PHY 324: Data Analysis

Meet Chaudhari

7th February 2023

Abstract

This report describes the theory and implementation of data analysis necessary to characterize a particle detector's response to detection events. It contains the methods of using various energy estimators such as amplitude of integral of a signal to convert the detector's raw voltage measurements to energy spectra. The process involves finding calibration factors for various energy estimators via a known calibration source of fixed energy and then calibrating real signals to find an energy spectrum. The data is processed and fitted to understand the goodness of fit via various statistical measures. This is implemented using python programming language and is quite useful when dealing with experimental data in physics and other fields.

1 Introduction

The aim of the report is to demonstrate a simple procedure to analyse raw data collected from an experimental setup of a particle detector to find a functional form of the energy spectrum of photons from an unknown source. For a particle with given energy that hits the detector, it's generally not possible to directly measure its energy, therefore it needs to be encoded in form of some other measureable quantity. When making detection observations from an unknown source of particles, it becomes important to convert the measured data to the energy of the particles. To accomplish this, it is necessary to record data from a source of particles with known energies (calibration source) and find an energy estimator which is a quantity that is very correlated to the total energy of an incident particle. By analysing the distribution of these energy estimators then, a calibration relation could be found, which would convert the measured calibration data to the energy of the known particles. We can apply it to the unknown source to find the energy spectrum of the emitted particles. This report demonstrates this entire process for a given detector and its data.

For the detector whose data is under consideration, the detector's sensors measure voltage fluctuations from 0V induced by incoming photons, which deposit energy to stationary electrons either completely via photon absorption or in form of momentum transfer through Crompton scattering [1]. These electrons then form an electric current and hence a voltage. The detector has a highly efficient trigger system as well which would communicate the sudden increase in voltage in sensors to the rest of the system for recording the voltage as a function of time properly. By requiring that the voltage fluctuations occur via exactly the same process as above for every detection and fixing the characteristics of the detector setup, we would expect the photon detections to follow a fixed voltage vs time function with variable amplitude. This functional form of each photon detection or pulse would be referred to as a "trace". Each trace is represented as 4096 measurements of voltage by a 1 MHz data acquisition system. This would mean 1 voltage measurement per $1\mu s$ and each trace is setup to have a $20\mu s$ rise time and $80\mu s$ fall time (see figure 1). For a pulse of amplitude A, the pulse would be described by:

$$V(t) = A \cdot C \cdot (e^{-\dfrac{t}{\tau_{rise}}} - e^{-\dfrac{t}{\tau_{fall}}}) \tag{1}$$

where

$$C = (\frac{\tau_{fall}}{\tau_{rise}})^{-\dfrac{\tau_{rise}}{\tau_{fall} - \tau_{rise}}} \cdot (\frac{\tau_{rise} - \tau_{fall}}{\tau_{fall}}) \text{ and } \tau_{rise} = 20\mu s, \tau_{fall} = 80\mu s \tag{2}$$

is a constant to ensure that the amplitude of $V(t)$ is exactly given by A. This would further mean that the energy of an incident photon would be related to the amplitude of the trace. Moreover,

[1]Experiment and Data, PHY324 Data Analaysis Project Instructions Manual

when the trigger system senses the beginning of the sharp rise in voltage indicating the beginning of the pulse, the data processing system of the detector shifts it to the $1\mu s$ point in the trace. It also adds a 'pre-pulse period' from $(0-1000)\mu s$ which is the quiescent voltage recorded $1\mu s$ before the trigger went on. This helps in quantifying the level of noise just before the detection of a photon, which would be quite important as noise might be time dependent.

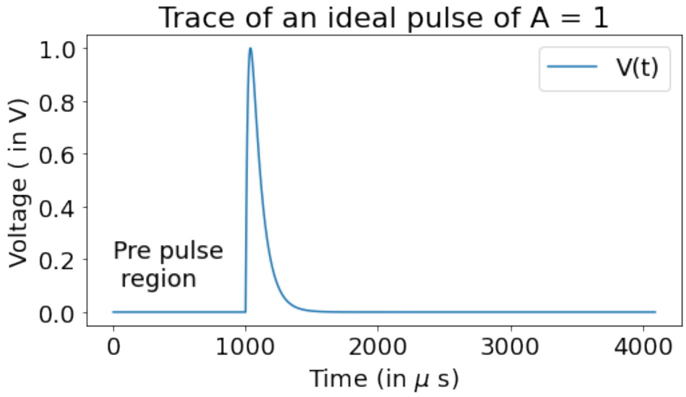

Figure 1: The voltage vs time form or 'trace' of an ideal pulse with amplitude 1. There are 4096 voltage measurement plotted at a 1 MHz rate i.e. one voltage measurement per 1μ s. The region from $(0-1000)\mu s$ is the pre-pulse region which is added by the detector's data processing system to place the onset / beginning of the pulse at $1000\mu s$. Real data in general would be very noisy.

The data analysis is performed with 3 sets of experimental raw data collected by the detector: signal data, calibration data, and noise data (Credit: PHY324 faculty). Each of the three files contains 1000 events or traces. The signal data contains the traces of photon detections from an unknown source. It is not yet known what kind of distribution is followed by the photon energies from this source. Whereas, the calibration data contains traces of photons emitted by a 10 keV source. These photon energies are known to follow a gaussian distribution with a mean of $\mu = 10$ keV. Finally, the noise data file contains 1000 traces of measurements of the background noise which could be caused by temperature fluctuations, photons from the environment, radioactive decay, electrical noise, etc. As an example of the kind of data contained in each file, a plot of the 10th event in each of the file is shown in figures 2, 3, 4 below. Observe how there seems to be more noise in the signal trace compared to that in the calibration file's trace. This 'extra-noise' might be an indication of a more complicated energy spectrum for the signal photons.

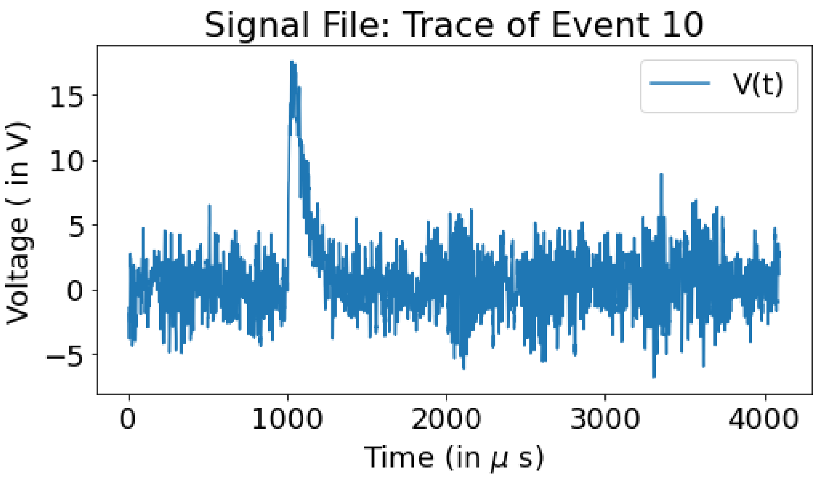

Figure 2: Trace of 10th event in the signal file. The signal file contains traces of photon detections of unknown energy. Observe how there is noise in the entire range, particularly signficant around $t = 3000\mu s$.

