

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 2 |
| 2 | Results | 4 |
| 2.1 | Qualitative results | 5 |
| 2.1.1 | Each sample area | 5 |
| 2.1.2 | General results from the spectra | 10 |
| 2.1.3 | Calibration | 11 |
| 2.2 | Quantitative results | 13 |
| 3 | Discussion | 16 |
| 3.1 | Analysis steps in HyperSpy | 16 |
| 3.1.1 | Loading the data and specifying the elements | 16 |
| 3.1.2 | Removing the background linearly | 18 |
| 3.1.3 | Quantification after linear background removal | 18 |
| 3.1.4 | Removing the background with model fitting | 19 |
| 3.1.5 | Quantification after model fitting | 20 |
| 3.1.6 | Calibrating the spectrum with the HyperSpy model | 20 |
| 3.2 | Peak and background modelling | 21 |
| 3.3 | Calibration | 22 |
| 3.4 | Background models | 22 |
| 3.5 | Analysis failure | 22 |
| 3.6 | Calibration decision | 22 |
| 3.7 | Choices in HyperSpy | 22 |

Chapter I

Introduction

The main goal of this project is to improve EDS analysis. There are multiple ways to do this. One way is to make the analysis more transparent, which would make it easier to understand and use. A second option is to improve the input parameters of the analysis by control checking the instrument with a known sample. A third way is to make the quantitative analysis more accurate, which would improve EDS analysis. In this project, the main focus have been trying to improve the transparency of the analysis. Thus, the problem statement was formulated as:

Main problem statement. How does different spectroscopy data processing influence the quantitative Cliff-Lorimer analysis in HyperSpy?

Most of the time used in this project was spent on trying to understand the different methods of data processing in EDS analysis. Towards the end the different methods were applied on the same data set using the Cliff-Lorimer quantification in HyperSpy to see how the different methods influence the quantitative analysis. Solving the main problem statement with an open-source Jupyter Notebook would increase the transparency of the analysis and allow users to both adjust their analysis and understand better the different steps in the analysis. The main problem statement was broken down into five sub-problems. The sub-problems and a short description follow below.

Sub-problem 1. How accurate is the out-of-the-box quantification in AZtec and HyperSpy?

Sub-problem 2. What are done with the data at the different steps in the analysis when using HyperSpy?

Sub-problem 3. How are the peaks and the background be modelled?

Sub-problem 4. How is the spectrum calibrated, and is AZtec different from HyperSpy?

Sub-problem 5. When does the analysis fail, both in AZtec and HyperSpy?

(**Question for Ton: Should I have a short paragraph here about the status of EDS analysis today?**)

Paragraph about why EDS in AZtec is for dummies.

(**Brynjart: Paragraph about Dispersion, offset, energy resolution?**)

(**Brynjart: Paragraph about Other parameters of EDS analysis?**)

(**Brynjart: Paragraph about Improving quantitative EDS analysis?**)

This remainder of this report is built up around the main problem statement and the sub-problems. The theory chapter contains the physics of X-rays and empirical adjustments in the analysis, a section on data processing, and a section about the hardware in an EDS setup. The method chapter explains how the data was collected, while the arguments for and against the different methods are presented in the discussion chapter. The result chapter contains qualitative and quantitative results. The qualitative results are presented as figures of spectra showing the elements in the sample, and also how well different calibrations fit with the theoretical values. The quantitative results are presented as tables with compositional results with different methods and adjustments. The different methods are using AZtec and two approaches in HyperSpy. The different adjustments are results with different calibrations, different background models, (**Brynjart: "and different peak models"?**) The discussion chapter follow the structure of the sub-problems, and discuss both the methods and the results of the analysis consecutively. The conclusion chapter summarizes the report with an answer to the main problem statement, and provides ideas for further work. The appendix contains the code used in the analysis, which is also available on GitHub (**Brynjart: Link to GitHub**).

Chapter 2

Results

The results are presented in this chapter. First qualitative then quantitative results are presented. All the spectra taken were qualitatively analyzed. Only the GaAs bulk spectra was quantitatively analyzed. The spectrum from the GaAs bulk wafer taken on 30 kV is shown in Figure 2.1. This plot was made with HyperSpy, which utilize Matplotlib for plotting. The plotting method in HyperSpy can add where the theoretical peak centers are. The lines added also show an estimate of the weight of the peak.

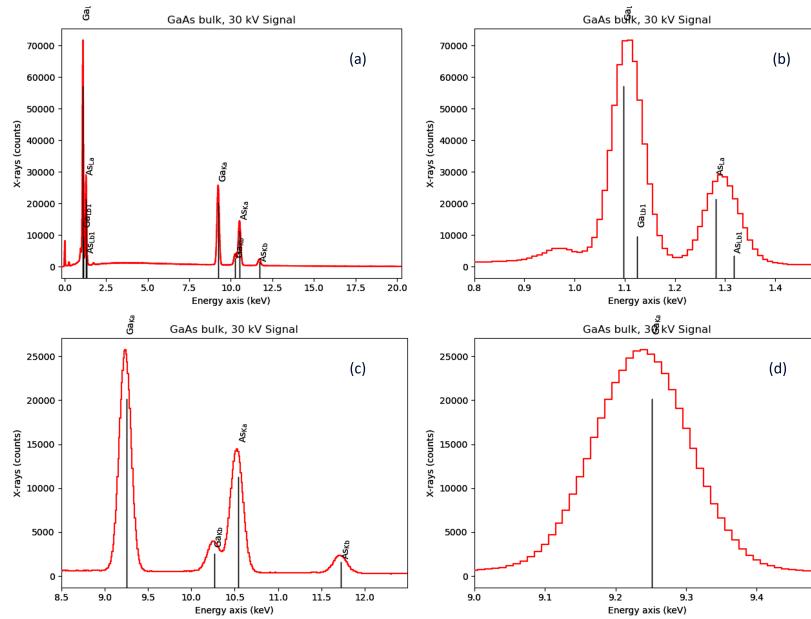


Figure 2.1: The GaAs spectrum taken at 30 kV. (a) is the whole spectrum. (b) is the zoomed in on the L-peaks. (c) is zoomed in on the K-peaks. (d) is zoomed in on the Ga K α peak. This plot has the calibration from AZtec, and it is clear that the line position is deviating from the center of the peaks.

2.1 Qualitative results

This qualitative section present first the spectra from each sampled area, then some general remarks and finally the calibration of the spectra. The spectra are presented as plots and the calibrations are presented as tables.

