PHY324: Data Analysis Project

Idil Yaktubay

13 February 2022

1 Introduction

The goal of this paper is to convert the raw data from a particle detection experiment to an energy spectrum, and fit this energy spectrum with computational methods. By illustrating the steps necessary to reach this goal, this paper gives an introduction to the data analysis methods used in extracting meaningful physical information from experimental data. These methods are highly useful in physics as they provide the means to draw quantitative conclusions from the findings of an experiment. Examples of such conclusions include the experimental value of an unknown constant, or the energy spectrum of an unknown energy source as in the case of our experiment.

2 Experimental Setup

In this experiment, a specific particle detector setup was used to collect data. In this setup, the incident particles can be detected when they excite and eject the electrons in the detector material through the photo-absorption effect or Compton scattering. The detector then converts the energy of the incident particles into a signal of voltage with respect to time. The relation between energy and voltage in this conversion is perfectly linear. In the signal, the detection of a particle manifests itself as a deviation from the regular pattern of voltage, referred to as a "pulse". When the detector senses an incident particle, it collects 4096 values of voltage that contain the pulse and regions before and after the pulse. This set of 4096 data points is referred to as a "trace". In our setup, the rise time τ_{rise} is 20 μs, the fall time τ_{fall} is 80 μs, and traces follow the function

$$y = A\left(\frac{\tau_{fall}}{\tau_{rise}}\right)^{\frac{-\tau_{rise}}{\tau_{fall}-\tau_{rise}}} \cdot \left(\frac{\tau_{rise}-\tau_{fall}}{\tau_{fall}}\right) \cdot (e^{-t/\tau_{rise}} - e^{-t/\tau_{fall}}) \tag{1}$$

where A is the amplitude of a trace and depends on the energy deposition. In a trace, the beginning of a pulse is located on the 1000th data point, and the regular voltage value is set at 0 V. Therefore, deviations from 0 V usually indicate energy detection, whether it is from noise or from an energy source of interest. Figure 1 depicts what a typical trace containing a pulse from our detector setup looks like.[1]

With this setup, three sets of data were collected. The fist data set is 1000 traces of noise, the second data set is 1000 traces from a known calibration source, the third data set is 1000 traces from (an) unknown energy source(s), with each trace containing 4096 data points. The calibration source is known to emit 10 keV photons that are all absorbed by the detector material. The reason behind collecting noise data and calibration data will become apparent in the next section.[1]

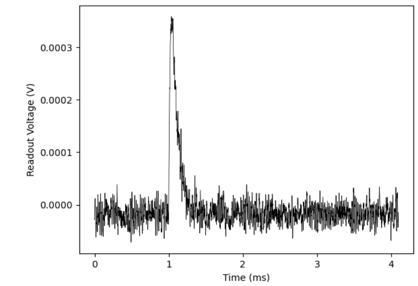

Figure 1: The 15th trace from the calibration data. The total number of samples is 4096, and the pulse begins on the 1000th sample. The base of the trace is set at 0 V, and data points outside the pulse fluctuate about 0 V due to noise. Apart from the noise, the trace follows equation 1.[2]

3 Methodology

The first step in extracting the energy spectrum of the unknown energy source is to calculate the calibration factor with which the detector converts incident energies to voltages. With the knowledge that the detector was exposed to the 10 keV photons from the calibration source as well as the noise from the environment, we expect that a distribution of the 1000 pulses from the calibration data in volts is a narrow Gaussian distribution centered at 10 keV converted to voltage by this calibration factor. We can construct this distribution by establishing various ways of quantifying the size of each pulse in volts from the calibration data.

