PHY324 - Data Analysis for Simulated Particle Detector

Chaitanya Kumar Mahajan

February 6, 2023

1 Introduction[1]

A typical particle detector sensor converts the incident particle's energy into a digital signal in a form of a sharp pulse, where the amplitude of peak indicates the energy of the particle. An ideal pulse in this detector has a 1ms pre-pulse region and a sharp continuous rise of $t_{\text{rise}} = 20\mu$s and gradual decrease of $t_{\text{fall}} = 80\mu$s. The form is then given by,

$$y(t) \propto \exp(-t/t_{\text{rise}}) - \exp(-t/t_{\text{fall}}) \tag{1}$$

And we assume that the amplitude of the pulse scaled linearly with energy of the particle.

However, as the detector is not ideal, noise in the data exists from various sources such as random temperature fluctuations, random/stray subatomic particles hitting the detector, vibrations around the system or even from the internal circuit in the detector. We can see an example of a real-signal in 1, which shows 5 pulses from independent trials, layered together.

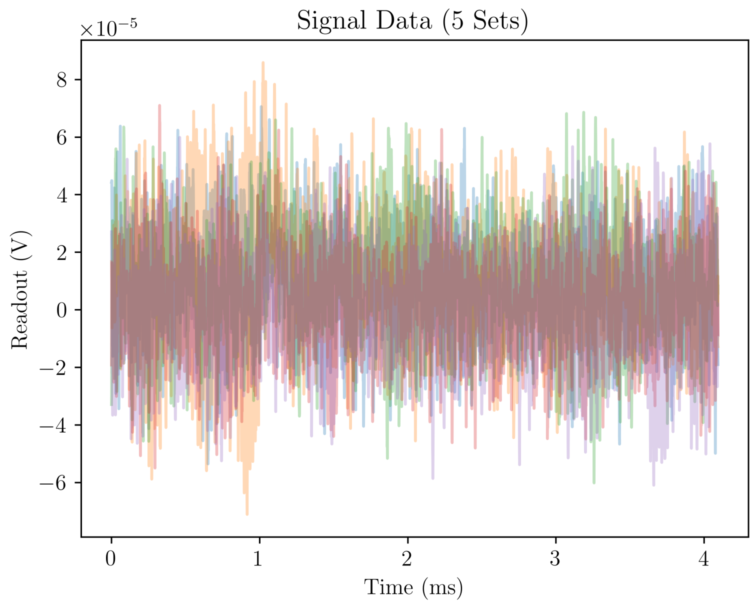

Figure 1: 5 noisy pulses from the calibration data-set used in this report. Each pulse is stored as 4096 voltage readings spanning 4ms. Note the sharp increase at $\sim$ 1ms and the shape of the pulse which is similar to equation 1.

Since the detector converts the particle energy into a voltage, we use a calibration data-set of known particle energy (in this report, $E = 10.00$keV), to convert the resulting voltage reading to the desired energy spectrum centered at the particle's energy, via "energy-estimators" (Section 1.1).

[1]Any equations and technical details is summarized from the PHY324 Data Analysis Lab Manual. https://q.utoronto.ca/courses/297235/files/24196590?module_item_id=4373299

In this report, we will demonstrate the process of analysing an unknown signal from a detector. In order to convert the voltage signals to an energy spectrum, we use an energy estimator to convert the signals into an voltage spectrum and multiply the spectrum by a conversion factor (determined by the energy estimator we chose).

1.1 Energy Estimator

Energy estimators are different methods to process a signal to estimate the energy of a signal. In this report we will use 6 estimators on the calibration data-set and determine the best estimator by χ^2 fitting the resulting histogram bins (where the bins follow a Poisson distribution, and are approximated by a Gaussian distribution) to a Gaussian distribution and calculating the energy resolution of the estimator. Here the resolution is the standard deviation of the Gaussian curve from its mean.

The following estimators will be used in this report.