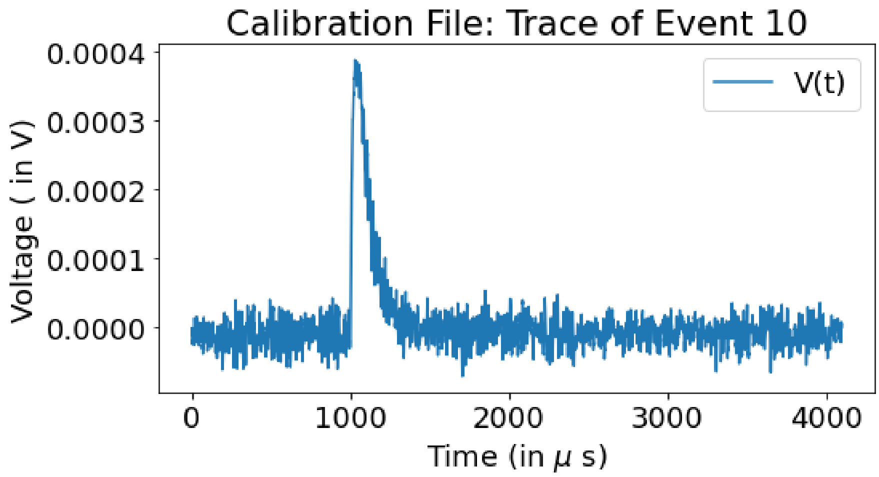

Figure 3: Trace of 10th event in the calibration file. The calibration file contains traces of 10 keV photon detections from a known source. Note that not all traces here correspond to a 10keV photon because the energy of photons is assumed to be a gaussian distribution with a mean of 10 keV.

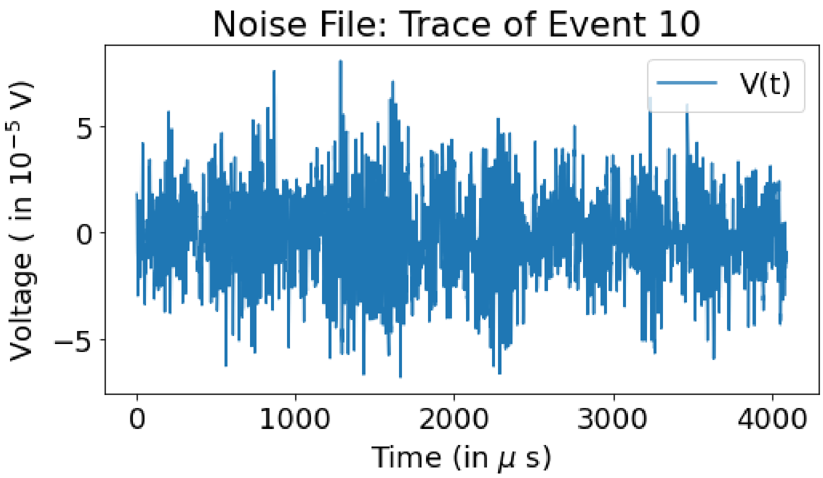

Figure 4: Trace of 10th event in the noise file. The noise file contains 1000 traces of background noise collected by the detector. It can be seen from the plot that the background noise has a mean of about 0 V. The causes of noise are generally spare photons absorbed due to temperature flucatuations, reflections from the environment, radioactive decay from the background, etc.

2 Summary

In the investigation described in this report, various kinds of energy estimators were analyzed to find the best energy estimator that can reconstruct the energy spectrum of a known calibration source emitting 10 keV photons. The estimators analyzed included three different types of amplitudes and three different kinds of integrals under the pulse in each trace of calibration file. It was found that the amplitude calculated by subtracting the minimum of a trace from its maximum was the best energy estimator. This was then applied to get an energy spectrum of the signal data form an unknown source.

3 Method

It is known before hand that the energy spectrum of the calibration source, whose photons' traces are stored in the calibration file, follows a Gaussian distribution with a mean of $\mu = 10$ keV. Hence, an energy estimator, a quantity related to trace that is expected to be very correlated with the energy of the photon detected, could be used to form a distribution or histogram of the calibration data file's traces. These energy estimators can be fitted with a gaussian, because the underlying calibration data follows a gaussian distribution, and the mean of the fitted gaussian function could be converted to the mean $\mu = 10$ keV via a calibration factor. By fitting with a gaussian again, we can conclude which energy estimator best converts the calibration data into the expected gaussian distribution with $\mu = 10$ keV.

As can be seen from equations 1 and 2, the amplitude of the voltage versus time plot, the trace, must store some information about the energy of a detected photon because a higher energy photon would deposit more energy to the electrons in the sensor, causing a higher current, and hence a higher voltage amplitude. Thus, one natural energy estimator from the system would be the amplitude of the trace. But in a trace plagued with background noise such as that in figure 3, there are multiple ways to estimate the amplitude of the trace; we describe 3 of them here. Furthermore, the integral under the trace can also be a valid energy estimator as a higher energy deposition by a photon would increase the height of the curve and would broaden the base of the pulse's peak. This will make the pulse take a longer time to decrease to the quiescent voltage and hence would increase the area under the pulse. Similar to amplitudes, there are also multiple ways of tweaking the integral to find a suitable energy estimator which will be discussed below. Note that although the noise in the trace would obviously make the energy estimator calculations for each trace inaccurate, assuming the background is nearly constant, the noise should not ideally affect the distribution of the estimators as it would change the calculated mean deviate from the true mean very slightly and we would have to callibrate the mean to 10 keV anyways.

