Automated Discovery of Temperature Dependent Structural Change
SMC Data Challenge 2017

Gabriel Santucci[1]

[1]Nucleon Decay and Neutrino Group, Stony Brook University Physics Department

August 2017

Abstract

In this study we look for phase transitions as a function of temperature by inspecting data from neutron diffraction. A phase transition is defined here as a structural change in the analyzed material. We characterize the intensity peak structure in the diffraction data at adjacent temperatures to conclude if a phase transition has occured.

1   Introduction

We are interested here in finding an algorithm to decide if a phase transition occurred in a material between two different temperatures. The data structure is given by different curves of Intensity (I) as a function of distance (d - characteristic spacing of the material in Angstroms) for different values of temperature (T). To define a phase transition we first study one particular peak in the intensity curves, namely the peak between 3.2 and 3.3 $\mathring{A}$. The characterization of this peak consists of calculating its center, a characteristic width and the area under it. Once we have the characterization procedure we move on characterize all the peaks in the data set for a given temperature value. Finally, one can define a phase transition by looking at the sets of peaks and its characteristics for two different adjacent curves (temperatures). We define a phase transition as a distinct change in structure of the intensity peaks.

2   Inspecting the Data

Before we inspect any particular peaks, let us look at different sets of curves for randomly chosen values of temperatures to understand how the peak structure is. Figure 1 shows the curves of intensity as a function of d-spacing for random values of T.
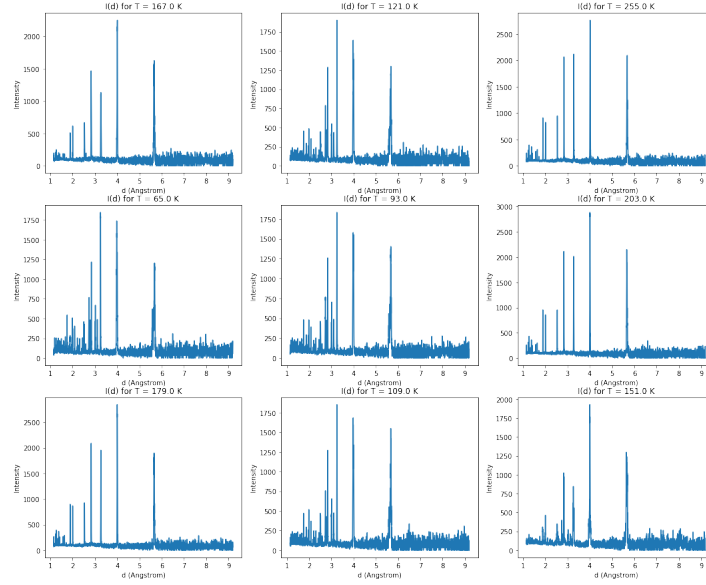
Figure 1: Intensity curves as a function of d-spacing for different values of temperature.

We can also look at the maximum intensity distribution as a function of temperature. Figure 2 shows the highest value of I for each temperature T. We can already see a big hint of a structure change at around T = 150 K. This particular data set has a unique maximum intensity value for each temperature. This allows us to uniquely map a temperature to an intensity value and vice-versa. So given a temperature value we can find the corresponding set of intensity values or the opposite. This is certainly not a general result valid for any data set, but it is valid for the particular one that we are working with here. Building this one-to-one map makes it faster to get a curve given a temperature and also scales a lot faster with more data, in case the new data also presents this feature.
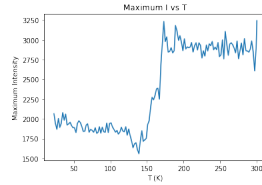

Figure 2: Maximum intensity value as a function of function temperature.