We used six different ways to quantify the size of each pulse from the calibration data. The first way was to subtract the minimum value of voltage, or amplitude, from the maximum amplitude for each trace. The second way was to subtract the trace baseline from the maximum amplitude of each trace. A trace baseline was approximated as the average of the first 1000 samples of each trace, or the average of the "pre-pulse region". The third way was to approximate the integral of each trace as the sum of the 4096 samples from each trace. The fourth way was to correct the issues with the previous integral approximation by summing the 4096 samples after subtracting the trace baseline. The fifth way was to further correct this approximation by limiting the bounds of the integral to the pulse onsets and ends. Finally, the last way was perform a chi-squared fit on each of the 1000 traces using equation 1, and use the maxima from these fits as our pulse sizes. This fit was performed on the data points including and after the 1000th data point of each trace to ensure that the trace is compatible with equation 1. The fixed uncertainty used in this fit was the average standard deviations of the traces from the noise data. One of these 1000 fits is shown on Figure 2.

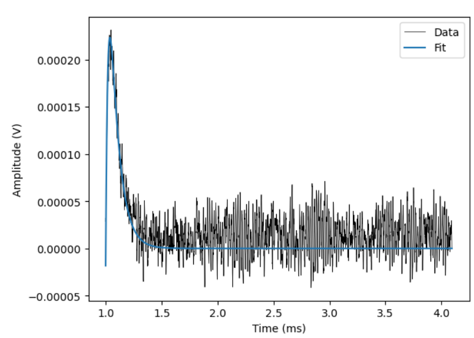

Figure 2: A fit of the 200th trace from the calibration data. The portion of the trace that starts from the pulse onset was fit to equation 1.[2]

To find the locations, or mean values, of the distributions corresponding to each one of the six energy estimators, we first constructed histograms of the pulse sizes from each energy estimator, then performed a chi-squared fit on each one of the histograms using the height of each bin as our data points,and $\sqrt{N}$ as our uncertainties where N is the corresponding count. The choice of uncertainties is justified because our data is random due to the detected noise. The Gaussian model used to fit the histograms had the form

$$G = ae^{\frac{-(x-\mu)}{2\sigma^2}} + b \tag{2}$$

where μ is the mean, σ is the standard deviation, a is the scaling factor, and b is the translation of the distribution. We found the calibration factor of each energy estimator by dividing 10 keV by the optimized distribution mean in mV. To convert these distributions to energy spectra, we simply multiplied the data points by the corresponding calibration factors, and performed chi-squared fits on the converted histograms using equation 2. Figures 3-8 depict the distributions of each energy estimator before and after energy calibration. Additionally, Table 1 contains useful information on the quality of the chi-squared fits.

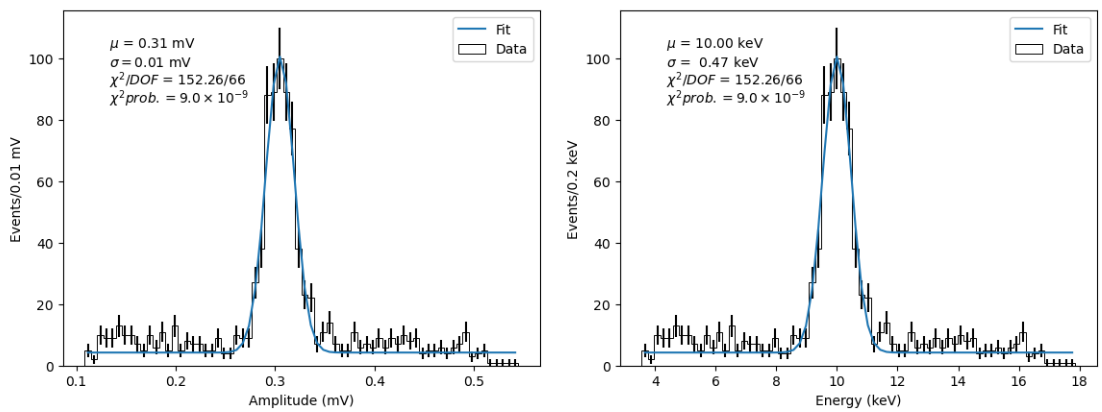

Figure 3: Histograms of the energy estimator corresponding to the max-min amplitude of the traces from the calibration data set, before (left) and after (right) energy calibration.[2]