2.1.1 Each sample area

Figure 2.2 to Figure 2.8 shows the spectra for the six different areas of the sample plotted with Plotly. The plots are available as an interactive HTML plots on the GitHub repository. ([Brynjart: Upload the HTML files to the GitHub repository](#)) The calibration used in these spectra is based on the calibration of Ga L α and As K α from the GaAs bulk wafer. The y-axis is normalized to the highest peak value in each spectrum, i.e. the highest peak is always 1. The spectra are scaled to show the region of interest for each sample area. Peaks which are taller than the y-axis are marked with a gray stippled line. The annotated energy on the plots are where the annotation line ends, and it not necessarily exactly the center of the peak when the peak is Gaussian fitted and nor exactly the theoretical line value. Even though the annotated values deviate a few eV, the goal in the qualitative analysis is to figure out what the peaks are. ([Brynjart: Move to discussion?](#))

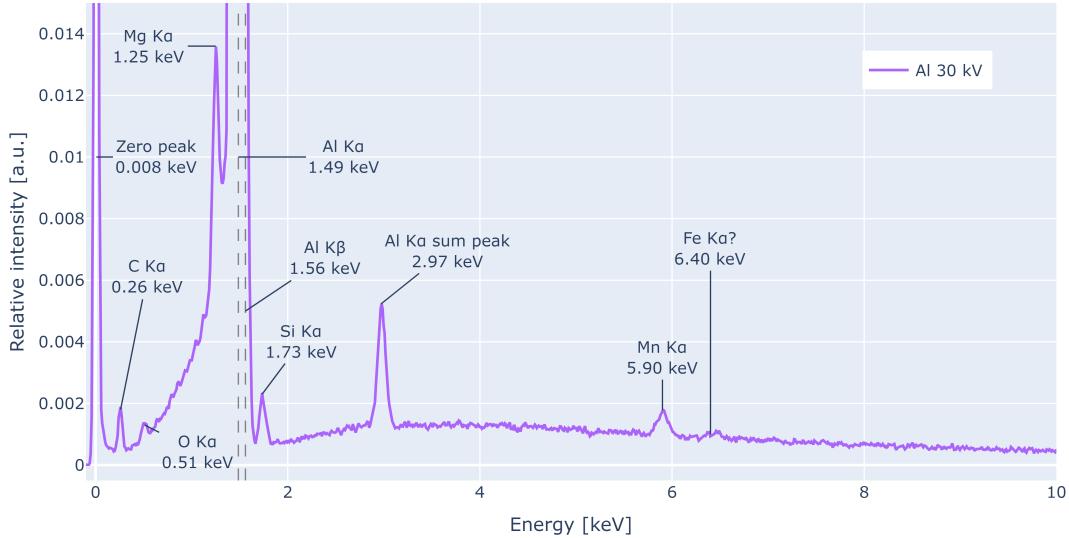


Figure 2.2: The spectrum of the FIB stub, which gave a strong Al signal. This was expected to be Fe, but very little Fe signal was found. The tallest peak is at 1.49 keV, and it is the Al K α peak with some signal at the K β peak at 1.56 keV. The relative weight for Al K β to K α is 0.013 (from HyperSpy). Fe K α at 6.40 keV has a question mark, because the FIB stub was initially expected to be Fe. The signal from Fe K α is barely a signal different from the background.

Figure 2.2 shows the spectrum of the FIB stub, which gave a strong Al signal. Only one spectrum was taken of the FIB stub, and it was taken at 30 kV. As for all the other spectra, this spectrum have a zero peak, a C K α peak at 0.260 keV and a O K α peak at 0.51 keV. Most of the other spectra also have a Si K α peak at 1.74 keV. The tallest peak with a relative

intensity of 1 is the combined peak of Al K α and K β at 1.49 keV. The contribution of Al K β to Al K α is 0.013, which is small, but still changes the peak shape. The FIB stub was initially expected to be Fe, but the spectrum lacks a Fe K α peak at 6.40 keV. There are some signal at 6.40 keV, but it is barely a signal different from the background. The peaks at 1.25 keV and 5.9 keV indicate that the FIB stub is an alloy of mostly Al with some Mg and Mn. The last peak in the spectrum is the Al K α sum peak at 2.97 keV. (**Question for Ton: When stating that some peaks are sum peaks, do you want me to state what other possibilities they could be, or can I just discuss that in the discussion?**) The background is increase rapidly and almost linearly from C K α to Al K α , and drops to 10% after the Al K α peak. (**Question for Ton: How much of the free text do you want me to repeat in the figure text?**)

Figure 2.3 shows the spectra taken from the pure Si wafer sample part. As in the other spectra, this spectrum have a zero peak, a C K α peak at 0.260 keV and a O K α peak at 0.51 keV. In addition, there is an unidentified peak at 0.080 keV. The tallest peak with a relative intensity of 1 is the Si K α peak at 1.73 keV, which have a contribution from the K β peak at 1.84 keV. The relative weight for Si K β to K α is 0.028. As in the Al spectrum, the Si spectra have a sum peak from the tallest peak. The sum peak is at 3.48 keV, which is the sum peak of the Si K α . In all four spectra the background drops significantly after the Si K α peak. The relative drop down is biggest in the 30 kV spectrum.

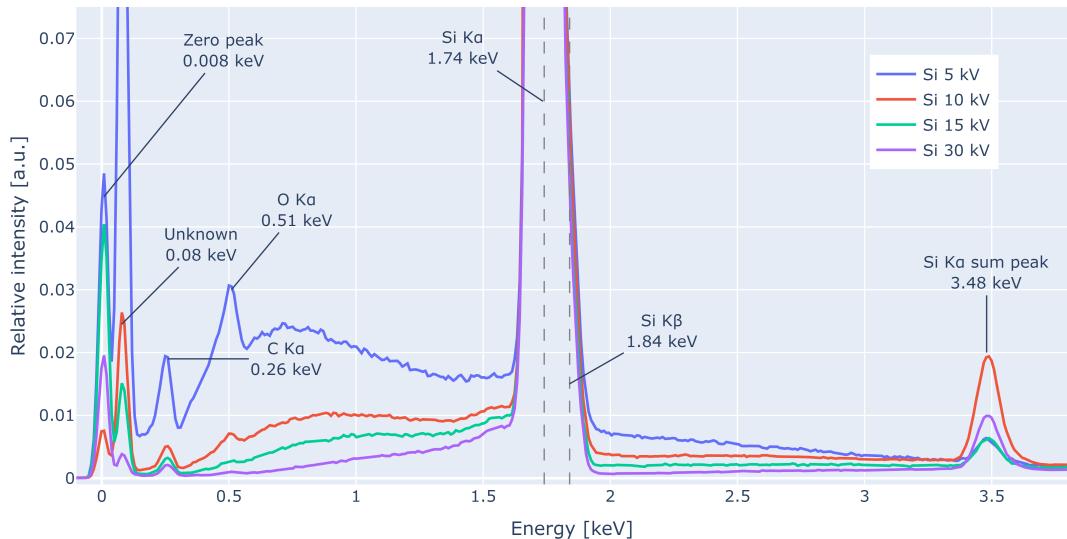


Figure 2.3: The spectra of the pure Si wafer sample part. All four spectra have one large peak at 1.73 keV, which is the Si K α peak with some signal at the K β peak at 1.83 keV. The relative weight for Si K β to K α is 0.028. The zero peak is marked at 0.008 keV. After the zero peak there is another sharp peak at 0.080 keV, which is not identified. The energies annotated are the end of the annotation line, which can deviate a few percent from the actual peak energy.

Figure 2.4 shows the spectra taken from the Cu-tape sample part. These spectra have a zero peak, a C K α , a O K α peak and a small Si K α peak. The tallest peak with a relative intensity of 1 is the C K α peak at 0.26 keV. The Cu K α and Cu K β peaks are only visible at the 30 kV spectrum. The height of the Cu K α peak is 0.017, which means that the C K α peak is more than 55 times taller. None of the spectra have a Cu L α peak, which should have been at 0.93 keV. The 30

kV spectrum have a signal at 2.29 keV, which could be from Mo L α , but no Mo K α signal is visible.