2

3 The 3.25 Å Peak

To be able to detect a phase transition we first need to be able to detect a peak and characterize its features. In this section we describe how to characterize a peak. In particular, we will look at the peak located between 3.2 and 3.3 Å. We characterize a peak by its center, width and area under it. To attribute these values for a given peak, we fit a gaussian function corrected for arbitrary area and background level. The unit area gaussian centered around x_0 with width σ is given by eq. 1

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-x_0)}{2\sigma^2}},\tag{1}$$

We can add a constant background to this function by adding a constant b and by multiplying $f(x)$ by A, we can interpret A as the area under the curve:

$$g(x) = Af(x) + b = \frac{A}{\sqrt{2\pi}\sigma} e^{-\frac{(x-x_0)}{2\sigma^2}} + b,\tag{2}$$

As we will see, this function is not a good hypothesis to fit peaks close to the phase transition temperature. The shape of those peaks are not gaussian at all, and a gaussian description is not adequate. A good way to evaluate how well the adopted function describes the data is to calculate the χ^2 of the fit and look at the residues distribution. We are not implementing this analysis here. Mainly because we are actually not interested in the actual value of the area, or width of these peaks, but we are actually interested in finding a way to detect a phase transition. Since the value of these variables will change drastically during a phase transition, this will actually be enough to detect it, so an accurate description of the shape of these peaks is not necessary. By using the function defined in 2 we have a very simple interpretation for our data and relatively fast algorithm that scales well with size. Figure 3 shows the fitted curve for T = 0 K and T = 150 K. We can see that far from a phase transition, the fit works fine, whereas close to the phase transition the fit fails. Nonetheless we can see that the fitted values for A, x_0 and σ are very different which allows us to see something is going on at that temperature. Figure 4 shows the values of A, x_0, σ and background levels as a function of temperature for the 3.25 Å peak. Also, we show the same curves normalized in figure 5 for visualization purposes. We can clearly see an interesting structure around 150 K. In all variables, the pattern followed by the curves is drastically changed by this phase transition.

4 Finding Peaks

To generalize the procedure of peak characterization we need to be able to detect all the peaks in the data set and study them. To do this, we follow the procedure described in [1]. We use a simplified form of this algorithm, adapted to our needs and the present data set.
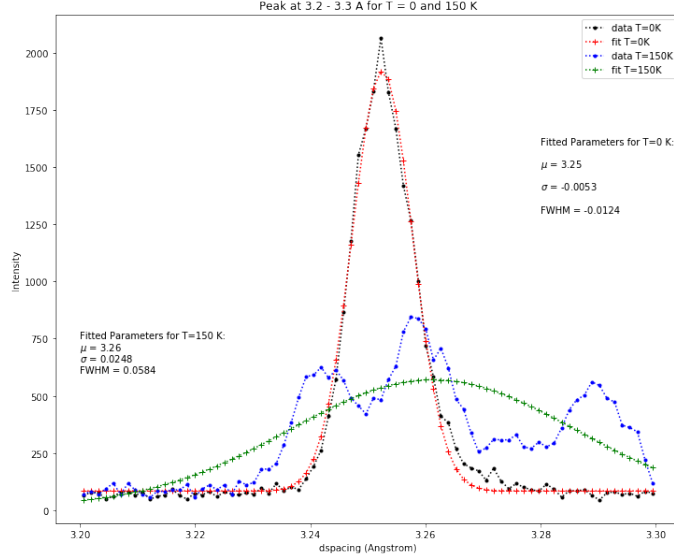
Figure 3: Gaussian fits for T = 0 and 150 K.

