect bandgap in the bulk form which becomes direct at the monolayer level [35]. The indirect-to-direct bandgap transition is the main reason for the interest in the use of TMDs for flexible electronics: it emphasizes the importance of the mechanical properties of these materials. It is much more difficult to uniformly deform 2D monolayers of a material compared to bulk samples, and therefore measuring on 2D systems can be challenging. One method that was developed to overcome this challenge involves bending a 2D monolayer suspended over a holey substrate and measuring the applied force and displacement ??, which made it possible to exfoliate monolayer flakes of MoS₂ which could handle 10% strain before breaking. TMDs are often combined with other 2D materials like graphene to make van der waals heterostructures, which need to be tuned in order to function as building blocks for many devices such as LEDs, solar cells, transistors, and photodetectors. This research field is still emerging and highly promising to have a big impact on future nanotechnology.

One example of a TMD exhibiting a pronounced dependence on its thickness is tungsten disulfide (WS₂), with an indirect-to-direct bandgap transition when going from bulk to bilayer or monolayer form. The effects of this transition are manifested as enhanced photoluminescence in monolayer WS₂, whereas only little emission is observed in the corresponding bulk form. WS₂ adopts a layered structure by stacking atomic layers of S-W-S in a sandwich-like configuration. Although the interaction between adjacent layers is a weak Van der Waals force, the dependence of the interlayer interaction on the stacking order of WS₂ is significant. Further applications of this material include storage of hydrogen and lithium for batteries [36].

A low-magnification TEM image of the WS₂ nanoflowers presented in [25] is displayed in the left panel of Fig. 3.1. These nanostructures are grown directly on top of a TEM substrate with holes in it. The right panel shows the magnification of a representative petal of a nanoflower, where the difference in contrast indicates terraces of varying thickness. Note that the black region corresponds to the vacuum, without substrate underneath. These WS₂ nanoflowers contain areas with different thicknesses, orientations and crystalline structures, therefore representing an ideal environment to investigate structural morphology in WS₂ with electronic properties at the nanoscale. What makes it even more interesting is that these nanoflowers display 3R/2H polytypism, an important is-

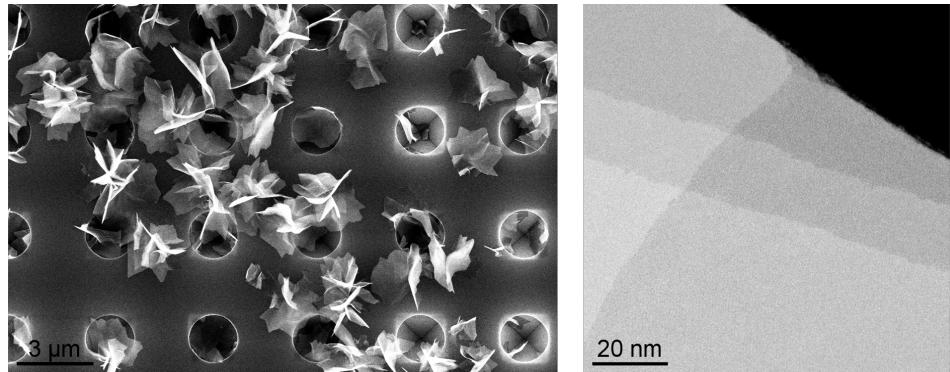


Figure 3.1. Left: low-magnification TEM image of WS_2 nanoflowers grown on top of a porous TEM substrate. Right: the magnification of a representative petal of a nanoflower, where the black region corresponds to the vacuum (no substrate) and the difference in contrast indicates terraces of varying thickness.

sue for the interlayer interactions within WS_2 : different stacking types tend to coexist, complicating the characterization of the physical properties [37]. One response of different stacking patterns to electric fields is spontaneous electrical polarization, leading to modifications on the electronic band structure and correspondingly on the band gap [?].

As mentioned before, one of the most interesting properties of TMDs that also occurs in WS_2 is the fact when the material is thinned down to a single monolayer, its indirect band gap of $E_{\text{bg}} \simeq 1.4$ eV switches to a direct band gap of approximately $E_{\text{bg}} \simeq 2.1$ eV. In general, it has been found that the type and magnitude of the bandgap of WS_2 depends quite sensitively on the crystalline structure and the number of layers that constitute the material. In Table 3.1 we collect representative results for the determination of the bandgap energy E_{bg} and its type in WS_2 , obtained by means of different experimental and theoretical techniques. For each reference we indicate separately the bulk results and those obtained at the monolayer level. We observe that for monolayers, the results for the measured value of E_{bg} are quite inconsistent, reflecting the challenges of its accurate determination.

Reference	Thickness	E_{bg} (eV)	Band gap type	Technique
[38]	bulk	1.4 ± 0.07	indirect	Gate-voltage dependence
[39]	ML	2.14	direct	Gate-voltage dependence
	bulk	1.40	indirect	
[40]	ML	2.03 ± 0.03	direct	DFT
	bulk	1.32 ± 0.03	indirect	
[41]	ML	1.76 ± 0.03	direct	Absorption edge coefficient fitting
	bulk	1.35	indirect	
[42]	ML	2.21 ± 0.3	direct	Bethe-Salpeter equation (BSE)

Table 3.1. Representative results for the determination of the bandgap energy E_{bg} and its type in WS_2 , obtained from a variety of experimental and theoretical techniques. For each reference we indicate separately the bulk results and those obtained for a monolayers

4 Fundamentals of Neural Networks

Artificial neural networks (NN) are based on the idea of simulating the functioning of neuron connections of the human brain. This machine learning technique is trained in a fashion similar to human learning with the goal to process complex inputs and conclude correct outputs [43]. A neural network is defined by a (usually large) number of neurons interconnected with strength parameters called weights.

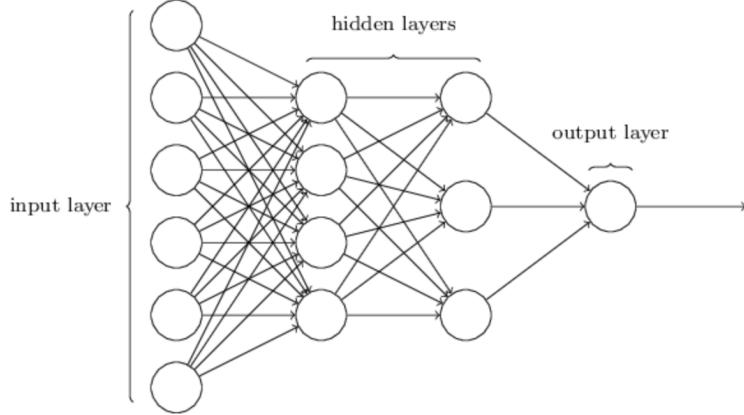


Figure 4.1. Schematic representation of a four-layer Neural Network with two hidden layers and one output neuron

Each neuron, represented in the above figure with a circle, is connected to a number of neurons in the previous layer and its output is the result of an activation function to its inputs:

$$z = \sum_j w_j x_j + b, \quad (4.1)$$

where x_j are the outputs of the preceding neurons, w_j are the corresponding weights and b is the bias (offset) of the neuron. It is the latter two that will be optimized by training. A nonlinear activation function $f(z)$ is applied to come to the output value for each neuron; this procedure is called forward propagation. Typical examples for $f(z)$ are the Rectified Linear Unit (ReLU) and sigmoid function, as depicted in Fig. 4.2. Note that these are two often-used, but not the only possible activation functions. The choice of non-linear activation function influences computational and training properties of the neural nets [44]. For example, choice for the ReLU function ensures absolute positivity.

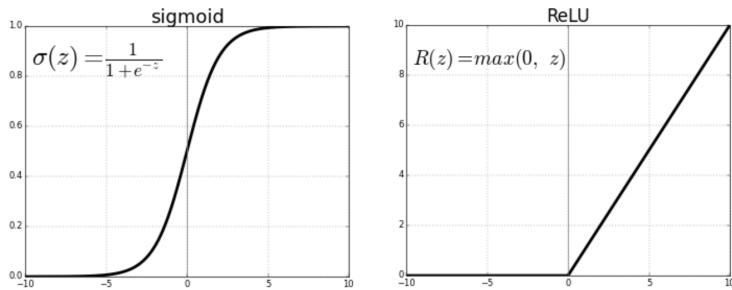


Figure 4.2. Sigmoid (left) and ReLU (right) activation functions. Both approach 0 as $z \rightarrow -\infty$ but note the different behaviour for $z \gg 0$.

For a certain set of inputs x_0 and a collection of weights and biases (w_j^l, b_j^l) , with w_j^l being the j -th neuron in the l -th layer. The neural network gives an output y that can be compared to the desired output y_{target} . A cost function $C(y, y_{target})$ is used to quantify

how 'far' we are from our desired output, and training the algorithm is done by minimizing the cost function with respect to all (w_j^l, b_j^l) of the system. This minimization is done by means of the gradient descent method, which means that after each iteration the weights and biases are adjusted a bit towards the minimum of the cost function. To make this minimization computationally tractable, the so-called back-propagation algorithm is being used [45], allowing us to approximate the derivative of the cost function by averaging over the training set. All together, training the network relies on the following four fundamental equations:

$$\begin{aligned}\delta_j^L &= \frac{\partial C}{\partial a_j^L} f'(z_j^L), \\ \delta^l &= (w^{l+1} \delta^{l+1}) \odot f'(z^l), \\ \frac{\partial C}{\partial b_j^l} &= \delta_j^l, \\ \frac{\partial C}{\partial w_{jk}^l} &= a_k^{l-1} \delta_j^l.\end{aligned}\tag{4.2}$$

Here, δ_j^l is the error and a_j^l the output of neuron j in layer l , C is the cost, b is the bias and w is the weight of each neuron. It is the error δ_j^L that represents the total cost of the network. The last two equations evaluate the gradient of the model parameters with the gradient of the cost function. With this info one can update the model parameters by

$$\begin{aligned}b^{l+1} &= b^l - \eta \frac{\partial C}{\partial b^l} \\ w^{l+1} &= w^l - \eta \frac{\partial C}{\partial w^l}\end{aligned}\tag{4.3}$$

where η is the pre-defined learning rate, a measure for the size of the step taken with each iteration. From equations (4.2), one can see that the choice of non-linear activation function ($f(z)$) influences the learning of the network. This can already be deduced from the shape of the functions in Fig. 4.2: since $\sigma(z)$ saturates for large inputs $z \gg 0$, $d\sigma/dz \rightarrow 0$ and the network loses sensitivity [44].

We use the definition *epoch* for each time that the entire set of training data is passed forward once through the neural network and the network has backpropagated the error and updated all of its parameters accordingly. The network needs much more than just one epoch to optimize by means of gradient descent, as this is an iterative process. After each epoch, the performance of the network is evaluated by calculating the error (δ_j^L) on the training data and the weights and biases are adjusted accordingly. As the number of epochs increases, the network parameters are adjusted repeatedly and where the network was first underfitting the inputs at the beginning, at a certain moment it goes to optimal fitting, before it enters the overfitting regime. Several methods can be applied to determine the optimal stopping point of the network, that is, to find the moment at which the network is neither under-, nor overfitting the training data. One of such is splitting the total set of experimental data into two sets: the training dataset is the one we use to train the model, usually 80% of the total set. The other 20% is what we call the validation set and it is used to provide an unbiased evaluation of the model fit on the training set. This split ratio is common for models with a moderate number of hyperparameters: increasing the size or increasing the tunability of the model usually goes with increasing the share of the validation set. The validation subset is left out of the training set on purpose and the model can not learn on this data points. After each epoch, the total performance of the system can be validated by feeding this subset to the network and calculating the total

error on this data. Tracing the cost function on both the training and the validation set gives insight in if the network is overfitting the training data. First, both the training and validation error will be decreasing, but at a certain point the network will start overfitting and the validation error slowly starts to increase. The optimal stopping point is defined as the global minimum of the error of the validation sample, computed over a large fixed number of iterations. A typical progress of the training and validation error over the course of the optimization can be observed in Fig. ??.

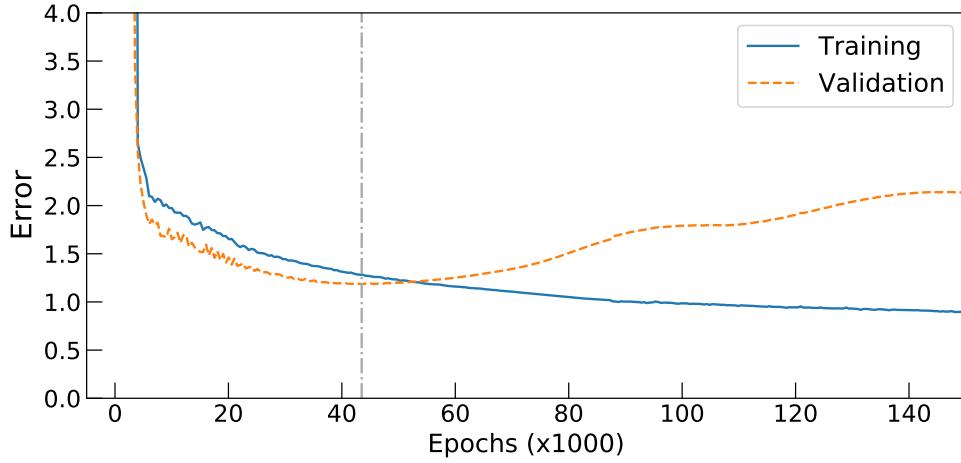


Figure 4.3. Progress of both the training and validation error over one training session as a function of the number of epochs. The optimal stopping point is where the validation error is at its absolute minimum, here after 43,500 epochs.

Once the optimal network parameters have been determined and stored, the network can be used to make predictions on any set of inputs.

5 Neural network determination of the ZLP

In this section we present our strategy to parametrise, predict and subtract the zero loss peak by means of machine learning. As mentioned in the introduction, our strategy will be inspired by the NNPDF method [46] originally developed in the context of high-energy physics for studies of the quark and gluon substructure of the proton [47]. The NNPDF approach has been successfully applied, among others, to the determination of unpolarised [21–24, 48] and polarised parton distributions functions of protons, nuclear parton distributions [49, 50], and the fragmentation functions of partons into neutral and charged hadrons [51, 52]. Neural networks benefit from the ability to parametrise multidimensional input data with arbitrarily non-linear dependencies: even with a single hidden layer, a neural network can reproduce arbitrary functional dependencies provided it has a large enough number of neurons. We can therefore apply a similar procedure for the determination of the functional dependence of the ZLP intensity.

We note that recently several applications of machine learning to transmission electron microscopy analyses in the context of material science have been presented, see *e.g.* [53–59]. Representative examples include the automated identification of structural information at the atomic scale [57] and the extraction of chemical information and defect classification [58]. For the readout of EEL spectra specifically, machine learning has been put forward for the prediction of spectral features in the core-loss regime [?]. To the best of our knowledge, this is the first time that neural networks are used as unbiased background-removal interpolators and that they are used combined with Monte Carlo sampling to construct an estimate of the model uncertainties.

In this section, we discuss the parametrisation of the ZLP in terms of neural networks. The ultimate goal is to create a model that is able to predict the contribution I_{ZLP} in the total intensity profile of any EEL spectrum recorded over a specimen, and subsequently to subtract this distribution from the spectrum to isolate the inelastic scattering contributions. In order to do so, we first need to develop a model that is able to predict the general shape of the zero loss peak as a function of its input parameters. For this we use in-vacuum zero loss peak recordings, which function as a baseline to create this generic, multidimensional model. In this regard we also explain the Monte Carlo replica method, which is used to estimate and propagate the uncertainties from the input data to the model predictions. After this we move on to the training strategy on sample spectra: we explain how the method is modified to use it on spectra recorded over WS₂ specimens and how one can select the hyper-parameters that appear in the model.

5.1 ZLP parametrisation

Without any loss of generality, we can decompose recorded the intensity profile in any EEL spectrum as

$$I_{\text{EEL}}(\Delta E) = I_{\text{ZLP}}(\Delta E) + I_{\text{inel}}(\Delta E), \quad (5.1)$$

where ΔE is the measured electron energy loss; I_{ZLP} is the zero loss distribution arising both from instrumental origin and from elastic interactions; and $I_{\text{inel}}(\Delta E)$ contains the contributions from the electrons that have undergone inelastic scattering with the specimen. It is the latter contribution that we're particularly interested in, but in order to get hold of it we need to disentangle it from the zero loss contribution. As shown by the representative example of Fig. 2.1, there are two limits for which one can straightforwardly separate the two contributions. The first region is for sufficiently high energy losses, where I_{ZLP} vanishes and $I_{\text{EEL}} \rightarrow I_{\text{inel}}$. Secondly, in the region close to zero, all emission can be associated to the ZLP such that $I_{\text{EEL}} \rightarrow I_{\text{ZLP}}$. It is the region in between that is of

particular interest, the ultra-low-loss region where I_{ZLP} and I_{inel} become of comparable order of magnitude.

Our goal is to construct a parametrisation of I_{ZLP} based on artificial neural networks, which we denote by $I_{\text{ZLP}}^{(\text{mod})}$, which allows us to extract the relevant inelastic contribution by subtracting the contribution of the ZLP from the total EEL spectra:

$$I_{\text{inel}}(\Delta E) \simeq I_{\text{EEL}}(\Delta E) - I_{\text{ZLP}}^{(\text{mod})}(\Delta E). \quad (5.2)$$

Isolating I_{inel} from the total spectrum makes us able to exploit the physical information contained in the low-loss region. Crucially, we aim to estimate and propagate all the relevant sources of uncertainty associated both to the input data and to methodological choices. This helps us to verify our results and to separate lucky findings from real insights.

As discussed in Sect. 2, the ZLP depends both on the value of the electron energy loss ΔE as well as on the operating conditions of the microscope, such as the electron beam energy E_b and the exposure time t_{exp} . Therefore we want to construct a multidimensional model which can theoretically take any number of relevant variables as input, in order to reproduce the predicted zero loss peak. This means that in general Eq. (5.2) can be written as

$$I_{\text{inel}}(\Delta E) = I_{\text{EEL}}(\Delta E, E_b, t_{\text{exp}}, \dots) - I_{\text{ZLP}}^{(\text{mod})}(\Delta E, E_b, t_{\text{exp}}, \dots), \quad (5.3)$$

where we note that the subtracted spectrum $I_{\text{inel}}(\Delta E)$ should depend only on the energy loss, but not on the microscope parameters. Ideally, the ZLP model should be able to accommodate as many input variables as possible. The output of the ZLP model is parametrised by means of multi-layer feed-forward artificial neural networks. This means that the predicted zero loss peak intensity can be expressed as

$$I_{\text{ZLP}}^{(\text{mod})}(\Delta E, E_b, t_{\text{exp}}, \dots) = \xi_1^{(n_l)}(\Delta E, E_b, t_{\text{exp}}, \dots), \quad (5.4)$$

where $\xi_1^{(n_l)}$ is the activation state of the single output neuron in the last of the n_l layers of the network. Here the n_I inputs are the variables $\{\Delta E, E_b, t_{\text{exp}}, \dots\}$ that represent the relevant information about the operating conditions during the recording of the spectra. Note that for this work, we have used spectra recorded under the known conditions $\{\Delta E, E_b, t_{\text{exp}}\}$. This set of inputs could potentially be extended by including extra variables, *e.g.* aperture width, aberration correction and temperature. The neural network is then trained by means of supervised learning and non-linear regression on these n_I inputs, using the known corresponding ZLP intensities $I_{\text{ZLP}}^{(\text{mod})}$ as training outcomes. Each iteration, the weights and thresholds of this neural network model are optimized from the minimization of the error on this training dataset.

