# Automated data processing and feature identification in EELS spectral images with machine learning

author list

[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

Spectral images in Electron Energy Loss Spectroscopy (EELS) are two-dimensional sets of spectra where each pixel corresponds a highly localised region of the analysed sample. Here we present a novel approach to automated data processing and feature identification in EELS spectral images based on machine learning. The constituent spectra are clustered into groups of pixels associated to sample regions with similar thickness using unsupervised learning, and then the zero-loss peak (ZLP) is subtracted in a model-independent manner by means of deep neural networks. The resulting spectral images are processed to determine local electronic properties such as the band gap and the dielectric function across the sample, which are then correlated with the local thickness and other relevant structural properties. In addition to providing unique information on the direct correlation between structural and electrical properties, our approach makes possible the automated identification of interesting features in the spectra (say a narrow peak) then determine how these features (peak position and width) vary within different regions of the sample. This strategy, which can be straightforwardly extended to higher-dimensional datasets, is implemented into a new release of the open source code `EELSfitter`.

*Keywords:* Transmission Electron Microscopy, Electron Energy Loss Spectroscopy, Neural Networks, Machine Learning, Transition Metal Dichalcogenides, Bandgap, Dielectric Function.

*corresponding author:* `s.conesaboj@tudelft.nl`

# Contents

# 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 break-throughs such as electron monochromators [1, 2] and aberration correctors [3], modern EELS analyses can study these properties with highly competitive spatial and spectral resolution. A particularly important region of EEL spectra is the low-loss region, defined by electrons that have lost a few tens of eV, $\Delta E \lesssim 50$ eV, 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 specimen is electron-transparent, as required for TEM inspection, 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. In EEL spectra, these electrons are recorded as a narrow, high intensity peak centered at energy losses of $\Delta E \simeq 0$, 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 from the interactions of the beam electrons with the sample. Therefore, relevant signals of low-loss phenomena such as excitons, phonons, and intraband transitions risk becoming drowned in the ZLP tail [10]. An accurate removal of the ZLP contribution is thus crucial in order to accurately chart and identify the features of the low-loss region in EEL spectra.

In monochromated EELS, the properties of the ZLP 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 is associated to interactions with the sample such as atomic scatterings, phonon excitation, and exciton losses. This implies that EEL measurements in vacuum can be used for calibration purposes but not to subtract the ZLP from spectra taken on specimens, since their shapes will in general differ.

Several approaches to ZLP subtraction [8, 12, 13] have been put forward in the literature. These are often based on specific model assumptions about the ZLP properties, in particular concerning its parametric functional dependence on the electron energy loss $\Delta E$, from Lorentzian [14] and power laws [6] to more general multiple-parameter functions [15]. Another approach is based on 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 [17–20]. These various methods are however affected by three main limitations. Firstly, their reliance on model assumptions such as the choice of fit function introduces 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. Thirdly, *ad hoc* choices such as those of the fitting ranges introduce a significant degree of arbitrariness in the procedure.

In this study we bypass these limitations by developing a model-independent strategy to realise a multidimensional determination of the ZLP with a faithful uncertainty estimate. Our approach is based on machine learning (ML) techniques originally developed in high-energy physics to study the quark and gluon substructure of protons in 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 end result is a faithful sampling of the probability distribution in the ZLP space which can be used to subtract its contribution to EEL spectra while propagating the associated uncertainties. One can also extrapolate the predictions from this ZLP parametrisation to other TEM operating conditions beyond those included in the training dataset.

This work is divided into two main parts. In the first one, we construct a ML model of ZLP spectra acquired in vacuum, which is able to accommodate an arbitrary number of input variables corresponding to different operation settings of the TEM. We demonstrate how this model successfully describes the input spectra and we assess its extrapolation capabilities for other operation conditions. In the second part, we construct a one-dimensional model of the ZLP as a function of $\Delta E$ from spectra acquired on two different specimens of tungsten disulfide ($WS_2$) nanoflowers characterised by a 2H/3R mixed polytypism [25]. The resulting subtracted spectra are used to determine the value and nature of the $WS_2$ bandgap in these nanostructures as well as to map the properties of the associated exciton peaks appearing in the ultra-low loss region.