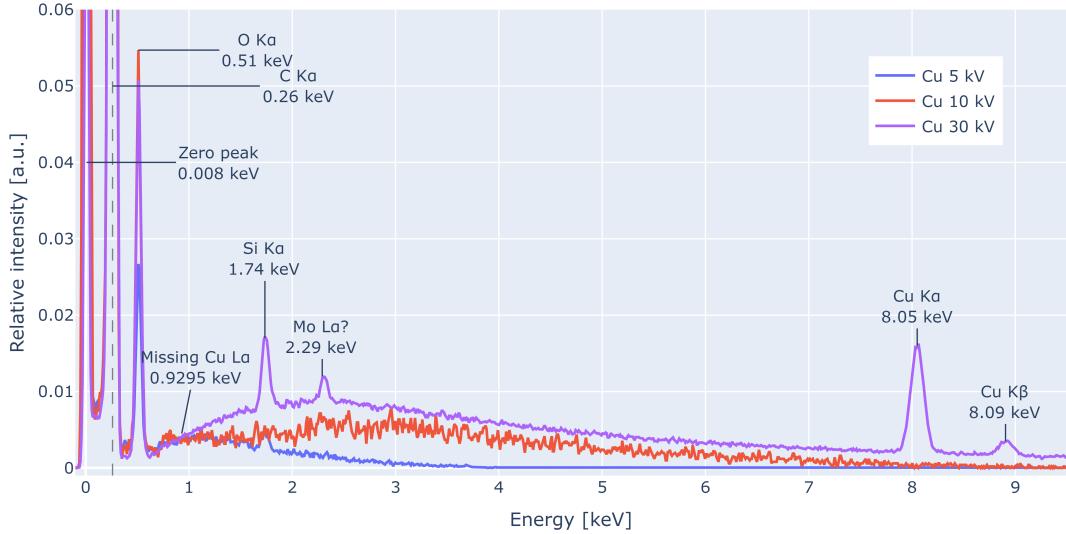


Figure 2.4: The spectra of the Cu sample part. The Cu sample was Cu-tape from the lab, but the Cu K β peak is only barely visible at the 30 kV spectrum. The highest peak in all three spectra is at 0.260 keV, which is the C K α peak, slightly off from the expected 0.277 keV. The plot is limited to 9.5 keV, because there are no peaks above that energy. The Mo L α peak at 2.29 keV is only visible in the 30 kV spectrum. The Mo L α is marked with a question mark, because there are no Mo K α signal.

[Figure 2.5](#) shows the spectra taken from the Mo-disk sample part. The four spectra have a zero peak, a C K α , a O K α peak. The Si K α give a very small signal, and is visible as the tiny peak before Mo Ll. The background of the 5 kV spectrum is very high, but that is due do its lower signal on the Mo L α , making the 0.18 keV peak the tallest and thus scaling up the background. The tallest peak in the 5 kV spectrum is the 0.18 keV peak, which is also visible in the other three spectra. The energy of 0.18 keV only match with the B K α line at 0.1833 keV. The tallest peak in the 10, 15 and 30 kV spectra is the Mo L α peak at 2.29 keV, which is contributed by the Mo L β_1 peak at 2.39 keV. The HyperSpy database have the energy and weight of two other Mo peaks visible in the spectrum, which are the Mo Ll and Mo L γ_3 peaks at 2.01 keV and 2.83 keV. The X-ray Booklet does not include these lines, but include some other lines which are not visible in these spectra. Only the 30 kV spectra have a signal at the Mo K α and Mo K β peaks at 17.46 and 19.62 keV. The peak at 4.6 keV is the sum peak of Mo L α and Mo L β_1 . As in the other spectra, the background drops significantly after the tallest peak.

[Figure 2.6](#) shows the spectra taken from the GaAs sample part, and [Figure 2.7](#) shows a zoomed in part of the spectra to enlarged the background and the sum peaks. All the spectra have a zero peak, a C K α peak, and an O K α peak. The Ga Ll peak at 0.96 keV is visible in all four spectra. The Si K α peak is small in all the spectra, but strongest in the 30 kV spectrum and weakest in the 5 kV spectrum. The tallest peak in all four spectra is the Ga L α peak at 1.1 keV. The As L α peak has a decreasing relative intensity from 5 to 30 kV. For the first 2 keV the background is relatively highest in the 5

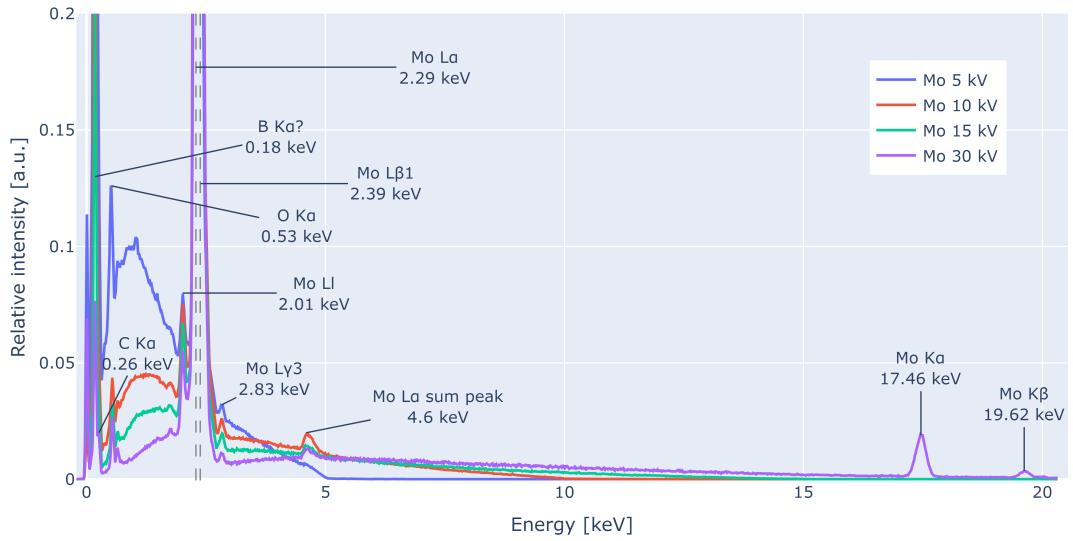


Figure 2.5: The spectra of the Mo sample part. The Mo $K\alpha$ peak at 17.47 keV is the 30 kV spectrum, with a high noise level. The high double peak is Mo $L\alpha$ at 2.29 keV and Mo $L\beta$ at 2.39 keV. The Mo $L\beta_1$ peak at 2.49 keV has the weight 0.327. In the Mo spectra the Si $K\alpha$ line is just barely visible as a very small peak, and is not annotated. The peak at 2.01 keV is the Mo Li and the peak at 2.83 keV is the Mo Ly_3 . Both are with name and energy from the HyperSpy database. Mo Li and Mo Ly_3 have weights at 0.041 and 0.011, respectively.

kV spectrum and decreases with increasing voltage. After 2 keV this trend reverses, and the background is highest in the 30 kV spectrum. The K peaks are visible in the 15 and 30 kV spectra. The sum peaks of the L peaks are visible in the 5, 10 and 15 kV spectra, but not in the 30 kV spectrum. The sum peaks of the K peaks are only visible in the 30 kV spectrum.