The number of hidden layers and neurons that is optimal is very task dependent and should therefore be decided arbitrarily, there is no general rule of thumb. We have chosen to use an n_I -10-15-5-1 architecture with three hidden layers, which corresponds to a total number of 289 (271) free parameters for $n_I = 3$ ($n_I = 1$) to be optimised. However, we have verified that results are fairly independent of this exact choice: predictions on the training data did not change significantly when the architecture was increased by a factor of two.

A schematic representation of this model is displayed in Fig. 5.1. The input is an n_I array containing ΔE and the rest of operation variables of the microscope, and the output is the value of the intensity of the ZLP distribution associated to those input variables. We use a sigmoid activation function for the three hidden layers and a ReLU for the final one. The choice of ReLU for the final layer guarantees that our model for the ZLP is positive-definite, as required by general physical considerations: the intensity count can

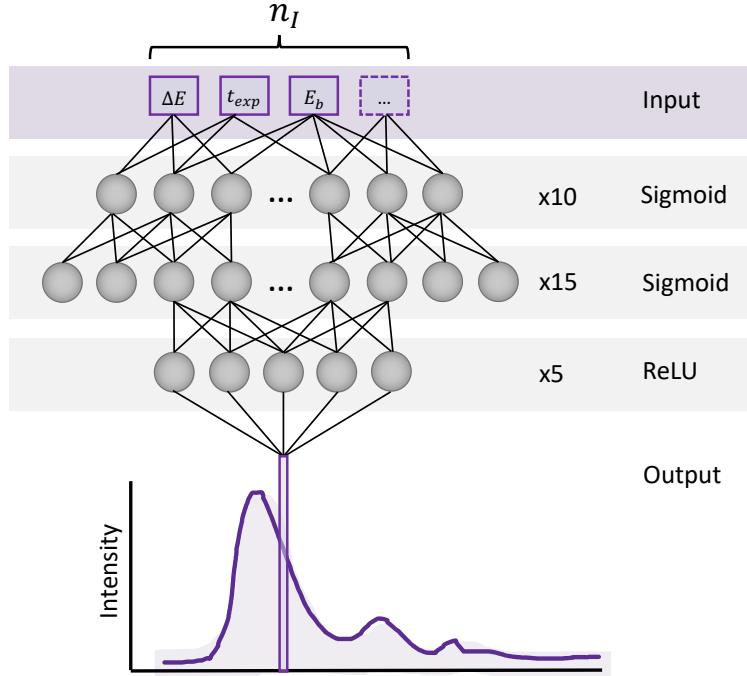


Figure 5.1. Schematic representation of our ML model for the ZLP, Eq. (5.4). The input is an n_I -dimensional array containing ΔE and other operation variables of the microscope such as E_b and t_{exp} . The output is the predicted value of the intensity of the zero-loss peak distribution associated to those specific input variables. The architecture is chosen to be n_I -10-15-5-1, with sigmoid activation functions in all layers except for a ReLU in the output neuron.

never be smaller than zero. We have adopted a redundant architecture to ensure that the ZLP parametrisation is sufficiently flexible, which means that this way we guarantee that the network can over-fit on the training inputs. However, the final results should be evaluated before the network starts overfitting, as described in Sect. ???. This means that we need to define a suitable regularisation strategy, which will be explained later onwards in Sect. 5.3.

5.2 Uncertainty propagation

We discussed in Sect. 2 how even for EEL spectra taken at identical operating conditions of the microscope, in general the resulting ZLP intensity profiles will be different. Also, the input data can be described by a large number of different neural network configurations, each with a different functional form of $I_{\text{ZLP}}^{(\text{mod})}$ but representing the data equally well. The Monte Carlo replica method can be used to estimate these two sources of uncertainties, introduced by the experimental data and the methodology, and to propagate them to physical predictions. The basic idea is twofold: first, it is useful to represent problems with a possibility of non-gaussian errors through the use of their central values and uncertainties, which are obtained from a Monte Carlo sample as their averages and standard deviations. Second, when a problem requires a reconstruction of discrete sampling without making assumptions on its functional form, neural networks are useful to work as unbiased interpolators. It is the combination of both techniques that explains the use of neural networks to separate a smooth signal from background signals, while the MC samples handles the fluctuations within the data.

The MC strategy is schematically summarized in figure ?? and it involves two stages. In the first, we generate an ensemble of replicas of the original training set. Then, each replica is passed through an individual neural network, which thereby outputs a parametrisation of the ZLP. Any physical observables can be calculated over the set of computed ZLP distributions.

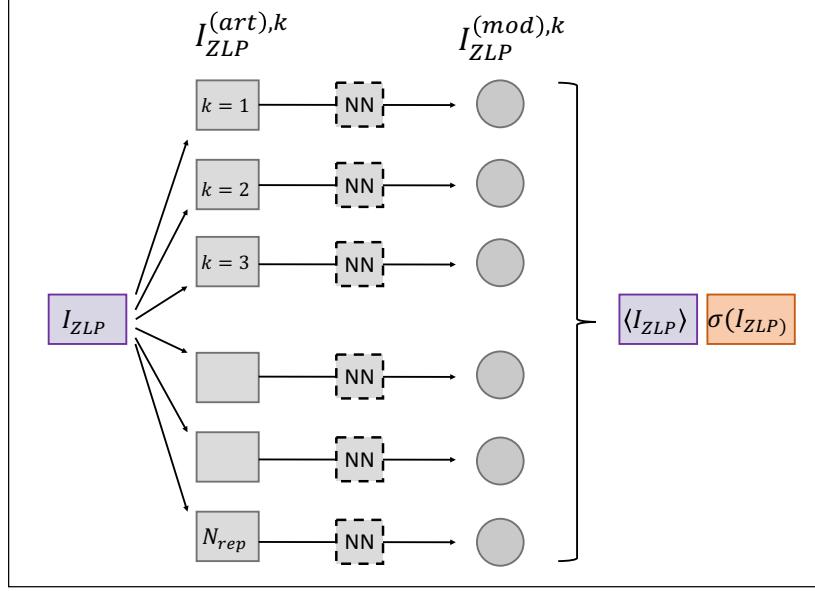


Figure 5.2. Representation of the Monte Carlo replica strategy. From the original set of training data, an ensemble of replicas is created from the experimental central values and uncertainties. On each replica, an individual neural network is trained and a ZLP parametrisation is obtained. The total set of predictions is then used to calculate physical observables from the corresponding central values and uncertainties.

Let us assume that we have n_{dat} independent measurements of the ZLP intensity, so our training dataset contains n_{dat} data points, all with different combinations of input parameters. The collective of inputs are given as $\{z_i\}$:

$$I_{\text{ZLP},i}^{(\text{exp})}(\{z_i\}) = I_{\text{ZLP},i}^{(\text{exp})}(\Delta E_i, E_{b,i}, t_{\text{exp},i}, \dots), \quad i = 1, \dots, n_{\text{dat}}. \quad (5.5)$$

The Monte Carlo method is based on the generation of a large number N_{rep} of Monte Carlo replicas of these original data points by means of a multi-Gaussian distribution, where we use the central values and covariance matrices from the input measurements. The strategy of the generation of Monte Carlo replicas goes as follows: we create for each original data point ($I_{\text{ZLP},i}^{(\text{exp})}$) an ensemble of artificial pseudo points (replicas):

$$I_{\text{ZLP},i}^{(\text{art})(k)} = I_{\text{ZLP},i}^{(\text{exp})} + r_i^{(\text{stat},k)} \sigma_i^{(\text{stat})} + \sum_{j=1}^{n_{\text{sys}}} r_{i,j}^{(\text{sys},k)} \sigma_{i,j}^{(\text{sys})}, \quad \forall i, \quad k = 1, \dots, N_{\text{rep}}. \quad (5.6)$$

where $\sigma_i^{(\text{stat})}$ and $\sigma_{i,j}^{(\text{sys})}$ represent the statistical and systematic uncertainties (the latter divided into n_{sys} fully point-to-point correlated sources) and $\{r_i^{(k)}\}$ are Gaussianly distributed random numbers. In the end, each k -th replica contains exactly as many data points as the original set.

In our case we have no information on experimental correlations between the measurements and for this reason we assume that there is only one single source of point-by-point systematic uncertainty, which is uncorrelated. Should in the future correlations become available, it would be straightforward to extend our model to that case. In other words, to each data point we associate an individual uncertainty and we discard covariances between datapoints, which means that we drop the last term in Eq. (5.6) and it reduces to

$$I_{\text{ZLP},i}^{(\text{art})(k)} = I_{\text{ZLP},i}^{(\text{exp})} + r_i^{(\text{stat},k)} \sigma_i^{(\text{stat})} \quad (5.7)$$

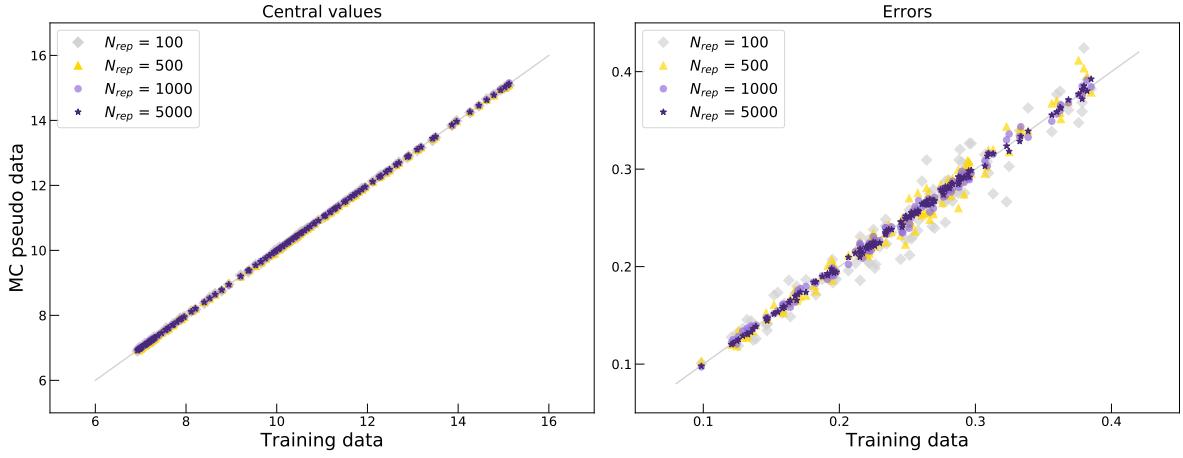


Figure 5.3. Comparison between the original experimental central values $I_{\text{ZLP},i}^{\text{exp}}$ (left panel) and the corresponding statistical uncertainties $\sigma_i^{(\text{stat})}$ with the results of averaging over a sample of N_{rep} Monte Carlo replicas generated by means of Eq. (5.6), for different values of N_{rep} .

These statistical errors on the training data can be derived by means of what is called equal width discretization (EWD) and it works as follows. The input measurements will be composed in general on subsets of EEL spectra taken with identical operation conditions. For example, we have one specific set of N_{sp} spectra all recorded with the same exposure time and beam energy. Since the range of ΔE over which the spectra have been recorded is usually not the same in each case, first of we uniformise a common binning in ΔE with n_{dat} entries. Then we evaluate the total experimental uncertainty in one of these bins as

$$\sigma_i^{(\text{exp})} = \left(\frac{1}{N_{\text{sp}} - 1} \sum_{l=1}^{N_{\text{sp}}} \left(I_{\text{ZLP},i}^{(\text{exp}),l} - \left\langle I_{\text{ZLP},i}^{(\text{exp})} \right\rangle_{N_{\text{sp}}} \right)^2 \right)^{1/2}, \quad i = 1, \dots, n_{\text{dat}}, \quad (5.8)$$

that is, $\sigma_i^{(\text{exp})}$ is the standard deviation in bin i calculated over the N_{sp} spectra. This uncertainty is separately evaluated for each set of microscope operation conditions for which data available. The value of the number of generated MC replicas, N_{rep} , should be chosen such that the set of replicas models accurately the probability distribution of original training data.

To verify that this is the case, Fig. 5.3 displays a comparison between the original experimental central values $I_{\text{ZLP},i}^{\text{exp}}$ (left) and the corresponding total uncertainties $\sigma_i^{(\text{exp})}$ (right panel) with the results of averaging over a sample of N_{rep} Monte Carlo replicas generated by means of Eq. (5.6) for different number of replicas. We find that $N_{\text{rep}} = 500$ is a value that ensures that both the central values and uncertainties are reasonably well reproduced, and we adopt it in what follows.

5.3 Training strategy

The training of the neural-network model for the ZLP peak differs between the cases of EEL spectra taken on vacuum, where by construction $I_{\text{EEL}}(\Delta E) = I_{\text{ZLP}}^{(\text{mod})}(\Delta E)$, and for spectra taken on samples. In fact, EEL spectra taken in the vacuum but close to the sample might still receive inelastic contributions coming from the specimen. When we talk about vacuum spectra, we consider exclusively those taken reasonably far from the surface of the analysed nanostructures. In the latter case, we need to find a training strategy to make predictions about the ZLP distribution in the low-loss regime, while we can not directly train on data in this region. As indicated by Eq. (5.2), in order to avoid biasing the results it is important to ensure that the model is trained only on the region of the spectra where the ZLP dominates over the inelastic scatterings. Then we

can make predictions in the extrapolation regime based on these training parameters. We now describe the training strategy that is adopted in both cases.

Training of vacuum spectra. For each of the N_{rep} generated Monte Carlo replicas, we train an independent neural network as described in Sect. 5.1. Fitting the neural networks parameters to the data is performed by minimisation of a figure of merit (error function), defined as:

$$E^{(k)}(\{\theta^{(k)}\}) = \frac{1}{n_{\text{dat}}} \sum_{i=1}^{n_{\text{dat}}} \left(\frac{I_{\text{ZLP},i}^{(\text{art})(k)} - I_{\text{ZLP},i}^{(\text{mod})}(\{\theta^{(k)}\})}{\sigma_i^{(\text{exp})}} \right)^2, \quad (5.9)$$

which is the χ^2 per data point comparing each replicated ZLP intensity with the corresponding model prediction. In this expression $E^{(k)}(\{\theta^{(k)}\})$ is the error for the values $\{\theta^{(k)}\}$ of the network weights and thresholds; $I_{\text{ZLP},i}^{(\text{art})(k)}$ is the k -th replica for the ZLP intensity and $I_{\text{ZLP},i}^{(\text{mod})}$ is the model prediction on this replica; and $\sigma_i^{(\text{exp})}$ again represents the error associated to this experimental data point.

The chi-squared method is the cornerstone of almost all fitting, as it is an intuitively reasonable measure of how well the predictions fit the data. If the model predictions are all within one standard deviation from the data, then the χ^2 per data point takes a value roughly equal to 1. In general, if χ^2/n_{dat} is of the order 1, we can say that the fit is a good approximation to the real data.

In order to speed up the neural network training process, prior to the optimisation all inputs and outputs are scaled to lie between [0.1, 0.9] before being feed to the network. This preprocessing facilitates that the neuron activation states will typically lie close to the linear region of the sigmoid activation function.

The contribution to the figure of merit from the input experimental data, Eq. (5.9), needs in general to be complemented with that of theoretical constraints on the model. For instance, when determining nuclear parton distributions [50], one needs to extend Eq. (5.9) with Lagrange multipliers to ensure that both the $A = 1$ proton boundary condition and the cross-section positivity are satisfied. In the case at hand, our model for the ZLP should satisfy the constraint that $I_{\text{ZLP}}(\Delta E) \rightarrow 0$ when $|\Delta E| \rightarrow \infty$, since far from $\Delta E \simeq 0$ the contribution from the ZLP is completely negligible. In order to implement this constraint, we add n_{pd} pseudo-data points to the training dataset "far" away from the ZLP and we modify the figure of merit Eq. (5.9) accordingly,

$$E^{(k)}(\{\theta^{(k)}\}) \rightarrow E^{(k)}(\{\theta^{(k)}\}) + \lambda \sum_{i'=1}^{n_{pd}} \left(I_{\text{ZLP},i'}^{(\text{mod})}(\{\theta^{(k)}\}) \right)^2, \quad (5.10)$$

where λ is a Lagrange multiplier whose value is tuned to ensure that the $I_{\text{ZLP}}(\Delta E) \rightarrow 0$ condition is satisfied without affecting the description of the training dataset. In other words, we add a contribution to the error function coming from these pseudo points in the higher-loss regime, making sure that the optimization of the error function also takes into account the fact that at higher losses the intensity should vanish. However, when calculating the overall fit quality after training, we do not include the second term on the right-hand side of Eq. (5.10) in the evaluation of the figure of merit, as these "fake" datapoints do not give any physical information about the original dataset whatsoever.

The pseudo-data points are added in the region $[\Delta E_{\text{pd}}^{(\text{min})}, \Delta E_{\text{pd}}^{(\text{max})}]$, and symmetrically for negative energy losses. The value of $\Delta E_{\text{pd}}^{(\text{min})}$ can be determined automatically

by evaluating the ratio \mathcal{R}_{sig} between the central experimental intensity and the total uncertainty in each data point,

$$\mathcal{R}_{\text{sig}}(\Delta E_i) \equiv \frac{I_{\text{ZLP}}^{(\text{exp})}(\Delta E_i)}{\sigma^{(\text{exp})}(\Delta E_i)}, \quad (5.11)$$

which corresponds to the statistical significance for the i -th bin of ΔE to differ from the null hypothesis (zero intensity) taking into account the experimental uncertainties. For sufficiently large energy loss this ratio approaches 1, which indicates that we are essentially fitting statistical noise. In order to avoid this and only fit data that is different from zero within errors, we determine the value of $\Delta E_{\text{pd}}^{(\min)}$ from the condition that $\mathcal{R}_{\text{sig}}(\Delta E_i) \simeq 1$.

We keep the training data in the region $\Delta E \leq \Delta E_{\text{pd}}^{(\min)}$ and the pseudo-data points are added for $[\Delta E_{\text{pd}}^{(\min)}, \Delta E_{\text{pd}}^{(\max)}]$. The value of $\Delta E_{\text{pd}}^{(\max)}$ can be chosen arbitrarily and can be as large as necessary to ensure that $I_{\text{ZLP}}(\Delta E) \rightarrow 0$ as $|\Delta E| \rightarrow \infty$.

We note that another important physical condition on the ZLP model, namely its positivity (since in EEL spectra the intensity is just a measure of the number of counts in the detector for a given value of the energy loss) is automatically satisfied since we use a ReLU activation function for the last layer. A further obvious requirement is that the best fit is independent of any assumptions made about the ZLP distribution. This requirement can be met by making sure the parametrisation is redundant: the size of the neural network used, and therefore the number of optimizable parameters, is much larger than the minimum required in order to reproduce the data. This redundancy can be verified *a posteriori*, that results are independent of the size and architecture of the neural network.