1. Estimate the amplitude of the pulse via, `max(signal) - min(signal)`, where `max, min` are the global maximum and minimum of the signal.
2. Estimate the amplitude of the pulse via `max(signal) - baseline(signal)`, where `baseline` is the average pre-pulse value of the signal.
3. Estimate the amplitude of the pulse via integrating over the entire pulse.
4. Estimate the amplitude of the pulse via integrating over the entire pulse and subtracting the `baseline` of the signal.
5. Estimate the amplitude of the pulse via integrating over a small region around the pulse peak. (In this report, we integrate over index 950 to 1250)
6. Estimate the amplitude by χ^2 fitting the signal to an ideal pulse of form 1 with $t_{\text{rise}} = 20\mu\text{s}$ and $t_{\text{fall}} = 80\mu\text{s}$. For the uncertainty in χ^2 calculation, we use the average standard deviation of all 1000 noise data-sets.

2 Calibration

Given a 1000 calibration data-sets (a sample of 10 shown in 2). We use the above mentioned energy estimators to generate histograms of the amplitude values and overlay it with a Gaussian curve, where the parameters (μ, σ, amplitude, baseline) are found via `scipy.optimize.curve_fit`. We calculate the χ^2 and χ^2 probability of each estimator and summarize it below along with the energy resolution and calibration factor (which converts the voltage reading of each estimator to energy reading centered at 10.00keV).

Here we show the process of plotting one estimator (Estimator 6 - χ^2) remaining are plotted in Appendix. Given the calibration data of a known form, namely Equation 1 with appropriate rise and fall time and the amplitude as the constant of proportionality, we can

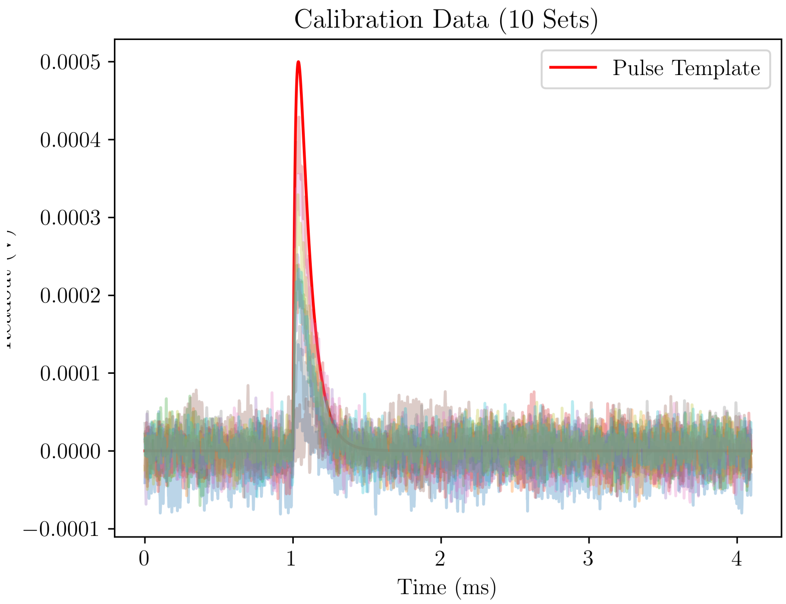

Figure 2: 10 noisy calibration pulse which come from a known 10.00keV particle. Each pulse is stored as 4096 voltage readings spanning 4ms, similar to the noise data-set. Overlaid with a scaled ideal pulse. Notice the sharp rise at 1ms and a gradual decrease of $t_{\text{fall}} = 80\mu$s. *Vertical axis reads - Readout (V)*

fit the pulse to the data using `scipy.optimize.curve_fit` to find the optimal amplitude parameter. Here we use the standard deviation of the average of each noise data-set, as the uncertainty in `curve_fit`.