This paper is organized as follows. First of all, in Sect. **??** we review the main features of EELS and present the $WS_2$ nanostructures that will be used as proof of concept of our approach. In Sect. 3 we describe the machine learning methodology adopted to model the ZLP features. Sects. 4 and **??** contain the results of the ZLP parametrisation of

spectra acquired in vacuum and in specimens respectively, which in the latter case allows us to probe the local electronic properties of the $WS_2$ nanoflowers. Finally in Sect. 5 we summarise and outline possible future developments. Our results have been obtained with an open-source PYTHON code, dubbed `EELSfitter`, whose installation and usage instructions are described in Appendix **??**. Furthermore, we discuss the possible role played by correlated uncertainties in the training dataset in Appendix **??**.

# 2 Theory overview

In this section we review the theory necessary to compute the local dielectric functions, thickness, and the deconvolution procedure, in order to make sure the paper is self consistent.

## 2.1 Deconvolution of the single scattering spectrum

In order to evaluate the local dielectric function of a material from its energy-loss spectrum, we need to deconvolute this spectrum and obtain the single scattering distribution.

When electrons go through the sample, the intensity of electrons that has no inelastic scattering is given by the zero-loss peak: $I_{ZLP}(E)$. The intensity of the electrons that do scatter, $I_{EEL}(E)$, is than dividable in the single scatter intensity, $I_1(E)$, the double scatter intensity, $I_2(E)$, the triple scatter intensity, $I_3(E)$, etc:

$$I(E) = I_{ZLP}(E) + I_{EEL}(E) = I_{ZLP}(E) + \sum_{n=0}^{\infty} I_n(E). \tag{2.1}$$

The integrated intensity of each n-scattering spectrum $N_n$ depends on the total integrated intensity $N$, assuming independed scattering events, through the bionomal distribution:

$$N_n = \frac{N}{n!} \left(\frac{t}{\lambda}\right)^n \exp\left[-t/\lambda\right]. \tag{2.2}$$

Here $t$ is the thickness of the sample, and $\lambda$ is the mean free path of electrons in the sample. END DISREGARD

Since we know the zero-loss-peak due to Lau [**?**], the response function of the instrument, $R(E)$, is easily deducted by:

$$R(E) = I_{ZLP}(E)/N_{ZLP}, \tag{2.3}$$

where $N_{ZLP}$ is the integrated intensity of $I_{ZLP}(E)$.

Now we will show how the total recorded spectrum is build up from single-scattering distribution $S(E)$, and the above defined response function $R(E)$.

The spectrum recorded due to the single scattering events, $J^1(E)$, is these two distributions convoluted:

$$I_1(E) = R(E)^* S(E) \equiv \int_{-\infty}^{\infty} R\left(E - E'\right) S\left(E'\right) dE'. \tag{2.4}$$

It can be easily seen, that as a double-scattering event, is a series of two single-scattering event, the double-scattering intensity is given by the self convolution of the single-scattering intensity, normalised to match eq. (2.2), and once again convoluted with the response function:

$$I_2(E) = R(E)^* S(E)^* S(E)/\left(2! N_{ZLP}\right). \tag{2.5}$$

For higher order scattering spectra, this generalises to:

$$I_n(E) = R(E)\left[^*S(E)\right]^n / \left(n! N_{ZLP}^{n-1}\right).\tag{2.6}$$

The complete recorded spectrum, neglecting any surface plasmons, is than given by (analogous to eq. (2.1)):

$$
\begin{aligned}
I(E) &= I_{ZLP}(E) + I^1(E) + I^2(E) + I^3(E) + \cdots \\
&= I_{ZLP}(E)^* \left[\delta(E) + S(E)/N_{ZLP} + S(E)^*S(E)/\left(2! N_{ZLP}^2\right)\right. \\
&\quad \left.+ S(E)^*S(E)^*S(E)/\left(3! N_{ZLP}^3\right) + \cdots\right] \\
&= I_{ZLP}(E)^* \left[\delta(E) + \sum_{n=1}^{\infty} \delta(E)\left[^*S(E)\right]^n / \left(n! N_{ZLP}^n\right)\right].
\end{aligned}
\tag{2.7}
$$