To fit the histograms for all the estimators, the range of the histogram and the number of bins were used to find the bin-centres on the horizontal axis. Then, the points (bin centre, counts) were plotted equipped with $\sqrt{\text{count}}$ uncertainty bars. This was done because each measurement of a photon's energy is a measurement of the mean of the underlying poisson distribution of calibration photon energies which we are approximating as a gaussian [2]. Since in a poisson ditribution standard deviation is $\sqrt{\text{mean}}$, we put $\sqrt{\text{count}}$ uncertainty bars. These points on the histogram are then fitted with a gaussian (see equation 3) using python scipy's *curve_fit* method.

$$V(t) = A \cdot e^{-\dfrac{(t - \text{ mean})^2}{2 \cdot \text{ width}^2}} + \text{ base} \tag{3}$$

from which the parameters A, mean, width, and base could be extracted to plot the model gaussian on individual histograms. The energy resolution of the detector after calibration is given by the *width* parameter. Since we know that the source majorly emits 10 keV photons, we expect a narrow gaussian peak at 10 keV after calibration. Thus, since more noise would broaden the gaussian peak, we want *width* to be comparatively smaller than 10 keV. Further, the reduced chi-squared is calculated as :

$$\chi^2_{red} = \frac{\chi^2}{\text{DOF}} = \sum_i^n \frac{(count - count\_fit)^2}{\text{DOF} \cdot count} \tag{4}$$

where DOF is the number of degrees of freedom = bin count - 4 (because 4 parameters are fitted for the gaussian in equation 3). An ideal χ^2_{red} is close to 1 whereas $\chi^2_{red} < 1$ implies overfitting and $\chi^2_{red} > 1$ implies poor model.

The chi-squared probabilities was directly calculated via python's *chi2.cdf* method using the formula:

$$\chi^2 \text{ prob.} = 1 - \text{chi2.cdf}(\chi^2\text{,DOF}) \tag{5}$$

The value of this probability should ideally be in the 0.05-0.90 range as it gives the probability of our fitted gaussian to be an accurate model of the distribution of energy estimator and hence the calibrated data.

4 Procedure and Results

Now that we know how to calculate these statistical properties, let's inspect our energy estimators.

4.1 Energy Estimator: Max - Min Amplitude

One way to estimate the amplitude of a trace is to subtract the minimum of the trace from the maximum. Thus, the max-min amplitude was calculated for each trace in the calibration data.

[2]PHY324 Data Analysis Slides (2023), Author: PHY324 faculty

These were then used to create a histogram (see figure 5) which was then fitted with a gaussian against a flat background to find the uncalibrated mean as $\mu_{1,\text{ uc}} = 0.31$ mV. The number of bins were chosen to be $n = 80$ because it led to the closest reduced chi-squared value to 1 while maintaing a reasonable chi-squared probability to be ≈ 0.2. Further, the histogram was created only in the range $(0.25, 0.36)$ because including further data points only lead to an increase in chi-squared value while essentially not altering the mean $\mu_{1,\text{ uc}}$.Then, the calibration factor was calculated as

$$f_1 = \frac{10^4 \text{ eV}}{\mu_{1,\text{ uc}}} \approx 32,743 \frac{\text{eV}}{\text{mV}} \approx 33,000 \frac{\text{eV}}{\text{mV}} \tag{6}$$

Multiplying this calibration factor f_1 to the max-min amplitude data, we get a calibrated histogram with energy in eV on the x-axis and counts on the y-axis (see figure 6)

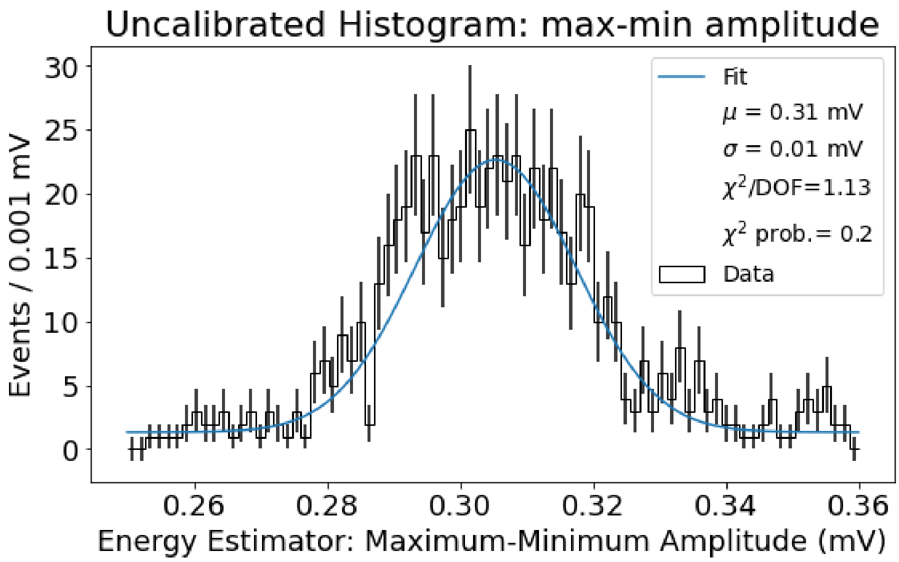

Figure 5: Uncalibrated histogram of maximum-minimum amplitudes of traces from the calibration data file and a gaussian fit. Visually, the distribution looks like a gaussian as expected. The number of bins used are 80. Some of the data is neglected in the histogram in favour of a good fit.

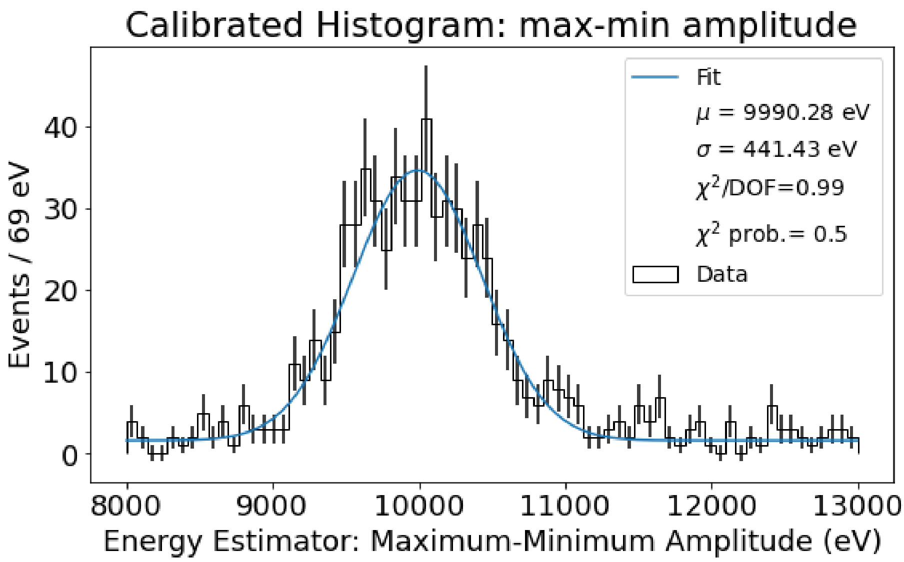

Figure 6: Calibrated histogram of photon energies corresponding to maximum-minimum amplitudes of traces from the calibration data file. A gaussian is fitted with a good reduced chi-squared value of 0.99 and chi-squared probability of 0.5 which means there is a 50% chance our fitted model is a good indicator of underlying distribution of calibration source.