Plotting the histogram of the amplitudes we χ^2 fit a Gaussian curve and calculate the mean, standard deviation, amplitude and baseline from `scipy.optimize.curve_fit`. We calculate the χ^2 value via,

$$\chi^2 = \sum_i^N \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2$$

and get $\chi^2 = 11.03$, where x_i is the measured value, μ_i is the mean of the i-th data-set and σ_i is the error in the i-th data-set. Given the number of bins in the histogram, the degree of freedom (DOF) is DOF = Num Bin − Num Params = 17. The CDF of $\chi^2(x = 11.03, k = 17) = 0.145$ and this gives a p-value of $p = 1 - 0.145 = 0.855 \approx 0.9$.

Plotting a histogram of the 1000 amplitude values, we see in figure 3, the process shown above. Notice the mean is $\mu = 0.21 \pm 0.02$mV where the uncertainty is the standard deviation of the fitted Gaussian curve. Since the calibration data-set had a particle of energy 10.00keV, we get a calibrating factor of,

$$\lambda = \frac{10.00\text{keV}}{0.21 \pm 0.02\text{mV}} \approx 47.0 \pm 4.5 \text{ keV/mV}$$

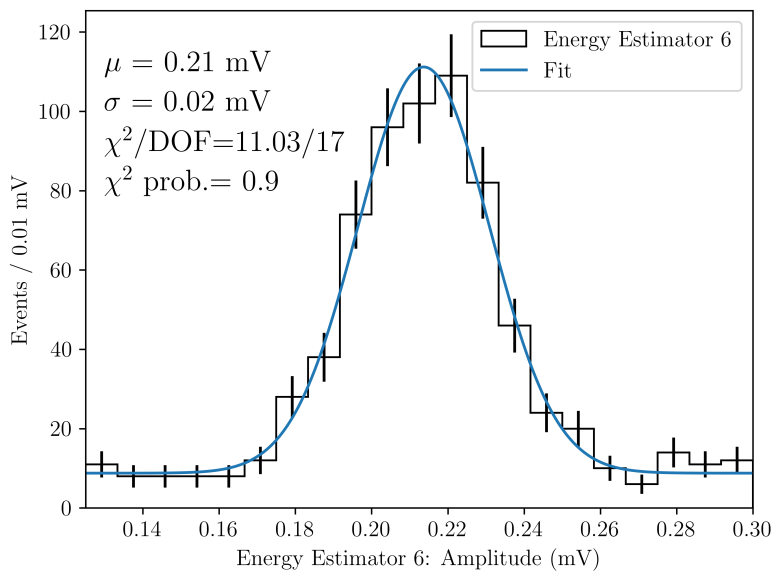

Figure 3: Histogram of χ^2 method to estimate amplitude of a 1000 data sets of 10.00keV calibration trials. Plotted with an optimized Gaussian curve of $\mu = 0.21$mV and $\sigma = 0.02$mV. The vertical uncertainty is $\sqrt{n}$ where n is the number of counts in each bin. We take this as the uncertainty since we assume each frequency is normally distributed.

Multiplying each amplitude by λ and plotting the resulting histogram and performing another χ^2 fit for a Gaussian curve, we get figure 4.

Working through the same process as above for the remaining estimators, we compile the results in table 1. The uncertainty in each estimator is calculated multiplying the percent uncertainty in the amplitude curve by the calibration factor in the energy spectrum. Notice

Energy Estimator	Calibration Factor (keV/mV)	Energy Resolution (keV)	Fit χ^2 Probability
Max - Min	32.78 ± 1.1	0.45	0.10
Max - Baseline	41.57 ± 1.7	0.50	0.10
Integral over Pulse	0.36 ± 0.55	15.16	0.80
Integral over Pulse - Baseline	0.37 ± 0.59	14.89	0.50
Integral over Pulse Region	0.39 ± 0.05	1.31	0.10
Chi-Square Fit	47.0 ± 4.5	0.86	0.60

Table 1: Six energy estimators (Section 1.1) used to calibrate the data-set along with the calibration factor, energy resolution and χ^2 probability of the fit.

that the Chi-Squared Fit has a p-value between 0.05 and 0.95 with an extremely fine energy resolution of 0.86keV. Based on these facts along with visual confirmation, we will use this estimator to process the signal in the next section.