In this work we adopt the **TensorFlow** neural-net libraries to assemble the architecture illustrated in Fig. 5.1. Before training, all weights and biases are initialized in a non-deterministic order by the built-in global variable initializer. The optimisation of the figure of merit Eq. (5.10) is carried out by means of stochastic gradient descent (SGD) combined with backpropagation. The specific SGD optimizer used is the Adam algorithm. The hyper-parameters of the optimisation algorithm such as the learning rate have been adjusted to ensure proper learning is reached in the shortest amount of time possible. Given that we have a extremely flexible parametrisation, one should be careful to avoid overlearning the input data. Here over-fitting is avoided by means of a cross-validation stopping criterion. We separate the input data into training a validation subsets, with a 80%/20% splitting which partition varies randomly for each Monte Carlo replica. We then run the optimiser for a very large number of iterations and store both the state of the network and the value of the figure of merit Eq. (5.9) restricted to the validation dataset, $E_{\text{val}}^{(k)}$ (which is not used for the training). We are not looking for the absolute minimum of the error function, but we rather search for an optimal stopping point to avoid overfitting. At this optimal stopping point, it reproduces the information contained in the dataset, but not its statistical fluctuations. This point can be determined by the formulation of a stopping criterion, making it possible to, once the network completed the training, choose the parametrization of the network weights right before it entered the overlearning regime. The optimal stopping point is determined for each replica by keeping track of the figure of merit, which typically evolves as depicted in Fig. ???. The specific network configuration that leads to the deepest minimum of $E_{\text{val}}^{(k)}$: is chosen *a posteriori*, that's why it is called look-back stopping, a method that has been widely used in the context of neural networks. The number of epochs should be chosen high enough to reach for each replica the absolute minimum of $E_{\text{val}}^{(k)}$, rather than a local minimum. For this work we need approximately 40,000 epochs to guarantee overlearning. This corresponds to a serial running time of 60 seconds per replica when running the optimization on a

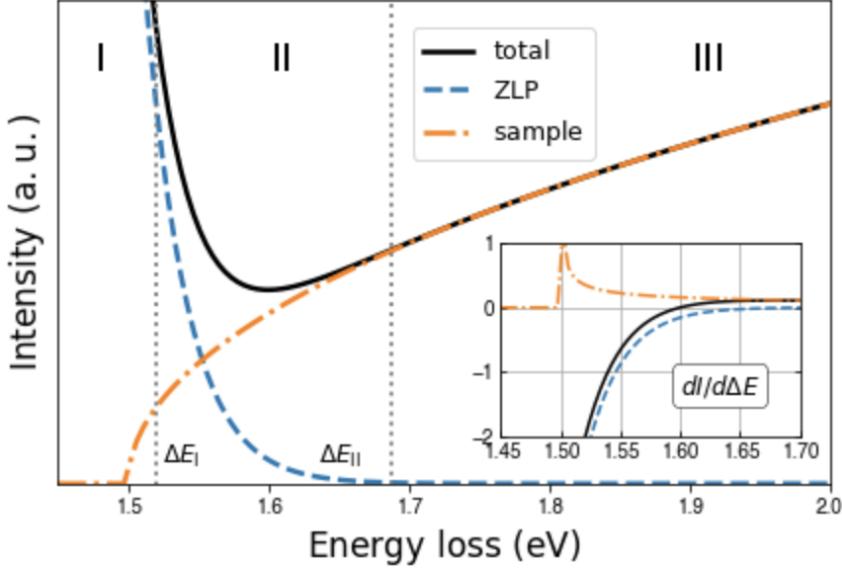


Figure 5.4. A toy model for a generic EEL spectrum and its derivatives (in the inset). We show the separate contributions from I_{ZLP} and I_{inel} as well as their sum. We indicate the two regions used for the model training (*I* and *III*), while the trained model is interpolated to region *II*, defined for $\Delta E_I \leq \Delta E \leq \Delta E_{II}$.

single CPU for a set of 500 datapoints. Once the training of all the N_{rep} neural network models for the ZLP has been carried out as specified above, we quantify the overall fit quality of the model by computing the χ^2 defined as

$$\chi^2 = \frac{1}{n_{\text{dat}}} \sum_{i=1}^{n_{\text{dat}}} \left(\frac{I_{\text{ZLP},i}^{(\text{exp})} - \langle I_{\text{ZLP},i}^{(\text{mod})} \rangle_{\text{rep}}}{\sigma_i^{(\text{exp})}} \right)^2, \quad (5.12)$$

which is the analog of Eq. (5.12) now comparing the average model prediction to the original experimental data values. A value $\chi^2 \simeq 1$ indicates that a satisfactory description of the experimental data, within the corresponding uncertainties, has been achieved. Note that in realistic scenarios χ^2 can deviate from unity, for instance when some source of correlation between the experimental uncertainties has been neglected or when the errors on the data points have been over- or underestimated.

Training of sample spectra. The training strategy in the case of EEL spectra taken on samples (rather than on vacuum) must be adjusted to account for the fact that the input data set, Eq. (5.1), receives contributions both from the ZLP and from inelastic scatterings. To avoid biasing the ZLP model, we should make sure to include only the former contributions in the training dataset. Else, if we were to train the neural network on data that contains also inelastic contributions, subsequent subtraction to the EEL spectra would make us lose important information.

We can illustrate the situation at hand with the help of a toy model for the low-loss region of EEL spectra, represented in Fig. 5.4. Let us use a general assumption that the ZLP is described by a Gaussian distribution with a standard deviation of $\sigma_{\text{ZLP}} = 0.3$ eV, and that the contribution from the inelastic interactions with the sample can be approximated in the low-loss regime by $I_{\text{inel}}(\Delta E) \propto (\Delta E - E_{\text{bg}})^b$ with $E_{\text{bg}} = 1.5$ and $b = 0.5$. The motivation for this choice will be explained in Sect. 7. We display the separate contributions from I_{ZLP} and I_{inel} , as well as the total intensity, with the inset showing the values of the corresponding derivatives, $dI/d\Delta E$.

The toy model of Fig. 5.4 is sufficiently general to draw a number of useful considerations concerning the relation between I_{ZLP} and I_{inel} in realistic spectra:

- The ZLP intensity, $I_{\text{ZLP}}(\Delta E)$, is a monotonically decreasing function and thus its derivative is always negative.
- The first local minimum of the total spectrum, $dI_{\text{EELS}}/d\Delta E|_{\Delta E_{\min}} = 0$, corresponds to a value of ΔE for which the contribution from the inelastic emissions is already significant.
- The value of ΔE for which I_{inel} starts to contribute to the total spectrum corresponds to the position where the intensity derivatives in-sample and in-vacuum start to differ. We note that a direct comparison between the overall magnitude of the sample and vacuum ZLP spectra is in general not possible, as explained in Sect. 2.

These considerations suggest that when training the ML model on EEL spectra taken on samples, the following categorisation should be adopted:

1. For small energy losses such that $\Delta E \leq \Delta E_I$ (region *I*), the model training proceeds in the same way as for the vacuum case via the minimisation of Eq. (5.9).
2. For $\Delta E \geq \Delta E_{II}$ (region *III*), we use instead Eq. (5.10) without the contribution from the input data, since for such values of ΔE one has that $I_{\text{inel}} \gg I_{\text{ZLP}}$. In other words, we only fit on the pseudo data and therefore the only information that the region *III* provides on the model is the one arising from the implementation of the constraint that $I_{\text{ZLP}}(\Delta E \rightarrow \infty) \rightarrow 0$.
3. The EELS data in region *II*, defined by $\Delta E_I \leq \Delta E \leq \Delta E_{II}$, is excluded from the training dataset, given that in this region the contribution to I_{EEL} coming from I_{inel} is already significant. The model predictions in this regime are obtained from an interpolation of the predictions obtained in regions *I* and *III*.

This classification introduces two new hyper-parameters of our model, ΔE_I and ΔE_{II} , that need to be specified before the training. They should satisfy $\Delta E_I \leq \Delta E_{\min}$ and $\Delta E_{II} \geq \Delta E_{\min}$, with ΔE_{\min} being the position of the first local minimum of I_{EEL} . Let's interpret this physically: ΔE_{\min} is the first local minimum, which means that the inelastic contributions are significantly present. Training the data for higher than this value, we are sure that our training data includes at least some amount of inelastic scatterings. Therefore, it is certain that we need to cut the training data already at a lower energy loss, so $\Delta E_I \leq \Delta E_{\min}$. As indicated by the toy spectra of Fig. 5.4, a suitable value for ΔE_I would be somewhat above the onset of the inelastic contributions: this way we maximise the amount of training data while ensuring that I_{EEL} is still dominated by I_{ZLP} .

We can use the derivatives of the spectra, $dI_{\text{EEL}}/d\Delta E$, to select suitable minimum and maximum values for ΔE_I . Using first and second derivatives is an often-used feature extraction method to achieve a relatively high accuracy with a low computational complexity. In an ideal microscope the electron beam would be perfectly monochromatic, correspondingly the ZLP would appear as a delta function in an EEL spectrum [12]. In practice the ZLP has a finite width defining the energy resolution of the system. Concerning ΔE_I , its minimum possible value is selected as the value where the derivative taken on the sample data start to differ significantly as compared to those spectra taken on vacuum. This value is obtained by looking at the ratio of the derivatives of the spectra compared to the vacuum derivatives:

$$R_{dI/d\Delta E} = \left(\frac{dI_{\text{ZLP}}/d\Delta E}{dI_{\text{EEL}}/d\Delta E} \right). \quad (5.13)$$

For low energy losses this ratio equals 1, which means that in the low-loss regime the spectra recorded in vacuum and on specimens behave similar. However, at some energy

loss $\Delta E_{I,min}$ the sample spectrum stops monotonically decreasing and the ratio deviates from 1. It is this point where the contributions from the sample start to change the shape of the ZLP distribution and we can use this measure to mark the transition from regime I to regime II in Fig. 5.4. We know that the hyper-parameter ΔE_I should satisfy $\Delta E_{I,min} \leq \Delta E_I \leq \Delta E_{min}$, which corresponds to the region where $R_{dI/d\Delta E} \neq 1$ and $R_{dI/d\Delta E} \geq 0$. The neural network will be trained on a range of different ΔE_I values within this interval, and the optimal choice will be determined *a posteriori* from the results. Also, generating results for different values of ΔE_I makes us able to cross-validate on the stability of our results regardless of the exact choice of this hyperparameter.

Another important difference as compared to the training of the vacuum spectra is that each of the sample spectra will have different values of ΔE_{min} and thus of ΔE_I . For this reason we calculate ΔE_{min} for each of the sample spectra and we use the highest of these as the maximum value for the hyper-parameter ΔE_I . As we determine the best choice of ΔE_I after training for each of the spectra separately, we are sure to capture all suitable results and select the best value for each individual spectrum.

Concerning ΔE_{II} , its minimum value should mark the region where $I_{ZLP}(\Delta E \rightarrow \infty) \rightarrow 0$. It is the value where we start adding pseudo-data, so $\Delta E_{II} = \Delta E_{pd,min}$. In order to implement this constraint, similar to the previous section (Eq. (??)) we look at the ratio $\mathcal{R}_{sig}(\Delta E_i)$ to determine the energy loss ΔE_{pd} at which the contributions from the vacuum-recorded ZLP vanish. As a measure, we use the energy loss value where the ratio $\mathcal{R}_{sig}(\Delta E_i)$ drops below 1, as in this regime we would be fitting statistical noise. We set the value of ΔE_{II} equal to this energy loss and add pseudo-data points for $\Delta E \geq \Delta E_{II}$. Note that in this region the intensity of the ZLP is several orders of magnitude smaller than the intensity of the elastic emissions and therefore the exact choice of ΔE_{II} does not listen too closely.

Now that we have determined our strategy for the parametrisation of the ZLP, we move on to present the results, first on vacuum recorded data and afterwards on specimen data.

6 Results I. Vacuum spectra

In this section we present the application of the previously presented strategy to the parametrisation of ZLP spectra acquired in vacuum. Applying our model to this case has a two-fold motivation. First of all, we aim to demonstrate that our model is flexible enough to effectively reproduce the input EELS measurements for a range of variations of the operation parameters of the microscope. Herefore, we modify the input parameters beyond those included in the training set and see how the model is able to interpolate and extrapolate on its inputs.

Second, it will allow us to provide a calibration prediction useful for the case of the in-sample measurements. Such calibration is necessary since, as explained in Sect. 5.3, some of the model hyper-parameters are determined by the comparison of the intensity derivatives between spectra taken in vacuum and those in sample.

In this section, first of all we present the input dataset and motivate the choice of training settings and model hyperparameters. Then we validate the model training by assessing the fit quality. Lastly, we study the dependence of the model predictions in its various input variables, and study the dependence of the model uncertainties upon the removal of a subset of the training dataset.

6.1 Training initialization

In Table 6.1 we collect the main properties of the EELS spectra acquired in vacuum to train the neural network model. For each set of spectra, we indicate the exposure time t_{exp} , the beam energy E_b , the number of spectra N_{sp} recorded for these operation conditions, the number N_{dat} of bins in each spectrum, the range in electron energy loss ΔE , and the average full width at half maximum (FWHM) evaluated over the N_{sp} spectra with the corresponding variance. All the spectra listed on Table 6.1 were acquired with a Titan TEM equipped with a Schottky field emitter. We point out that since here we are interested in the low-loss region, ΔE_{max} does not need to be too large, and in any case the large ΔE behaviour of the model is fixed by the constraint implemented by Eq. (5.10).

Set	t_{exp} (ms)	E_b (keV)	N_{sp}	N_{dat}	ΔE_{min} (eV)	ΔE_{max} (eV)	FWHM (meV)
1	100	200	15	2048	-0.96	8.51	47 ± 7
2	100	60	7	2048	-0.54	5.59	50 ± 4
3	10	200	6	2048	-0.75	5.18	26 ± 3
4	10	60	6	2048	-0.40	4.78	34 ± 2

Table 6.1. Summary of the main properties of the EELS spectra acquired in vacuum to train the neural network model. For each set of spectra, we indicate the exposure time t_{exp} , the beam energy E_b , the number of spectra N_{sp} recorded for these operation conditions, the number N_{dat} of bins in each spectrum, the range in electron energy loss ΔE , and the average FWHM evaluated over the N_{sp} spectra with the corresponding standard deviation

The energy resolution of these spectra, quantified by the value of their FWHM, ranges from 26 meV to 50 meV depending on the specific operation conditions of the microscope, with a standard deviation between 2 and 7 meV. The value of the FWHM varies only mildly with the value of the beam energy E_b but grows by around a factor two for spectra collected at larger exposure times t_{exp} . A total of almost 7×10^4 independent measurements will be thus used for the ZLP model training on the vacuum spectra. As will be highlighted

in Sects. ?? and ??, one of the advantages of our ZLP model is that it can extrapolate its predictions to other operation conditions beyond those used for the training.

Following the strategy presented in Sect. 5, first of all we combine the N_{sp} spectra corresponding to each of the four sets of operation conditions and determine the statistical uncertainty associated to each energy loss bin by means of Eq. (??). The resulting set of training points can be observed in Fig. 6.1.

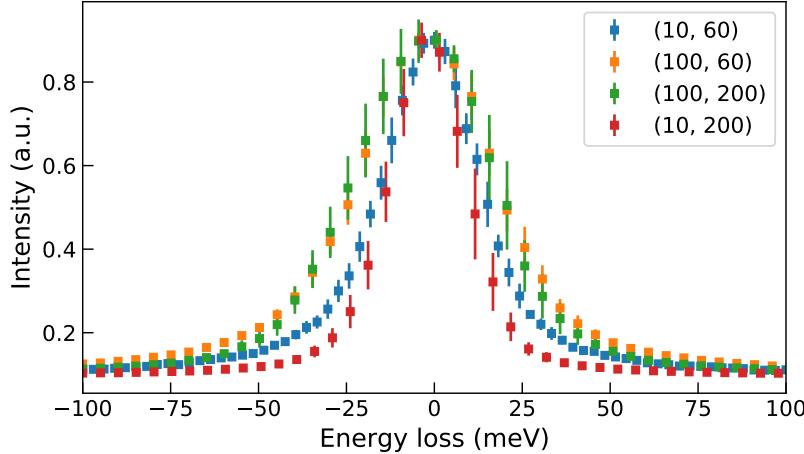


Figure 6.1. Graphical representation of the four sets of training inputs that have been calculated as means and standard deviations over the N_{sp} four each set of operating conditions $\{t_{\text{exp}}(\text{ms}), E_b(\text{keV})\}$.

For each of the training sets, we need to determine the value of $\Delta E_{\text{pd}}^{(\min)}$ that defines the lower limit of the range for which we add the pseudo-data that imposes the correct $\Delta E \rightarrow \infty$ limit of the model. This value is determined by inspecting the ratio between the central experimental value of the total EELS intensity, $I_{\text{EEL},i}^{(\text{exp})}$, and its corresponding total uncertainty defined in Eq. (??).

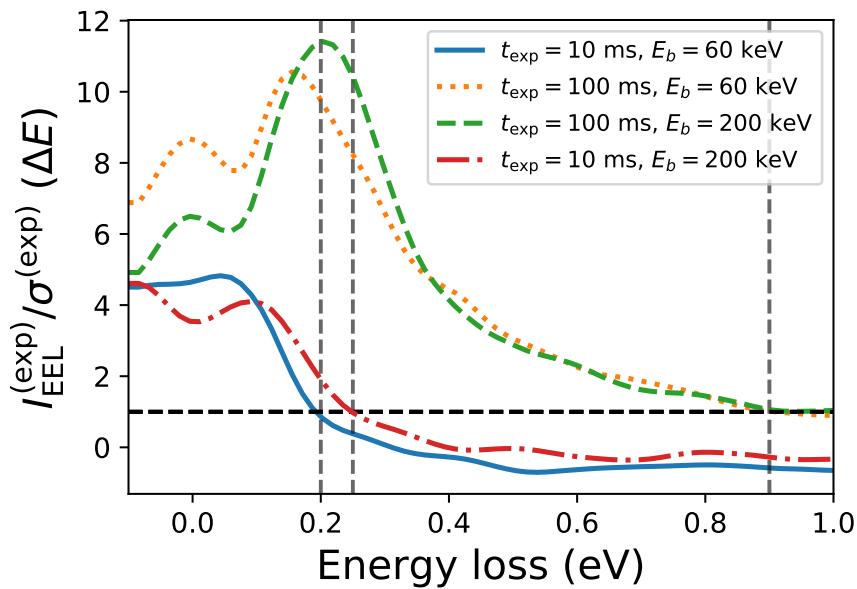


Figure 6.2. The ratio between the central experimental value of the total EELS intensity, $I_{\text{EEL},i}^{(\text{exp})}$, and the corresponding total uncertainty defined in Eq. (??). Results are shown for the four combinations of t_{exp} and E_b listed in Table 6.1. The vertical dashed lines mark the values of ΔE for which this ratio becomes smaller than unity, which indicates when the input data starts to be dominated by the statistical noise.

Fig. 6.2 displays this ratio for the four combinations of t_{exp} and E_b listed in Table 6.1. The vertical dashed lines indicate the values of ΔE for which this ratio becomes smaller than unity. For larger ΔE , the EELS spectra become consistent with zero within uncertainties and can thus be discarded and replaced by the pseudo-data constraints. Thus the cross-over value of ΔE where the ratio satisfies $I/\sigma \simeq 1$ is a reasonable choice for $\Delta E_{\text{pd}}^{(\text{min})}$. The total uncertainty of the pseudo-data points is chosen to be

$$\sigma_j^{(\text{pd})} = \frac{1}{10} I_{\text{EEL}}^{(\text{exp})} \left(\Delta E = \Delta E_{\text{pd}}^{(\text{min})} \right), \quad j = 1, \dots, N_{\text{pd}}. \quad (6.1)$$

The factor of 1/10 is found to be suitable to ensure that the constraint is enforced without distorting the training to the experimental data. We note from Fig. 6.2 that $\Delta E_{\text{pd}}^{(\text{min})}$ will depend on general on the operation conditions. We find that for our training samples $\Delta E_{\text{pd}}^{(\text{min})} \simeq 200$ meV for $t_{\text{exp}} = 10$ ms and $\simeq 900$ meV for 100 ms, roughly independent on the value of E_b .