Since a convolution in energy domain translates to a multiplication in the frequency domain, it makes sense to take the Fourier transform (FT) of the above equation. Eq. (2.7) than becomes, using the taylor expansion of the exponential function:

$$
\begin{aligned}
i(\nu) &= z(\nu)\left\{1 + s(\nu)/N_{ZLP} + [s(\nu)]^2/\left(2! N_{ZLP}^2\right) + [s(\nu)]^3/\left(3! N_{ZLP}^3\right) + \cdots\right\} \\
&= z(\nu) \sum_{n=0}^{\infty} \frac{s(\nu)^n}{n! N_{ZLP}^n} \\
&= z(\nu) \exp\left[s(\nu)/N_{ZLP}\right],
\end{aligned}
\tag{2.8}
$$

where $i(\nu)$ is the FT of the intensity spectrum, $z(\nu)$ is the FT of the zero-loss peak, and $s(\nu)$ is the FT of the single-scattering distribution.

The single scattering distribution can than be retrieved by rewriting eq. (2.8), and taking the inverse Fourier transform:

$$s(\nu) = N_{ZLP} \ln\left(\frac{i(\nu)}{z(\nu)}\right),\tag{2.9}$$

$$
\begin{aligned}
S(E) &= \mathcal{F}^{-1}\left\{s(\nu)\right\} \\
&= \mathcal{F}^{-1}\left\{N_{ZLP} \ln\left[\frac{i(\nu)}{z(\nu)}\right]\right\} \\
&= \mathcal{F}^{-1}\left\{N_{ZLP} \ln\left[\frac{\mathcal{F}\{I(E)\}}{\mathcal{F}\{I_{ZLP}(E)\}}\right]\right\}.
\end{aligned}
\tag{2.10}
$$

However, eq. (2.10) only works for an "ideal" spectrum. Any noise on the spectrum will blow up, as noise normally preveals itself at higher frequencies, and $i(\nu)$ tends towards zero for high frequencies. Therefor, it is advised to calculate not $S(E)$, but $I_1(E)$, by convoluting once more with $I_{ZLP}(E)$, see eq. (2.4). [**?**]

**Discussion points**   What is the most official way to calculate the errors in R(E), from the errors in the ZLP? Just calculate for each ZLP and take the std of those, or can you use the error of the ZLP to calculate it at once? Because it comes back in the $N_{ZLP}$ as well.

### 2.1.1 Analytical analysis of possible recorded spectra

To be able to analyse a deconvolution program, it is usefull to create some toymodels which represent different possible $I(E)$, of which we know what the expected outcome is, so we can validate the program written.

**Single scattering distribution as Gaussian**  One of the interesting approximations of a spectrum to review as toy model, is a spectrum in which the zero loss peak is a gaussian such that:

$$I_{ZLP}(E) = \frac{N_{ZLP}}{\sqrt{2\pi}\sigma_{ZLP}} \exp\left[-(x - \mu_{ZLP})/(2\sigma_{ZLP}^2)\right], \tag{2.11}$$

and the single scattering distribution is a gaussian, given by:

$$S(E) = \frac{A_S}{\sqrt{2\pi}\sigma_S} \exp\left[-(x - \mu_S)/(2\sigma_S^2)\right]. \tag{2.12}$$