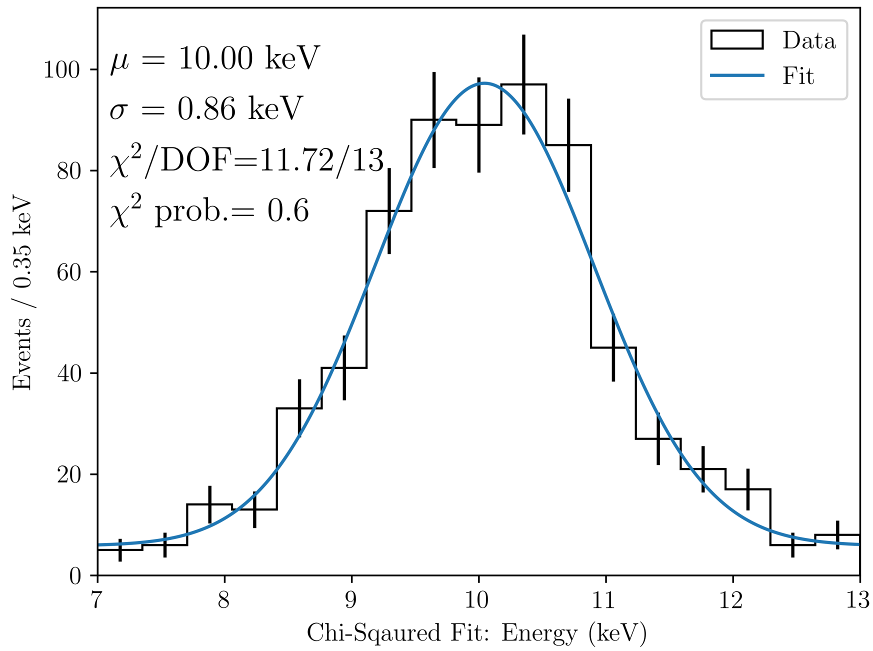

Figure 4: Calibrated energy estimator for χ^2 estimator on a 1000 data-sets of 10.00keV particles. Plotted with an optimized Gaussian curve of $\mu = 10.00$keV and an energy resolution of $\sigma = 0.86$keV. Along with the χ^2 probability of the fit where DOF = 13. The vertical uncertainty is $\sqrt{n}$ where n is the number of counts in each bin. We take this as the uncertainty since we assume each frequency is normally distributed.

3 Signal

Applying the calibration factor of $\lambda = 47.0 \pm 4.5$keV/mV to the signal data, we plot the histogram seen in figure 5. To account for the uncertainty in the calibration factor, we ensure that the bin sizes are relatively larger, in order for the range of values to remain in the same bin. i.e. the converted data up to the uncertainty remains in the same bin as it would have, if we ignored the uncertainty. Although this method is not ideal, it serves for this report.

In figure 5 we also plot a function of form,

$$y(t) = \frac{A}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2} + B \tag{2}$$

where A, B are the unit-less scaling factor and the baseline in keV respectively. Parameter k is the mean in keV and Γ is the gamma function. Notice that Equation 2 is a modified chi-squared distribution and has mean of k and variance of $2k$.[2]We chose a chi-squared distribution as it had the highest χ^2 probability for various numbers of bins out of the function forms we tried to fit, which included a Gamma distribution, Poisson distribution, Skew-Normal distribution and even sum of Gaussian distributions. Out of these forms, the best fit was the chi-squared distribution, and the optimal parameters were found to be,

$$A = 319 \pm 19, \quad k = 2.49 \pm 0.10 \text{ keV}, \quad B = 6.2 \pm 2.5 \text{ keV}$$

Since the variance is $2k$ the standard deviation is $\sigma = \sqrt{2k} \approx 2.23 \pm 0.04$keV. Notice the increased counts in the 5-7keV range, this may be an attribute of the non-negligible energy resolution of the chosen estimator, however more detailed analysis would be needed to confirm this, which we will not consider in this report.

