

# Contents

<b>1 Results</b>	<b>2</b>
1.1 Qualitative results . . . . .	2
1.1.1 Peaks and background . . . . .	2
1.1.2 Calibration . . . . .	8
1.2 Quantitative results . . . . .	9
<b>2 Discussion</b>	<b>15</b>
2.1 Introduction . . . . .	15
2.2 Analysis steps in HyperSpy . . . . .	15
2.2.1 Loading the data and specifying the elements . . . . .	15
2.2.2 Removing the background linearly . . . . .	16
2.2.3 Quantification after linear background removal . . . . .	16
2.2.4 Removing the background with model fitting . . . . .	17
2.2.5 Quantification after model fitting . . . . .	17
2.2.6 Calibrating the spectrum with the HyperSpy model . . . . .	18
2.3 Peak and background modelling . . . . .	18
2.4 Calibration . . . . .	19
2.5 Background models . . . . .	19
2.6 Analysis failure . . . . .	19
2.7 Calibration decision . . . . .	20
2.8 Choises in HyperSpy . . . . .	20

# Chapter I

## Results

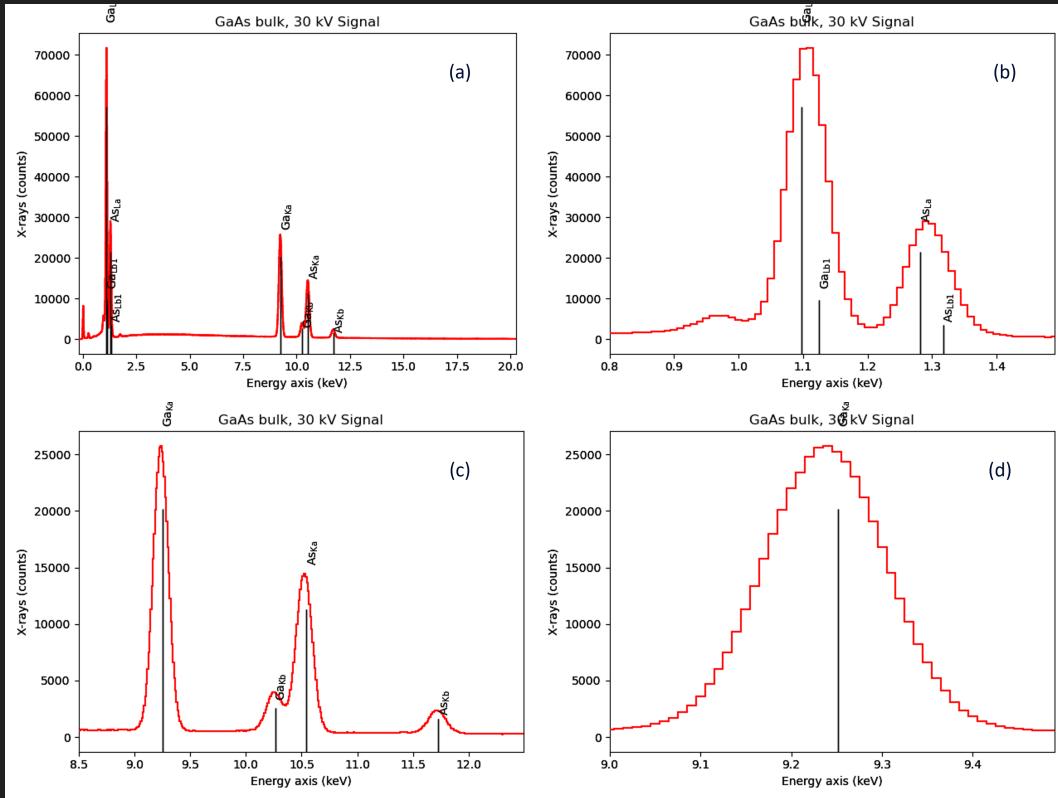
The results are presented in this chapter. First qualitative then quantitative results are presented. All the spectra taken were qualitatively analyzed. Only the GaAs bulk spectra was quantitatively analyzed. The spectrum from the GaAs bulk wafer taken on 30 kV is shown in Figure 1.1. This plot was made 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.

