PHY324: Using Energy Estimators to Determine the Energy Spectrum of a Light Signal Incident on a Voltage Detector

By: Gurmanjot Singh

Due: 20 January 2023

1 Introduction

The process of extracting a reliable and accurate energy spectrum from a signal is usually not discussed very precisely in the treatment of related theoretical subjects in physics. However, it can be a very arduous task, dependent on the nature of the signal and the signal's detector, and how we quantify the magnitude of energy in a signal.

In particular, the measurement of voltage signals as a result of energy deposition from incident particles–mainly photons and their absorption or scattering from the sensor's electrons–is of particular interest. For an idealized detector in this scenario, this corresponds to sampling every non-trivial pulse or perturbation's resulting voltage change over 4096 samples with a microsecond increment, due to a trigger system that detects the onset a millisecond prior.

The exact structure of this pulse, due to the nature of the experiment, is a spike and fall that occurs for 20 microsenconds and 80 microseconds respectively, given by a function of the form:

$$A(t) = A_0(e^{-t/(20\mu s)} - e^{-t/(80\mu s)})$$

for some normalization A_0, starting at $1000\mu s$ due to the idealized trigger system of our detector. Note that this specialized shape of the pulses will be important for purposes of energy estimation we will specify later.

This gives us a set of pulses corresponding to the complete signal. Based on how we quantify the "energy" of each pulse from the voltage, we can construct an energy spectrum for this signal as a histogram of this classification. if we then fit a Gaussian to this histogram, we get a rough estimate of the actual energy of the signal–as the Gaussian's mean value–and the resolution of this estimate–as the standard deviation of that Gaussian.

Based on how good this fit is in the chi-squared sense, we can then decide on the best way of extracting the energy data of the signal out of the voltage pulses, specifically for that signal. If we assume that all other signals of a similar type behave this way, we can take one signal as a calibration to this energy estimate, and then apply a scaling factor to make sure the observed voltage readout is the same as the known calibration energy.

We can then use that to measure the energies of other more interesting signals, as well as their spectra. This procedure is exactly the goal of our experiment, as will be outlined in the next section.

2 Methods

In this experiment, we will have a detector that has already pre-sampled 1000 photon pulse detections from a calibration source. It is important to note that these pulses have noise in them from other sources of voltage deposition, although it will be much smaller than the actual pulse amplitudes.

Not only that, but it is possible for some other sources of photons–due to radioactive matter and the such–could give rise to non-calibration pulses within this signal. In the experiment, we have a well-shielded apparatus to minimize the events that occur due to this. Hence, we assume that the number of samples in this data that are not calibration sourced are negligible.

With this data, we perform six different energy estimations of the pulses: (i) Max-min amplitude, (ii) Max-min amplitude from a chosen baseline, (iii) integration as a traced sum over the pulse data, (iv) the same integration but with a baseline, (v) integration in a limited range, (vi) fitting the pulse to the standard pulse shape.

For (i), we are simply taking the maximum deviation from 0 voltage to be the corresponding energy estimate, which is just the maximum of the absolute value of the data set.

For (ii), we do the same thing as in (i), but instead of from 0 voltage, we take an average baseline value of the pre-pulse and set that to be the origin.

For (iii), we take the average over the total pulse–like a time integral–which gives us another way of quantifying the "size" of the pulse.

For (iv), we do the same as in (iii), but once again apply the baseline difference to the signal and then take the integral as the estimate.

For (v), we instead restrict the domain of the integration, assuming that not all the data in the pulse is relevant, only the data near the actual spike. In this case, we take an average over a sample from 999μs to 1100μs as the approximate range in which the pulse stays visible over the noise (recall that the pulse goes up for 20μs and down for 80μs).

Finally, for (vi), we just perform a fit of each pulse to the expected pulse shape given by $A(t)$, with fitting parameter A_0. The parameter A_0 is then our effective estimate.

All of these will be done using Python code specifically designed to perform these calculations on the given calibration data of this experiment. In the resulting histograms, we will decide on the bin number and range in order to properly reflect the structure of the data.

Specifically, we find that the fitting procedure does not work if the range of the histogram is much larger than the actual range of the data, since then points where there is no data influence the actual fit. Not only that, but the outliers on the edges that are much less frequent than even the noise end up skewing the fit significantly, usually causing it to be approximated as flat or some other unjustified curvature (the Gaussian peaks downwards instead of upwards or doesn't peak at all). Hence, we will always set the range to traverse from the most consistent lowest reading to the most consistent highest reading.