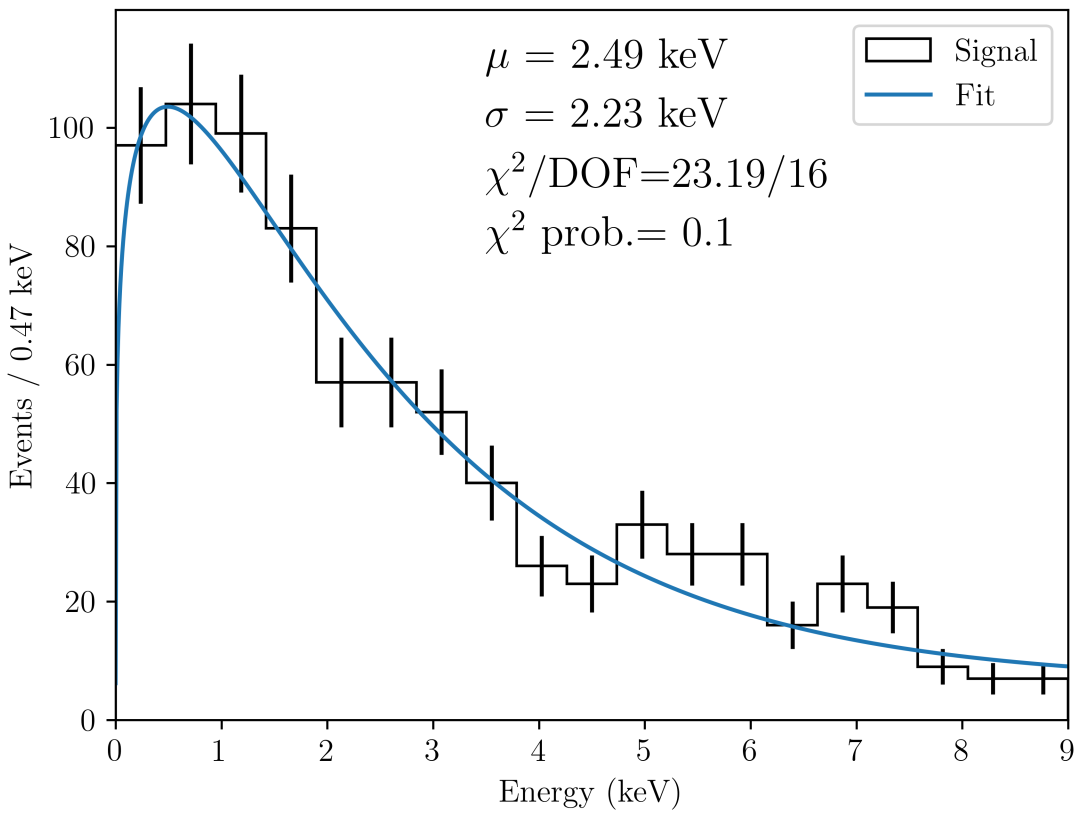

Figure 5: Energy spectrum of the signal data-set created via chi-squared estimator plotted along with a chi-squared distribution. When plotting the fit curve, we ignore the negative energies as they are non-physical and constituted of a negligible portion of the data-set. The vertical uncertainty is $\sqrt{n}$ where n is the number of counts in each bin. We take this as the uncertainty since we assume each frequency is normally distributed.

Thus the particle measured by the detector had energy 2.5 ± 2.3keV.

[2]Modern Mathematical Statistics with Applications 3ed - Jay L. Devore & Kenneth N. Berk & Matthew A. Carlton

4 Conclusion

As we see, data analysis is a useful and critical portion of particle detectors as there are various methods to convert a signal into an energy spectrum. In this report we considered six such estimators and chose the best one based on its energy resolution and p-value for a calibration data-set where the particle was known to have an energy of 10.00keV. Applying the calibration factor to an unknown signal, we found the corresponding energy spectrum and found that the chi-squared distribution had the best fit. Since the chi-squared is a known distribution, we can directly calculate its mean and standard deviation and found them to be 2.5 ± 2.3keV. The large uncertainty can be a factor of the uncertainty in the calibration factor or even due to the distribution we used.[3]

5 Appendix

Here we plot the voltage histograms for the calibration data-set for five energy estimators (χ^2 estimator shown in Section 2). Along side each we also plot the calibrated curve centered at 10.00keV.