### 1.1 Qualitative results

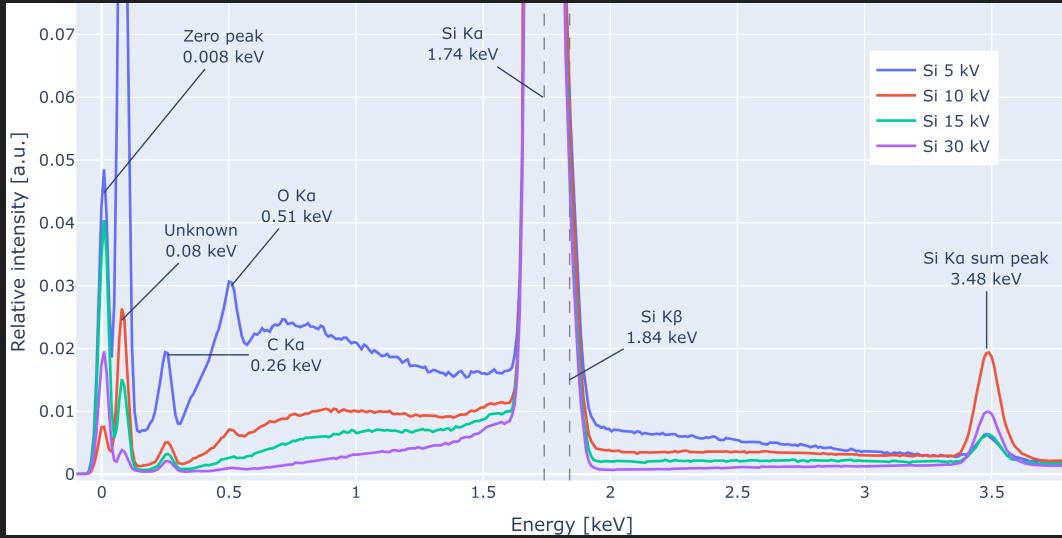
Figure 1.3 to Figure 1.2 shows the spectra for the six different areas of the sample plotted with Plotly. The plots are available as an interactive HTML plots on the GitHub repository. ([Brynjart: Upload the HTML files to the GitHub repository](#)) The calibration used in these spectra is based on the calibration of Ga L $\alpha$  and As K $\alpha$  from the GaAs bulk wafer. The y-axis is normalized to the highest peak value in each spectrum.