By combining eq. (2.1), (**??**), and eq. (2.6) you obtain for complete recorded spectrum $I(E)$:

$$
\begin{aligned}
I(E) &= \sum_{n=0}^{\infty} \frac{A_n}{\sqrt{2\pi}\sigma_n} \exp\left[-\frac{(x - \mu_n)^2}{2\sigma_n^2}\right], \\
A_n &= \frac{1}{n! N_{ZLP}^n} N_{ZLP} A_S^n, \\
&= \frac{A_S^n}{n! N_{ZLP}^{n-1}}, \\
\mu_n &= \mu_{ZLP} + n\mu_S, \\
\sigma_n &= (\sigma_{ZLP}^2 + n\sigma_S^2)^{1/2}.
\end{aligned}
\tag{2.13}
$$

This means that for an $I(E)$ constructed as the equation above, with our program, we hope to retrieve $S(E)$ as given by (2.12).

## 2.2 Dielectric function

The dielectric function, also called permittivity, is a measure for the polarisability of a material. From the dielectric function, a multitude of other characteristics can be calculated. Since the dielectric function can be extracted from the electron energy loss spectrum through Kramer-Kronig analysis, for an image of spectra, the dielectric function can be calculated for each pixel. [**?**]

From [**?**]:

-Cold field STEM?

Uses deconvolution and Kramer-Kronig relations to determine optical properties of wide-gap materials, specifically III-nitrate compounds (important for optoelectronics).

These properties are tested by comparing them to experimental measurements with synchrotron ellipsometry and theoretical studies.

The resolution is proven by determining the optical features of single layers in a heterostructure and in individual defects within wurtize GaN.

The low loss region of EELS is interesting, as it provides information on excitations of electron in the outer shell, which determine the optical properties of the material. Within the low loss region of the EELS, one can distinguish valence electron exitations, single electron excitations (interband transmissions?, depend on critical points in the bandstructure) and collective plasma oscillations (i.e. plasmon excitations?).

An alternative method to determine the optical features of material is through optical measurements. These offer a significantly higher energy resolution: 0.001eV-0.5eV where state of the art STEM reach just 0.5eV. The spatial resolution of optical measurements however, is significantly worse mainly due to the higher de Broglie wavelength of photons in respect to electrons [**?**]: order microns, whereas STEM can reach sub nanometer.

" By the use of subnanometer electron probes the spatial resolution of the measurements depends on the physical localization of the scattering process itself and thus is in the range of nanometers for low energy excitations." ??

In the low loss region of the EELS, the recorded energy loss function depends on the interjoint density of states between valence and conducting electrons. Peaks arise at critical areas where nesting occurs: where the Fermi surface of the valence electrons is (near) parralel to the Fermi surface of the conducting electrons. In the dielectric formulation, this energy loss function can be described as the response of the material to the passing electron probe, which is proportial to the imaginary part of the reciprocal of the dielectric function $\varepsilon$. Through deconvolution of the signal and Kramer-Kronig relations the complete dielectric function can subsequently be retrieved, see Section 2.4.

### 2.2.1 Relations optical properties to dieelectric function

From [?]:

## 2.3 Calculation of the dielectric function

## 2.4 Kramer-Kronig relations

The Kramer-Kronig relations are two functions that relate the imaginary part of an complex function to the real part and vice versa. The relations hold as long as the complex function is analytic in the upper half-plane. The relations for function $\chi(\omega) = \chi_1(\omega) + i\chi_2(\omega)$, with $\omega$ a complex variable are given by [?]:

$$\chi_1(\omega) = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty}\frac{\chi_2(\omega')}{\omega'-\omega}d\omega', \tag{2.14}$$

and:

$$\chi_2(\omega) = -\frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty}\frac{\chi_1(\omega')}{\omega'-\omega}d\omega'. \tag{2.15}$$

Here $\mathcal{P}$ denotes the Cauchy principal value of the integral. For causal functions, due to (anti)symmetries arrising from its causality, these can be rewritten to [?]:

$$\chi_1(\omega) = \frac{2}{\pi}\mathcal{P}\int_{0}^{\infty}\frac{\omega'\chi_2(\omega')}{\omega'^2-\omega^2}d\omega', \tag{2.16}$$

and:

$$\chi_2(\omega) = -\frac{2}{\pi}\mathcal{P}\int_{0}^{\infty}\frac{\omega\chi_1(\omega')}{\omega'^2-\omega^2}d\omega'. \tag{2.17}$$

Since the single scattering spectrum of a medium can be related to the imaginary part of the complex permittivity, the Kramer-Kronig relations can be used to retrieve energy dependence of the real permittivity [?].