From figure 5, we see that reduced chi-squared $\chi^2_{red} = 1.13$ with χ^2 prob. $= 0.2$. This indicates that there is only 20% chance of calculated mean $\mu_{1,\text{ uc}} = 0.31$ mV is a good value which would enable us to convert the distribution to the expected energy spectrum of a narrow gaussian with $\mu = 10$ keV.

On the other hand, from figure 6, the chi-squared probability is really good ($= 0.99$) because it is very close to 1. This indicates that the model is a good fit for the underlying distribution of calibration source. However, the mean of the calibrated data is not exactly $= 10$ keV, which I suspect has to do with errors in operations with floating point numbers in python and the consequent rounding errors. The results of the calibration are summarised in 1.

4.2 Energy Estimator: Max - baseline Amplitude

The second way to estimate the amplitude of a trace is to subtract the mean of the pre-pulse region of each trace from its maximum. Since the background noise in the pre-pulse region of the corresponding trace is likely to be constant across the trace which only lasts for about $4\mu s$, it is a good measure of the error in each event in the trace. Therefore, subtracting it from the maximum should give a more accurate sense of trace amplitude. Thus, the max-baseline amplitudes were calculated for each trace in the calibration data to create a histogram (see figure 7). Again, it was fitted with a gaussian to find the uncalibrated mean as $\mu_{2,\text{ uc}} = 0.24$ mV. The number of bins were chosen to be $n = 45$ by trial-error to find the closest reduced chi-squared value to 1 for the calibrated data. Then, the calibration factor was calculated as

$$f_2 = \frac{10^4 \text{ eV}}{\mu_{2,\text{ uc}}} \approx 41,643 \frac{\text{eV}}{\text{mV}} \approx 42,000 \frac{\text{eV}}{\text{mV}} \tag{7}$$

Multiplying this calibration factor f_2 to the max-baseline amplitude data, we get a calibrated histogram with energy in eV on the x-axis and counts on the y-axis (see figure 8)

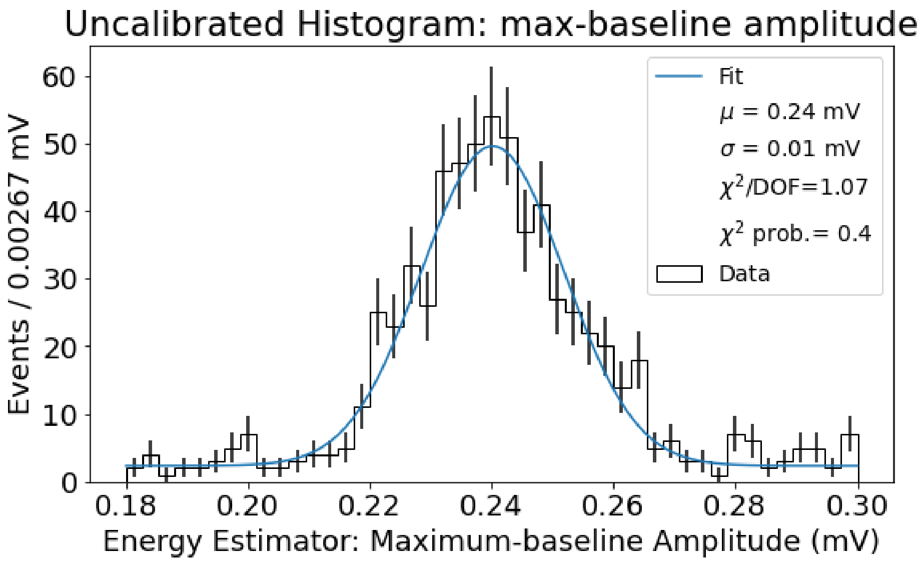

Figure 7: Uncalibrated histogram of maximum-baseline amplitudes of traces from the calibration data file and a gaussian fit. The number of bins used are 45. Some of the data is neglected in the histogram in favour of a good fit. The reduced chi-squared value of 1.07 and chi-squared probability of 0.4 mean that the calculated mean $\mu_{2,\text{ uc}} = 0.24$ mV is quite accurate to yield us an accurate calibration factor to get our calibrated data of 10 keV gaussian peak as expected.

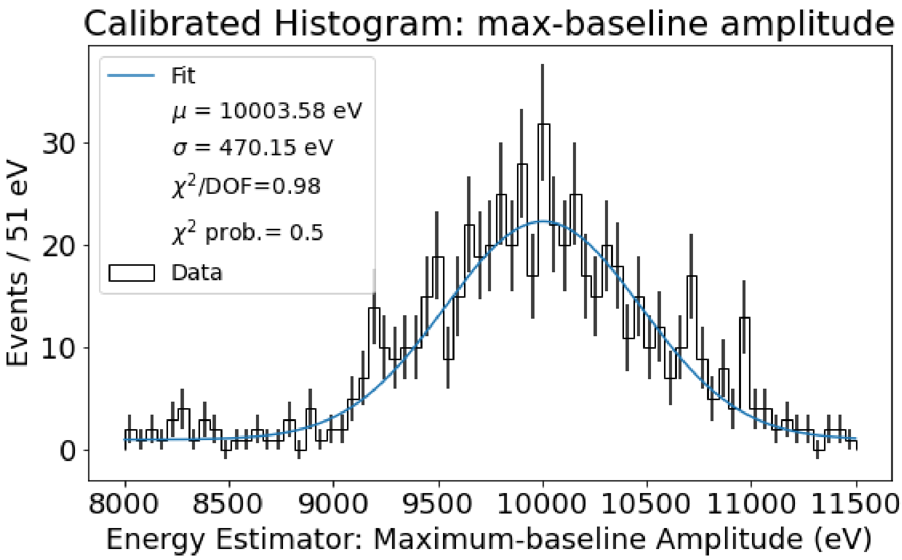

Figure 8: Calibrated histogram of photon energies corresponding to maximum-baseline amplitudes of traces from the calibration data file. A gaussian is fitted with a good reduced chi-squared value of 0.98 and chi-squared probability of 0.5 which means there is a 50% chance our fitted model is a good indicator of underlying distribution of calibration source.