#### 1.1.1 Peaks and background

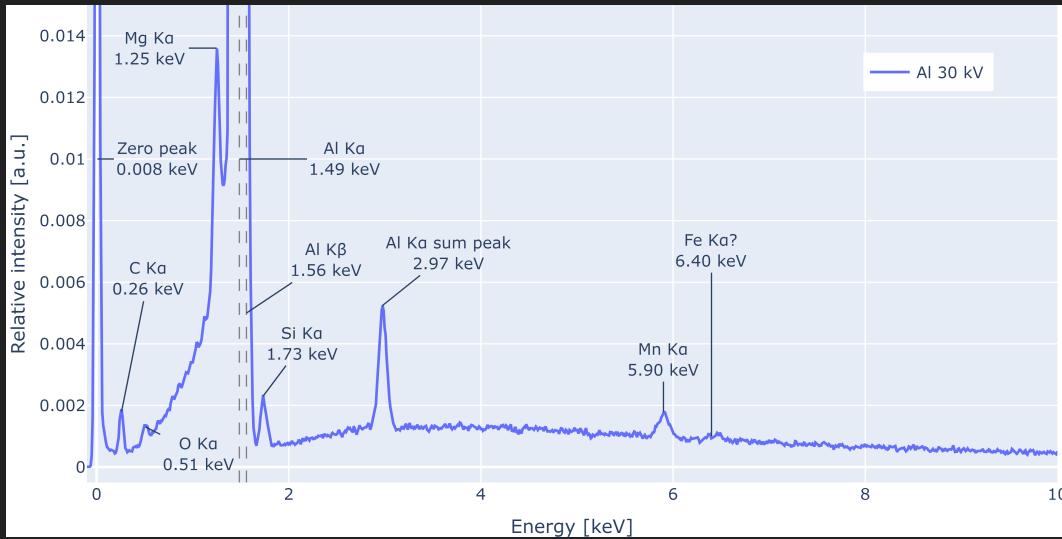
All the spectra have peaks with high peak-to-background ratio. The zero peak from the Oxford detector is visible in all the spectra as the first peak, with a center at 0.00924 keV. The GaAs, NW and Mo spectra show clearly that the peaks broaden with higher E, since they have peaks at low and middle to high energy. In all the spectra, the highest peaks are below 5 keV. 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. The FIB stub have a small peak at 1.25 keV, which could be from Mg K $\alpha$  at 1.253 keV. The Al FIB stub spectrum also have a small peak at 5.90 keV, which could be from Mn K $\alpha$  at 5.899 keV. Another discovery was that the Cu-tape does not give a good Cu signal. The high peak in the Cu-tape spectra at 0.260 keV are from C and not the Cu L $\alpha$  peak. Only the Cu-tape taken at 30 kV has a Cu peak, but it is very small. Even though the



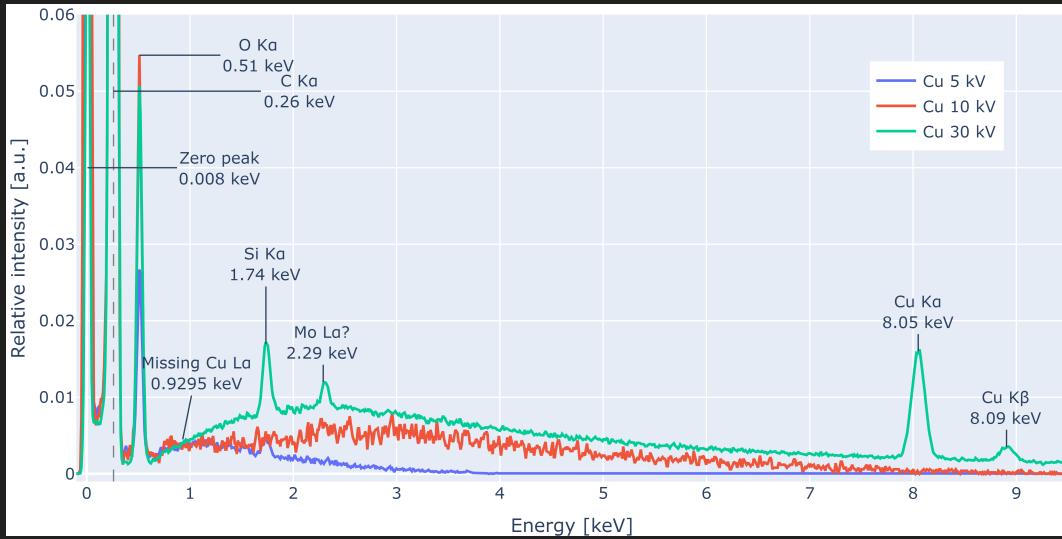
**Figure 1.1:** The GaAs spectrum taken at 30 kV. This plot was made with HyperSpy, which use Matplotlib. The theoretical peak centers are added as lines. (a) is the whole spectrum. (b) is the zoomed in on the L-peaks. (c) is zoomed in on the K-peaks. (d) is zoomed in on the Ga K $\alpha$  peak. This plot has the calibration from AZtec, and it is clear that the line position is deviating from the center of the peaks.