### 2.4.1 Spectrum analysis

If one ignores the instrumental broadening, surface-mode scattering and the retardation effects, the single scattering spectrum is approached by the single scattering distribution, which in place can be obtained from the recorded energy loss spectrum by the Fourier log method. [?]

$$I_1(E) \approx S(E) = \frac{2N_{ZLP}t}{\pi a_0 m_0 v^2} \text{Im}\left[\frac{-1}{\varepsilon(E)}\right] \int_0^\beta \frac{\theta d\theta}{\theta^2 + \theta_E^2}$$

$$= \frac{N_{ZLP}t}{\pi a_0 m_0 v^2} \text{Im}\left[\frac{-1}{\varepsilon(E)}\right] \ln\left[1 + \left(\frac{\beta}{\theta_E}\right)^2\right] \tag{2.18}$$

In this equation is $J^1(E)$ the single scattering distribution, $S(E)$ the single scattering spectrum, $N_{ZLP}$ the zero-loss intensity, $t$ the sample thickness, $v$ the velocity of the incoming electron, $\beta$ the collection semi angle, $\alpha$ the angular divergence of the incoming beam, and $\theta_E$ the characteristic scattering angle for energy loss $E$. In this equation $\alpha$ is assumed small in comparison with $\beta$. If this is not the case, additional angular corrections are needed. Furthermore, $\theta_E$ is given by:

$$\theta_E = E/(\gamma m_0 v^2). \tag{2.19}$$

Furthermore, it should be noted that to retrieve $\text{Re}\left[1/\varepsilon(E)\right]$ from $\text{Im}\left[-1/\varepsilon(E)\right]$, equation (2.16) should be rewritten to [**?**]:

$$\text{Re}\left[\frac{1}{\varepsilon(E)}\right] = 1 - \frac{2}{\pi}\mathcal{P}\int_0^\infty \text{Im}\left[\frac{-1}{\varepsilon(E')}\right]\frac{E'dE'}{E'^2 - E^2}. \tag{2.20}$$

### 2.4.2 Step 1: rescaling intensity

The first step of the K-K analysis is now to rewrite Eq. (2.18) to:

$$I_{1,ac}(E) = \frac{I_1(E)}{\ln\left[1 + \left(\frac{\beta}{\theta_E}\right)^2\right]} = \frac{N_{ZLP}t}{\pi a_0 m_0 v^2} \text{Im}\left[\frac{-1}{\varepsilon(E)}\right]. \tag{2.21}$$

As $\theta_E$ scales linearly with $E$, see eq. (2.19), the intensity in on the left side of the equation above now relatively increases for high energy loss with respect to low energy loss.

**Discussion points**  I assume $\beta$ and $v$ are known, and that we do not take a distribution for $v$?

### 2.4.3 Step 2: extrapolating

Since the upcoming integrals all extend to infinity, but the data acquisition is inherently up to a finite energy, the spectra need to be extrapolated. An often used form is $AE^{-r}$, where $r = 3$ if you follow the Drude-model, or $r$ can be deducted from experimental data.