The input experimental measurements listed in Table 6.1 are used to generate a sample of $N_{\text{rep}} = 500$ Monte Carlo replicas and train an individual neural network model to each of these replicas. The end result of the procedure is a set of model replicas,

$$I_{\text{ZLP}}^{(\text{mod})(k)}(\Delta E, E_b, t_{\text{exp}}), \quad k = 1, \dots, N_{\text{rep}}, \quad (6.2)$$

which can be used to provide a prediction for the intensity of the ZLP for arbitrary values of ΔE , E_b , and t_{exp} . Eq (6.2) provides the sought-for representation of the probability density in the space of ZLP models. For this sample of replicas one can evaluate statistical estimators such as averages, variances, and correlations (as well as higher moments) by means of the usual expressions, for instance

$$\langle I_{\text{ZLP}}^{(\text{mod})}(\{z_1\}) \rangle = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} I_{\text{ZLP}}^{(\text{mod})(k)}(\{z_1\}), \quad (6.3)$$

$$\sigma_{I_{\text{ZLP}}^{(\text{mod})}}(\{z_1\}) = \left(\frac{1}{N_{\text{rep}} - 1} \sum_{k=1}^{N_{\text{rep}}} \left(I_{\text{ZLP}}^{(\text{mod})(k)} - \langle I_{\text{ZLP}}^{(\text{mod})} \rangle \right)^2 \right)^{1/2}, \quad (6.4)$$

$$\rho(\{z_1\}, \{z_2\}) = \frac{\langle I_{\text{ZLP}}^{(\text{mod})}(\{z_1\}) I_{\text{ZLP}}^{(\text{mod})}(\{z_2\}) \rangle - \langle I_{\text{ZLP}}^{(\text{mod})}(\{z_1\}) \rangle \langle I_{\text{ZLP}}^{(\text{mod})}(\{z_2\}) \rangle}{\sigma_{I_{\text{ZLP}}^{(\text{mod})}}(\{z_1\}) \sigma_{I_{\text{ZLP}}^{(\text{mod})}}(\{z_2\})} \quad (6.5)$$

where as in the previous section $\{z_l\}$ denotes a possible set of input variables for the model. We now discuss some of features of this ZLP vacuum model, here $\{z_l\} = (\Delta E, E_b, t_{\text{exp}})$.

6.2 Fit quality

To begin with, we would like to quantify the overall fit quality of the model and demonstrate that it is flexible enough to describe all the available input datasets. In Table 6.2 we indicate the values of the final χ^2 per data point, Eq. (5.12), as well as the average values of the error Eq. (5.9) over the training and validation subsets, for each of the four sets of spectra listed in Table 6.1 as well as for the total dataset. We recall that for a satisfactory training one expects $\chi^2 \simeq 1$ and $\langle E_{\text{tr}} \rangle \simeq \langle E_{\text{val}} \rangle \simeq 2$ [60]. From the results of this table we can see that these expectations are satisfied, both for the individual sets and for the total dataset.

Then Fig. 6.3 displays the χ^2 distributions evaluated for the training and validation sets of the $N_{\text{rep}} = 500$ replicas of the sample trained on the spectra listed in Table 6.1.

Note that the training/validation partition differs at random for each replica. The χ^2_{tr} distribution peaks at $\simeq .7$, indicating that a satisfactory model training has been achieved, however that the errors on our data points have been slightly overestimated. We emphasize that the stopping criterion for the neural net training adopted here never considers the numerical values of the error function and determines proper learning entirely from the global minima of $E_{\text{val}}^{(k)}$. From Fig. 6.3 we also observed that χ^2_{tr} distribution peaks at a slighter higher value, $\simeq 1$, and is broader than its corresponding training counterpart. These results confirm both that a satisfactory model training that prevents overlearning has been achieved as well as an appropriate estimate of the statistical uncertainties affecting the original EEL spectra.

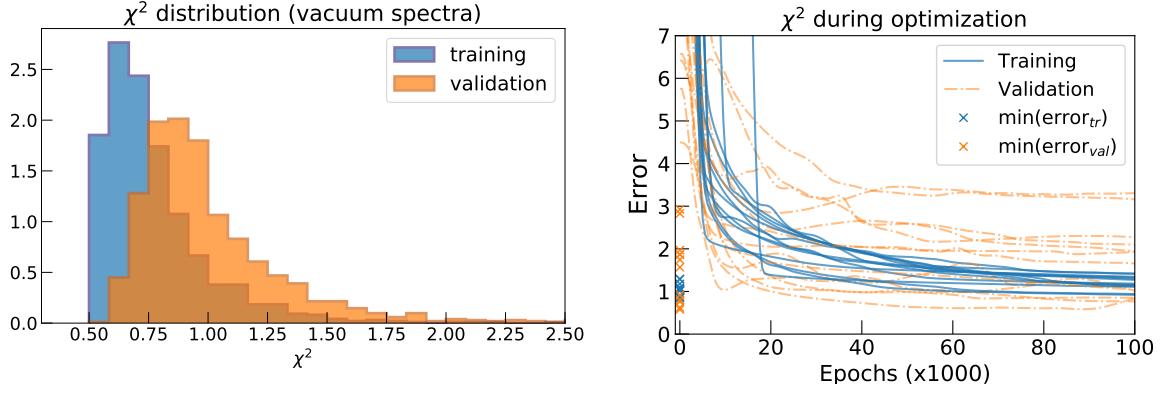


Figure 6.3. Left: The χ^2 (per data point) distribution evaluated for the training and validation sets of the $N_{\text{rep}} = 500$ replicas of the sample trained on the spectra listed in Table 6.1. Right: the progress of the training and validation error over the course of the optimization. Number of replicas is taken small for clarity. The final χ^2 is marked on the y-axis and shows how the distribution of training errors is more narrow and centered at a lower value compared to the validation errors.

Set	χ^2	$\langle E_{\text{tr}} \rangle$	$\langle E_{\text{val}} \rangle$
1	0.998	1.702	1.970
2	0.733	1.408	1.767
3	0.697	1.391	1.800
4	0.593	1.201	1.761
Total	0.771	1.470	1.853

Table 6.2. The values of the final χ^2 per data point, Eq. (5.12), as well as the average values of the error Eq. (5.9) over the training and validation subsets, for each of the four sets of spectra listed in Table 6.1 as well as for the total dataset.

From Table 6.2 we observe that the χ^2 of the central fit is about half the size the error, which measures the quality of the fit for each MC replica. This is what we would expect a fit which correctly represents the fluctuations of the underlying set of experimental data: given that replicas fluctuate about the experimental measurements, these measurements themselves fluctuate about their "true" underlying values.

6.3 Dependence on the electron energy loss

Having demonstrated that our neural network model provides a satisfactory description of all the input EEL spectra, we now present its predictions for specific choices of the input parameters. First of all, we investigate the dependence of the results as a function of the electron energy loss ΔE . Fig. 6.4 displays the central value and 68% confidence level uncertainty band for the ZLP model as a function of electron energy loss ΔE evaluated using Eqns. (6.3) and (6.4). We display results corresponding to three different values of E_b and for both $t_{\text{exp}} = 10$ ms (left) and $t_{\text{exp}} = 100$ ms (right panel). We emphasize that only beam energies of 60 and 200 keV were included in the dataset, so the network is trained on these data. It has never seen data with $E_b = 120$ keV, and thus our prediction in this case arises purely from the model interpolation. It is interesting to note how both the overall normalisation and the shape of the predicted ZLP depend on the specific operation conditions.

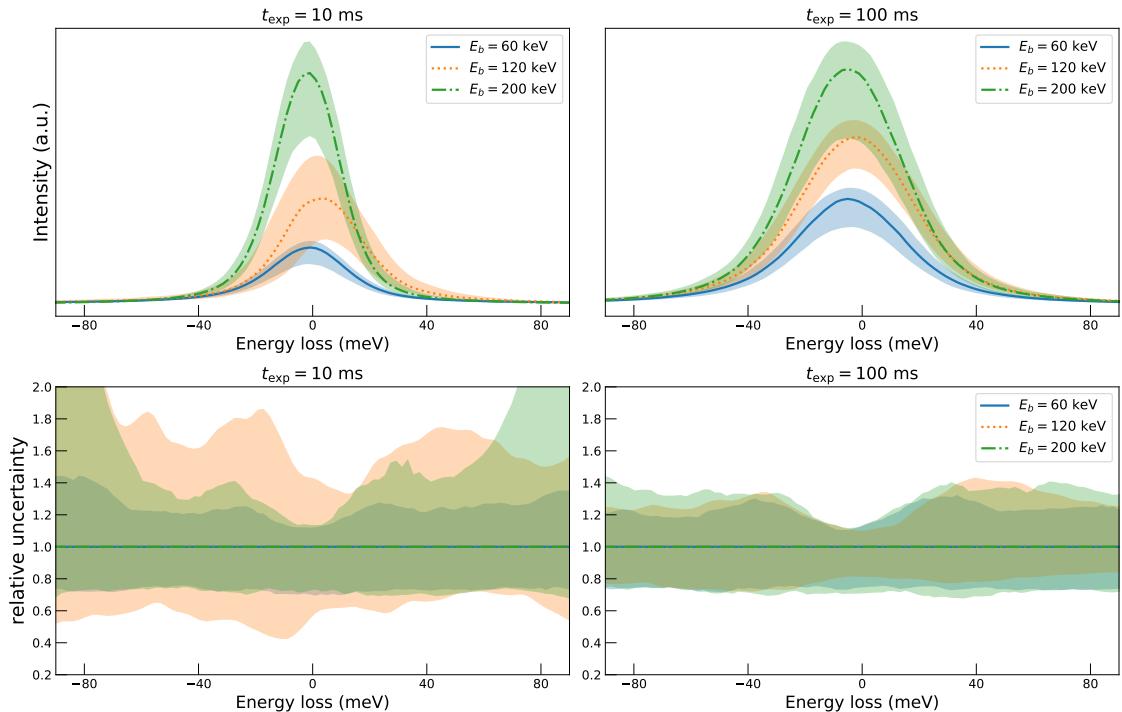


Figure 6.4. Top: the central value and 68% confidence level uncertainty band for the ZLP model as a function of electron energy loss ΔE evaluated using Eqns. (6.3) and (6.4). We display results corresponding to three different values of E_b and for both $t_{\text{exp}} = 10$ ms (left) and $t_{\text{exp}} = 100$ ms (right panel). Note that no training data with $E_b = 120$ keV has been used and thus our prediction in that case arises purely from the model interpolation. Bottom: the corresponding relative uncertainty as a function of ΔE for each of the three values of E_b .

In the bottom panel of Fig. 6.4 we show the corresponding relative uncertainty as a function of ΔE for each of the three values of E_b . The relative uncertainty is calculated as the absolute error in each point divided by the predicted intensity. Recall that in this work we allow for non-Gaussian distributions and thus the central value is the mean of the distribution while the error band in general will be asymmetric. Looking at the predictions for $t_{\text{exp}} = 10$ ms, we see how the model prediction at $E_b = 120$ keV typically exhibits larger uncertainties and is therefore less stable than the predictions for the two values of E_b for which we have training data. In the case of $t_{\text{exp}} = 100$ ms instead, the model predictions exhibit very similar uncertainties for the three values of E_b , which furthermore depend only mildly on ΔE . From these outcomes we can conclude that the network is well able to make predictions on inputs it has never seen before.

It is interesting to assess how the model results change once a subset of the data points is excluded from the fit. In other words, we remove data points in a certain region $[\Delta E_{\text{cut}}^{(\min)}, \Delta E_{\text{cut}}^{(\max)}]$ and train the model on the resulting dataset. Afterwards, we let the model make predictions on the full energy range and see what the effect of such a cut is. The reason to do this goes hand in hand with the fact that we will need this model later onwards for the prediction on sample data. As explained in Sect. 5 and as illustrated in Fig. 5.4, when training the model on sample spectra, a region with $\Delta E_I \leq \Delta E \leq \Delta E_{II}$ will be removed from the training dataset to avoid the contamination from the inelastic contributions. Fig. 6.5 displays the predicted central value and uncertainty in the model predictions for $I_{\text{ZLP}}(\Delta E)$ as a function of the energy loss for $E_b = 200$ keV and $t_{\text{exp}} = 10$ ms (left) and 100 ms (right panel). We show results for three different sets of training: first of all, one without any cut in the training dataset, and then for the case where the data points with $\Delta E \geq \Delta E_{\text{cut}}$ are removed from the training dataset. We consider two values of ΔE_{cut} , namely 50 meV and 100 meV, indicated with vertical dash-dotted lines. In both cases, data points are removed up until $\Delta E = 800$ meV. The pseudo-data points that enforce $I_{\text{EEL}}(\Delta E) \rightarrow 0$ are present in all three cases in the region $800 \text{ meV} \leq \Delta E \leq 1 \text{ eV}$.

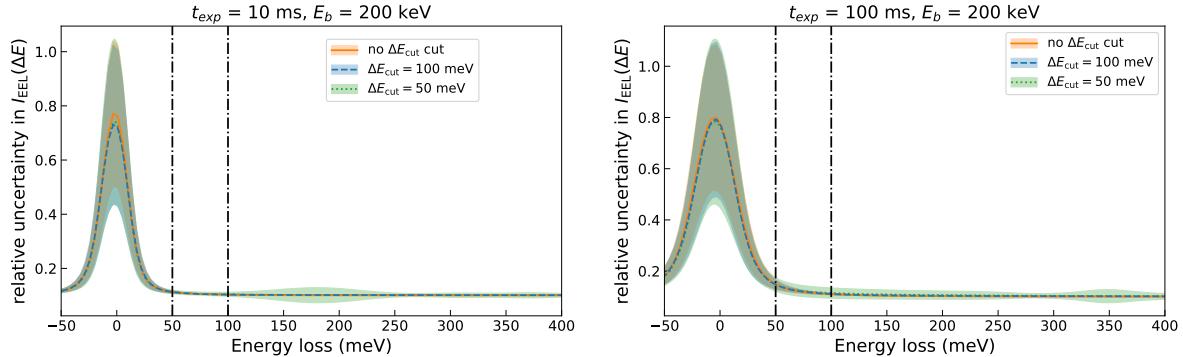


Figure 6.5. The central values and 68% uncertainty bands in the model predictions for $I_{\text{EEL}}(\Delta E)$ as a function of the energy loss for $E_b = 200$ keV and $t_{\text{exp}} = 10$ ms (left) and 100 ms (right panel). We show results for three different sets of trainings: without any cut in the training dataset, and for the case where the data points with $\Delta E \geq \Delta E_{\text{cut}}$ are removed from the training dataset for two different values of ΔE_{cut} . Note that the same pseudo-data points that enforce $I_{\text{EEL}}(\Delta E) \rightarrow 0$ are present in all three cases.

At first sight, the removal of data from the training dataset doesn't introduce large uncertainties for the ZLP predictions: they are little noticed when the data is cut at 50 meV, and for 100 meV the predictions seem to identical to the full range training. To emulate the effects of such cut, one needs to look at the *relative* uncertainty rather than the absolute, that is, normalizing the uncertainty in each point to the predicted intensity. The results can be observed in Fig. 6.6. We show the relative uncertainty in the model predictions using two different x and y scales, to visualize the results as clearly as possible.

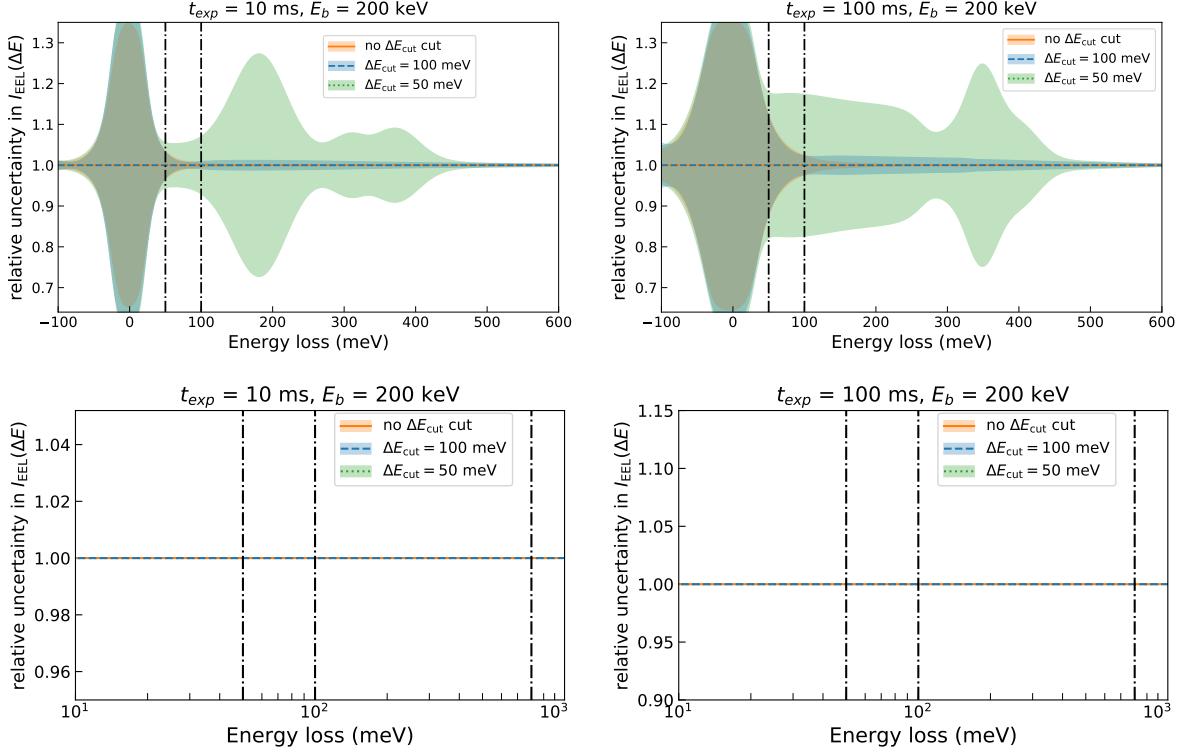


Figure 6.6. Top: The relative uncertainty in the model predictions for $I_{\text{EEL}}(\Delta E)$ as a function of the energy loss for $E_b = 200 \text{ keV}$ and $t_{\text{exp}} = 10 \text{ ms}$ (left) and 100 ms (right panel). Bottom: same results with different x and y scales for better visualization of the differences in relative uncertainties.

From this comparison we can observe how the model predictions become significantly more unstable once a subset of the training data is cut away, as expected due to the effect of the information loss. While for the cut $\Delta E_{\text{cut}} = 100 \text{ meV}$ the increase in model uncertainty is only moderate as compared with the baseline fit where no cut is performed (since for this value of ΔE uncertainties are small to begin with), rather more dramatic effects are observed for a value of the cut $\Delta E_{\text{cut}} = 50 \text{ meV}$. This comparison highlights how ideally we would like to keep as many data points in the training set for the ZLP model, provided of course we can verify that the possible contributions to the spectra related to inelastic scatterings from the sample can be neglected.

6.4 Dependence on beam energy and exposure time

As indicated in Table 6.1, the training dataset contains spectra taken at two values of the electron beam energy, $E_b = 60 \text{ keV}$ and 200 keV . The left panel of Fig. 6.7 displays model predictions for the FWHM of the zero loss peak (and its corresponding uncertainty) as a function of the beam energy E_b for two values of the exposure time, $t_{\text{exp}} = 10 \text{ ms}$ and 100 ms . The vertical dashed lines indicate the values of E_b for which training data is available. This comparison illustrates how the model uncertainty vary in the data region (near $E_b = 60 \text{ keV}$ and 200 keV), the interpolation region (for E_b between 60 and 200 keV), and the extrapolation regions (for E_b below 60 keV and above 200 keV). For $t_{\text{exp}} = 100 \text{ ms}$, we observe that the model interpolates reasonably well between the measured values of E_b and then uncertainties increase markedly in the extrapolation region above $E_b = 200 \text{ keV}$.