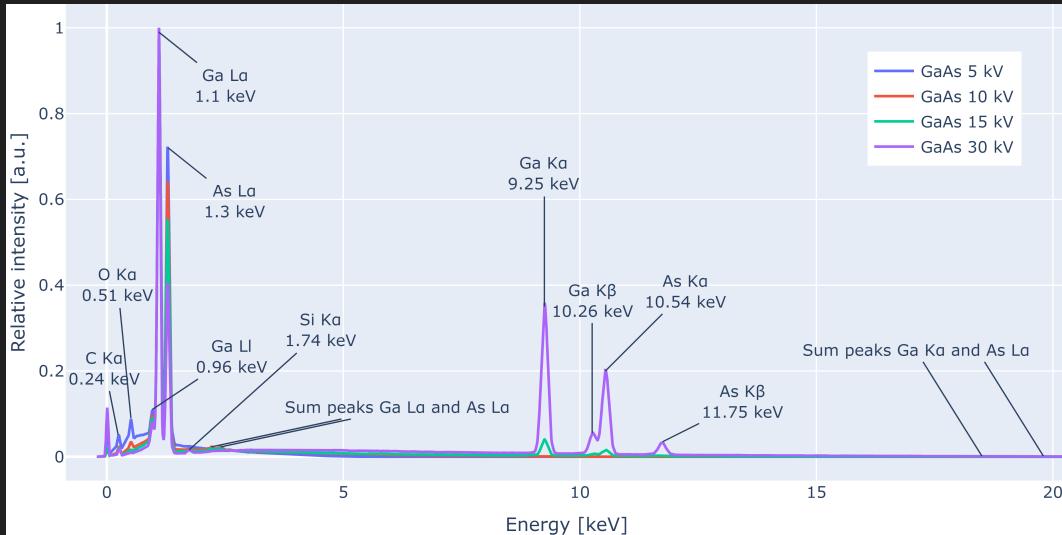
**Figure 1.2:** The spectra of the pure Si wafer sample part. All four spectra have one large peak at 1.73 keV, which is the Si K $\alpha$  peak with some signal at the K $\beta$  peak at 1.83 keV. The relative weight for Si K $\beta$  to K $\alpha$  is 0.028. The zero peak is marked at 0.008 keV. After 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.



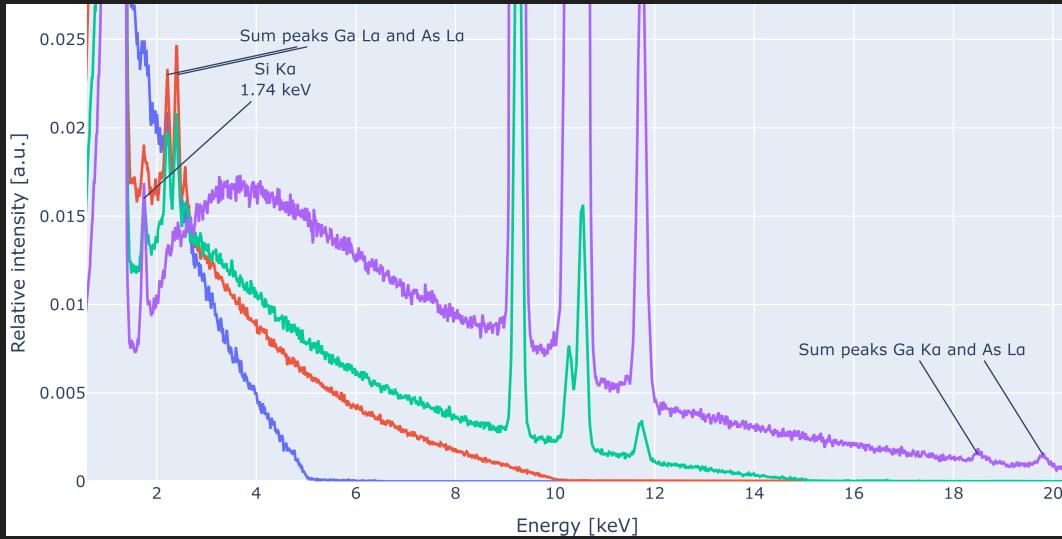
**Figure 1.3:** The spectra of the Al sample part. This was expected to be Fe, thus the label is wrong. The peak at 1.49 keV is the Al K $\alpha$  peak with some signal at the K $\beta$  peak at 1.56 keV. The relative weight for Al K $\beta$  to K $\alpha$  is 0.013 (from HyperSpy). Fe K $\alpha$  at 6.40 keV has a question mark, because the FIB stub was initially expected to be Fe. The signal from Fe K $\alpha$  is barely a signal different than the background.