### 2.4.4 Step 3: normalisation and retrieving $\text{Im}\left[\frac{1}{\varepsilon(E)}\right]$

Taking $E' = 0$ in (2.20), one obtains:

$$1 - \text{Re}\left[\frac{1}{\varepsilon(0)}\right] = \frac{2}{\pi}\int_0^\infty \text{Im}\left[\frac{-1}{\varepsilon(E)}\right]\frac{dE}{E}. \tag{2.22}$$

Now dividing both sides of Eq. (2.23) by the energy, and subsequently integrating them over energy results in a comparable integral:

$$\int_0^\infty I_{1,ac}(E)\frac{dE}{E} = \frac{N_{ZLP}t}{\pi a_0 m_0 v^2}\int_0^\infty \text{Im}\left[\frac{-1}{\varepsilon(E)}\right]\frac{dE}{E}. \tag{2.23}$$

Combining the two leads to:

$$\frac{\int_0^\infty I_{1,ac}(E)\frac{dE}{E}}{\frac{\pi}{2}\left(1 - \text{Re}\left[\frac{1}{\varepsilon(0)}\right]\right)} = \frac{N_{ZLP}t}{\pi a_0 m_0 v^2} \equiv K, \tag{2.24}$$

in which $K$ is the proportionality constant, used to estimate the absolute thickness if the zero-loss integral and the indicent energy are known. This formula requires $\text{Re}\left[\frac{1}{\varepsilon(0)}\right]$ to be known, as is the case in for example metals ($\text{Re}\left[\frac{1}{\varepsilon_{metal}(0)}\right] \approx 0$). If this is not the case, other options to estimate $K$ will be discussed later on.

This value of $K$, which is constant over $E$, can than in turn be used to retrieve the function of $\text{Im}\left[-\frac{1}{\varepsilon(E)}\right]$ from the observed single scattering energy distribution $J^1(E)$ with eq. (2.23).

### 2.4.5 Step 4: retrieving $\text{Re}\left[\frac{1}{\varepsilon(E)}\right]$

Having retrieved $\text{Im}\left[-\frac{1}{\varepsilon(E)}\right]$ from the steps above, one can now use eq. (2.20) to obtain $\text{Re}\left[\frac{1}{\varepsilon(E)}\right]$, where one must pay attention to avoid including $E = E'$ in the discrete integral over the spectrum, as this is a singularity. To avoid this singularity in a discrete signal, a couple of approaches are possible:

- In the integral (for discrete signals: summation) in eq. (2.20), simply exclude the $E = E'$ values.

- Shift the values of $\text{Re}\left[\frac{1}{\varepsilon(E)}\right]$ to values at $E_i'' = (E_i + E_{i+1})$, to make sure to avoid $E'' = E'$ in the summation.

- The dielectric function in the energy domain relate to the dielectric response function $1/\varepsilon(t) - \delta(t)$ through:

$$\text{Re}\left[\frac{1}{\varepsilon(E)}\right] = \mathcal{C}\left\{\frac{1}{\varepsilon(t)} - \delta(t)\right\} = \mathcal{F}\{p(t)\}, \tag{2.25}$$

and

$$\text{Im}\left[\frac{-1}{\varepsilon(E)}\right] = \mathcal{S}\left\{\frac{1}{\varepsilon(t)} - \delta(t)\right\} = i\mathcal{F}\{q(t)\}, \tag{2.26}$$

where $p(t)$ and $q(t)$ are the even and odd parts respectively of the dielectric response function, and $\mathcal{C}$ and $\mathcal{S}$ are the cosine and sine Fourier transforms respectively. Since the dielectric response function is a response function and therefor causal, it is 0 for $t < 0$. This results in:

$$p(t) = \text{sgn}[q(t)]. \tag{2.27}$$

Combining all this means that one can also obtain $\text{Re}\left[\frac{1}{\varepsilon(E)}\right]$ from $\text{Im}\left[-\frac{1}{\varepsilon(E)}\right]$ by:

$$\text{Re}\left[\frac{1}{\varepsilon(E)}\right] = \mathcal{C}\left\{\text{sgn}\left[\mathcal{S}^{-1}\left\{\text{Im}\left[\frac{-1}{\varepsilon(E)}\right]\right\}\right]\right\}. \tag{2.28}$$