Figure 2.8 shows the spectra of the nanowire sample part. This is the sample part with the most peaks, and contains signal from C, O, Ni, Cu, Ga, As, Si, Mo and Sb. One of the peaks is at 0.389 keV, which could be N $K\alpha$ peak, but it could also be other elements. The Ni signal is both from the $L\alpha$ at 0.85 keV and the $K\alpha$ at 7.49 keV. The Sb signal is from the $L\alpha$ at 3.6 keV, $L\beta_1$ at 3.8 keV, and Sb $L\beta_1$ at 4.1 keV. The Mo signal is both from the $L\alpha$ at 2.29 keV and the $K\alpha$ at 17.47 keV, but the K peak is very weak and thus not included in the plot. The Cu signal is both from the $L\alpha$ at 0.92 keV and the $K\alpha$ at 8.05 keV. The tallest peak in all four spectra is the Ga $L\alpha$ peak at 1.1 keV. In the 5 kV spectrum the C $K\alpha$ peak at 0.26 keV is equally high as the Ga $L\alpha$ peak. The As and Ga signals and their ratios are very similar to the GaAs spectrum signal, but the sum peaks are not visible in the NW spectra. The Ga and As $L\alpha$ sum peaks could have been visible, but the sum peak signal coincides with the Mo $L\alpha$ peak at 2.2 keV. The 10, 15 and 30 kV signal have an unidentified peak at 2.00 keV. The unidentified peak could be a sum peak or another artifact. (**Question for Tom:** Any ideas for the 2.00 keV peak? It does not match well with any peaks that I've found.)

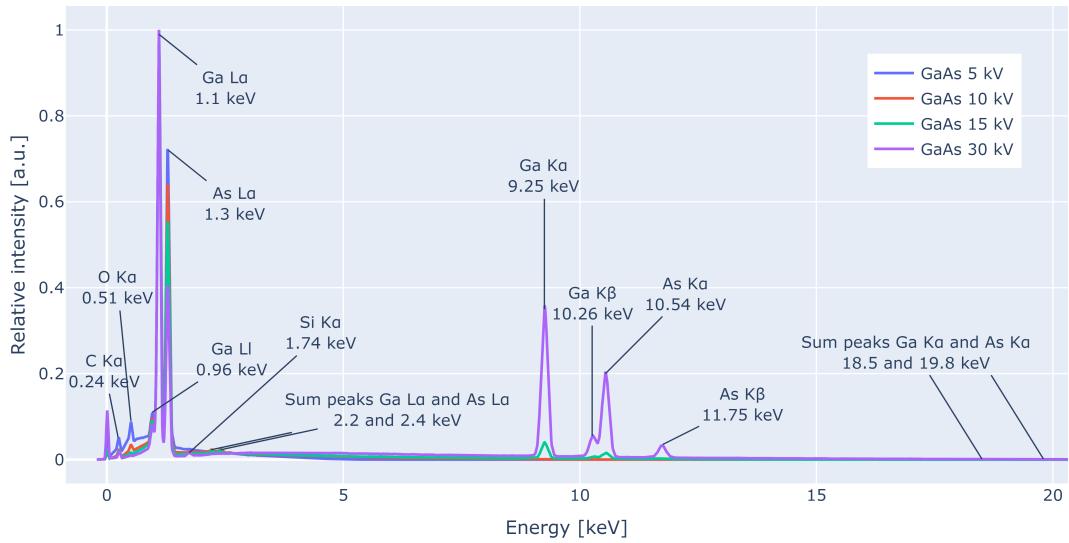


Figure 2.6: The spectra of the GaAs wafer sample part. This is bulk GaAs, where the ratio of Ga to As is 1:1. Both the K-peaks and the L-peaks of Ga and As are visible. There is a peak at 0.51 keV, which is the O K α peak. There is a peak at 0.24 keV, which is the C K α peak.

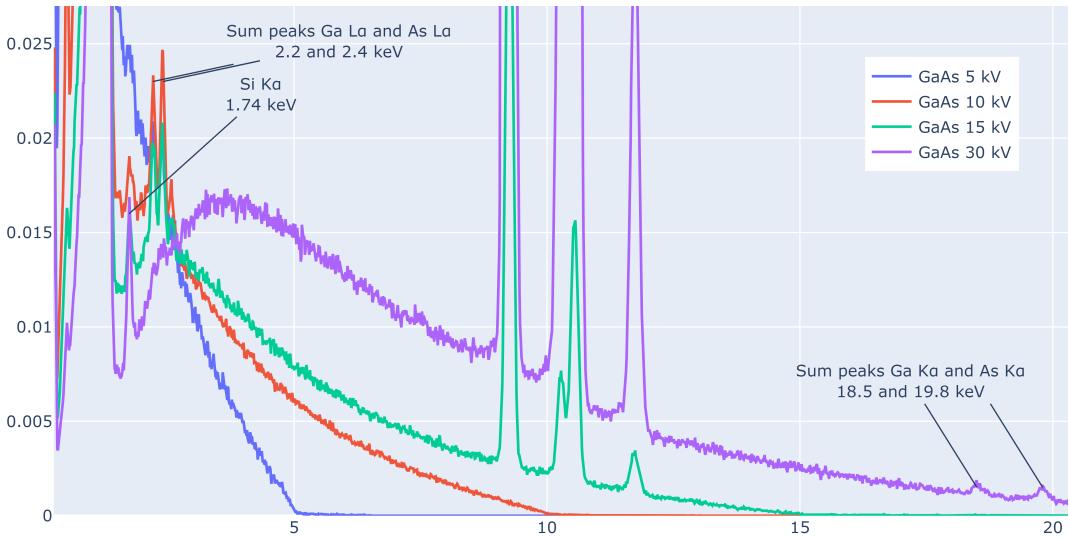


Figure 2.7: A zoomed in view of the GaAs sample part. The background is easier to see, and the sum peaks are visible. The sum peaks are the sum of the K α and peaks, and the L α peaks.

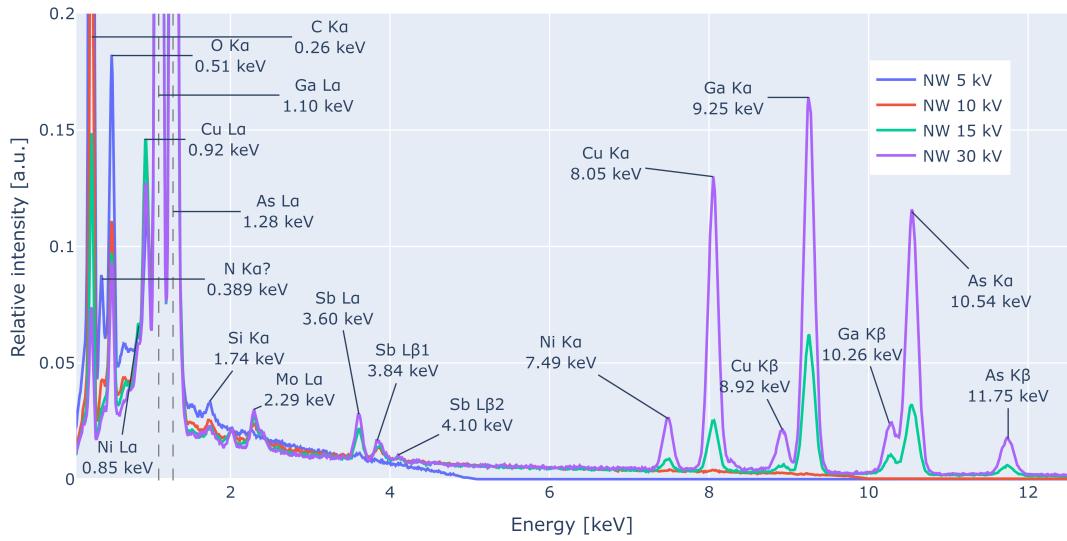


Figure 2.8: The spectra of the nanowire sample part. This spectrum has the most peaks, and contains C, O, Ni, Cu, Ga, As, Si, Mo and Sb, and maybe N. The line at 0.389 keV could be N K α peak, but it could also be other elements. The peak at 0.92 keV which is labeled as Cu K α is also getting a contribution from the Ga L β peak, at 0.95 keV. The 30 kV spectrum had a small signal from the Mo K α peak at 17.47 keV, but that signal was weak. The 10, 15 and 30 kV signal have an unidentified peak at 2.00 keV.