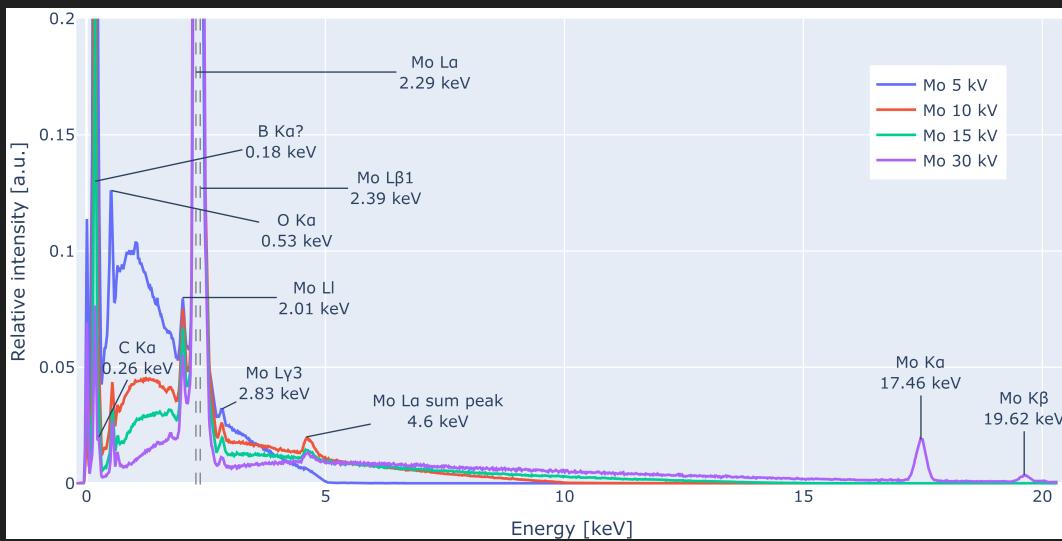
**Figure 1.4:** The spectra of the Cu sample part. The Cu sample was Cu-tape from the lab, but the Cu  $K\beta$  peak is only barely visible at the 30 kV spectrum. The highest peak in all three spectra is at 0.260 keV, which is the C  $K\alpha$  peak, slightly off from the expected 0.277 keV. The plot is limited to 9.5 keV, because there are no peaks above that energy. The Mo  $L\alpha$  peak at 2.29 keV is visible at the 30 kV spectrum, but not in the 5 or 10 kV spectra. The Mo  $L\alpha$  is marked with a question mark, because there are no signal whatsoever at the Mo  $K\alpha$  peak at 17.48 keV. The peaks which are taller than the plot is marked with a vertical grey stippled line.