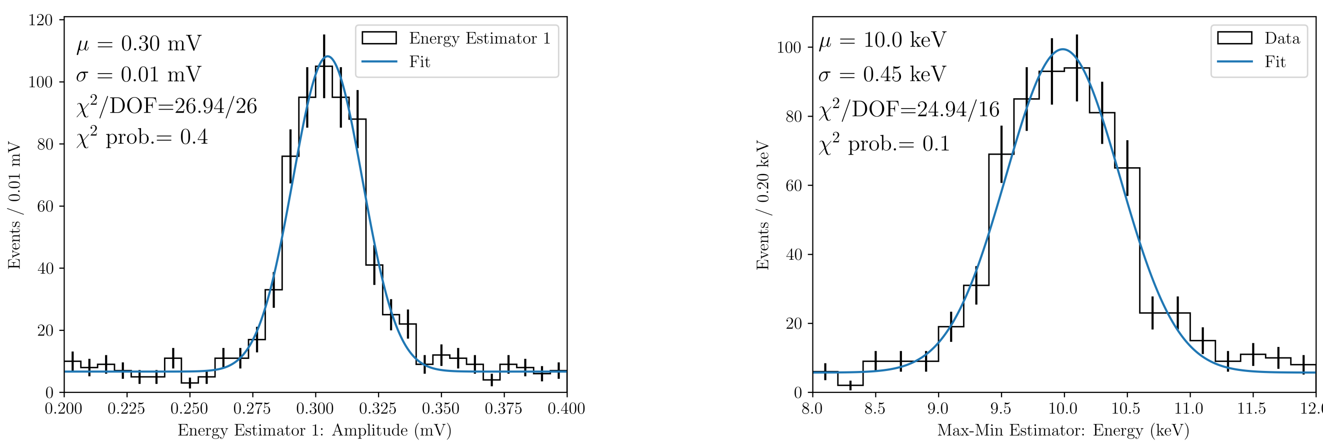

Figure 6: Initial voltage histogram for `max(signal) - min(signal)` estimator (Left) and its corresponding energy spectrum centered at 10.00keV with a resolution of 0.45 keV (Right). The vertical uncertainty is $\sqrt{n}$ where n is the number of counts in each bin. We take this as the uncertainty since we assume each frequency is normally distributed.

[3] Although this report used a chi-squared distribution, we found that a sum of 4-5 Gaussian distributions, each with its own mean and standard deviation, was also a close fit for the energy spectrum.