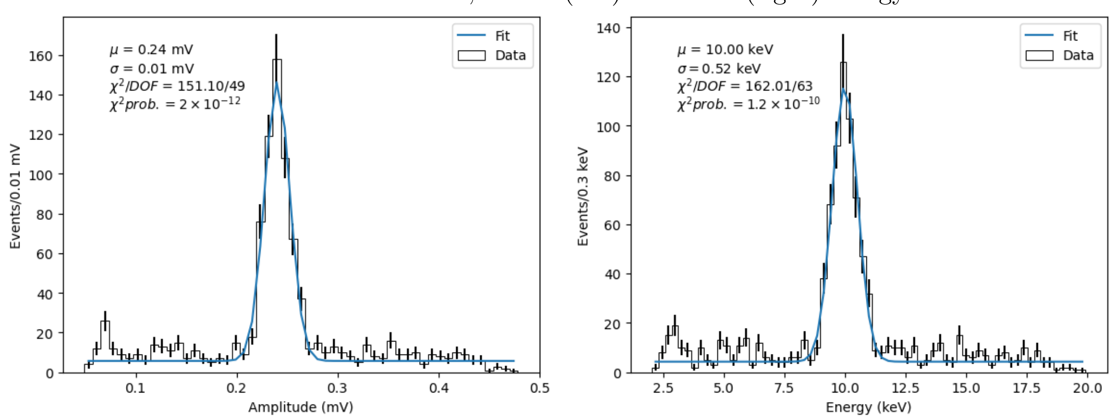

Figure 4: Histograms of the energy estimator corresponding to the max-baseline amplitude of the traces from the calibration data set, before (left) and after (right) energy calibration.[2]

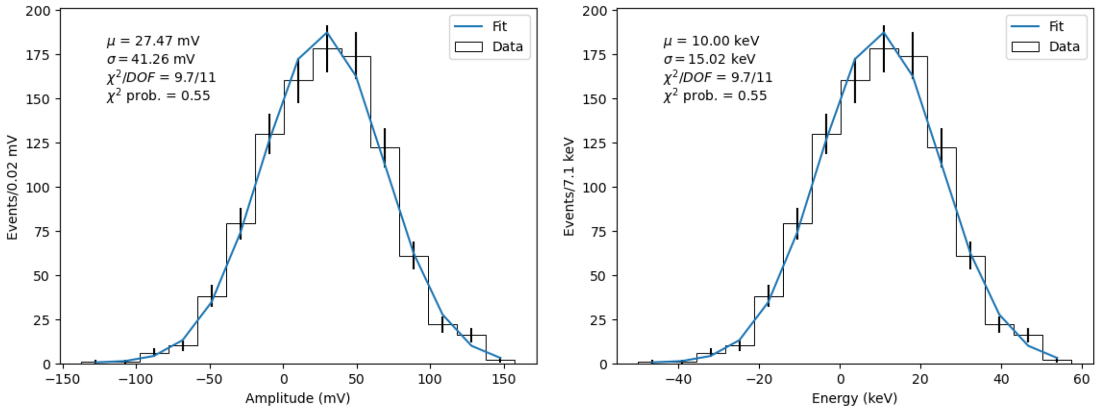

Figure 5: Histograms of the energy estimator corresponding to the integral of the traces from the calibration data set before subtracting the baseline and limiting the integral bounds, before (left) and after (right) energy calibration.[2]