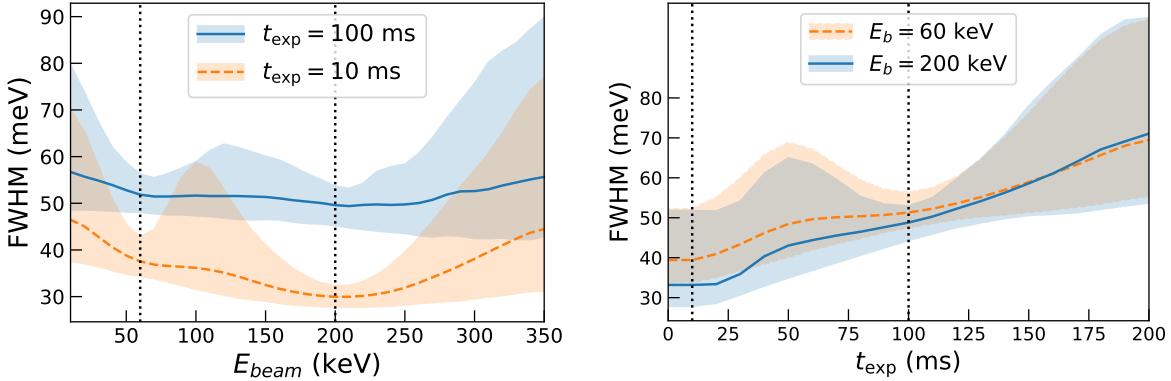


Figure 6.7. The model predictions for the FWHM of the zero loss peak (and its corresponding uncertainty) as a function of the beam energy E_b for two values of the exposure time (left panel) and as a function of t_{exp} for two values of E_b (right panel). The vertical dashed lines indicate the values of the corresponding parameter for which training data is available.

For this comparison one can observe that as expected the uncertainty in the prediction for FWHM of the ZLP is the smallest close to the values of E_b for which one has training data. The uncertainties increase but only in a moderate way in the interpolation region, indicating that the model can be reliable applied to predict the features of the ZLP other values of the electron energy beam (assuming that all other operation conditions of the microscope are unchanged). The errors increase rapidly in the extrapolation region, which is a characteristic feature of these neural network models. Indeed, as soon as the model departs from the data region there exist a very large number of different functional form models for $I_{\text{ZLP}}(\Delta E)$ that can describe equally well the training dataset, and hence a blow up of the extrapolation uncertainties is generically expected.

The network was trained on data with exposure times of 10 and 100 ms, so interpolation and extrapolation is possible. Similar to the predictions for varying beam energy, also for exposure time the uncertainties grow bigger as the value deviates more from the training inputs. The right panel of Fig. 6.7 displays a similar model comparison as in the left one but now as a function of t_{exp} for two values of E_b (right panel). We observe that the FWHM increases roughly in a linear manner with the exposure time, indicating that lower values of t_{exp} allow for an improved spectral resolution. Further, also in this case we find that the model uncertainties grow rapidly in the extrapolation region beyond that covered for the training dataset.

6.5 Stability and reliability checks

In this section, we assess the stability of the results and the reliability of the error estimate regarding the neural network parametrisation.

To begin with, we study the dependence on the architecture of the neural networks. As previously mentioned, we have chosen a redundant architecture as to assure that the network is sufficiently large to parametrise the ZLP. In order to check that this is indeed the case, we have both increased and decreased the number of neurons, so that we can assess the stability of results for both cases. Throughout this work, the default architecture is 3-10-15-5-1. The “big” architecture was chosen as a 3-15-25-10-1, increasing the number of free parameters from 289 to 731. The “small” network corresponds to an architecture of 3-5-8-6-1, which introduces 129 free parameters to be optimised.

We have followed a similar procedure as presented in Sect. 6.3 to investigate the dependence of the results as a function of the energy loss, corresponding to $E_b = 60$, 120 and 200 keV. The central value and 68% confidence level band for the predicted

ZLP distributions can be observed in Fig. 6.8 below, which can be directly compared to Fig. 6.4.

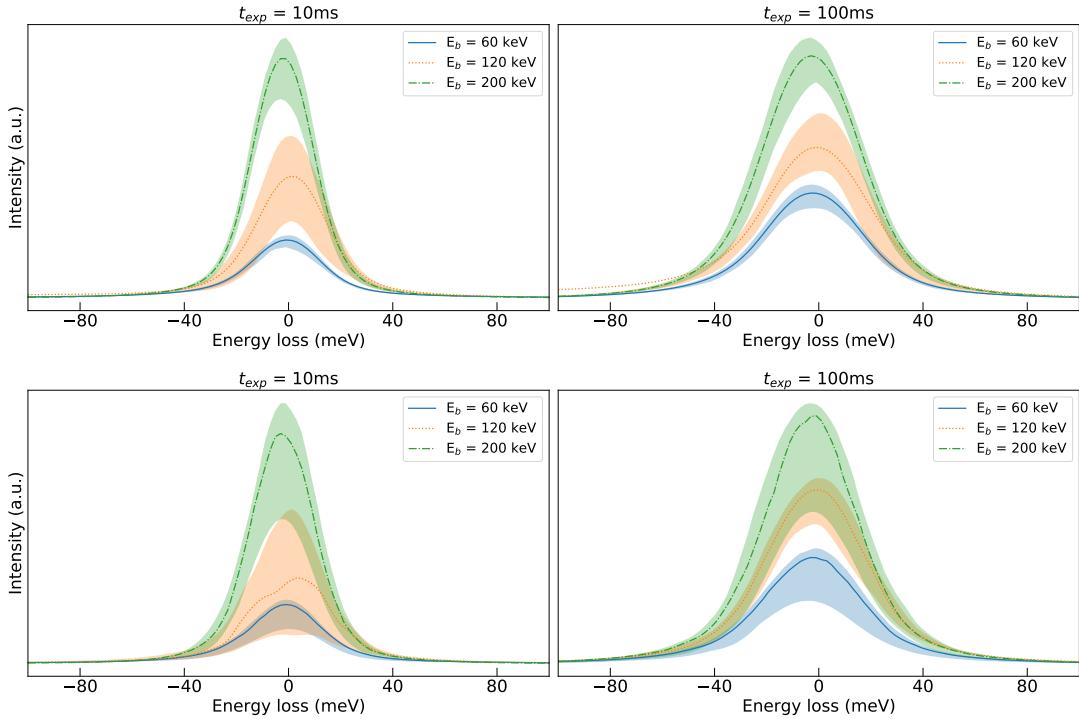


Figure 6.8. Similar to Fig. 6.4 but now for a network with twice (up) and half (down) the architecture. We see that the when doubling the network, predictions are almost identical, exact that uncertainty bands are slightly smaller. However, decreasing the network by a factor of 2 results in predictions that are a bit off from our default network. Note that the predictions made for $E_b=120$ keV are purely interpolation results; the training set contains only data for beam energies of 60 and 200 keV.

It can be observed that results are well reproduced for both smaller and bigger graphs. Particularly for the large network, both the central values and the uncertainty bands are in almost perfect correspondence with the results presented in Sect. 6.3. This confirms the fact that our default network is sufficiently large and flexible to parametrise the ZLP: since increasing the size doesn't affect the obtained results, the number of optimizable parameters is much larger than the minimum required in order to reproduce the data.

This becomes more evident when we substantially decrease the size of the neural network. Predictions on the training data ($E_b=60$ and 200 keV) are still consistent, however the uncertainty bands have grown remarkably bigger, indicating that the neural network outcomes are less stable than before. When we look at the prediction for the interpolation ($E_b=120$ keV), the central values differ (however only moderately) from the ones obtained on bigger networks. This difference indicates that the predicted outcomes indeed depend on the size of the network, and simultaneously it confirms the sufficient redundancy of the architecture used as default.

Next, we check that the results are stable among different subsets of 100 out of the total of 500 Monte Carlo replicas. The results can be observed in Fig. 6.9. We find remarkable stability: the two sets yield to almost identical representations, both for the training data as for the interpolation regime.

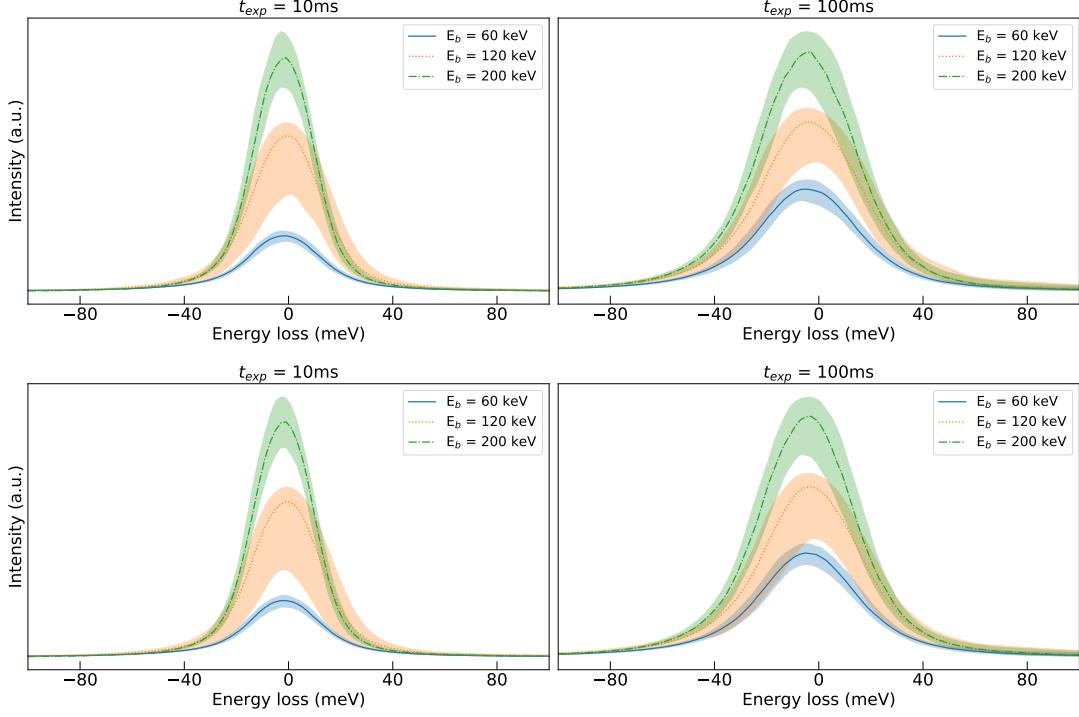


Figure 6.9. Similar to Fig. 6.4 but now for two subsets of 100 replicas instead of the full set of 500 replicas.

A final check is to see how uncertainties change with the number of replicas. We have produced the same plot but now based on three different sizes of MC ensembles: $N_{rep}=25$, 100 and 400. Again, we have created plots similar to the ones before, now using each time a different size of replicas. We would expect that using a small subset introduces larger errors compared to large ensembles of replicas. It is clear that for the smallest subset ($N_{rep}=25$), in the peaks the predictions are already well produced, however uncertainties in the tails are significantly larger. The difference between using either 100 or 400 replicas is hardly noticeable, which confirms our statement given in Sect. 5.2 that 500 replicas is (more than) enough to obtain a faithful representation.

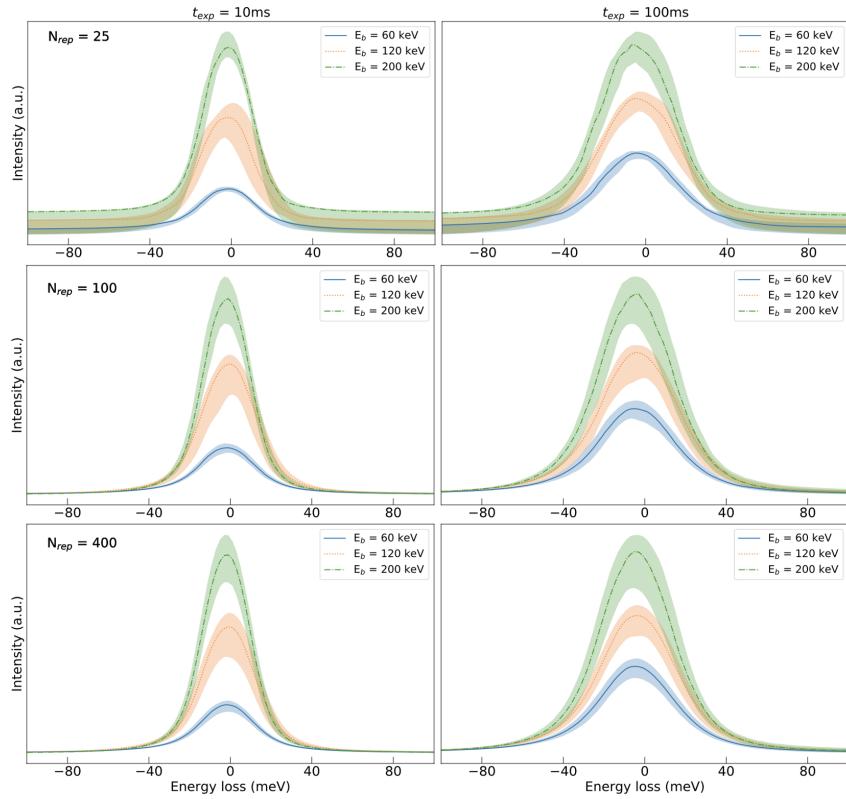


Figure 6.10. Predictions for three different number of replicas. See how uncertainties change..

7 Results II. Sample spectra

Following the discussion of the vacuum ZLP analysis, we now present the application of our machine learning strategy to parametrise the ZLP arising in spectra recorded on specimens, specifically for EELS measurements acquired in different regions of the WS₂ nanoflowers presented in Sect. 3. The resulting ZLP parametrisation will be applied to isolate the inelastic contribution in each spectrum. We will use these subtracted spectra first to determine the bandgap type and energy value from the behaviour of the onset region and second to identify excitonic transitions at very low energy losses.

We start this section by presenting the training dataset, which consists of two groups of EEL spectra recorded in thick and thin regions of the WS₂ nanoflowers respectively. Then we discuss the subtraction procedure, the choice of hyper-parameters, and the error propagation to the physical predictions. The resulting subtracted spectra provide the information required to extract the value and type of the bandgap and to characterise excitonic transitions for different regions of these polytypic WS₂ nanostructures.

7.1 Training dataset

Low-magnification TEM images and the corresponding spectral images of two representative regions of the WS₂ nanoflowers, denoted as sample A and B respectively, are displayed in Fig. 7.1. These spectral images have been recorded in the regions marked by a green square in the associated TEM images, and contain an individual EEL spectrum in each pixel. We indicate the specific locations where EEL spectra have been recorded, including the in-vacuum measurements acquired for calibration purposes. Note that in sample B the differences in contrast are related to the material thickness, with higher contrast corresponding to thinner regions.

These two samples are characterised by rather different structural morphologies. While sample A is a relatively thick region of WS₂, sample B corresponds to a region composed of thin petals, which is only one or a few monolayers thick. In other words, sample A is composed by bulk WS₂ while in sample B some specific regions are much thinner, down to the few monolayers level. This thickness information has been determined by means of the **Digital Micrograph** software.

One of the main goals of this study is demonstrating that our ZLP-subtraction method leads to a satisfactory performance for spectra recorded with different microscopes and operating conditions. With this motivation, the EELS measurements acquired on specimens A and B have been obtained varying both the microscopes and their settings. The TEM and EELS measurements acquired in specimen A are based on a JEOL 2100F microscope with a cold field-emission gun and equipped with an aberration corrector, operated at 60 kV. A Gatan GIF Quantum energy filter was used for the EELS analysis. The corresponding measurements on specimen B were recorded instead using a JEM ARM200F monochromated microscope operated at 60 kV, equipped with a GIF quantum ERS filter. See Methods at the end of this work for more details.

In Table 7.1 we collect the most relevant properties of the spectra collected in the locations indicated in Fig. 7.1 using the same format as in Table 6.1. Note that since the spectra from samples A and B have been acquired with different microscopes, features of the ZLP such as the FWHM are expected to be different. From this table one can observe how the ZLP for the spectra acquired on sample A exhibit a FWHM about five times larger as compared to those of sample B. This difference in resolution can be understood from the fact that the EELS spectra from sample B, unlike those from sample A, were recorded with a TEM equipped with a monochromator.

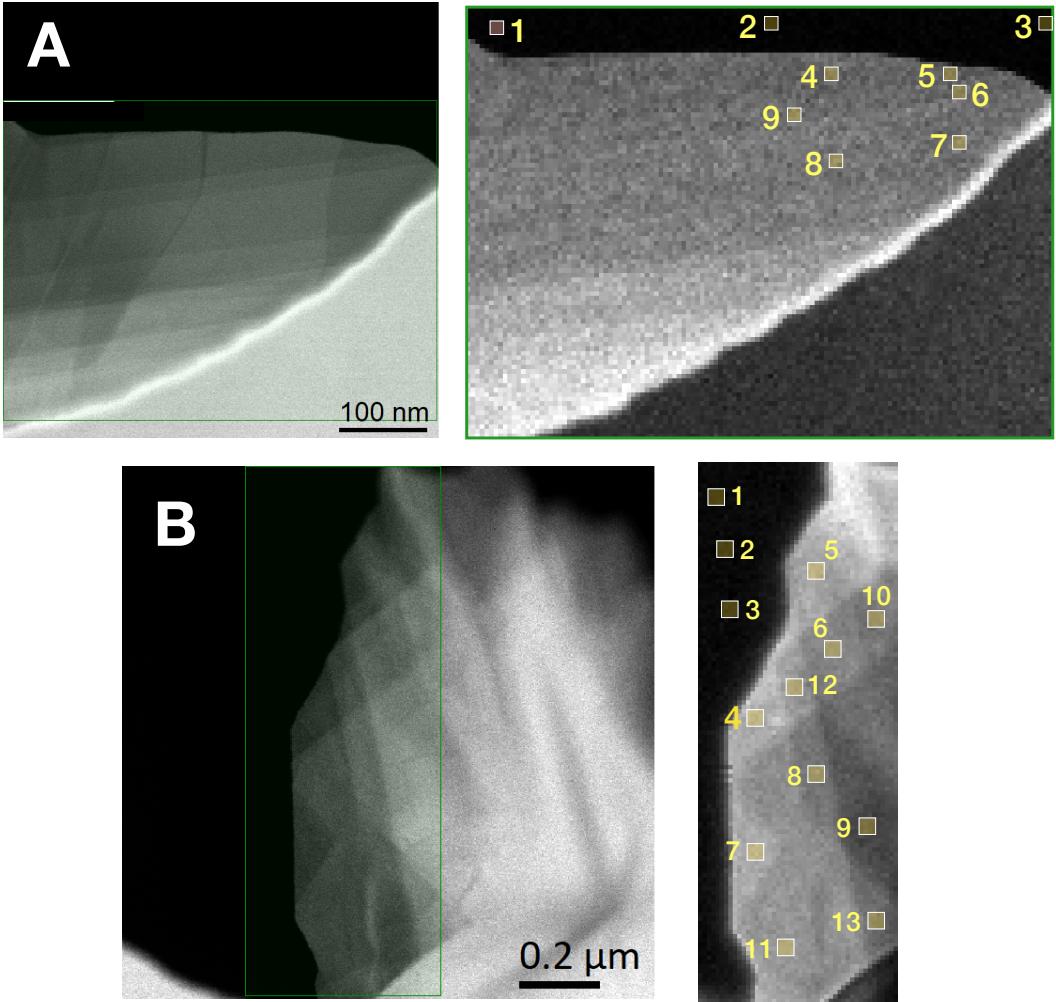


Figure 7.1. Low-magnification TEM images (left) and the corresponding spectral images (right panels) of two different regions of the WS₂ nanoflowers, denoted as sample A (upper) and sample B (lower panels) respectively. The spectral images have been recorded in the regions marked by a green square in the associated TEM images, and contain an individual EEL spectrum in each pixel. We indicate the locations where representative EEL spectra have been selected. In the left panel of sample B, the difference in contrast is correlated to the material thickness, with higher contrast indicating thinner regions of the nanostructure. The morphological differences between the two samples are discussed in the text.