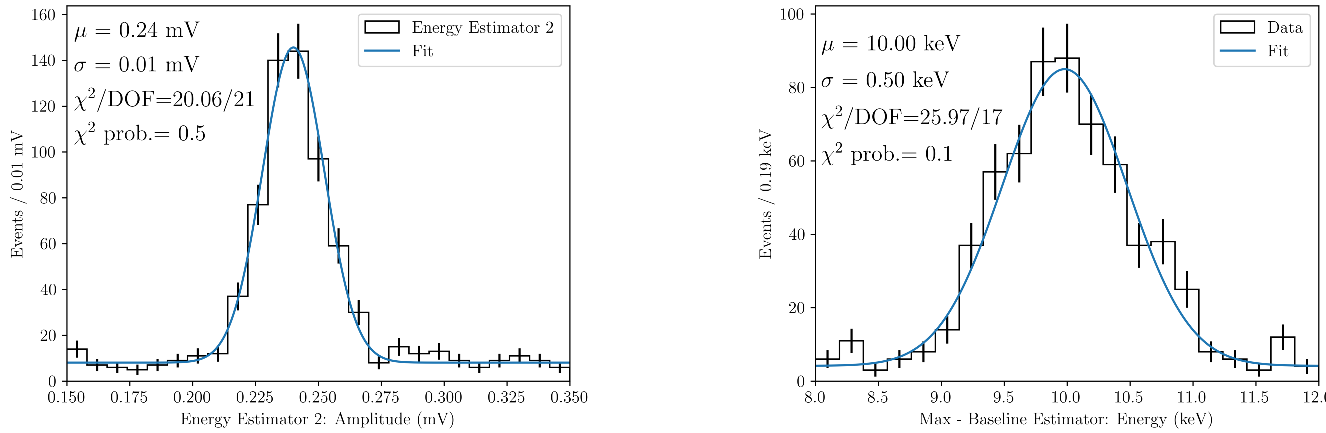

Figure 7: Initial voltage histogram for `max(signal) - baseline(signal)` estimator (Left) and its corresponding energy spectrum centered at 10.00keV with a resolution of 0.50 keV (Right). The vertical uncertainty is $\sqrt{n}$ where n is the number of counts in each bin. We take this as the uncertainty since we assume each frequency is normally distributed.

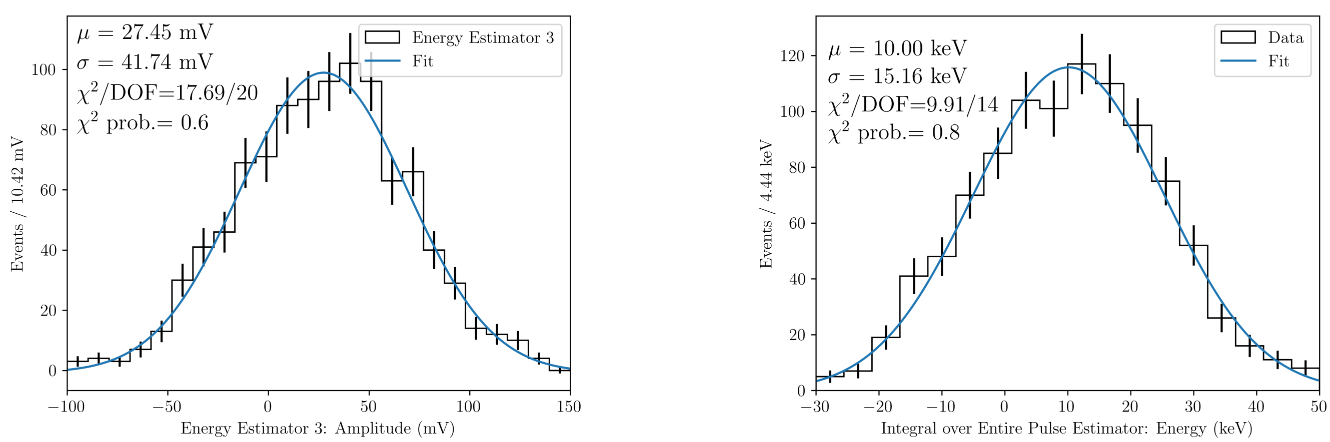

Figure 8: Initial voltage histogram for `sum(signal)` estimator (Left) and its corresponding energy spectrum centered at 10.00keV with a resolution of 15.16 keV (Right). The vertical uncertainty is $\sqrt{n}$ where n is the number of counts in each bin. We take this as the uncertainty since we assume each frequency is normally distributed.