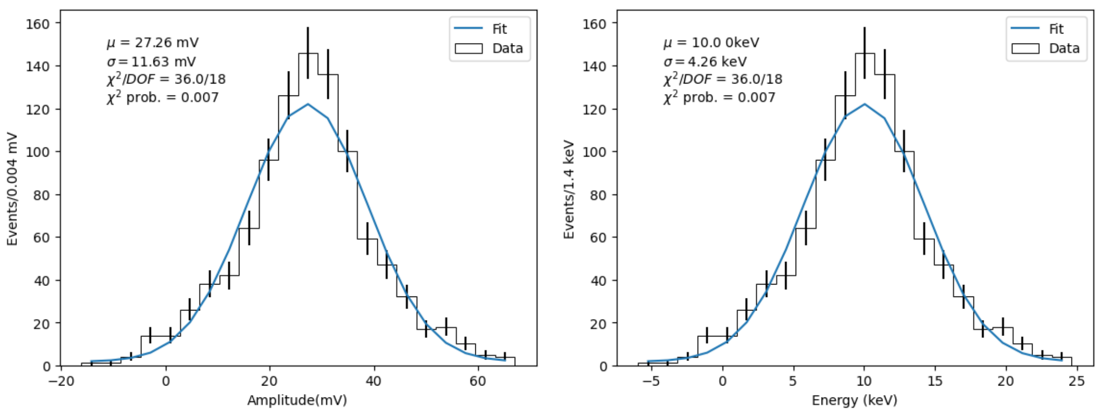

Figure 6: Histograms of the energy estimator corresponding to the integral of the traces from the calibration data after subtracting the baseline and before limiting the integral bounds, before (left) and after (right) energy calibration.[2]

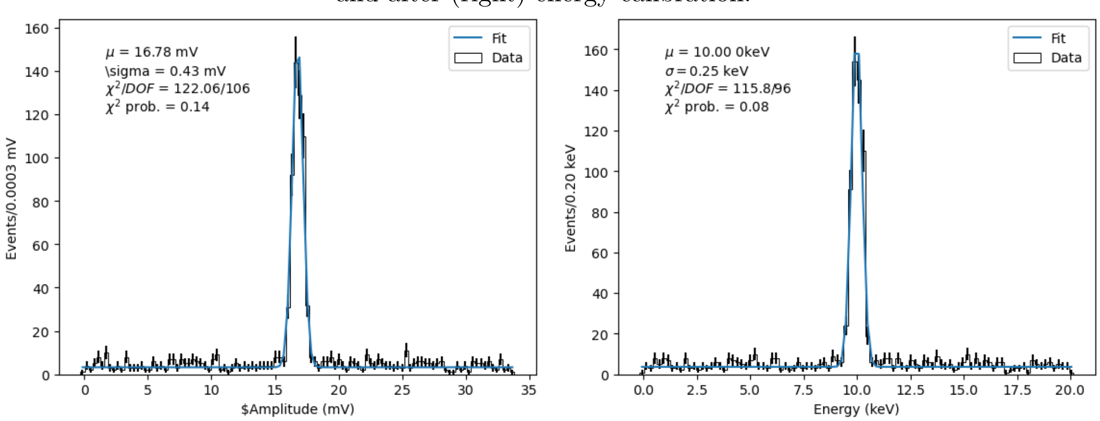

Figure 7: Histograms of the energy estimator corresponding to the integral of the traces from the calibration data after subtracting the baseline and limiting the integral bounds, before (left) and after (right) energy calibration.[2]