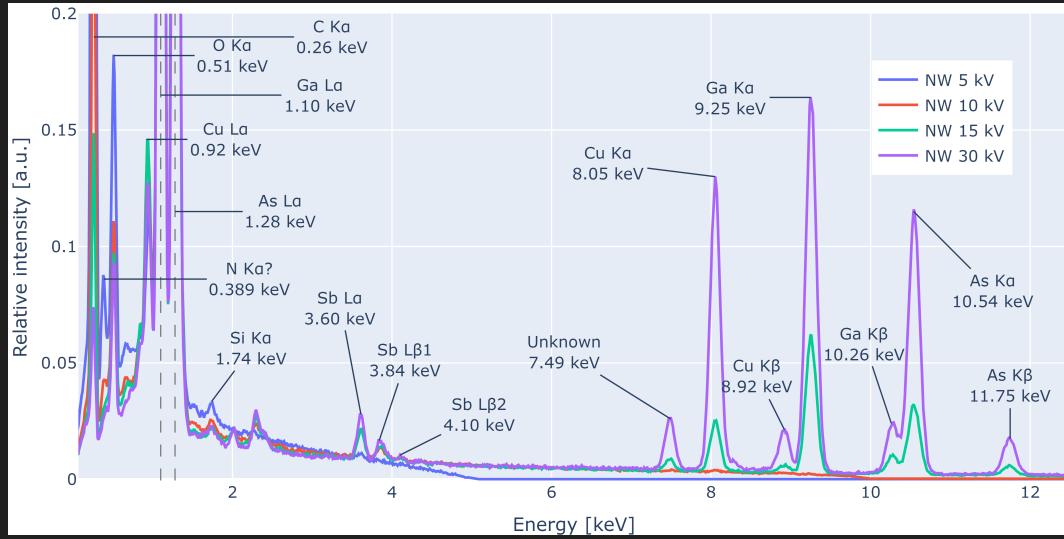
**Figure 1.5:** The spectra of the GaAs sample part. Both the K-peaks and the L-peaks of Ga and As are visible. There is a peak at 0.51 keV, which is the O  $K\alpha$  peak. There is a peak at 0.260 keV, which is the C  $K\alpha$  peak.



**Figure 1.6:** A zoomed in view of the GaAs sample part. The background is easier to see, and the sum peaks are visible. The sub peaks are on ([Brynjart: Add energies!](#))



**Figure 1.7:** The spectra of the Mo sample part. The Mo K $\alpha$  peak at 17.47 keV is barely visible at the 30 kV spectrum, and has a high noise level. The high doublet peak is Mo L $\alpha$  at 2.29 keV and Mo L $\beta$  at 2.39 keV. In the Mo spectra the Si K $\alpha$  line is just barely visible as a very small peak, and is thus not annotated. The peak at 2.01 keV is the Mo Ll peak, with name and energy from the HyperSpy database. The peak at 2.83 keV is the Mo M $\gamma$ 3 peak, with name and energy from the HyperSpy database. Mo Ll and Mo M $\gamma$ 3 have weights at 0.041 and 0.011, respectively. The Mo L $\beta$ 1 peak at 2.49 keV has the weight 0.327.



**Figure 1.8:** The spectra of the nanowire sample part. This spectrum have the most peaks, and contains Ga, As, Cu, Sb, Mo, C, and O. The line at 0.389 keV could be N K $\alpha$  peak, but it could also be other elements. The peak at 0.92 keV which is labeled as Cu K $\alpha$  is also getting a contribution from the Ga L $\beta$  peak, at 0.95 keV. The sum peaks of the L $\alpha$  peaks are not annotated, as they are the same as in the GaAs plot, i.e. at 2.2 adn 2.4 keV. These sum peak are also on top of the Mo L $\alpha$  peak at 2.29 keV, and they are not distinguishable. The 30 kV spectrum had a small singal from the Mo K $\alpha$  peak at 17.47 keV, but that signal was weak. The 10, 15 and 30 kV signal have an un-annotated and un-identified peak at 2.00 keV.