2.1.2 General results from the spectra

All the spectra have peaks with high peak-to-background ratio. In all the spectra, the highest peak is below 5 keV. The GaAs, NW and Mo spectra show clearly that the peaks broaden with higher E, since they have peaks at low and middle to high energy. The width of the peaks are quantified with FWHM in the quantitative section below. When doing the qualitative analysis, it became clear that the FIB stub was not made of Fe as expected, but rather of Al with a peak at 1.48 keV. All the 5 kV spectra decrease more or less linearly from 1 to 5 keV. Even though the Cu spectra at 30 kV has the Cu K α peak, the Cu L α peak is completely missing. ([Brynjart: Add transition sentence.](#))

Some peaks in the spectra are overlapping, which shifts the shape of the peaks. An example of this is the As K α peak and the Ga K β peak in the NW and the GaAs bulk wafer spectra. These peaks are overlapping, but also far enough apart that the peaks are still distinguishable. Another example of overlapping peaks is the Mo L α peak and the Mo L β_1 , which are overlapping so much that they are hard to distinguish. Since they are harder to distinguish, the peak fitting makes one Gaussian for the two peaks, which is off on both peak centers. Overlapping peaks makes counting the signal from specific peaks harder. ([Brynjart: Figure of double peaks?](#))

The signal from the background is another factor which makes counting more difficult. In general, background in the acquired spectra is low, but with different shapes. In the GaAs, Si and Mo the height of the background decrease with higher acceleration voltage. The background in the Cu spectra increase with higher acceleration voltage. The most

similar background signal over different voltages are in the NW spectra. The values of the background radiation is in general very low and almost flat above the highest peak in the spectra. The Al spectrum background is high before the high Al $K\alpha$ peak, and much lower after the high peak. Both the value and the shape of the background is different before and after the peak. The same behavior is clearly true in the Si spectra. The peak where the background change is the highest peak, Si $K\alpha$. The background values in Si 30 kV are 10 times higher before than after the highest peak. The background shape in Si 30 kV is almost linear from 0.6 to 1.6 keV, drops to 10% height from 1.6 to 1.9 keV, and then follows the expected background shape from 1.9 keV. The expected background shape is illustrated in ([Brynjart: Make a drawing of the background.](#)) All the other spectra show the same behavior with their highest peak and the peaks effect on the background as the Al and Si spectra. In general, the background signals are low, but their different shapes and heights makes it harder to fit the peaks of the characteristic X-ray lines. ([Brynjart: Figure of background? And figure of fit of background before and after a tall peak?](#)) ([Question for Ton: The last sentence is meant to be a transition/finishing sentence, but might be too much discussion.](#))

In addition to the characteristic peaks and the background, there are also artifacts and strays in the spectra. All the spectra have signal from the $K\alpha$ line of C and O, which could be contamination and oxide layers. The C and O signal is higher at lower acceleration voltage. All the spectra have a Si peak at 1.74 keV. For all but the Si spectra, the Si $K\alpha$ peak at 1.74 keV is a stray from outside the beam or the Si escape peak from the detector. Some spectra have additional signals from elements outside the area of the main beam, like the Mo, Sb and Cu peak in the NW spectra. The Sb peaks in the NW spectra are at 3.60, 3.85, 4.10 and 4.35 keV, being the $L\alpha_1$, $L\beta_1$, $L\beta_2$ and $L\gamma_1$ peaks. All four Mo spectra have a peak at 0.175 keV, which match best with B $K\alpha$ at 0.183 keV. ([Brynjart: Figure of strays?](#)) ([Brynjart: Finishing sentence for the general observations.](#))

2.1.3 Calibration

Different calibrations were explored. The initial calibration is the one from AZtec, and is the one used in the spectra in [Figure 2.1](#). This calibration has a left shift for the L-peaks and a right shift for the K-peaks. The second type of calibration is the one given by the model fit in HyperSpy. The third type is from the self-made model fit, using the distance between two high intensity and far apart peaks to calibrate the energy scale. The third type is both calculated with the Ga $L\alpha$ and As $K\alpha$ peaks in the GaAs 30 kV spectrum, and with Mo $L\alpha$ and Mo $K\alpha$ peaks in the Mo 30 kV spectrum.

Values for the four calibrations are given in [Table 2.1](#). The deviations are a few percent, and the accuracy of the different calibrations give on specific peaks are given in [Table 2.2](#). Here accuracy is the difference between the theoretical peak position and the peak center in the spectrum, given in percent. For almost all the peaks, the deviation is greatest for the AZtec calibration. One exception is the C $K\alpha$ peak, which deviates a lot less for the AZtec calibration. The difference between the HyperSpy calibration and the self-made calibration on the GaAs and Mo spectra are small. In the following qualitative section, the effect of the different calibrations on the spectra are explored.

Table 2.1: Different calibration values. The dispersion is calculated with ???. The offset is calculated with ???. The own calibration was done on Ga L α and As K α from the 30 kV measurement on the GaAs wafer. The HyperSpy calibration was done by making a model and fitting it to the data on the 30 kV GaAs spectrum.

| Calibration method | Dispersion, [keV/channel] | Zero offset [channels] |
|--|------------------------------|---------------------------|
| AZtec | 0.010000 | 20.000 |
| HyperSpy | 0.010028 | 21.079 |
| Calibration on Ga L α and As K α | 0.010030 | 21.127 |
| Calibration on Mo L α and Mo K α | 0.010040 | 21.076 |

Table 2.2: Peak accuracy of the different calibration methods on 30 kV spectra. The other acceleration voltages gave similar results. The accuracy here is the deviation from the theoretical peak to the measured peak, given in percent. The measured peak is the fitted center of the peak. The C K α is fitted well, but deviates much more than all the other peaks. The self-made calibration was done on two spectra: GaAs and Mo. The HyperSpy calibration was done on the GaAs spectrum.

| Peak | Theoretical [keV] | AZ dev. [%] | HS dev. [%] | Ga L α & As K α [%] | Mo L α & Mo K α [%] |
|---------------|-------------------|-------------|-------------|-----------------------------------|-----------------------------------|
| As L α | 1.2819 | 1.000 | 0.439 | 0.422 | 0.560 |
| As K α | 10.5436 | -0.202 | -0.025 | -0.010 | 0.093 |
| Ga L α | 1.098 | 1.044 | 0.342 | 0.318 | 0.463 |
| Ga K α | 9.2517 | -0.153 | 0.009 | 0.024 | 0.128 |
| Cu L α | 0.9295 | 1.767 | 0.888 | 0.857 | 1.010 |
| Cu K α | 8.0478 | -0.114 | 0.031 | 0.045 | 0.150 |
| Mo K α | 17.4793 | -0.325 | -0.108 | -0.090 | 0.011 |
| Mo L α | 2.2932 | 1.047 | 0.859 | 0.858 | 0.979 |
| Si K α | 1.7397 | 0.167 | -0.175 | -0.182 | -0.055 |
| Al K α | 1.4865 | 0.200 | -0.247 | -0.259 | -0.127 |
| Cu K α | 8.0478 | -0.116 | 0.029 | 0.043 | 0.148 |
| C K α | 0.2774 | -2.955 | -6.583 | -6.738 | -6.464 |