Set	t_{exp} (ms)	E_b (keV)	N_{sp}	N_{dat}	ΔE_{\min} (eV)	ΔE_{\max} (eV)	FWHM (meV)
A	1	60	6	1918	-4.1	45.5	470 ± 10
B	190	60	10	2000	-0.9	9.1	87 ± 5

Table 7.1. Same as Table 6.1 now for the EEL spectra taken on specimens A and B. Note that the location on the WS₂ nanoflowers where each spectra has been recorded was indicated in Fig. 7.1.

In the following we will present the results for spectra that are representative for each of the two samples. The full set of spectra is available together with `EELSfitter`, the code used to produce the results of this analysis, whose installation and usage instructions are presented in Appendix A.

7.2 Subtraction procedure

Following the strategy presented in Sect. 5, again we combine the N_{sp} spectra recorded over each sample and we determine the experimental central values and uncertainties for the training points. Next, we need to determine the choice for hyperparameters ΔE_I and ΔE_{II} , as we only want to train on data that is different from zero within uncertainties. As explained in Sect. 5.3, we need the location of the first local minimum on all the spectra to set a bound for ΔE_I and ΔE_{II} . The representative plots can be observed in Fig. 7.2 and here it becomes clear how much information is contained within the intensity derivatives of the spectra. One can observe how the derivatives recorded in vacuum (for both specimens at locations #1, #2 and #3) are monotonically decreasing and slowly converge to zero, whereas the derivatives recorded on different regions of a sample behave very similar, capturing fluctuations that are typical for the corresponding sample.

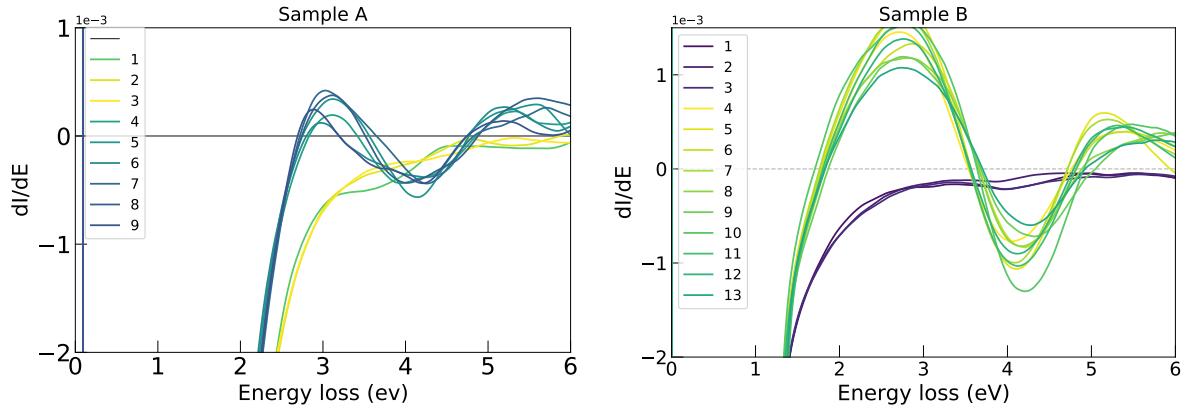


Figure 7.2. Plot of the first intensity derivatives of the spectra recorded over Sample A (left) and B (right) at the locations indicated in Fig. (7.1).

Set	$\Delta E _{\min}$ (eV)	ΔE_I (eV)	ΔE_{II} (eV)
A	2.70 ± 0.06	1.8	12
B	1.80 ± 0.04	1.4	6

Table 7.2. The mean value and uncertainty of the first local minima, $\Delta E|_{\min}$, averaged over the spectra corresponding to samples A and B from Fig. 7.1. We also indicate the corresponding values of the hyper-parameters ΔE_I and ΔE_{II} defined in Fig. 5.4 used for the training of the neural network model.

In Table 7.2 we have collected the mean value and uncertainty of the first local minimum, $\Delta E|_{\min}$, averaged over the spectra corresponding to samples A and B from Fig. 7.1. From the uncertainties in $\Delta E|_{\min}$, we see that the location of the first minimum is relatively stable among all the spectra belonging to a given set. This indicates that the onset of the inelastic contributions I_{inel} does not change significantly as we move to different regions of the sample. We also indicate the corresponding values of the hyper-parameters ΔE_I and ΔE_{II} defined in Fig. 5.4. Recall that only the data points with $\Delta E \leq \Delta E_I$ is used for the training of the neural network model. For $\Delta E \geq \Delta E_{II}$ instead, the training set includes only the pseudo-data that implements the $I_{\text{ZLP}}(\Delta E) \rightarrow 0$ constraint. The values for ΔE_{II} were determined from the vacuum recorded spectra following the same procedure as explained in Sect. 6 and a plot similar to Fig. 6.2 can be observed in Fig. 7.3, where

we show the ratio between the central experimental value of the vacuum EEL intensity and their corresponding uncertainty.

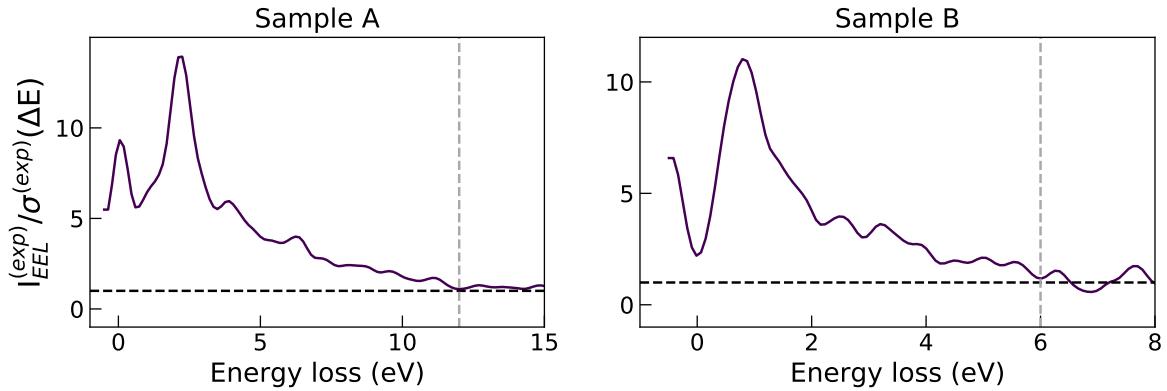


Figure 7.3. The ratio between the central experimental value of the total EELS intensity of the vacuum recorded peaks, $I_{vac}^{(exp)}$, and their corresponding uncertainty defined in Eq. 5.11. Results are shown for the mean of the spectra recorded over Sample A (left) and B (right). For both samples, spectrum positions #1, #2 and #3 in Fig. 7.1 were used to construct the vacuum central values. The vertical dashed lines mark the values of ΔE for which this ratio becomes smaller than unity, which indicates when the input data starts to be dominated by the statistical noise.

We note that the values of ΔE_{II} for this part are significantly higher than the ones found in Fig. 6.2. This could be ascribed to the fact that the vacuum spectra from sample A and B are recorded in proximity to a sample, and therefore effects from the sample are still present, although at a reduced rate.

The model training is performed for a range of ΔE_I values subject to the condition that $\Delta E_I \leq \Delta E_{min}$. The optimal values of ΔE_I listed in Table 7.2 are determined as follows. We evaluate the ratio between the derivative of the intensity distribution acquired on the specimen over the same quantity recorded in vacuum,

$$\mathcal{R}_{der}^{(j)}(\Delta E) \equiv \left\langle \frac{dI_{EEL}^{(exp)(j)}(\Delta E)/d\Delta E}{dI_{EEL}^{(exp)(j')}(\Delta E)/d\Delta E} \right\rangle_{N'_{sp}}, \quad (7.1)$$

where j' labels one of the N'_{sp} vacuum spectra and the average is taken over all available values of j' . This ratio allows to identify a suitable value of ΔE_I by establishing for which energy losses the shape (rather than the absolute value) of the intensity distributions recorded on the specimen starts to differ significantly from their vacuum counterparts. A sensible choice of ΔE_I could for instance be given by $\mathcal{R}_{der}(\Delta E_I) \simeq 0.9$, for which derivatives differ at the 10% level. Note also that the value of the energy loss satisfying $\mathcal{R}_{der}(\Delta E) = 0$ in Eq. (7.1) corresponds to the position of the first local minimum of the spectrum.

For both Sample A and B, the ratios calculated by means of Eq. (7.1) can be observed in Fig. ?? for two representative spectrum locations: we have used loc #4 in both cases. Similar results are obtained for different positions.

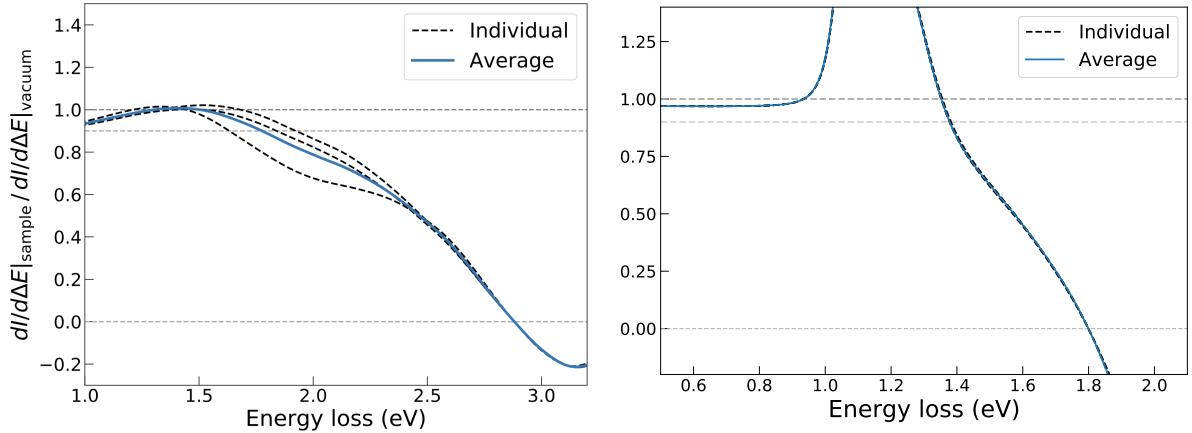


Figure 7.4. Plot of the first intensity derivatives of the spectra recorded over Sample A (left) and B (right) at the locations indicated in Fig. (7.1).

In specimen A it is straightforward to find that $\mathcal{R}_{\text{der}}(\Delta E_I) \simeq 0.9$ at an energy loss of 1.8 eV. For specimen B, $\mathcal{R}_{\text{der}} \simeq 0.9$ at $\Delta E = 1.4$ eV. In this case however, the ratio first starts to increase before it decreases again. This happens consistently for all three vacuum spectra: the three individual ratios can hardly be distinguished. It should be noted that the derivatives also differ at the 10% level when $\mathcal{R}_{\text{der}} \simeq 1.1$, which corresponds to $\Delta E = 1$ eV. Still, we use the point where $\mathcal{R}_{\text{der}} \simeq 0.9$ for the determination of ΔE_I . The reason to favor this choice over $\mathcal{R}_{\text{der}}(\Delta E_I) \simeq 1.1$ is from the physical interpretation: after this point, the derivative of the intensity distributions acquired on samples starts to approximate zero much faster than the derivatives in vacuum. When the effect of the sample kicks in, the intensity profile should flatten or increase rather than continue decreasing, which means the derivatives ratio should be *smaller* than unity. For this reason we choose $\Delta E_I = 1.4$ eV as hyper-parameter for specimen B. Note that the validity of this choice will be checked later on.

Now that we have verified our choices of ΔE_I and ΔE_{II} for both samples, we can move to the results of the training sessions. The end result of the neural network training is a set of $N_{\text{rep}} = 500$ replicas parametrising the ZLP, $I_{\text{ZLP}}^{(\text{mod})(k)}(\Delta E)$. Taking into account that we have N_{sp} individual spectra in each sample, the ZLP subtraction is performed individually for each Monte Carlo replica,

$$I_{\text{inel}}^{(\text{exp})(j,k)}(\Delta E) \equiv I_{\text{EELS}}^{(\text{exp})(j)}(\Delta E) - I_{\text{ZLP}}^{(\text{mod})(k)}(\Delta E), \quad \forall N_{\text{rep}}, \quad j = 1, \dots, N_{\text{sp}}, \quad (7.2)$$

from which statistical estimators can be evaluated as usual. For instance, the mean value for our model prediction for the j -th spectrum can be evaluated by averaging over the set of replicas,

$$\left\langle I_{\text{inel}}^{(\text{exp})(\text{(exp)}j)} \right\rangle (\Delta E) = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} I_{\text{inel}}^{(\text{mod})(j,k)}(\Delta E), \quad j = 1, \dots, N_{\text{sp}}, \quad (7.3)$$

and likewise for the corresponding uncertainties or correlations. For large values of ΔE the model prediction reduces to the original spectra, since in that region the ZLP contribution vanishes,

$$I_{\text{inel}}^{(\text{exp})(j,k)}(\Delta E \gg \Delta E_I) \rightarrow I_{\text{EEL}}^{(\text{exp})(j)}(\Delta E), \quad \forall j, k. \quad (7.4)$$

For very small values of the energy loss, the contribution to the total spectra from inelastic scatterings is negligible and thus the subtracted model prediction Eq. (7.2) should be zero. However, this will not be the case in practice since the neural-net model is trained on the N_{sp} ensemble of spectra, rather than on individual ones, and thus the expected

$\Delta E \rightarrow 0$ behaviour will only be achieved within uncertainties rather than at the level of central values. Therefore, in the zero-loss regime the subtracted spectrum will not vanish completely. To achieve the desired $\Delta E \rightarrow 0$ limit, we apply a matching procedure as follows. We introduce another hyper-parameter, $\Delta E_0 < \Delta E_1$, such that one has for the k -th ZLP replica associated to the j -th spectrum the following behaviour:

$$\begin{aligned} I_{\text{ZLP}}^{(\text{mod})(j,k)}(\Delta E) &= I_{\text{EELS}}^{(\text{exp})(j)}(\Delta E), \quad \Delta E < \Delta E_0, \\ I_{\text{ZLP}}^{(\text{mod})(j,k)}(\Delta E) &= I_{\text{EELS}}^{(\text{exp})(j)} + \left(\xi_1^{(n_l)(k)}(\Delta E) - I_{\text{EELS}}^{(\text{exp})(j)}(\Delta E) \right) \times \mathcal{F}, \quad \Delta E_0 < \Delta E \leq \Delta E_1, \\ \mathcal{F} &= \exp \left(-\frac{(\Delta E - \Delta E_1)^2}{(\Delta E_0 - \Delta E_1)^2 \delta^2} \right), \\ I_{\text{ZLP}}^{(\text{mod})(j,k)}(\Delta E) &= \xi_1^{(n_l)(k)}(\Delta E), \quad \Delta E > \Delta E_1. \end{aligned} \quad (7.5)$$

In Eq. (7.5), $\xi_1^{(n_l)(k)}$ indicates the output of the k -th neural network that parametrises the ZLP and δ is a dimensionless tunable parameter. This matching procedure might look complex at first sight, however it states just the following:

- For $\Delta E < \Delta E_0$, the modeled ZLP is exactly the same as the original spectrum;
- Between ΔE_0 and ΔE_1 , the transition sets in to the ZLP model prediction. $\mathcal{F}(\Delta E)$ represents a matching factor that ensures that the ZLP model prediction smoothly interpolates between $\Delta E = \Delta E_0$ (where $\mathcal{F} \ll 1$ and the original spectrum should be reproduced) and $\Delta E = \Delta E_1$ (where $\mathcal{F} = 1$ leaving the neural network output unaffected);
- At energy loss higher than ΔE_1 , the modeled ZLP is exactly the network prediction.

Here we adopt $\Delta E_0 = \Delta E_1 - 0.5$ eV, having verified that results are fairly independent of this choice. Taking into account the matching procedure, we can slightly modify Eq. (7.2) to

$$I_{\text{inel}}^{(\text{mod})(j,k)}(\Delta E) \equiv I_{\text{EELS}}^{(\text{exp})(j)}(\Delta E) - I_{\text{ZLP}}^{(\text{mod})(j,k)}(\Delta E), \quad \forall N_{\text{rep}}, \quad j = 1, \dots, N_{\text{sp}}. \quad (7.6)$$

The ensemble of ZLP-subtracted spectra $\{I_{\text{inel}}^{(\text{mod})(j,k)}\}$ can then be used to estimate the bandgap of the specimen in the region where they were acquired. Different approaches have been put forward to estimate the value of the bandgap from subtracted EEL spectra, *e.g.* by means of the inflection point of the rising intensity or a linear fit to the maximum positive slope [61]. Here we will adopt the approach of [12] where the behaviour of $I_{\text{inel}}(\Delta E)$ in the onset region is modeled as

$$I_{\text{inel}}(\Delta E) \simeq A (\Delta E - E_{\text{BG}})^b, \quad \Delta E \geq E_{\text{BG}}, \quad (7.7)$$

and vanishes for $E < E_{\text{BG}}$, where both the bandgap value E_{BG} as well as the parameters A and b are extracted from the fit. The exponent b is expected to be $b \simeq 1/2$ ($3/2$) for a semiconductor material characterised by a direct (indirect) bandgap. For each of the N_{sp} spectra and the N_{rep} replicas we fit to Eq. (7.6) the model Eq. (7.7) within a range taken to be $[\Delta E_1 - 0.5$ eV, $\Delta E_1 + 0.7$ eV]. One ends up with N_{rep} values for the bandgap energy and fit exponent for each spectra,

$$\left\{ E_{\text{BG}}^{(j,k)}, b^{(j,k)} \right\}, \quad k = 1, \dots, N_{\text{rep}}, \quad j = 1, \dots, N_{\text{sp}}, \quad (7.8)$$

from which again one can readily evaluate their statistical estimators. In the following, we will display the median and the 68% confidence level intervals for these parameters to account for the fact that their distribution will be in general non-Gaussian.

7.3 Bandgap analysis of sample A

We present first the results of the bandgap analysis of sample A, taking location #4 and #5 in Fig. 7.1 as representative spectra; compatible results are found for the rest of locations. As mentioned above, this region is characterised by a sizable thickness where WS_2 is expected to behave as a bulk material. Fig. 7.5 displays the original and subtracted EEL spectra together with the predictions of the ZLP model, where the bands indicate the 68% confidence level uncertainties and the central value is the median of the distribution. The inset shows the result of the polynomial fits using Eq. (7.7) to the subtracted spectrum together with the corresponding uncertainty bands.

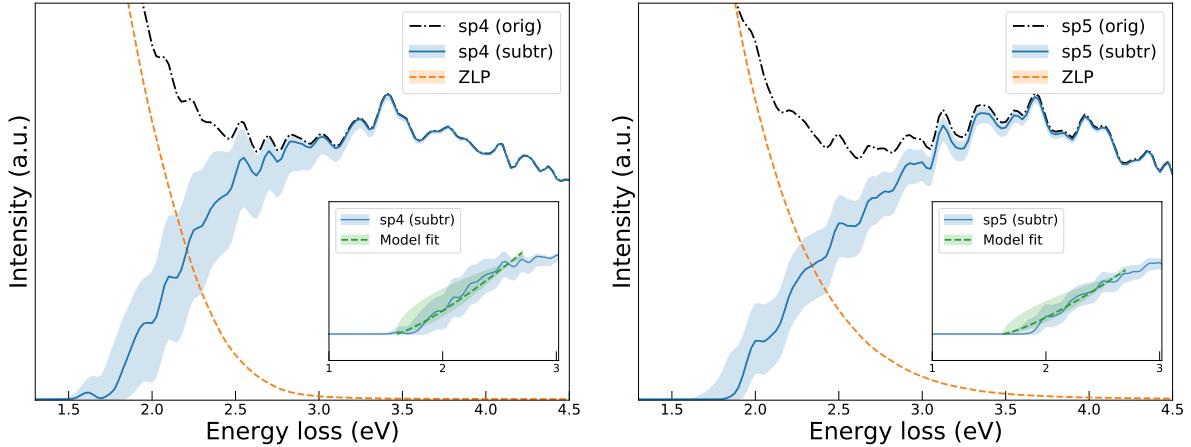


Figure 7.5. The original and subtracted EEL spectrum corresponding to location #4 (left) and #5 (right) of sample A in Fig. 7.1, together with the predictions of the ZLP model, where the bands indicate the 68% confidence level uncertainties. The inset displays the result of fitting Eq. (7.7) to the onset region of the subtracted spectrum.