For this method of amplitude calculation, the $\chi^2_{red} = 1.07$ for uncalibrated data in figure 7 indicates that the mean $\mu_{2,\text{ uc}}$ would provide a better calibration factor because the distribution fits very well with the fitted gaussian. This means that the calibrated data in figure 6 will be more accurately fitted by a gaussian with $\mu = 10$ keV. This is verified by observing that the mean of the gaussian fitted to the calibrated data set in 6 has a mean of $\mu_{(2,c)} = 10003.58$ eV ≈ 10 keV with a $\chi^2_{red} = 0.98 \approx 1$. Both of these things indicate that this energy estimator is really good. The results of the calibration are summarised in 1.

4.3 Energy Estimator: Pulse-fit Amplitude

The last method I tried to estimate the amplitude of a trace is by fitting each trace in the calibration data to a pre-defined pulse shape. The functional form of a trace pulse of amplitude A is given by the equations 1 and 2. Thus, to fit it, I first calculated the averaged standard deviations of all traces in the noise data and used that as the uncertainty in for each voltage measurement in the calibrated data traces. This way, it took into account the fact that the value of each voltage would be in the 2 averaged standard deviations range about 68% of the time. Then, using these uncertainties and the voltage data for each trace, the best fit amplitudes were calculated for all the 1000 events.

The corresponding histogram is shown in figure 9. Again, it was fitted with a gaussian to find the uncalibrated mean as $\mu_{3,\text{ uc}} = 0.21$ mV. The number of bins were chosen to be $n = 50$ by trial-error to find the closest reduced chi-squared value to 1 for the calibrated data. Then, the calibration factor was calculated as

$$f_3 = \frac{10^4 \text{ eV}}{\mu_{3,\text{ uc}}} \approx 46,743 \frac{\text{eV}}{\text{mV}} \approx 47,000 \frac{\text{eV}}{\text{mV}} \tag{8}$$

Multiplying this calibration factor f_3 to the pulse fitted amplitude data, we get a calibrated histogram with energy in eV on the x-axis and counts on the y-axis (see figure 10)

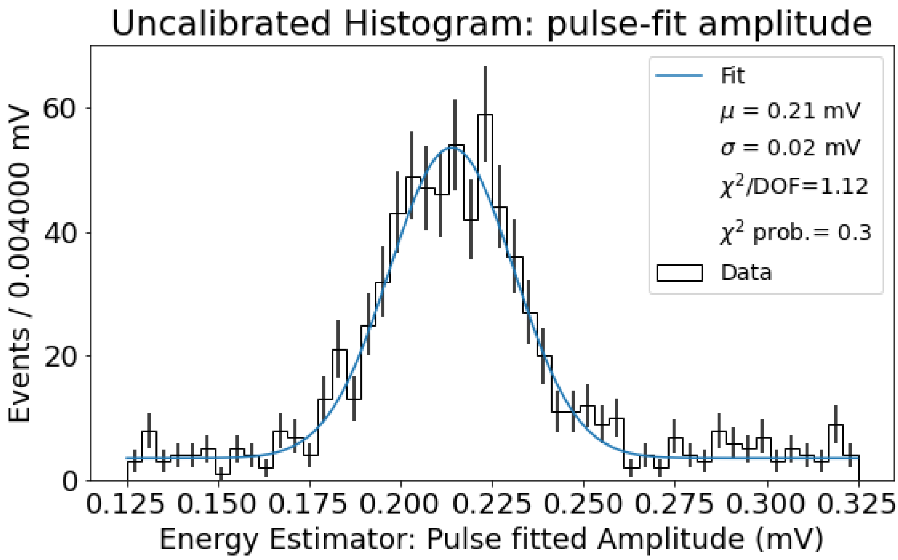

Figure 9: Uncalibrated histogram of pulse fitted amplitudes of traces from the calibration data file and a gaussian fit. The number of bins used are 50. Some of the data is neglected in the histogram in favour of a good fit. The reduced chi-squared value is 1.12 and chi-squared probability is 0.3 for the calculated mean $\mu_{3,\text{ uc}} = 0.21$ mV

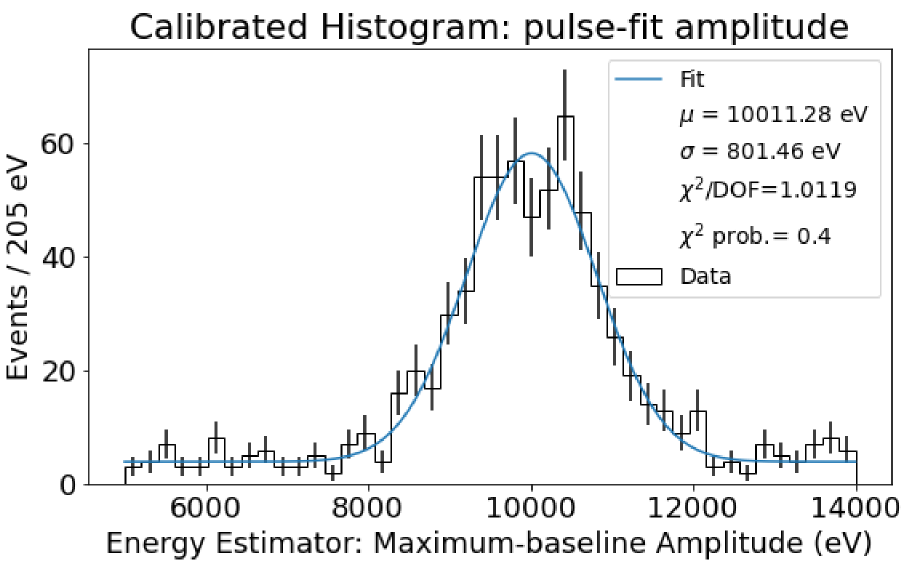

Figure 10: Calibrated histogram of photon energies corresponding to pulse fitted amplitudes of traces from the calibration data file. A gaussian is fitted with an excellent reduced chi-squared value of 1.01 and chi-squared probability of 0.4 which means there is a 40% chance our fitted model is a good indicator of underlying distribution of calibration source.

From the calibrated data in figure 6, we can see that it is very accurately fitted by a gaussian with $\mu_{\text{ 3, c}} = 10011\text{eV} \approx 10$ keV. The confidence in the pulse fitted amplitude being a very good estimator is increased while observing a $\chi^2_{red} = 1.01 \approx 1$ which is the closest χ^2_{red} we have got so far. However, it must also be noted that the resolution of the detector under this energy estimator is about $\sigma = 800$ eV which needs to be taken into consideration while finalizing the estimator.

Now, let's investigate the integral or area under the pulse estimators.