### 2.4.6 Step 5: retrieving $\varepsilon$

The dielectric function can subsequently be obtained from:

$$\varepsilon(E) = \varepsilon_1(E) + i\varepsilon_2(E) = \frac{\text{Re}[1/\varepsilon(E)] + i\,\text{Im}[-1/\varepsilon(E)]}{\{\text{Re}[1/\varepsilon(E)]\}^2 + \{\text{Im}[-1/\varepsilon(E)]\}^2}. \tag{2.29}$$

## 2.5 Dielectric function

The dielectric function, also called permittivity, is a measure for the polarisability of a material. From the dielectric function, a multitude of other characteristics can be calculated. Since the dielectric function can be extracted from the electron energy loss spectrum through Kramer-Kronig analysis, for an image of spectra, the dielectric function can be calculated for each pixel. [?]

From [?]:

-Cold field STEM?

Uses deconvolution and Kramer-Kronig relations to determine optical properties of wide-gap materials, specifically III-nitrate compounds (important for optoelectronics).

These properties are tested by comparing them to experimental measurements with synchrotron ellipsometry and theoretical studies.

The resolution is proven by determining the optical features of single layers in a heterostructure and in individual defects within wurtize GaN.

The low loss region of EELS is interesting, as it provides information on excitations of electron in the outer shell, which determine the optical properties of the material. Within the low loss region of the EELS, one can distinguish valence electron exitations, single electron excitations (interband transmissions?, depend on critical points in the bandstructure) and collective plasma oscillations (i.e. plasmon excitations?).

An alternative method to determine the optical features of material is through optical measurements. These offer a significantly higher energy resolution: 0.001eV-0.5eV where state of the art STEM reach just 0.5eV. The spatial resolution of optical measurements however, is significantly worse mainly due to the higher de Broglie wavelength of photons in respect to electrons [?]: order microns, whereas STEM can reach sub nanometer.

" By the use of subnanometer electron probes the spatial resolution of the measurements depends on the physical localization of the scattering process itself and thus is in the range of nanometers for low energy excitations." ??

In the low loss region of the EELS, the recorded energy loss function depends on the interjoint density of states between valence and conducting electrons. Peaks arise at critical areas where nesting occurs: where the Fermi surface of the valence electrons is (near) parralel to the Fermi surface of the conducting electrons. In the dielectric formulation, this energy loss function can be described as the response of the material to the passing electron probe, which is proportian to the imaginary part of the reciprocal of the dielectric function $\varepsilon$. Through deconvolution of the signal and Kramer-Kronig relations the complete dielectric function can subsequently be retrieved, see Section 2.4.

### 2.5.1 Relations optical properties to dieelectric function

From [?]:

## 2.6 Benchmarking

Here we present some benchmarking comparing our calculations of things like the deconvoluted spectrum, the dielectric function, and the thickness of the same with independent software packages, in parrticular with the Egerton software and with `HyperSpy`. This way we validate that our theory calculations are kosher.

## 2.7 The role of surface effects

In the previous discussion we consider the calculation of the local dielectric function using only bulk effects. But surface effects might be important for some of the nanostructures

that we are considering, so we should say something about surface effects in the calculation of the dielectreic function here.

# 3 Methodology

In this section we present our strategy to parametrise and subtract in a model-independent manner the zero-loss peak that arises in the low-loss region of EEL spectra by means of machine learning.

## 3.1 Clustering

Here we explain the $k$-means clustering to classify the individual spectra into groups. Definition of the figure of merit, determination of the number of clusters, how to treat outliers within clusters. Maybe compare with other methods for clustering?

## 3.2 Neural network training

Here we explain that we use a NN with two inputs to train the spectra data, motivate the choice of preprocessing, validation of the training.

We also explain how we evaluate the total errors and the covariance matrix, which is used to define the figure of merit. Show that once we include the covariance matrix we obtain $\chi^2/n_{\text{dat}} \simeq 1$.