2.2 Quantitative results

To do the quantitative analysis, HyperSpy needs k-factors. The k-factors for Ga and As are given in [Table 2.3](#). These k-factors are from the GaAs bulk wafer, and HyperSpy have estimated them theoretically. (**Question for Ton: Shall I list the other k-factors for the other sample areas? I do not think I will use them, since I've only quantified the GaAs bulk wafer. But the other k-factors are results too. Eventually including NW data too, but I do not know that ratio.**)

Table 2.3: K-factors for Ga and As, extracted from AZtec. All the k-factors are theoretically estimated. AZtec provides either the k-factor for the L α or the K α line, and selects automatically based on the energy of the incident electrons.

| V _{acc} [kV] | Line | Ga k-factor | As k-factor | Ga / As |
|-----------------------|------------|-------------|-------------|---------|
| 5 | L α | 1.21 | 1.086 | 0.898 |
| 10 | L α | 1.31 | 1.223 | 0.934 |
| 15 | L α | 1.331 | 1.259 | 0.946 |
| 30 | K α | 4.191 | 3.268 | 0.780 |

The initial quantification was done on the data from the GaAs wafer in AZtec. The results are presented in [Table 2.4](#). The wafer is a 1:1 alloy of gallium and arsenic, so the atomic percent of Ga and As should be 50% and 50% respectively.

Table 2.4: Initial quantification of the GaAs wafer. The ratio in the wafer is 1:1, so the correct ratio is 50% and 50%, because the results are in atomic percent. (**Brynjar: The idea was to use both AZtec and HyperSpy, but it did not make sense to use HyperSpy with the discussion I have now. Thus it is only AZtec.**)

| V _{acc} | Ga | As |
|------------------|------|------|
| 5 kV | 50 % | 50 % |
| 10 kV | 50 % | 50 % |
| 15 kV | 50 % | 50 % |
| 30 kV | 50 % | 50 % |

To better understand the ratios between Ga and As, the areas under the peaks in the spectra were counted. Table [Table 2.5](#) gives the ratios between the areas under the peaks for 5, 10, 15 and 30 kV. The table compares L α peaks, K α peaks, K β peaks and the sum of the peaks. The table also lists the FWHM of the peaks.

One of the adjustments explored was the affect of the calibration on the quantification. Using different calibrations in [Table 2.1](#) gave different quantification results when using CL in HyperSpy. The results are presented in [Table 2.6](#). The same method was used on all spectra, but the quantification on 10 and 15 kV are obviously wrong.

Table 2.5: Ratios of Ga and As on the GaAs wafer. The spectrum was calibrated with GaAs 30 kV, but different calibrations did not change the ratios significantly. K β at 15 kV was too low to be detected and is therefore not included in the table. The ratio column is Ga sum divided by As sum. The ratio * k column is the ratio multiplied by the respective k-factor in Table 2.3. The empty cells are for peaks where there was no k-factor or for the sum of the peaks, which have no center and no FWHM.

| Peak | Peak*k | Ratio | Ga value [keV] | As value [keV] | Ga FWHM [eV] | As FWHM [eV] | Ga sum [eV] | As sum |
|-----------------------------------|--------|-------|-------------------|-------------------|-----------------|-----------------|----------------|--------|
| 5 kV | | | | | | | | |
| L α | 1.151 | 1.282 | 1.101 | 1.288 | 74.010 | 80.921 | 75.462 | 58.844 |
| 10 kV | | | | | | | | |
| L α | 1.349 | 1.444 | 1.100 | 1.287 | 73.841 | 80.827 | 76.222 | 52.770 |
| 15 kV | | | | | | | | |
| L α | 1.579 | 1.669 | 1.100 | 1.287 | 73.830 | 81.137 | 77.001 | 46.146 |
| K α | | 2.445 | 9.253 | 10.536 | 155.080 | 181.951 | 6.013 | 2.459 |
| L α +K α | | 1.708 | | | | 83.014 | 48.605 | |
| 30 kV | | | | | | | | |
| L α | | 2.279 | 1.098 | 1.287 | 72.309 | 80.849 | 76.465 | 33.546 |
| K α | 1.301 | 1.678 | 9.253 | 10.542 | 157.799 | 168.238 | 58.718 | 34.994 |
| K β | | 1.603 | 10.276 | 11.736 | 171.804 | 185.034 | 8.821 | 5.503 |
| L α +K α | | 1.972 | | | | 135.184 | 68.540 | |
| L α +K α +K β | | 1.945 | | | | 144.004 | 74.042 | |

Table 2.6: Quantification with different calibration methods. The quantification is done by in HyperSpy with Cliff-Lorimer method on the 30 kV GaAs spectrum. AZ is the AZtec calibration. HS is the HyperSpy calibration. GaAs is the calibration using Ga L α and As K α peak centers. The accuracy of the quantification is the deviation from 50%, because the sampled area is 1:1 GaAs wafer.

| Acceleration voltage | Element | Line | AZ | HS | GaAs |
|----------------------|---------|------|--------|--------|--------|
| 5 | As | L | 44.81 | 44.29 | 44.19 |
| 5 | Ga | L | 55.19 | 55.71 | 55.81 |
| 10 | As | L | 100.00 | 100.00 | 100.00 |
| 10 | Ga | L | 0.00 | 0.00 | 0.00 |
| 15 | As | L | 5.23 | 4.39 | 5.87 |
| 15 | Ga | L | 94.77 | 95.61 | 94.13 |
| 30 | As | K | 56.25 | 57.14 | 59.02 |
| 30 | Ga | K | 43.75 | 42.86 | 40.98 |

Chapter 3

Discussion

The discussion is presented in this chapter. Producing code from scratch is a time-consuming process, but it is also a learning process. While developing the code, the author learned a lot about EDS analysis and new ideas emerged. The sections below follow the structure of the sub-problems of the main problem statement formulated in [Chapter 1](#).

3.1 Analysis steps in HyperSpy

[Sub-problem 1](#) was to do qualitative analysis of the GaAs wafer in AZtec and HyperSpy. The analysis was done as an out-of-the-box analysis, i.e. just following the steps in the documentation. AZtec has a GUI for analysis, but it is not possible to see what is done with the data at the different steps. HyperSpy have documentation online, and the following subsections explain how the analysis was done in HyperSpy.