4.4 Energy Estimator: Area under entire range

Since the voltage measurements are separated by 1μs, we can perform a simple integral by summing all the 4096 voltage measurements. The resulting discrete integral would have units mV$\cdot\mu s$. The corresponding histogram is shown in figure 11. It was fitted with a gaussian to find the uncalibrated mean as $\mu_{4,\text{ uc}} = 28.17$ mV$\cdot\mu s$. The range of histogram was limited to $(-100, 150)$ mV$\cdot\mu s$ because the base of the gaussian was almost 0 and including more data points deviated the χ^2_{red} value and the χ^2 probability from their ideal numbers. The calibration factor was calculated as

$$f_4 = \frac{10^4 \text{ eV}}{\mu_{4,\text{ uc}}} \approx 355 \frac{\text{eV}}{\text{mV}\cdot\mu s} \tag{9}$$

Multiplying this calibration factor f_4 to the summed integral data, we get a calibrated histogram with energy in eV on the x-axis and counts on the y-axis (see figure 12)

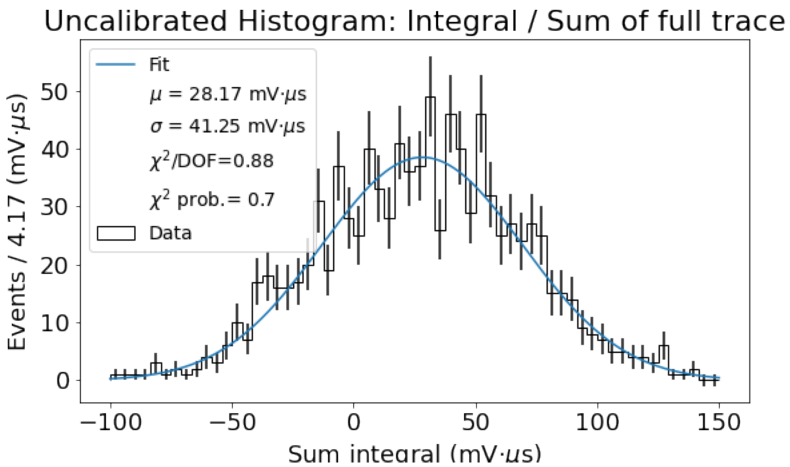

Figure 11: Uncalibrated histogram of integral under the pulse calculated by summing all the voltage values in a given trace. The number of bins used are 60. Some of the data is neglected in the histogram because the base of the gaussian was nearly 0. The reduced chi-squared value is 0.88 and chi-squared probability is 0.7 for the calculated mean $\mu_{4,\text{ uc}} = 28.17$ mV$\cdot\mu s$

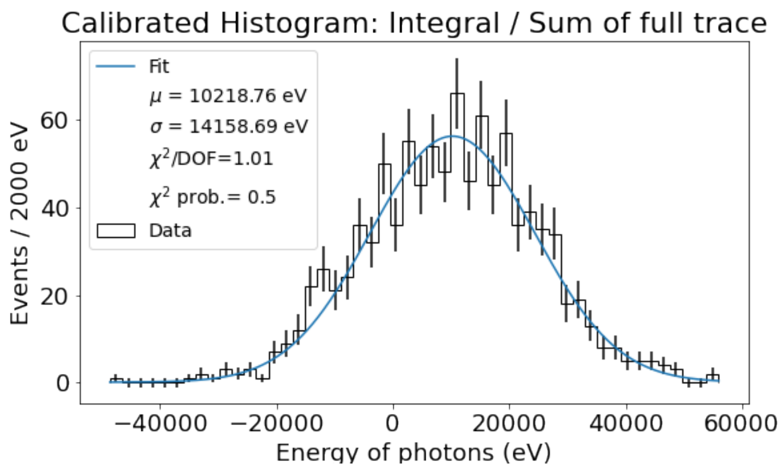

Figure 12: Calibrated histogram of photon energies corresponding to sum over traces from the calibration data file. A gaussian is fitted with an excellent reduced chi-squared value of 1.01 and chi-squared probability of 0.5 which means there is a 40% chance our fitted model is a good indicator of underlying distribution of calibration source.

Note from figure 12 that although the χ^2_{red} is close to 1 by like 1%, the resolution of the detector is around $\sigma \approx 14 \cdot 10^3$ mV$\cdot\mu s$ which is way higher than what we need. Thus this is not a good energy estimator.

4.5 Energy Estimator: Area under entire range - baseline

A similar integral can be calculated for the entire trace's range by just subtracting the average baseline, which is the average of the pre-pulse region, from each voltage point in a given trace. The corresponding histogram is shown in figure 13. The uncalibrated mean was $\mu_{5,\text{ uc}} = 27.20$ mV$\cdot\mu s$. The range of histogram was limited to $(-20, 80)$ mV$\cdot\mu s$ because the gaussian approaches 0 from those points in either direction. The calibration factor was calculated as

$$f_5 = \frac{10^4 \text{ eV}}{\mu_{5,\text{ uc}}} \approx 367 \frac{\text{eV}}{\text{mV}\cdot\mu s} \tag{10}$$

Calibrating the summed integral data, we get a calibrated histogram with energy in eV on the x-axis and counts on the y-axis (see figure 14)

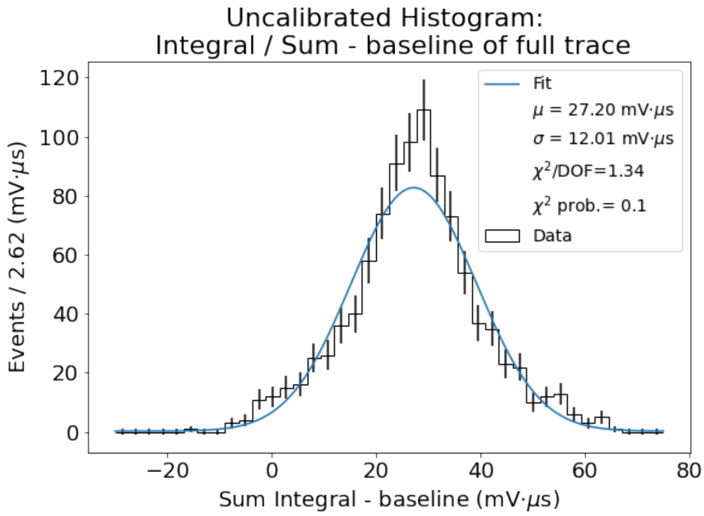

Figure 13: Uncalibrated histogram of integral under the pulse minus the pre-pulse average baseline for a given trace. The number of bins used are 40. The reduced chi-squared value is 1.34 and chi-squared probability is 0.1 for the calculated mean $\mu_{4,\text{ uc}} = 27.20$ mV$\cdot\mu s$