Cu spectra at 30 kV has the Cu K $\alpha$  peak, the Cu L $\alpha$  peak is completely missing. (**Brynjart: add transition sentence.**)

Some of the peaks in the spectras are overlapping, which shifts the shape of the peaks. An example of this is the As K $\alpha$  peak and the Ga K $\beta$  peak in 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 $\alpha$  peak and the Mo L $\beta_1$ , 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 spesific 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 aquired spectra is low, but with different shapes. In the GaAs, Si and Mo the height of the background decrease with higher acceleration voltage. In the Cu spectra the background increase with higher acceleration voltage. In the NW spectra the background is fairly similar, except for the 5 kV spectrum where the background is cut off at 5 keV. All the 5 kV spectra decrease more or less linearly from 1 to 5 keV. The background is very low and almost flat above the highest peak in the spectra. For example, the Cu 10 kV spectrum have its highest peak at 0.5 keV, and the background is almost flat above 0.5 keV. The background in the Si 30 kV spectrum increase a lot up to the high Si K $\alpha$  peak at 1.7 keV, and then the background drops down right after the peak. The background in Si 30 kV is 10 times higher before than after the peak, and does also have a different shape before and after the peak at 1.7 keV. The background shape

in Si 30 kV is almost linear from 0.6 to 1.6 keV, drops down 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 affect on the background as the Si 30 kV spectrum. 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 a Si peak at 1.74 keV. Some of the spectra have peaks from areas outside the impacted 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. NW have clear C K $\alpha$  and O K $\alpha$  peaks at 0.260 and 0.515 keV. GaAs and Mo have some C and O signal. The C and O signal is higher at lower acceleration voltage. All four Mo spectra have a peak at 0.175 keV, which match with B K $\alpha$  or Be K $\alpha$ .

Another artifact present in most spectra is sum peaks. The Si spectra have a sum peak at 3.49 keV, which is the sum of Si K $\alpha$  at 1.74 keV. The Al spectrum have its highest peak at 1.48 keV and a sum peak at 2.98 keV. The Mo spectrum have a sum peak at 4.65 keV, which is the sum of Mo L $\alpha$  at 2.293 keV and Mo L $\beta_1$  at 2.395 keV. The GaAs spectrum at 30 kV have two small sum peak signals at 18.5 and 19.5 keV, while the NW spectrum at 30 kV with lower DT does not have these peaks.

### 1.1.2 Calibration

Different calibrations were explored. The initial calibration is the one from AZtec, and is the one used in the spectra in Figure 1.1. 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 1.1. The deviations are a few percent, and the accuracy of the different calibrations give on specific peaks are given in Table 1.2. 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 presented.

**Table 1.1:** Different calibration values. The AZtec calibration is referred to as the uncalibrated value. The dispersion is calculated with ???. The offset is calculated with ???. The own calibration was done on Ga L $\alpha$  and As K $\alpha$  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 $\alpha$ and As K $\alpha$	0.010030	21.127
Calibration on Mo L $\alpha$ and Mo K $\alpha$	0.010040	21.076

## 1.2 Quantitative results

To do the quantitative analysis, HyperSpy needs k-factors. The k-factors for Ga and As are given in Table 1.3. 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 1.2:** Peak accuracy of the different calibration methods. The accuracy here is the deviation from the theoretical peak to the measured peak. The measured peak is the Gaussian fitted center of the peak. The Mo L $\alpha$  deviates much because the peak is not well fitted. The C K $\alpha$  is fitted well, but deviates much more than all the other peaks. All results are from the 30 kV measurements. The other acceleration voltages gave similar results. The self made calibration was done on two data sets. One was done on Ga L $\alpha$  and As K $\alpha$  from the 30 kV measurement on the GaAs wafer. The other was done on the more far apart peaks Mo L $\alpha$  and Mo K $\alpha$  from the 30 kV measurement on the Mo wafer. The HyperSpy calibration was done by making a model and fitting it to the data on the 30 kV GaAs spectrum. AZ is short for AZtec. HS is short for HyperSpy. All deviations are in percentage difference from the theoretical peak value.