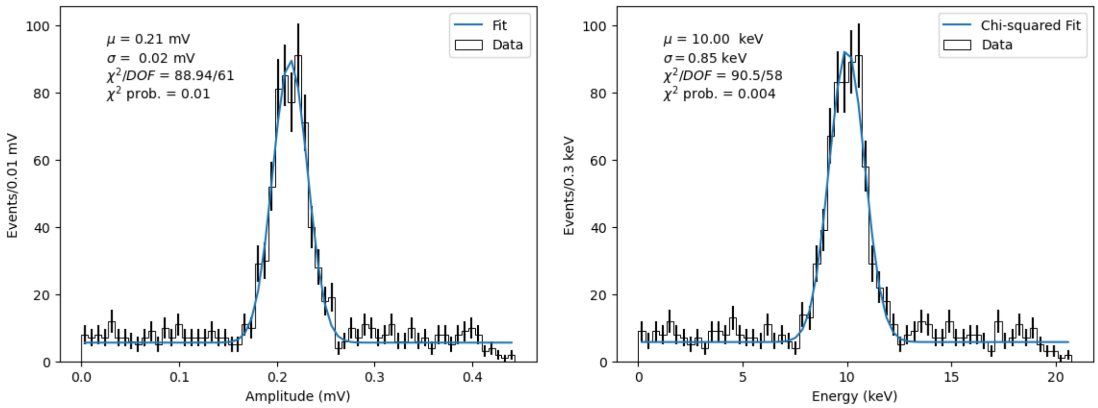

Figure 8: Histograms of the energy estimator corresponding to the chi-square fit traces from the calibration data, before (left) and after (right) energy calibration.[2]

Energy Estimator	**Calibration Factor**	**Energy Resolution**	**Fit χ^2 probability**
Max-min Amplitude of Trace	32.8 keV/mV	0.47 keV	9.0×10^{-9}
Max-baseline Amplitude of Trace	41.6 keV/mV	0.52 keV	1.2×10^{-10}
1) Integral (No baseline subtraction or bound limitation)	0.36 keV/mV	15.02 keV	0.55
2) Integral (With baseline subtraction and no bound limitation)	0.37 keV/mV	4.26 keV	0.007
3) Integral (With baseline subtraction and bound limitation)	0.60 keV/mV	0.25 keV	0.08
Chi-square Fit of Traces	47.6 keV/mV	0.85 keV	0.004

Table 1: Energy estimator evaluation for the energy estimators depicted on figures 3-8.[2]

As we can see from Table 1, the fits with the two highest χ^2 probabilities are the third and first integrals. The corresponding reduced χ^2 values are 0.9 and 1.2. These values are very close to the ideal value of 1 for a high quality fit. However, we notice that the energy resolution, or the standard deviation, of the calibrated distribution of the first integral is significantly large. Therefore, since the detector is known to be predominantly exposed to 10 keV electrons, we can conclude that the narrow distribution of the third integral and its ideal χ^2 and χ^2_{red} values indicate that it is the best performing energy estimator out of the six. Hence, the conversion of the unknown signal to an energy spectrum can be done with the calibration factor of 0.60 keV/mV.

The first step of analyzing the unknown signal was to determine the energy estimators of each of the 1000 traces by subtracting the baseline from and summing each set of 4096 points corresponding to each trace. Then, these values were converted to energy values by multiplying them by the calibration factor of 0.60 keV/mV. A histogram, or the energy spectrum of these energy values was constructed. The spectrum is depicted by figure 9 below.

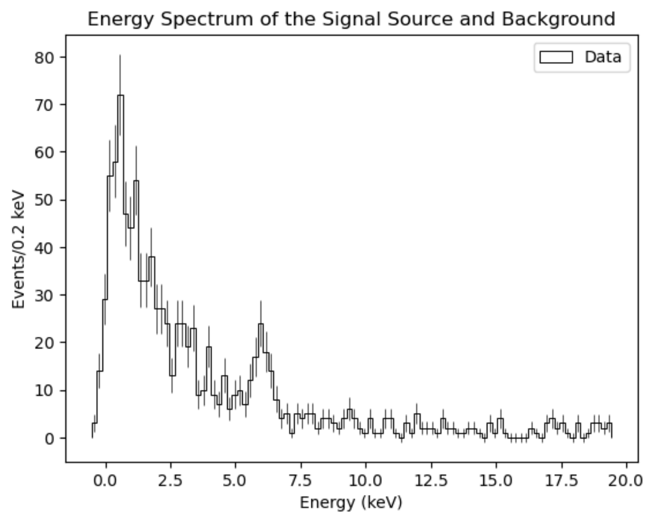

Figure 9: Energy spectrum of the unknown energy source(s) and background noise after calibration and before performing a chi-squared fit. The spectrum was calibrated with a calibration factor of 0.60 keV/mV as calculated previously from the third integral energy estimator.[3]