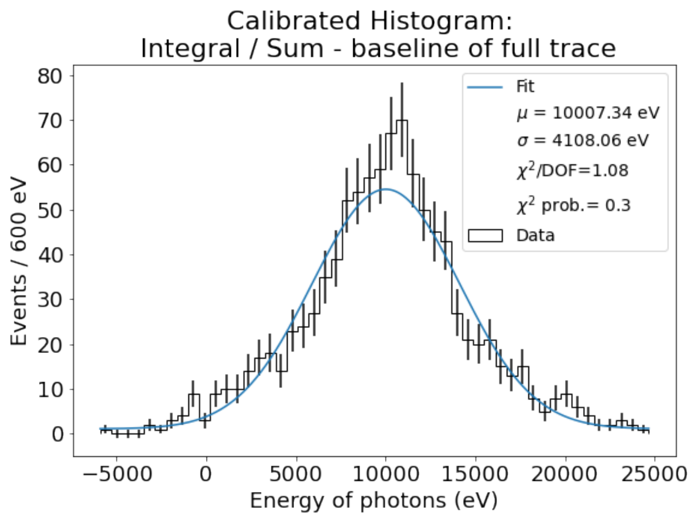

Figure 14: Calibrated histogram of photon energies corresponding to sum over traces minus the pre pulse baseline from the calibration data file. A gaussian is fitted with a reduced chi-squared value of 1.08 and chi-squared probability of 0.3 which means there is a 30% chance our fitted model is a good indicator of underlying distribution of calibration source.

Again, note from figure 14 that the energy resolution of the detector under this estimator is too high.

4.6 Energy Estimator: Area under limited range

Finally, the integral can be calculated for a specified range in all the traces so that we can ignore the effect of random noise on our integral away from the main pulse. I decided to use the range $(900\mu s, 1200\mu s)$ because the onset of the pulse is at $1000\mu s$ and each pulse more or less approaches the quiescent voltage by about $1200\mu s$. The voltages in this range for each trace were summed up and the results are shown in the histograms 15 and 16 where we have used the calibration factor for the uncalibrated mean $\mu_{6,\text{ uc}} = 24.33$ mV$\cdot\mu s$:

$$f_6 = \frac{10^4 \text{ eV}}{\mu_{6,\text{ uc}}} \approx 411 \frac{\text{eV}}{\text{mV}\cdot\mu s} \tag{11}$$

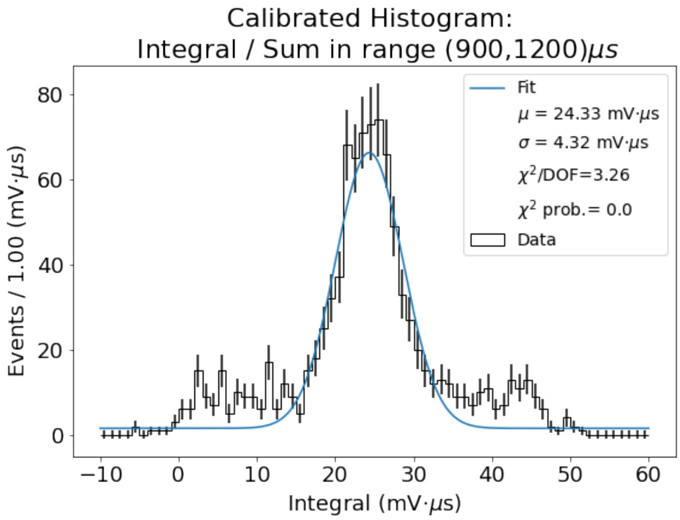

Figure 15: Uncalibrated histogram of integral under the pulse in the range $(900\mu s, 1200\mu s)$ for each pulse. The number of bins used are 70. The reduced chi-squared value is 3.26 which makes this model very unfit to descrbe our distribution. The chi-squared probability is 0 implying our model is completely non-ideal

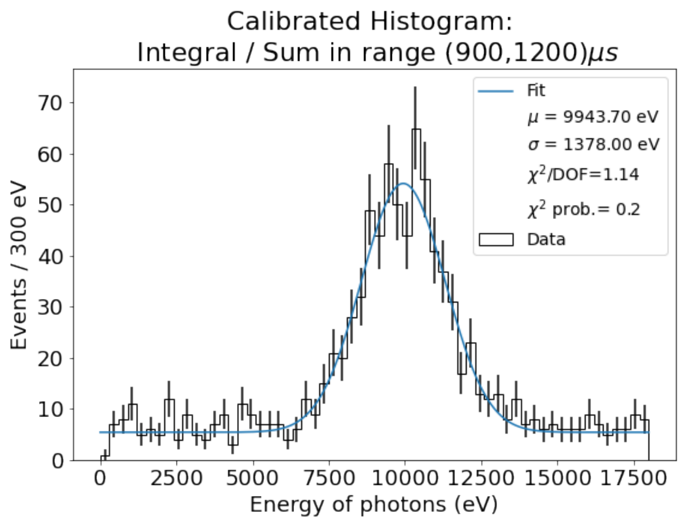

Figure 16: Calibrated histogram of photon energies corresponding to integral under the pulse in the range $(900\mu s, 1200\mu s)$ for all traces in calibration file. A gaussian is fitted with a reduced chi-squared value of 1.14 and chi-squared probability of 0.2 which means there is a 20% chance our fitted model is a good indicator of underlying distribution of calibration source.

Here, it is quite ambiguous whether this fit is good or not because the uncalibrated distribution has a really high reduced chi-squared value of 3.26 while the calibrated distribution has about 1.34. I think this integral is not a good energy estimator because it fails to fit a sharp gaussian model as seen in figure 15.

Npw that we have seen all energy estimators, we can tabulate their χ^2_{red}, χ^2 probabilities, and energy resolutions σ to find the energy estimator which was the most accurate in modelling our known calibration data as a narrow gausssian peak at $\mu = 10$ keV. This details are shown in table 1

Table 1: Summary of the qualities of all 6 energy estimators. The table lists the name of the energy estimators used, the calibration factors derived, the energy resolution of the detector under the estimators, the χ^2 probabilities, and χ^2_{red}

Energy Estimator	**Calibration Factor**	**Energy Resolution in eV)**	χ^2 **probability**	$\boldsymbol{\chi^2_{\mathrm{red}}}$
max-min amplitude	$33000\ \frac{\text{eV}}{\text{mV}}$	441	0.5	0.99
max-baseline amplitude	$42000\ \frac{\text{eV}}{\text{mV}}$	470	0.5	0.98
pulse fit amplitude	$47000\ \frac{\text{eV}}{\text{mV}}$	801	0.4	1.01
Sum over entire range	$355\ \frac{\text{eV}}{\text{mV}\cdot\mu s}$	14158	0.5	1.01
Sum - baseline	$367\ \frac{\text{eV}}{\text{mV}\cdot\mu s}$	4108	0.3	1.08
Sum over limited range	$411\ \frac{\text{eV}}{\text{mV}\cdot\mu s}$	1378	0.2	1.14