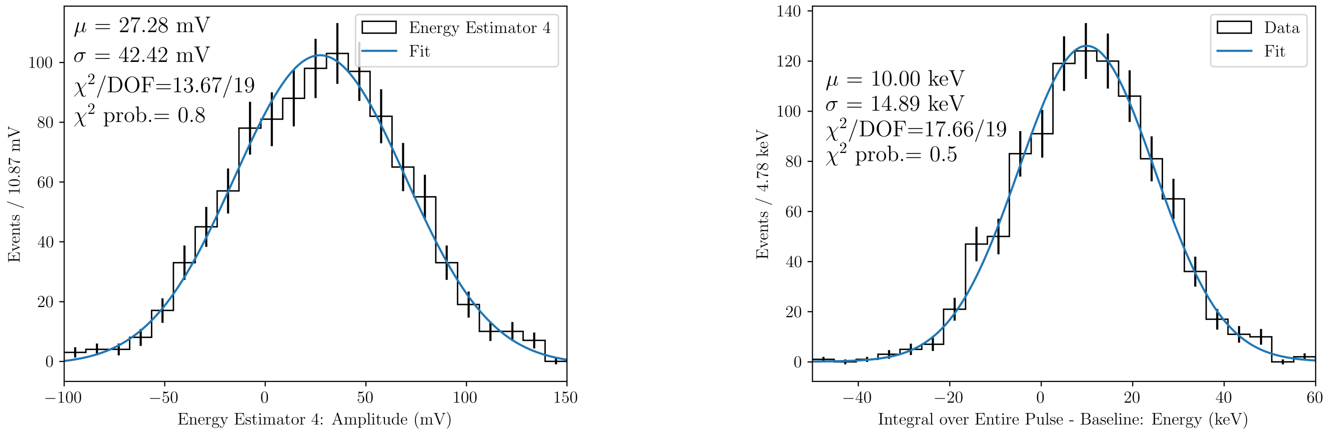

Figure 9: Initial voltage histogram for `sum(signal) - baseline(signal)` estimator (Left) and its corresponding energy spectrum centered at 10.00keV with a resolution of 14.89 keV (Right). The vertical uncertainty is $\sqrt{n}$ where n is the number of counts in each bin. We take this as the uncertainty since we assume each frequency is normally distributed.

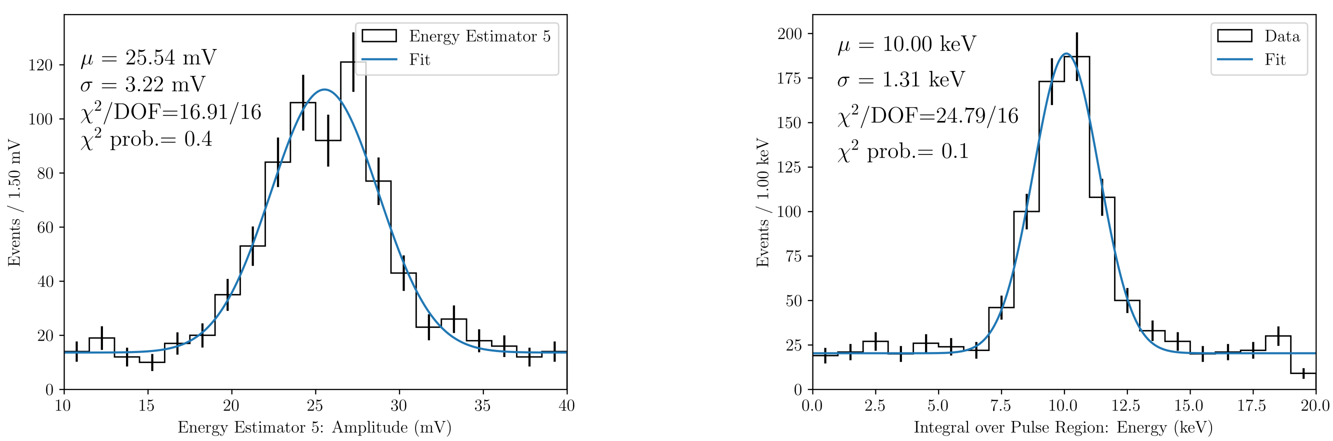

Figure 10: Initial voltage histogram for `sum(signal) - baseline(signal)` estimator (Left) and its corresponding energy spectrum centered at 10.00keV with a resolution of 1.31 keV (Right). The vertical uncertainty is $\sqrt{n}$ where n is the number of counts in each bin. We take this as the uncertainty since we assume each frequency is normally distributed.