We explain how as opposed to the previous paper now we don't need to generate replicas, since we have sufficient spectra per cluster to pick one at random.

We need to revisit the choice of $\Delta E_{\text{I}}$ and of $\Delta E_{\text{II}}$, in particular since we don't have vacuum anymore but some material. So we need to use a different, more general criteria.

Diffferences in the treatment of the EELS data as compared to the previous paper, for example now we don't need the binning

## 3.3 Calculation of dielectic function

Here we explain how the subtracted ZLP spectra are used to evaluate the dielectric function, and its uncertainties. Explain how we determine interesting properties of the dielecttic function which can be represented as a 2D image

## 3.4 Automated feature classification

We explain how we use ML to indentify relevant features in the spectra, say peaks, and how we can study how their features vary along the sample, We can use multivariate techniques to identify anomalies in the data.

## 3.5 Comparison with Laurien's paper

Check that with the new code we reproduce in a reasonable manner the results of Laurien's paper

# 4 Results

We now move to discuss the application of the strategy presented in the previous section to the parametrisation of ZLP spectra acquired in vacuum.

Some of the plots that we need for this section are

A bunch of fits (by "fit" I mean the full model say with 500 "replicas" at least) for different values of DeltaEI

A heat map plot showing the calculation bandgap across the spectral image, using the approach in Laurien paper

A heat map plot for the thickness, now computed with the deconvoluted spectra just as with the sum of intensities, checking that the results of the two approaches are consistent

A plot of the dielectric function (with uncertainties) for different locations in the spectral image (representative).

A 2D plot demonstrating that the model interpolates in a sensible manner in the "intensity input"

A heat map with the crossing of the x axis for the real (or was it the imaginary) part of the dielectric function across the image

Some explicit test of the stability of the model fitting, for example comparing the results of two fits trained on different random subsets of spectra

A fit that includes the full covariance matrix in the definition of the chi2, rather than only the diagonal component

## 4.1 Direct correlation of structural and electrical properties

Then at some point we can produce correlation plots, for example assessing whenever there is a correlation between thickness and change in the location of the bandgap etc. We should find measures that allow us to identify in an automated way once we have these kind of correlations. We can define for example a local correlation coefficient between two distinct featires of the sample

# 5 Summary and outlook

In this work we have presented a novel, model-independent strategy to parametrise and subtract the ubiquitous zero-loss peak that dominates the low-loss region of EEL spectra.

Say somethihg about the connection of our approach with indirect Dark Matter searches and how we can efficiently test new materials that can eventually be used as dark matter detectors.

## Acknowledgments

## Funding

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Methods

The EEL spectra used for the training of the vacuum ZLP model presented in Sect. 4 were collected in a ARM200F Mono-JEOL microscope equipped with a GIF continuum spectrometer and operated at 60 kV and 200 kV. For these measurements, a slit in the monochromator of 2.8 $\mu$m was used. The TEM and EELS measurements acquired in Specimen A for the results presented in Sect. ?? were recorded in a JEOL 2100F microscope with a cold field-emission gun equipped with aberration corrector operated at 60 kV. A Gatan GIF Quantum was used for the EELS analyses. The convergence and collection semi-angles were 30.0 mrad and 66.7 mrad respectively. The TEM and EELS measurements acquired for Specimen B in Sect. ?? were recorded using a JEM ARM200F monochromated microscope operated at 60 kV and equipped with a GIF quantum ERS. The convergence and collection semi-angles were 24.6 mrad and 58.4 mrad respectively in this case, and the aperture of the spectrometer was set to 5 mm.

## 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`].

[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].

[22] R. D. Ball et al., *Parton distributions with LHC data*, Nucl.Phys. **B867** (2013) 244, [arXiv:1207.1303].

[23] **NNPDF** Collaboration, R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, [arXiv: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].

[25] S. E. van Heijst et al., *Fingerprinting 2H/3R Polytypism in WS2 Nanoflowers from Plasmons and Excitons to Phonons*, . in preparation.