As can be seen from table 1, the properties of all the 6 energy estimators vary quite a bit. We can clearly observe that the integral / area based energy estimators have much more detector energy resolution than the amplitude based energy estimators. Having a lower energy resolution is more necessary for a good energy estimator because our calibration data should have narrow gaussian peak. Having higher energy resolutions means the gaussains fitted to the calibrated distributions are broader than expected. Thus, we must choose one of the amplitude based energy estimators. Among those three, the max-min amplitude estimator is the best because it has the lowest energy resolution of about $\sigma = 441$ eV, a good χ^2 probability of 0.5 indicating there is a 50% chance the gaussian fitted to the calibrated data is a good fit for the distribution of photon energies deduced from the max-min amplitude. Finally, it also has a reduced chi-squared valuen of $\chi^2_{red} = 0.99$ which differs by the ideal value of 1 by just 1%. Due to these reasons, I decided to use the max-min amplitude as the energy estimator to analyze the real signal data.

The max-min energy estimator was applied to the signal data and the corresponding historgram was generated as shown in figure 17

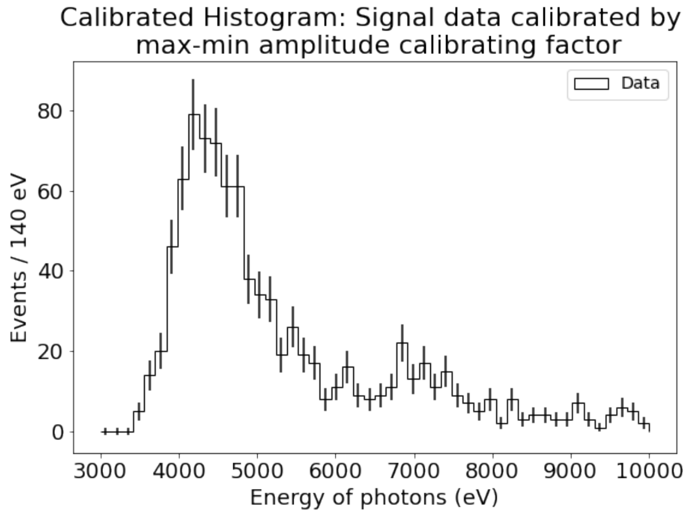

Figure 17: Calibrated histogram of photon energies corresponding to signal traces. The data was calibrated using the max-min amplitude estimator. Clearly, it shows a functional form of an increasing exponential followed by a gaussian and followed by a decreasing expoenential of lower rate.

Given that the above energy spectrum shows a functional form appears to have a trend of an increasing exponential followed by a gaussian and followed by a decreasing expoenential of lower rate, I decided to fit that function to it and plotted it. This is shown in figure 18

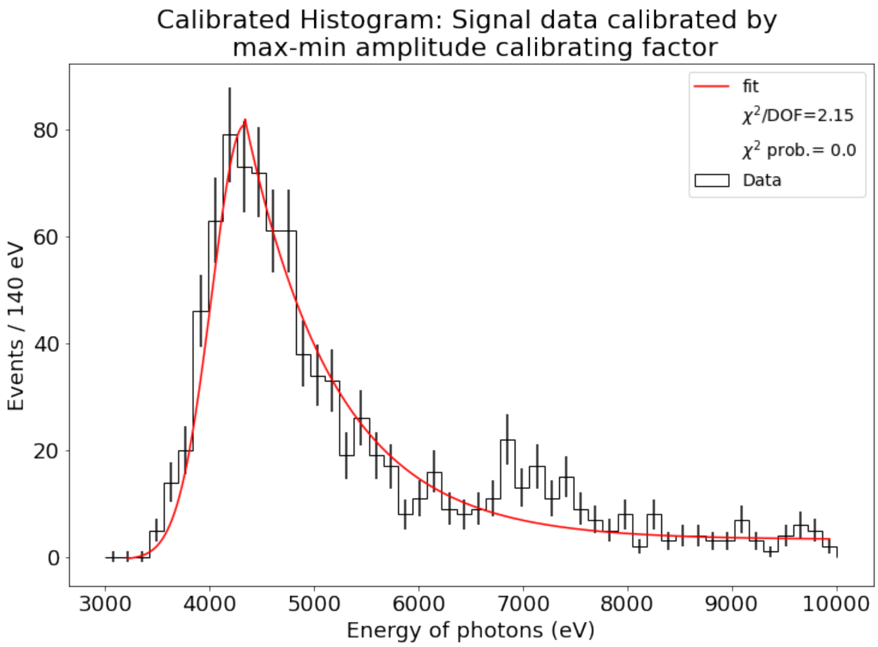

Figure 18: Calibrated histogram of photon energies corresponding to signal traces with an increasing exponential followed by a gaussian and then a decreasing exponential plotted to show how the functional form depends on these functions.

The fitted functional form calculated was hence:

$$\text{fit }(x) = e^{0.0088*(x-2000)} \tag{12}$$

$$+81 * e^{-(x-4336)**2/(2*319**2)} - 0.46$$

$$+78.7 * e^{-319*(x-0.012)} + 3.25$$

As can be seen from the graph, the reduced chi-squared value for this functional fit to the signal data is $\chi^2_{red} = 2.15$ which although is greater than 1 and shows an under fit, it is quite close to it. Thus, the functional form described in equation 12 is a good model. Clearly, because of the inaccuracy of the fit, the χ^2 probability is 0.0 which also makes sense as this is only an approximate model. Further, as can be seen from equation 12, the amplitude of the gaussian is almost double of what it should be. This arises from the sensitivity and difficulty of setting the $p0$ parameter in python's curve fit.

5 Conclusion

We analysed different kinds of energy estimators including amplitude and integrals under the traces in the calibration file to choose the best energy estimator as the max-min amplitude. This was decided based on statistical properties such as reduced chi squared and chi squared probabilities. Finally, we applied this to the traces in the signal file which consists of photon detections from an unknown source. The distribution for these signal traces was calibrated using the calibration factor of 33000 $\frac{\text{eV}}{\text{mV}}$ found for the max-min amplitude. It was finally observed that the energy spectrum of the signal data follows a functional form of an increasing exponential followed by a gaussian and followed by a decreasing expoenential of lower rate. This had a $\chi^2_{red} = 2.15$ which shows that the functional form is not that bad.

6 References

1. PHY324 Data Analaysis Project Instrunctions Manual (2023), Author: PHY324 faculty
2. PHY324 Data Analysis Slides (2023), Author: PHY324 faculty