As we can see on figure 9, the shape of the spectrum looks nothing like a Gaussian distribution. However, it looks like a linear combination of equation 1 and equation 2. Therefore, we chose to fit the energy spectrum to the function described by

$$a(e^{\frac{-(x-b)}{c}} - e^{\frac{-(x-b)}{d}}) + E(e^{\frac{-(x-\mu)^2}{2\sigma^2}}) + F. \tag{3}$$

The uncertainties were used as $\sqrt{N}$ for each bin as before, due to the randomness of the data. Figure 10 depicts this fit, and table 2 provides information on it.

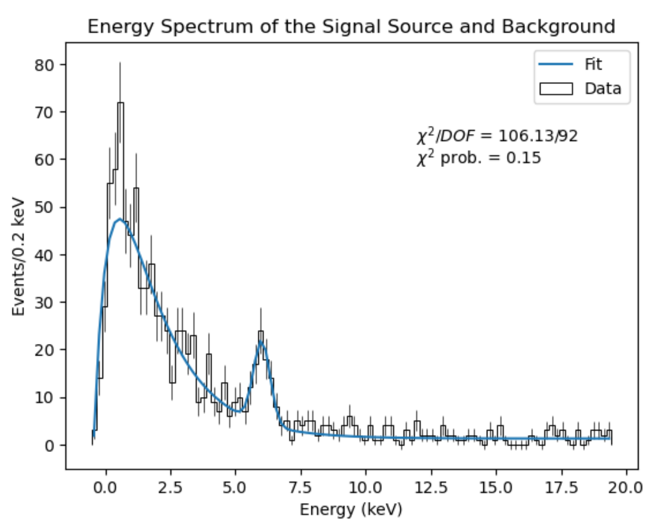

Figure 10: Energy spectrum of the unknown energy source(s) and background noise after calibration and performing a chi-squared fit. The fit was performed using equation 3.[3]

Parameter	Optimized Value	Uncertainty (±
a	120	30
b	-0.44	0.01
c	1.8	0.2
d	0.6	0.1
E	17	3
F	1.2	0.2
μ	6.01	0.05
σ	0.32	0.05

Table 2: The optimized values of the parameters from equation (3) with their uncertainties. These values were obtained by using the `curve_fit()` function from the `scipy.optimize` library of Python.[3]

As we see from figure 10, the χ^2 probability and the χ^2_{red} of the fit are 0.15 and 1.2, respectively. This indicates that the fit quality is very high. Therefore, we have converted the raw data from unknown energy source to an energy spectrum, and successfully performed a fit.

4 Summary

In this paper, we walked through the steps in converting the raw data from a particle detector to an energy spectrum and performed a fit on the spectrum, with the motivation of providing the reader an introduction on the data analysis methods used in physics experiments. We explored various energy estimators to calibrate a detector that displays energy detection in units of voltage, and established a metric to choose which energy estimator provides the most ideal calibration. We found that the fifth calibration method, which involved the integrals of traces after an appropriate baseline subtraction and integral bound limitation, was the most ideal energy estimator. This is due to the fact that such integrals contain a minimal amount of noise within their bounds, and take into account the most number of amplitude values. With the chosen energy estimator, we calibrated the unknown signal, and were able to perform a high-quality fit on its energy spectrum. Hence, the chosen energy estimator was shown to be successful in characterizing the energy spectrum of an unknown source.

5 References

1. (2022). *PHY324 Data Analysis Project. Winter 2022*: University of Toronto.
2. (2022). "`calibration.pkl`" file from PHY324 Quercus course page.
3. (2022). "`noise.pkl`" file from PHY324 Quercus course page.
4. (2022). "`signal.pkl`" file from PHY324 Quercus course page.