The way our algorithm works is by declaring all the points in the data set as a possible peak and then start removing peaks that do not satisfy some criteria, namely:

1. Promote all the data points to peak candidates;

2. Remove all the points that are below a certain threshold (defined by the user) - mph;

3. Remove all the points that are close to a higher peak, where the distance is defined by the user - mpd;

Once we detect all the peaks in a data set, we can use the algorithm defined in the previous section (gaussian fit) to construct a list with all the peaks and their variables (number of peaks, areas, centers and widths). Figure 6 shows the detected peaks in the intensity profile for T = 0 K. Ideally we would have systematically tried different values of thresholds for height and distance as described in the algorithm. But in this first study we simply used reasonable values without optimization. A complete optimization of these hyper-parameters is necessary to improve results. But as discussed before, drastic changes in the material structure will still be detected even without further optimization. To find the height threshold we first calculate the average intensity of the data set and the standard deviation. We define the threshold as:

$$mph = \mu + 1.5 \cdot \sigma, \tag{3}$$

where μ is the mean value of the data set and σ, the standard deviation. For the minimum

4
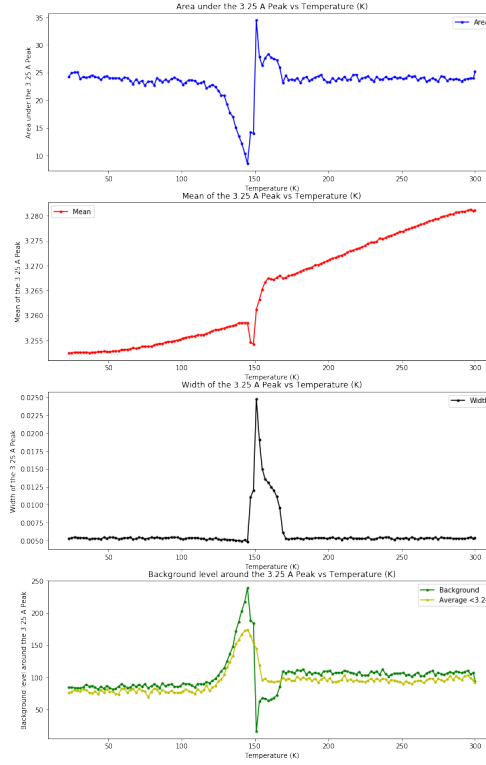
Figure 4: Area under the curve, center of the peak, peak width and background level as a function of temperature.

distance we used mpd = 50, i.e, tow peaks have to be at least 50 points apart in the data set.

5 Phase Transitions

Finally, we are ready to detect a phase transition in the material. Having an algorithm to find all the peaks in the data and a way to characterize them, we can make a comparison between all the peaks found in two different adjacent temperatures and determine if a phase transition occurred or not. Unfortunately, we did not perform a deep study of different ways of detecting a phase transition. In a complete study we can try at least making a peak by peak comparison, but here we propose the following simple criteria:

1. Compare the number of peaks in the two data sets.

2. If the number of peaks is different, compare the average width of all the peaks.
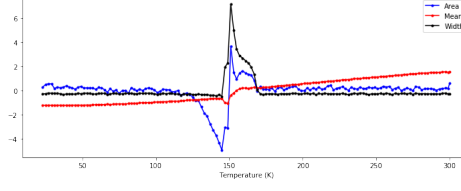
Figure 5: Normalized area under the curve, center of the peak, peak width and background level as a function of temperature.
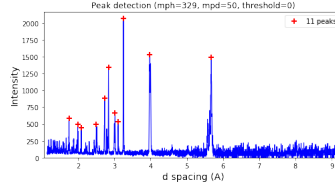

Figure 6: Intensity as a function of d-spacing for T = 0 K. The red points indicate the detected peaks in the data.

3. A phase transition occurs if: $|w1 - w2| > 0.005$.

6   Conclusions

In this study we have defined how to detect a phase transition in a material given intensity data. The process consists of detecting all the peaks in the data for two adjacent temperatures and calculating defining quantities for every peak, namely, their areas, centers and widths. For sets with different number of peaks in the data and with different average width (difference > 0.005) we declare that a phase transition has occurred.

7   Appendix

The code used to generate the plots in this study can be found in gcsantucci github repository under SMC DataChallenge Diffraction [2].

References

[1] http://nbviewer.jupyter.org/github/demotu/BMC/blob/master/notebooks/DetectPeaks.ipynb

[2] GitHub Repository

6