One can observe how the ZLP model uncertainties are small at low ΔE (due to the matching condition) and large ΔE (where the ZLP vanishes), but become significant in the intermediate region where the contributions from I_{ZLP} and I_{inel} become comparable. It is worth emphasizing that these (unavoidable) uncertainties are neglected in most ZLP subtraction methods, and this is therefore the power of this method. The validity of our choice for the hyperparameter ΔE_I (Table 7.2) can be verified *a posteriori* by evaluating the ratio

$$\mathcal{R}_{\text{abs}}^{(j)}(\Delta E_I) \equiv \left\langle I_{\text{ZLP}}^{(\text{mod})(j)} \right\rangle_{\text{rep}} / \left. I_{\text{EEL}}^{(\text{exp})(j)} \right|_{\Delta E = \Delta E_I}, \quad (7.9)$$

which in this case turns out to be $\mathcal{R}_{\text{abs}} = 0.98$. It is important to verify that $\mathcal{R}_{\text{abs}}(\Delta E_I)$ is not too far from unity, indicating that the training dataset has not been contaminated by the inelastic contributions. By requiring that $\mathcal{R}_{\text{der}}^{(j)}(\Delta E_I) \simeq 0.9$ we obtain the value $\Delta E_I = 1.8$ eV, which is used as baseline for the analysis. It should be noted that this choice is not unique, for example requiring $\mathcal{R}_{\text{der}}^{(j)}(\Delta E_I) \simeq 0.8$ instead would have led to $\Delta E_I = 2.0$ eV. It is therefore important to assess the stability of our results when the hyper-parameter ΔE_I is varied around its optimal value.

With this motivation, we have performed the training over the EEL spectra for a range of ΔE_I values to assess the stability of our results. In Fig. 7.6 we display the values of the exponent b and the bandgap energy E_{BG} obtained from spectrum #4 (left panel in Fig. 7.5) for variations of ΔE_I around its optimal value (1.8 eV, indicated by the horizontal dashed line) by an amount of ± 0.2 eV. The central value and the error band for each value of ΔE_I is evaluated as the median and the 68% CL interval over the $N_{\text{rep}} = 500$ Monte Carlo replicas. We observe that the fit parameters for both b and E_{BG} are stable with respect to variations of ΔE_I , with any shift in the central value contained

within the uncertainty bands. We can therefore conclude that our approach is robust with respect to the choice of its hyper-parameters.

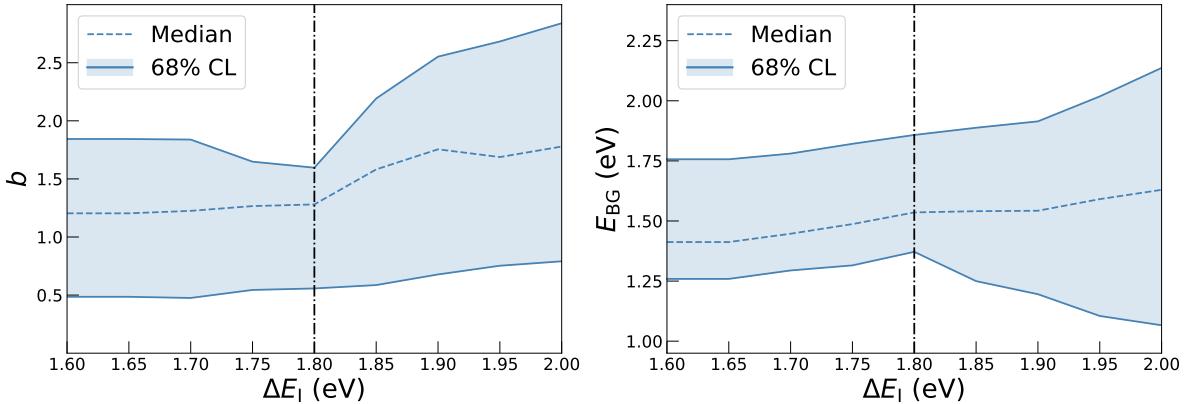


Figure 7.6. The values of the exponent b (left) and the bandgap energy E_{BG} (right panel) from the model Eq. (7.7) obtained from the subtracted spectrum sp14 as ΔE_I is varied by ± 0.2 eV around its optimal value, indicated by the horizontal dashed line.

The final values for E_{BG} and b obtained in the analysis for spectrum 4 and 5 are

$$E_{BG}^{(4)} = 1.6_{-0.2}^{+0.3} \text{ eV}, \quad b^{(4)} = 1.3_{-0.7}^{+0.3} (\#4), \quad (7.10)$$

$$E_{BG}^{(5)} = 1.6_{-0.2}^{+0.2} \text{ eV}, \quad b^{(5)} = 1.3_{-0.5}^{+0.3} (\#5). \quad (7.11)$$

We thus find that for this specific region of the WS_2 nanoflowers the model fit to the subtracted EEL spectrum exhibits a clear preference for an indirect bandgap (where $b \simeq 1.5$), though a direct one ($b \simeq 0.5$) cannot be excluded within uncertainties. This result is consistent with the theoretical expectations of the local electronic properties of bulk WS_2 . Further, the value of E_{BG} is consistent with previous determinations in the same material at the bulk level, such as those collected in Table 3.1. Consistent results are obtained for other locations of Fig. 7.1 where spectra have been recorded. To demonstrate this, we show in Fig. 7.7 the fitted values for E_{BG} and b for all spectra in Sample A, all evaluated using $\Delta E_I=1.8$ eV for the model training. The uncertainty of the bandgap fit varies between the spectra, however what is important is the fact that central values are stable within uncertainty bands. This implies that we find a consistent favor for an indirect bandgap for each of the locations on Sample A, which is what we would expect from theoretical expectations of bulk WS_2 . To the best of our knowledge these results represent the first EELS bandgap analysis of WS_2 nanostructures whose crystalline structure is based on mixed 2H/3R polytypes.

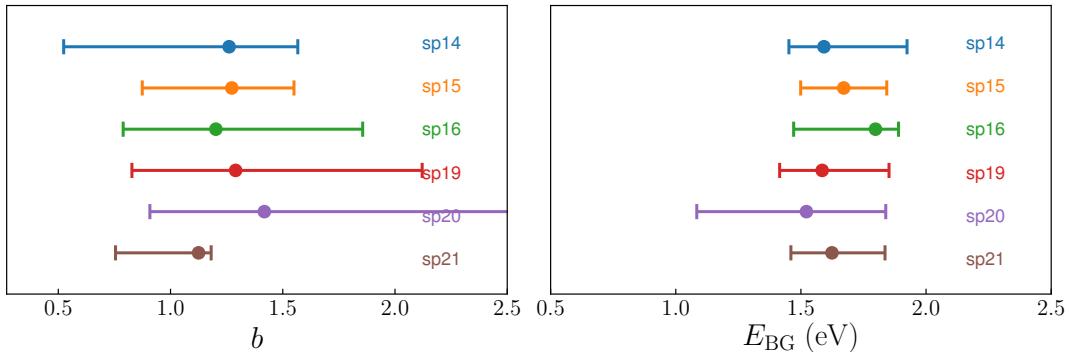


Figure 7.7. The values of the exponent b (left) and the bandgap energy E_{BG} (right panel) from the model Eq. (7.7) obtained from all the subtracted spectra in Sample A, for ΔE_I at its optimal value (1.8 eV).

7.4 Bandgap analysis of sample B

We now discuss the results of the bandgap analysis of sample B, taking location sp4 in Fig. 7.1 as representative spectrum; compatible results are found for the rest of locations. In Fig. 7.8 we present the analogous result to that of Fig. 7.5.

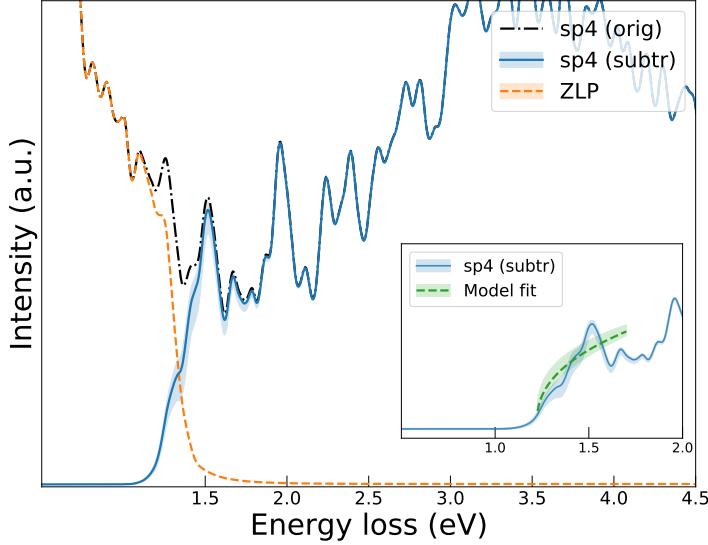


Figure 7.8. Same as Fig. 7.5 now for sp4 (sample B) from Fig. 7.1.

Fig. 7.9 displays the the ZLP-subtracted spectra from sample B corresponding to locations sp4, sp5, and sp6 from Fig. 7.1 together with the corresponding model uncertainties. Results are shown for two values of the hyperparameter ΔE_1 , 1.45 eV (left) and 1.55 eV (right panel). Note how features of the subtracted spectra such as the peaks at $\Delta E \simeq 1.5$, 1.7 and 2.0 are common across the three spectra, demonstrating that there are genuine physical features of the ultra-low-loss region rather than statistical fluctuations. The research on fundamental optical properties of TMDs has led to the discovery of different types of exciton transitions in the ultra-low-loss region of WS₂ nanostructures. The origin of these peaks can be attributed to the formation of an electron-hole pair mitigated by the dielectric screening from the surrounding lattice [62]. In reduced dimensions as in single layers of TMDs, exciton peaks arise with binding energies up to ten times larger than for bulk structures. In the optical spectra of TMDs, two strongly pronounced resonances denoted by A and B excitons are often observed, appearing at binding energies of 300-500 meV below the true band gap [63]. This is in accordance with the features observed in Fig. 7.9 at $\Delta E \simeq 1.5$ and 1.7 eV, which is exactly 300-500 meV below the true bandgap value expected for 2D structures of WS₂. The presence of these low loss features makes the bandgap analysis on this sample non-trivial, since one can only fit Eq. (7.7) to the onset region of the bandgap excitation. Nevertheless, the ZLP-subtracted spectra in Fig. 7.9 show that we have been able to cleanly resolve exciton features down to 1.5 eV within appropriate uncertainty.

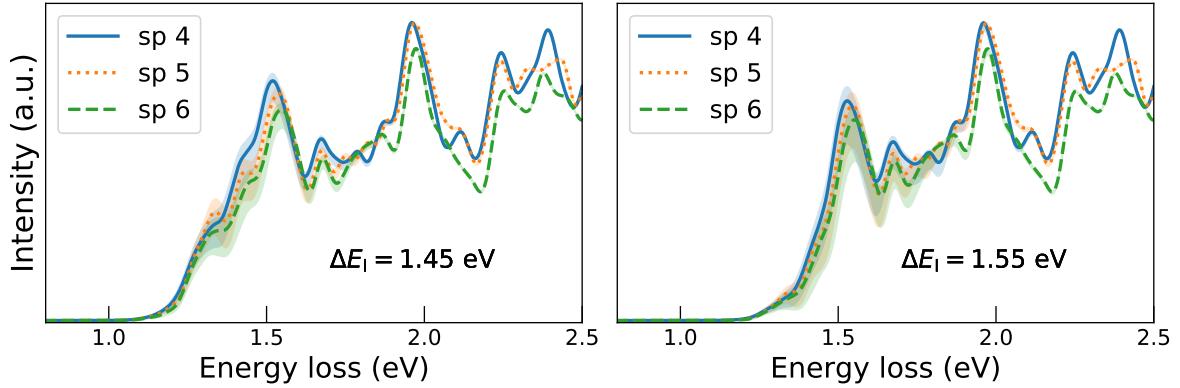


Figure 7.9. The ZLP-subtracted spectra from sample B corresponding to locations sp4, sp5, and sp6 from Fig. 7.1 together with the corresponding model uncertainties. Results are shown for two values of the hyperparameter ΔE_i , 1.45 eV (left) and 1.55 eV (right panel). Note how features of the subtracted spectra such as the peaks as $\Delta E \simeq 1.5$, 1.7 and 2.0 are common across the three spectra.

8 Summary and outlook

In this work we have presented a novel strategy to parametrise and remove the zero-loss peak that appears in the low-loss region of electron energy loss spectroscopy measurements. Our strategy is based on machine learning techniques developed in the context of high-energy physics, in particular for studies of the quark and gluon substructure of protons. An important advantage of this method is the faithful estimation of the input data and methodological uncertainties and their propagation to the subtracted spectra without the need to rely on any kind of approximation.

We have demonstrated how, in the case of vacuum spectra, the model is flexible enough to accomodate several input variables, corresponding to different operation conditions, without the need to assume any *ad-hoc* functional form for the ZLP parametrisation. This flexibility allows us for example to extrapolate the expected FWHM of the ZLP corresponding to other operation conditions of the microscope, beyond those included in the training set. When applied to EEL spectra recorded on samples, we have shown that our approach allows to cleanly disentangle the ZLP contribution from that of the inelastic emissions from the sample, and produce subtracted spectra that account for all relevant sources of uncertainty. This way it becomes possible to fully exploit the valuable physical information contained in the ultra low-loss region of these spectra.

As a proof of concept we have applied this ZLP subtraction strategy to EEL spectra recorded in different regions of WS₂ nanoflowers and used the results both to estimate the local value of the bandgap energy E_{BG} and to assess whether this bandgap is direct or indirect. Our analysis displays preference for either direct or indirect depending on the specific WS₂ nanoflower sample being considered. Comparing with previous results in the literature ...

The approach presented in this work could be extended in several directions. First of all, it would be interesting to test its robustness when additional operation conditions of the microscope are included as input variables, and to assess to which extend ZLP models obtained with an specific microscope can be generalised to an altogether different TEM. Further, a strong cross-check of our method would be provided by comparing our predictions for other operation conditions of the microscope, such as the FWHM as a function of the beam energy E_b reported in Fig. 6.7 with actual measurements and verifying whether or not there is agreement within the uncertainties of the prediction.

Concerning the physical interpretation of the low-loss region of EEL spectra, our method could be applied to study the bandgap and other local electronic properties of different types of nanostructures built upon 2D layered materials, such as MoS₂ nanowalls and nano-pillars and WS₂/MoS₂ arrays or heterostructures. In addition to the bandgap characterisation, one might consider the implications of our approach for the study of other phenomena relevant for the interpretation of the low-loss region such as plasmons, excitons, phonon interactions, and intra-band transitions. One could further exploit the subtracted EEL spectra produced with our method to evaluate the complex dielectric function and its associated uncertainties by means of the Kramers-Kronig relation. Such phenomenological studies of the local electronic properties would be compared with *ab initio* calculations such as Density Functional Theory, based on the same underlying crystalline structure of the analysed samples. We recall that the results presented in this work are to the best of our knowledge the first EELS bandgap analysis of WS₂ nanostructures based on mixed 2H/3R polytypes.

Another possible generalisation of our method would be to the study of spectral TEM images, where each pixel in the image contains an individual EEL spectrum (possibly extended to 3D images). Here machine learning methods would be useful in order to identify relevant features of the spectra (peaks, edges, shoulders) in a fully automated way without having to process each spectrum individually, and then assess how these

features vary as we move along different regions of the nanostructure. Such application would combine two of the most topical uses of machine learning, regression on the one hand and classification on the other hand.

Acknowledgments

The work of J. R. has been partially supported by the Dutch Research Council (NWO). We are grateful to Emanuele R. Nocera and Jake Ethier for assistance in installing the code in the Nikhef cluster.

Supplementary Material

A Installation and usage of EELSfitter

In this appendix we provide some instructions about the installation and the usage of the `EELSfitter` code developed in this work. The code is available from its GitHub repository

<https://github.com/LHCfitNikhef/EELSfitter>

and is composed by a number of Python scripts. The code requires a working installation of Python3 and the following libraries: NumPy, TensorFlow (v2), pandas, SciPy and scikit-learn.

`Load_data.py` This script reads the spectrum intensities and create data-frames to be used for training the neural network. It reads out the EEL spectra intensities, automatically selects the energy loss at which the peak intensity occurs and shifts the dataset such that the peak intensity is centered at $\Delta E = 0$. Further, for each spectrum it returns the normalized intensity by normalizing over the total area under the spectrum. The output is two datasets, `df` and `df_vacuum` which contain the information on the in-sample and in-vacuum recorded spectra respectively. The user needs to upload the spectral data in .txt format to the 'Data' folder and make sure that the vacuum and in-sample spectra are added to the appropriate one. For each of the spectra the minimum and maximum value of the recorded energy loss need to be set manually in `Eloss_min` and `Eloss_max`.

`Fitter.ipynb` This script is used to run the neural network training on the data that was uploaded using `load_data.py`. It involves a number of pre-processing steps to determine the hyper-parameters ΔE_I and ΔE_{II} and then it automatically prepares and cuts the data before it is fed to the neural network to start the training. It is structured as follows:

- *Importing libraries and spectral data* from the `load_data.py` script.
- *Evaluate ΔE_I from the intensity derivatives.* In order to determine the value for the hyper-parameter ΔE_I , a dataframe `df_dx` is created and it calculates the derivatives of each of the in-sample recorded spectra, stored as `df_dx['derivative y*']`, where * is any of the in-sample recorded spectra. The first crossing of any of the derivatives with zero is determined and stored as the value of ΔE_I .
- *Evaluate ΔE_{II} for the pseudo-data.* It calculates the mean over all vacuum spectra, `df_mean`, and the ratio of the intensity to the experimental uncertainty for each value of ΔE , `df_mean['ratio']`. The value of ΔE_{II} is then determined as the energy loss at which this ratio drops below 1 and is stored together with the value of ΔE_I as the hyper-parameters for training. However, if one wishes to use other values for these parameters, for instance for cross-validating the best value for ΔE_I , these can also be adjusted manually.
- *Experimental data processing.* The next step is to keep only the data points with $\Delta E \leq \Delta E_I$ and dropping the points with higher energy losses. Experimental central values and uncertainties are calculated by means of equal width discretization, for which the number of bins has to be set as `nbins`. The default value is 32, which means that 32 training inputs are spread equally over the range $[\Delta E_{min}, \Delta E_I]$. Note that the logarithm of the intensity is used as training inputs, because this facilitates the optimization of the neural network (I_{EEL} being a steeply falling function

of ΔE). The code translates this back to the original intensity values after the training. N_{pd} pseudo datapoints are added in the range $[\Delta E_{\text{II}}, \Delta E_{\text{max}}]$, where ΔE_{max} is the maximum energy loss value of the recorded spectra. The values for N_{pd} and ΔE_{max} should be changed manually by setting them in `max_x` and `N_pseudo`. The output is a dataframe `df_full` containing all training data and pseudo data points, corresponding to a total of $N_{\text{in}} (= n_{\text{bins}} + N_{\text{pd}})$ training inputs.