Each subsection starts with some code lines, followed by an explanation of what the lines do. All variables inside crocodile need to be set by the user, e.g. `<element_list>` would be set to `['Ga', 'As']` for the GaAs wafer. An example notebook with quantification of the GaAs wafer is attached in APPENDIX. ([Brynjart: Make a notebook with GaAs quantification in HyperSpy, with the data somehow.](#))

3.1.1 Loading the data and specifying the elements

```
s = hs.load(<filepath>, signal="EDS_TEM")  
s.set_elements(<element_list>)
```

The first step in the analysis is to load the data as a HyperSpy `signal` type, and specifying the signal as TEM. The `signal` type is a class in HyperSpy that contains the data and the metadata, and it has methods for analysis. The `signal` type must be specified as TEM, because the `signal` type for SEM is very limited and does not have a method

for quantification. When using .emsa files from AZtec, as is done in this project, the metadata contains some relevant and some irrelevant information. The information relevant later in this project is: acceleration voltage, dispersion, zero offset, energy resolution Mn $K\alpha$. After loading, it is possible to plot the data with `s.plot()`.

Already at this point there is a big problem with using HyperSpy on the acquired data: the analysis methods are not implemented for SEM but for TEM. As explained by Skomedal in her master's thesis [3], the quantification with the Cliff-Lorimer method of TEM EDS data use approximations which are valid for thin samples. The approximation is that the sample is thin enough to ignore absorption and fluorescence, which is a very crude approximation to use on bulk SEM EDS data. Still, the results show that it is possible to quantify the elements in the GaAs wafer with the Cliff-Lorimer method and get plausible results, but also that the analysis breaks down occasionally. Implementing quantification from the SEM type signal in HyperSpy have been discussed on the repositories GitHub issues page¹. The topic have been discussed in 2020, 2017 and 2015. People have been working on quantification of SEM EDS data, but the work have not been finished yet.

The quantification done with the CL method in HyperSpy on the three different calibrations in Table 2.6 are all within 10% of the expected composition. Carter and Williams [4, p. 612 and 648] state that the accuracy of EDS is at best \pm 3-5%, but that it could be reduced to around \pm 1.7% with very long acquisition times and a careful analysis. Carter and Williams claim that quantitative errors less than \pm 5-10% takes a lot of time and effort, and compositions with error below 5% should be regarded extra carefully and with suspicion. The best result, i.e. the result closest to 50%, from the HyperSpy CL quantification is strangely from the AZtec calibration with a 6.25% error. It is strange that the AZtec calibration is the best, because that is the calibration which misses most on the line accuracy in Table 2.2. This implies that the calibration is not the most important factor for the quantification, at least not when the calibration is around \pm 1%. The different calibrations are discussed further in Section 3.3. In the 30 kV GaAs sample the signal-to-noise ratio is good, the sample time was long and the peaks are well resolved, which make the input data good and allows for better quantification. The CL method is known to work on SEM data, but normally one would also do the ZAF correction, which is described in chapter 19.10.3 of Goldstein [2]. The ZAF correction adjust the signal for the atomic number effect, the absorption in the bulk of the sample and the X-ray fluorescence of new lines with lower energy than the initial characteristic X-ray. Since Ga and As are number 31 and 33 in the periodic table, it could be that the ZAF corrections are small and would not have a big effect on the results. When doing the quantification it was tested to include O and C, but that changed the results from plausible to completely wrong. The quantification analysis also broke down when using the 10 and 15 kV spectra, but the reason for that was not clear. The breaking down of the analysis is discussed further in Section 3.5.

Results in this project show that the CL method in HyperSpy for TEM EDS data sometimes yield plausible results on SEM EDS data and sometimes break down, thus the method should be used with caution. The goal of this project was not to do or to implement cutting edge quantification of EDS data from a SEM sample, but rather to understand

¹<https://github.com/hyperSpy/hyperSpy/issues/2332>

the analysis steps and what factors that could affect the results. Thus, using the quantification method from the TEM signal type was regarded as good enough for this project.

(Question for Ton: I am happy with this first part of the discussion (08.12.22). How would you rate it?)

3.1.2 Removing the background linearly

```
bw = s1.estimate_background_windows(windows_width=<number>)

iw = s1.estimate_integration_windows(windows_width=<number>)
```

The next step is to remove the background, which with the above code is done by a linear fit. The background can be removed through model fitting, which is covered in [Section 3.1.4](#). The variable `windows_width` sets how wide the windows are for the background and integration, measured in FWHMs. A good starting value for `windows_width` is 2, but it should be tested by the user with a plot to see if the background will be removed correctly. The estimated windows can be plotted with:

```
s.plot(xray_lines=True, background_windows=bw, integration_windows=iw)
```

Plotting has proven to be a very useful tool for understanding what the different steps in the analysis do. Unfortunately, the implemented plotting in HyperSpy is limited. For example, when plotting a modelled spectrum with Ga and As, the model contains all the peaks in the HyperSpy library and all the peaks are plotted, but it is not possible to see what the different peaks are. In other words, the plot of a model can show all the independent components, but adding a legend is not an option. Another limitation of the plotting is the interactivity when saving the plot. Matplotlib requires the code to run in the background for it to be interactive, while Plotly can save the plot as an interactive HTML file for later inspection. One of the advantages with the plotting in HyperSpy is how accesable it is to plot in the different steps, which the user can use to understand and verify the analysis. The HyperSpy plotting was probably not made with the intention of being used for publication, but rather for quick and easy plotting of the data. If that is the case, the plotting serves its purpose well and has proven to be very useful for this project.

One of the more helpful insights from plotting in this project was that removing the background is not trivial. This is especially true for background signals modelled as a polynomial. Linear background removal works well for some peaks, for example the K-lines of Ga and As, but not for the L-lines. The background around the K-lines are more linear, because of the exponential decay of the background. Around the L-lines the background is affected by the absorption of the Bremsstrahlung lower than 2 keV and the observation of higher background levels below the highest peak in the spectrum. The background removal is discussed further in [Section 3.4](#).

3.1.3 Quantification after linear background removal

```
s_i = s.get_lines_intensity(background_windows=bw, integration_windows=iw)
```

```

k_factors = [<k-factor 1>, <k-factor 2>]

quant = s.quantification(s_i, method='CL', factors=k_factors)

print(f'E1: {quant[0].data[0]:.2f} \%, E1: {quant[1].data[0]:.2f} \%')

```

The quantification is done with the four lines of code above, where the last one prints the results. The first line gets the intensity of the peak corresponding to the lines of the specified element. HyperSpy selects automatically which lines to use for quantification. To see which lines are used, the `s_i` variable can be printed. The second line sets the k-factors. The k-factors in this project have been the one from AZtec, which are theoretically estimated. The third line does the quantification, where the method is specified. The method is the Cliff-Lorimer method, described in detail in Mari Skomedal's master thesis [3, Sec. 2.2.3]. HyperSpy has a method for quantification with the zeta factor method. The zeta method requires the value for the beam current, which was not measured in this project.²

The k-factors are essential for the Cliff-Lorimer quantification, and the k-factors listed in Table 2.3 give some insight into how AZtec calculates the k-factors. From the results it is clear that AZtec at least adjust the k-factor for element and for voltage. It could be that factors like the beam current and time live could affect the k-factor, but this was not investigated in this project. The L-line k-factors are start at 1.1 for 5 kV and increase to 1.2 for 15 kV. The As k-factor for the L-line at 5, 10 ad 15 kV is 11%, 7% and 6% greater than for Ga. The K-line k-factor are 3.3 for Ga and 4.2 for As, implying either a high sensitivity for higher voltages or two different models for calculating the k-factors. Another possibility is that the K-lines and L-lines have a separate model for calculating the k-factors. This can be tested by extracting the k-factors for the K-lines from the 15 kV spectrum, as the 15 kV spectrum do have peaks for both the K-lines and L-lines. The k-factors for Ga are in all four cases higher than for As, which makes sense, as all the Ga lines are higher and have more counts than the As lines. Quantification of the ratio between the Ga and As peaks are done in Table 2.5, where it is clear that the k-factor push the ratio towards 1. The peak ratio times the k-factor for 5 kV is 1.1, and the next closest to 1 is 1.3 for 30 kV. This result imply that the theoretically calculated k-factors could be better for high and low voltages.