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



The initial quantification was done on the data from the GaAs wafer in AZtec and in HyperSpy as out-of-the-box as possible. The results are presented in Table 1.4. 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 1.4:** 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: Put in the actual results here. Use both HyperSpy linear and model fitted results?)

$V_{acc}$	AZtec		HyperSpy	
	Ga	As	Ga	As
5 kV	50 %	50 %	50 %	50 %
10 kV	50 %	50 %	50 %	50 %
15 kV	50 %	50 %	50 %	50 %
30 kV	50 %	50 %	50 %	50 %

To better understand the ratios between Ga and As, the areas under the peaks in the spectra were counted. Table Table 1.5 gives the ratios between the areas under the peaks for 5, 10, 15 and 30 kV. The table compares L $\alpha$  peaks, K $\alpha$  peaks, K $\beta$  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 the calibrations in Table 1.1 gave different quantification results when using Cliff-Lorimer in HyperSpy. The results are presented in Table 1.6. The quantification on 10 and 15 kV are obviously wrong, but the same method was used for all the quantifications.

**Table 1.5:** 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. ([Brynjár: Delete?](#)) K $\beta$  at 15 kV was too low to be detected and is therefore not included in the table.

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

**Table 1.6:** Quantification with different calibration methods. The quantification is done by in HyperSpy with Cliff-Lorimer method. The CL method is for this samples, while the GaAs wafer used here is a bulk sample. AZ is the AZtec calibration. HS is the HyperSpy calibration. GaAs is the calibration on the GaAs 30 kV spectrum. The accuracy of the quantification is the deviation from 50%, because the sampled area is 1:1 GaAs wafer.

Vacc	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 2

## Discussion

### 2.1 Introduction

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 ??.

### 2.2 Analysis steps in HyperSpy

(**Question for Ton: Is this interesting to write about? The problem here is that the text is both method, some results and kind of discussion. What do I do with that? I want to keep it, but also restructuring it.**)

The next sub-problem was to find out what is done with the data at the different steps in the analysis when using HyperSpy. In these steps it is assumed that the user have done qualitative analysis and want to do quantitative analysis on a set of elements. The analysis in AZtec is done as a black box, so it is not possible to see what is done with the data at the different steps. 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.**)

#### 2.2.1 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()`.

### 2.2.2 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 2.2.4. 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)
```

### 2.2.3 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 [1, 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.<sup>1</sup>

---

<sup>1</sup>Results from the zeta methon can be converted to the cross section method, see the "EDS Quantification" documentation in HyperSpy.

#### 2.2.4 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)  
  
m.add_family_lines(<list_of_element_lines>)  
  
m.plot()
```

The lines above removes the zero peak, create a model from the `signal s`, 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 a 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.

#### 2.2.5 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 2.2.3, 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.

### 2.2.6 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.

## 2.3 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  $\text{FWHM} = \text{std} * 2 * \sqrt{2 * \ln 2}$ <sup>2</sup>. 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 is 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

---

<sup>2</sup>FWHM defined at: [https://en.wikipedia.org/wiki/Full\\_width\\_at\\_half\\_maximum](https://en.wikipedia.org/wiki/Full_width_at_half_maximum)

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.

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

## 2.4 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 1.1](#) shows calibration from AZtec, HyperSpy, and the self produced Python script.

## 2.5 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 piecewise polynomial. The spline is a piecewise 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.

## 2.6 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?**)

## **2.7 Calibration decision**

Why I selected the Ga La and As Ka peaks for calibration. Less extrapolation. Peaks are far apart. The peaks need a good Gaussian fit. Need a nice curve. High peak to background. Should be a sample that is easily available. Cu tape would be nice, since it is in all labs and samples. Mo is far apart, but looks bad. Mo is also harder to fit automatically, because of the close double peak.

## **2.8 Choises in HyperSpy**

Commercial packages does quantify SEM signal, and they are kinda good at it. Should be more accurate on NW sample, but is it? Reference the discussion on GitHub?

# Bibliography

- [1] 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.