- *Initialize the NN model.* where the code defines the neural network architecture and prepares the data inputs to feed them to the neural network for training. The function `make_model()` allows to define the number of hidden layers and nodes per layer. The default architecture is 1-10-15-5-1.
- *Initialize data for NN training.* Here the code prepares the recorded spectra to be used as inputs for the neural network. First, we initiate placeholders for the variables `x`, `y` and `sigma` which allow us to create our operations and build our computation graph, without needing the data itself. The dimension of the placeholder is defined by `[None, dim]` where 'dim' should be set to the dimension of the corresponding variable. In this case the input is one-dimensional, so `dim=1`. These placeholders are used to define `predictions`, which is in fact a placeholder that is used later to make predictions on inputs `x`. Also, we define a vector `predict_x` that is used to make a direct prediction after training on each of the replicas. It consists of N_{pred} data points in the energy loss range `[pred_min, pred_max]`.
- *Create the Monte Carlo replicas.* The final step to be taken before we can start training is the creation of sample of N_{rep} Monte Carlo replicas of the original EEL spectra, following the procedure described in Sect. ???. This is done automatically using the experimental intensities `train_y` and uncertainties `train_sigma` for a total of `Nrep` replicas. The output is an $(N_{\text{in}}, N_{\text{rep}})$ vector containing all the MC replicas.
- *Train the neural networks.* The final part of the script, where the NN training is carried out, is based on the function `function_train()` that implements the strategy presented in Sect. 5.3. The cost function, optimizer and learning rate are defined here, together with a 'saver' used to save the network parameters after each optimization. We start a loop over `Nrep` replicas to initiate a training session on each of the individual replicas in series. For each iteration, the k -th replica is selected from the sample of N_{rep} replicas. The data is split into 80% training and 20% validation data, this partition is done at random for each replica. The resulting `train_y` and `test_y` arrays are used as training and validation labels. The total number of training epochs per session is defined in `training_epochs`. The script displays intermediate results after each number of epochs defined by `display_step`. Running the session object over the optimizer and cost function requires knowledge about the values of `x` and `sigma`, which are defined inside the `feed_dict` argument. After each epoch the average training validation costs are evaluated and the network parameters updated accordingly.

Once the maximum number of epochs had been reached, the optimal stopping point is determined by taking the absolute minimum of the validation cost and restoring the corresponding network parameters by means of the 'saver' function. From this network graph, one can directly output the prediction on the values of `train_x` and the results are stored in the array `predictions_values`. It is also possible to make predictions on any input vector of choice by feeding the vector `predict_x` to the network, which outputs an array `extrapolation`.

The datafiles that are stored upon successfully executing this script are the following:

- `Prediction_k` contains the energy loss `train_x`, the MC training data `train_y` and the ZLP prediction made on the array `train_x`, where k is the k -th replica.
- `Cost_k` contains the training and validation error for the k -th replica stored after each display step. The minimum of the validation array is used to restore the optimal neural network parameters.
- `Extrapolation_k` contains the arrays `predict_x` and the ZLP predictions made on these values.

These text files can be retrieved later to make new ZLP predictions without the need to repeat the training procedure. Further, we store the optimal network parameters after each training session in the folder 'Models/Best_models'. These can be loaded at a later stage to make predictions for an arbitrary set of input variables.

Running the loop over all replicas in series, using an input array of ~ 50 training points and a total number of training epochs of 25000 per session, takes approximately 20 seconds per optimization (~ 200 replicas per hour).

`predictions.ipynb` This script is used to analyse the predictions from the trained neural networks that have been stored in the text files indicated above.

- *Import libraries and spectral data* from the `load_data.py` script.
- *Create dataframes with all individual spectra.* In order to later subtract all the predictions from the original individual spectra, we create a datafile `original` which contains the intensity values for each of the original input spectra restricted to the region between `E_min` and `E_max`.
- *Load result files.* In order to import the files that were stored during the NN training, one should input to this script the right directions to find the prediction .txt files by adjusting the lines `path_to_data` and `path_predict`, `path_cost` and `path_extrapolate`, containing the predictions, cost function data and the extrapolation predictions respectively.
- *Post-selection criteria.* Here one select the datafiles that satisfy suitable post-fit selection criteria, such as the final error function being smaller than a certain threshold. Once these datasets have been selected and stored in an array called `use_files`, we move on to the evaluation of the ZLP predictions.
- *Subtraction.* At this step the code uses the function `matching()` to implement the matching procedure described in Sect. 7. It also automatically selects the values of ΔE_I and ΔE_{II} for the training session. If the user aims to extract the bandgap properties from the onset of I_{inel} , the `bandgap()` function can be used to fit Eq. 7.7 to the onset region.

Here the code loops over the N_{rep} replicas and reads each prediction from the extrapolation data file `predict_x`. For each replica k , the code creates a datafile containing the original spectra intensities (`original['x*']` and `original['y*']`), the predicted ZLP for this replica (`prediction y`) and the predicted ZLP after matching with each spectrum (`match *`). For each replica we subtract the matched spectrum from the original spectrum to obtain the desired subtraction: `dif * = original * - match *`. This is done for each of the total of the replicas and all these results are stored in the `total_replicas` dataframe. This file is saved in 'Data/results/replica_files' such that a user can retrieve them at any time to calculate the statistical estimators such as prediction means and uncertainties.

- Evaluate the subtracted spectra. Here the code creates a `mean_rep` file that contains all the median predictions and the upper and lower bounds of the 68% confidence intervals for the predicted ZLP, matched spectra and the subtracted spectra, for each of the original recorded spectra originally given as an input. A graphical representation of the result is then produced, showing the original spectrum, the matched ZLP and the ZLP-subtracted spectrum including uncertainty bounds.

We emphasize that the `predictions_pretrained_net.ipynb` script is similar to the `predictions.ipynb` script, but can be executed stand-alone without the need to train again the neural networks, provided that the model parameters corresponding to some previous training with the desired input settings are available. The item **load result files** is now replaced by **create result files**, which can be done by importing the pre-trained nets from the `Models` folder.

Bibliography

References

- [1] M. Terauchi, T. M., K. Tsuno, and M. Ishida, *Development of a high energy resolution electron energy-loss spectroscopy microscope*, Journal of Microscopy **194** (2005) 203–209.
- [2] B. Freitag, S. Kujawa, P. Mul, J. Ringnalda, and P. Tiemeijer, *Breaking the spherical and chromatic aberration barrier in transmission electron microscopy*, Ultramicroscopy **102** (2005) 209–214.
- [3] M. Haider, S. Uhlemann, E. Schwan, H. Rose, B. Kabius, and K. Urban, *Electron microscopy image enhanced*, Nature **392** (1998) 768–769.
- [4] J. Geiger, *Inelastic Electron Scattering in Thin Films at Oblique Incidence*, Phys. Stat. Sol. **24** (1967) 457–460.
- [5] B. Schaffer, K. Riegler, G. Kothleitner, W. Grogger, and F. Hofer, *Monochromated, spatially resolved electron energy-loss spectroscopic measurements of gold nanoparticles in the plasmon range*, Micron **40** (2008) 269–273.
- [6] R. Erni, N. D. Browning, Z. Rong Dai, and J. P. Bradley, *Analysis of extraterrestrial particles using monochromated electron energy-loss spectroscopy*, Micron **35** (2005) 369–379.
- [7] B. Rafferty and L. Brown, *Direct and indirect transitions in the region of the band gap using electron-energy-loss spectroscopy*, Physical Review B **58** (1998) 10326.
- [8] M. Stöger-Pollach, *Optical properties and bandgaps from low loss EELS: Pitfalls and solutions*, Nano Lett. **39** (2008) 1092–1110.
- [9] R. Egerton, *Electron energy-loss spectroscopy in the TEM*, Rep. Prog. Phys **72** (2009) 1.
- [10] F. García de Abajo, *Optical excitations in electron microscopy*, RevModPhys **82** (2010) 209–256, [[arXiv:0903.1669](https://arxiv.org/abs/0903.1669)].
- [11] J. Park, S. Heo, et al., *Bandgap measurement of thin dielectric films using monochromated STEM-EELS*, Ultramicroscopy **109** (2008) 1183–1188.

- [12] B. Rafferty, S. Pennycook, and L. Brown, *Zero loss peak deconvolution for bandgap EEL spectra*, *Journal of Electron Microscopy* **49** (2000) 517–524.
- [13] R. Egerton, *Electron Energy-Loss Spectroscopy in the Electron Microscope*. Plenum Press, 1996.
- [14] A. Dorneich, R. French, H. Müllejans, et al., *Quantitative analysis of valence electron energy-loss spectra of aluminium nitride*, *Journal of Microscopy* **191** (1998) 286–296.
- [15] K. van Benthem, C. Elsässer, and R. French, *Bulk electronic structure of SrTiO₃: Experiment and theory*, *Journal of Applied Physics* **90** (2001).
- [16] S. Lazar, G. Botton, et al., *Materials science applications of HREELS in near edgestructure analysis an low-energy loss spectroscopy*, *Ultramicroscopy* **96** (2003) 535–546.
- [17] R. Egerton and M. Malac, *Improved background-fitting algorithms for ionization edges in electron energy-loss spectra*, *Ultramicroscopy* **92** (2002) 47–56.
- [18] J. T. Held, H. Yun, and K. A. Mkhoyan, *Simultaneous multi-region background subtraction for core-level EEL spectra*, *Ultramicroscopy* **210** (2020) 112919.
- [19] C. S. Granerød, W. Zhan, and Øystein Prytz, *Automated approaches for band gap mapping in STEM-EELS*, *Ultramicroscopy* **184** (2018) 39–45.
- [20] K. L. Fung, M. W. Fay, S. M. Collins, D. M. Kepaptsoglou, S. T. Skowron, Q. M. Ramasse, and A. N. Khlobystov, *Accurate EELS background subtraction, an adaptable method in MATLAB*, *Ultramicroscopy* **217** (2020) 113052.
- [21] **The NNPDF Collaboration**, R. D. Ball et al., *A determination of parton distributions with faithful uncertainty estimation*, *Nucl. Phys.* **B809** (2009) 1–63, [[arXiv:0808.1231](https://arxiv.org/abs/0808.1231)].
- [22] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl.Phys.* **B867** (2013) 244, [[arXiv:1207.1303](https://arxiv.org/abs/1207.1303)].
- [23] **NNPDF Collaboration**, R. D. Ball et al., *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040, [[arXiv:1410.8849](https://arxiv.org/abs/1410.8849)].
- [24] **NNPDF Collaboration**, R. D. Ball et al., *Parton distributions from high-precision collider data*, *Eur. Phys. J.* **C77** (2017), no. 10 663, [[arXiv:1706.00428](https://arxiv.org/abs/1706.00428)].
- [25] S. E. van Heijst, M. Masaki, E. Okunishi, H. Kurata, and S. Conesa-Boj, *A nanoscale tulip garden: Fingerprinting plasmons and excitons in the petals of 2h/3r - ws2 nanoflowers*, .
- [26] O. Krivanek, N. Dellby, J. Hachtel, and M. Hotz, *Progress in ultrahigh energy resolution eels*, *Ultramicroscopy* **203** (2019) 60–67.
- [27] L. Gu, V. Srot, and W. Sigle et al., *Band-gap measurements of direct and indirect semiconductors using monochromated electrons*, *Physical Review B - Condensed Matter and Materials Physics* **75** (2007).
- [28] U. Bangert, A. Harvey, D. Freundt, and R. Keyse, *Highly spatially resolved electron energy-loss spectroscopy in the bandgap regime of GaN*, *Journal of Microscopy* **188** (1998) 237–242.

- [29] J. Hachtel, A. Lupini, and J. Idrobo, *Exploring the capabilities of monochromated electron energy loss spectroscopy in the infrared regime*, *Scientific Reports* **8** (2018) 5637.
- [30] H. Tenailleau and J. M. Martin, *A new background subtraction for low-energy EELS core edges*, *Journal of Microscopy* **116** (1992) 297–306.
- [31] B. Reed and M. Sarikaya, *Background subtraction for low-loss transmission electron energy-loss spectroscopy*, *Ultramicroscopy* **93** (2002) 25–37.
- [32] M. Bosman, M. Watanabe, D. Alexander, and V. Keast, *Mapping chemical and bonding information using multivariate analysis of electron energy-loss spectrum images*, *Ultramicroscopy* **106** (2006) 1024–1032.
- [33] G. Kothleitner and F. Hofer, *EELS performance measurements on a new high energy resolution imaging filter*, *Micron* **34** (2003) 211–218.
- [34] M. Chhowalla, H. S. Shin, G. Eda, L.-J. Li, K. P. Loh, and H. Zhangs, *The chemistry of two-dimensional layered transition metal dichalcogenide nanosheets*, *Nature Chemistry* **5** (2013) 263–275.
- [35] A. Splendiani, L. Sun, T. Li, et al., *Emerging Photoluminescence in Monolayer MoS₂*, *Nano Lett.* **10** (2010) 1271–1275.
- [36] R. Bhandavat, L. David, and G. Singh, *Synthesis of Surface-Functionalized WS₂ Nanosheets and Performance as Li-Ion Battery Anodes*, *J. Phys. Chem. Lett.* **3** (2012) 1523–1530.
- [37] W. Na, K. Kim, J.-U. Lee, and H. Cheong, *Davydov splitting and polytypism in few-layer MoS₂*, *2D Materials* **6** (2018) 015004.
- [38] D. Braga, I. G. Lezama, H. Berger, and A. F. Morpurgo, *Quantitative Determination of the Band Gap of WS₂with AmbipolarIonic Liquid-Gated Transistors*, *NanoLetters* **12** (2012) 5218.
- [39] S. Jo, N. Ubrig†, H. Berger, A. B. Kuzmenko†, and A. F. Morpurgo, *Mono- and Bilayer WS₂ Light-Emitting Transistors*, *Nano Letters* **14** (2014) 2019–2025.
- [40] J. G. et al, *Electronic Properties of Bulk and Monolayer TMDs: Theoretical Study Within DFT Framework (GVJ-2e Method)*, *Physica Status Solidi A* **214** (2007).
- [41] K. Kam and B. Parkinson, *Detailed photocurrent spectroscopy of the semiconducting group VIB transition metal dichalcogenides*, *J. Phys. Chem.* **86** (1982) 463–467.
- [42] H. Shi, H. Pan, Y.-W. Zhang, and B. I. Yakobson, *Quasiparticle band structures and optical properties of strained monolayer MoS₂ and WS₂*, *PHYSICAL REVIEW B* **87** (2013) 155304.
- [43] E. Greplova et al., *Quantum parameter estimation with a neural network*, .
- [44] J. Rojo, “Machine learning: a new toolbox for theoretical physics.” University Study Guide, 2019.
- [45] R. Hecht-Nielsen, *Neural Networks for Perception*. Elsevier, 1992.
- [46] J. Rojo, *Machine Learning tools for global PDF fits*, in *13th Conference on Quark Confinement and the Hadron Spectrum*, 9, 2018. [arXiv:1809.04392](https://arxiv.org/abs/1809.04392).

- [47] J. Gao, L. Harland-Lang, and J. Rojo, *The Structure of the Proton in the LHC Precision Era*, *Phys. Rept.* **742** (2018) 1–121, [[arXiv:1709.04922](https://arxiv.org/abs/1709.04922)].
- [48] **The NNPDF** Collaboration, L. Del Debbio, S. Forte, J. I. Latorre, A. Piccione, and J. Rojo, *Neural network determination of parton distributions: The nonsinglet case*, *JHEP* **03** (2007) 039, [[hep-ph/0701127](https://arxiv.org/abs/hep-ph/0701127)].
- [49] **NNPDF** Collaboration, R. Abdul Khalek, J. J. Ethier, and J. Rojo, *Nuclear parton distributions from lepton-nucleus scattering and the impact of an electron-ion collider*, *Eur. Phys. J.* **C79** (2019), no. 6 471, [[arXiv:1904.00018](https://arxiv.org/abs/1904.00018)].
- [50] R. Abdul Khalek, J. J. Ethier, J. Rojo, and G. van Weelden, *nNNPDF2.0: Quark Flavor Separation in Nuclei from LHC Data*, [arXiv:2006.14629](https://arxiv.org/abs/2006.14629).
- [51] **NNPDF** Collaboration, V. Bertone, S. Carrazza, N. P. Hartland, E. R. Nocera, and J. Rojo, *A determination of the fragmentation functions of pions, kaons, and protons with faithful uncertainties*, *Eur. Phys. J.* **C 77** (2017), no. 8 516, [[arXiv:1706.07049](https://arxiv.org/abs/1706.07049)].
- [52] **NNPDF** Collaboration, V. Bertone, N. Hartland, E. Nocera, J. Rojo, and L. Rottoli, *Charged hadron fragmentation functions from collider data*, *Eur. Phys. J. C* **78** (2018), no. 8 651, [[arXiv:1807.03310](https://arxiv.org/abs/1807.03310)].
- [53] O. M. Gordon and P. J. Moriarty, *Machine learning at the (sub)atomic scale: next generation scanning probe microscopy*, *Machine Learning: Science and Technology* **1** (2020) 023001.
- [54] Y. Zhang, A. Mesaros, and K. e. Fujita, *Machine learning in electronic-quantum-matter imaging experiments*, *Nature* **570** (2020) 484–490.
- [55] B. Jany, A. Janas, and F. Krok, *Retrieving the Quantitative Chemical Information at Nanoscale from Scanning Electron Microscope Energy Dispersive X-ray Measurements by Machine Learning*, *Nano Letters* **17** (2017) 6520–6525.
- [56] M. Ziatdinov, O. Dyck, A. Maksov, X. Li, X. Sang, K. Xiao, R. R. Unocic, R. Vasudevan, S. Jesse, and S. V. Kalinin, *Deep Learning of Atomically Resolved Scanning Transmission Electron Microscopy Images: Chemical Identification and Tracking Local Transformations*, *ACS Nano* **11** (2017) 12742–12752.
- [57] S. R. Young, D. C. Rose, T. P. Karnowski, S.-H. Lim, and R. M. Patton, *Optimizing deep learning hyper-parameters through an evolutionary algorithm*, in *Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments, MLHPC ’15*, (New York, NY, USA), Association for Computing Machinery, 2015.
- [58] M. Ziatdinov, O. Dyck, A. Maksov, X. Li, X. Sang, K. Xiao, R. R. Unocic, R. Vasudevan, S. Jesse, and S. V. Kalinin, *Deep learning of atomically resolved scanning transmission electron microscopy images: Chemical identification and tracking local transformations*, *ACS Nano* **11** (2017), no. 12 12742–12752, [<https://doi.org/10.1021/acsnano.7b07504>]. PMID: 29215876.
- [59] K. de Haan, Z. S. Ballard, Y. Rivenson, Y. Wu, and A. Ozcan, *Resolution enhancement in scanning electron microscopy using deep learning*, *Scientific Reports* **9** (2019), no. 1 12050.
- [60] S. Forte, L. Garrido, J. I. Latorre, and A. Piccione, *Neural network parametrization of deep-inelastic structure functions*, *JHEP* **05** (2002) 062, [[hep-ph/0204232](https://arxiv.org/abs/hep-ph/0204232)].

- [61] S. Schamm and G. Zanchi, *Study of the dielectric properties near the band gap by VEELS:gap measurement in bulk materials*, *Ultramicroscopy* **96** (2003) 559–564.
- [62] A. Hanbicki, M.Currie, G.Kioseoglou, A.L.Friedman, and B.T.Jonker, *Measurement of high exciton binding energy in the monolayer transition-metal dichalcogenides WS₂ and WSe₂*, *Solid State Communications* **204** (2016) 16–20.
- [63] B. Kaviraj and D. Sahoo, *Physics of excitons and their transport in two dimensional transition metal dichalcogenide semiconductors*, *RSC Advances* (2019).