3.1.4 Removing the background with model fitting

Another way to remove the background is to fit a model to the data. This step would be done right after loading the data. If the raw data contains a zero peak, as is the case for most Oxford instrument EDS detectors, the zero peak needs to be removed before fitting the model. The zero peak is removed by skipping the first n channels, where n=30 works well with the data from the GaAs wafer. The model fitting is done with the following code:

```

s = s.isig[<zero_peak>:]

m = s.create_model(auto_background=False)

m.add_polynomial_background(order=12)

```

²Results from the zeta method can be converted to the cross section method, see the page "EDS Quantification" in the HyperSpy documentation.

```
m.add_family_lines(<list_of_element_lines>

m.plot()
```

The lines above removes the zero peak, create a model from the `signals`, adds a 12th order polynomial, add the lines of the elements in the `signal`, and plot the model. This model is not fitted, it is just a generated spectrum with the lines of the elements. Eventually, the method `create_model()` can take the boolean argument `auto_add_lines=True`, which will automatically detect the elements in the sample. The model consists of a number of components, which can be accessed with `m.components`. The components are all the Gaussian peaks in the spectrum, in addition to the background as a 12th order polynomial. The order of the polynomial can be changed, but it should be tested by the user to see if it is a good fit. Further, the model must be fitted.

```
m.fit()

m.plot()
```

The first line fits the model to the data to the components and the second line plots the model. HyperSpy have an own option for fitting only the background. Since the background is one of the components in `m`, it is fitted with the code line above.

3.1.5 Quantification after model fitting

```
m_i = m.get_lines_intensity()

k_factors = [<k-factor 1>, <k-factor 2>]

quant = s.quantification(s_i, method='CL', factors=k_factors)

print(f'E1: {quant[0].data[0]:.2f} \%, E1: {quant[1].data[0]:.2f} \%)
```

The quantification after model fitting is done in the same way as in [Section 3.1.3](#), but with intensity from the model instead of the signal. When modelling GaAs, the model can add the intensity from both K-lines and L-lines. Since AZtec only gives the k-factors for either the K-lines or the L-lines, the user must remove the lines without k-factors before quantification.

3.1.6 Calibrating the spectrum with the HyperSpy model

```
m.calibrate_energy_axis(calibrate='scale')

m.calibrate_energy_axis(calibrate='offset')
```

The two lines above calibrates the spectrum with the HyperSpy model and updates the dispersion and zero offset. The metadata in the `signals` is updated with the new calibration. Thus, doing the previous step with quantification after model fitting can give a more correct quantification.

3.2 Peak and background modelling

The next sub-problem was to find out how the peaks and the background are modelled in a way that is easy to understand. The model was built without HyperSpy, with the idea of making every step easier to understand. The model was used to be able to remove the background and be able to calibrate the spectrum. The model was compared to the HyperSpy model. The model could be used to quantify the elements in the sample, but this was not done in this project. ([Brynjar: Do I want to do this?](#))

The first step in creating a model is to identify the peaks. The peaks are assumed to be Gaussian curves. The initial way of identifying peaks was that the user manually identified the peaks. Later the peaks were identified with the function `find_peaks()` from the `scipy.signal` package. Different peak prominence were tested, and the peak prominence of 0.01 gave the best results.

The second step is to make a Gaussian in each peak and one polynomial for the background. To do the fitting, the components need an initial guess. The background needs a coefficient for each order of the polynomial. Each Gaussian need to have a mean, a standard deviation, and a height. The mean is the peak position. The standard deviation is the width of the peak, where $\text{FWHM} = \text{std} * 2 * \sqrt{2 * \ln 2}$ ³. The height is the amplitude of the peak. The easiest way to get the initial guesses for the Gaussians is to normalize the data and set all three parameters to 1. In the normalization the highest peak was set to 1, and the rest of the peaks were scaled accordingly. The best way to get the initial guesses for the background is to clip out the peaks with linear interpolation and fit a polynomial. The initial guesses for the background are then the coefficients of the polynomial. With the initial guesses, the whole model is ready to be fitted.

The third step is to fit the model to the data. Using the `curve_fit()` function from the `scipy.optimize` package, the model is fitted to the data. The function `curve_fit()` uses the Levenberg–Marquardt algorithm to fit the model to the data. The function `curve_fit()` returns the optimal parameters for the model. Fitting both the Gaussians and the background at the same time makes the fitting more stable. One of the first iterations, where the user manually inputted the peaks, the fitting tended to partially fail. The issue was that the fitting only was done on the peaks. To minimize the error in the fitting, one of the Gaussian curves with a low amplitude was moved and got a huge standard deviation, which compensated the background. This was fixed by fitting both the Gaussians and the background at the same time. Doing this made the fitting both better, and it failed less often.

([Brynjar: Issue: fitting e.g. Mo with two clear peaks, but not with enough prominence to be found by the peak finder.](#))

³FWHM defined at: https://en.wikipedia.org/wiki/Full_width_at_half_maximum

3.3 Calibration

The next sub-problem was to calibrate the data with a self produced Python script. With a fitted model of the spectrum, the calibration can be done. Calibration can both be done on raw data with channels on the x-axis and on poorly calibrated data with energy on the x-axis. The dispersion is calculated with [??](#). Table [Table 2.1](#) shows calibration from AZtec, HyperSpy, and the self produced Python script.

3.4 Background models

The next sub-problem was to find out how different background models affect the quantitative analysis done in Hyper-Spy, and how well different order polynomials fit the background. The background models were tested on the spectrum of GaAs, and later also on ([Brynjart: TODO: other spectra. Also make a table here with results](#)). The background was modelled as a polynomial of different orders. To quantify the different background models, the residuals were calculated. The residuals are the difference between the data and the model. ([Brynjart: Use root-mean-square error?](#)) The TABLE XXXX ([Brynjart: make table](#)) shows the residuals for the different order background models. The best orders were visually inspected. A later idea was to model the background as a spline, which is a piece wise polynomial. The spline is a piece wise polynomial with a smooth transition between the pieces. The spline was not tested in this project, but it could be a good alternative to the polynomial background model.

3.5 Analysis failure

The next sub-problem was to find out when the analysis fails, both in AZtec and HyperSpy.

([Question for Ton: Section about normalization too?](#))

3.6 Calibration decision

3.7 Choices in HyperSpy

Bibliography

- [1] C.B. Carter and D.B. Williams. *Transmission Electron Microscopy: Diffraction, Imaging, and Spectrometry*. Springer International Publishing, 2016.
- [2] Joseph I. Goldstein, Dale E. Newbury, Joseph R. Michael, Nicholas W.M. Ritchie, John Henry J. Scott, and David C. Joy. *Scanning Electron Microscopy and X-Ray Microanalysis*. Springer New York, New York, NY, 2018.
- [3] Mari Sofie Skomedal. Improving quantitative EDS of III-V heterostructure semiconductors in low voltage STEM. Master's thesis, Norwegian University of Science and Technology, 2022.