As a special note, this will modify the fitting probabilities favourably, so it may not reflect the same result as if we were to just brute force fit the data when keeping outliers. This will help us tremendously when comparing estimation methods, as we will get the same result, just stronger. Not to say, the actual signal's analysis will be easier.

Moreover, we choose the bin number to make sure all the peaks are preserved and to minimize the variance or noise in the voltage readings. This is done by trial and error, just looking for a good number that does not reduce the quality of the data peaks while also getting rid of inconsistent pileup in data readings around those peaks.

We can then determine an energy estimator dependent calibration factor C which will scale these to convert into the energy domain, by setting the mean voltage of the Gaussian fit we perform to match up with the actual energy of the calibration signal, 10keV. Since the histogram itself should be identical after the scaling, just with a squished x axis, we really only need to know the calibration factor in each case. Note that the resulting mean (evidently) and standard deviation in the energy domain will also be scaled by C, which we will do when providing a table of results. This is important, as it does not give us anymore predictive power to translate the scale of the histograms we produce in the understanding of this experiment.

The uncertainties on these histograms event frequency is given by the square root of the bar size/frequency, as a standard way of defining error bars on histograms. The uncertainty in the voltage measurement is integrated into the noise, and thus is part of the fitting analysis on deciding which of these energy estimations provide the most accurate noise-independent energy spectrum that fits well into a Gaussian. That is, the resolution/standard deviation of the fit will be the effective uncertainty in the energy measurement.

Once we have fit all of the resulting voltage spectra to this Gaussian, we will look at the calculated energies and their uncertainties, as well as the chi-squared fit probabilities of each one. We then conclude that the one with the highest probability is the best estimate of the actual energy of similar type signals.

We then use this estimator on an unknown signal with a potentially non-Gaussian distributed energy spectrum, and determine a good functional form for this spectrum. The utility of this method will be discussed with the results in the next section.

3 Results

As previously mentioned, we will be estimating the energy spectrum of a calibration signal incident on our detector, a sample of which is given as below: We
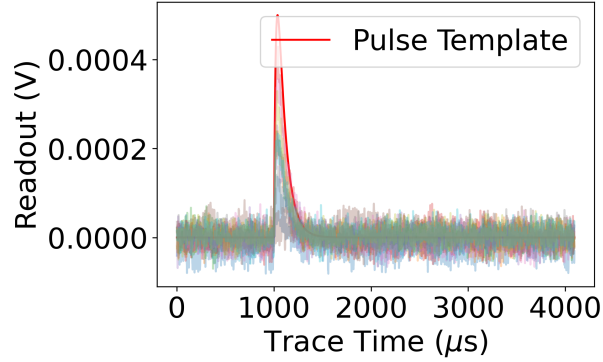

Figure 1: First 10 pulses in signal's profile with ideal pulse.

note that the approximate behaviour of pulses, ignoring the noise, indeed follows the curve of $A(t)$ with appropriate fitting. We now proceed to use the min-max energy estimator on this set of data, and extract the chi-squared probability, the calibration factor, and other quantities describing it: We see that this method
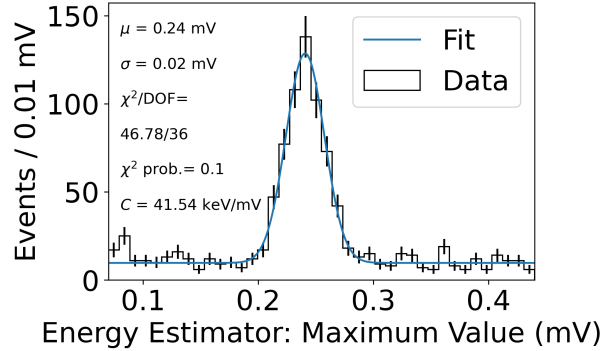

Figure 2: Result of max-min estimation.

results in a relatively clean peak, with a few unclean deviations from the flat background of the Gaussian. Because of this, the chi-squared probability is just 0.1. Other than that, it looks well-fit, with a chi-squared of around 48.8, although it could be closer to the degrees of freedom.

We now consider the same estimate but with a baseline average subtracted out of the data, giving us the following:
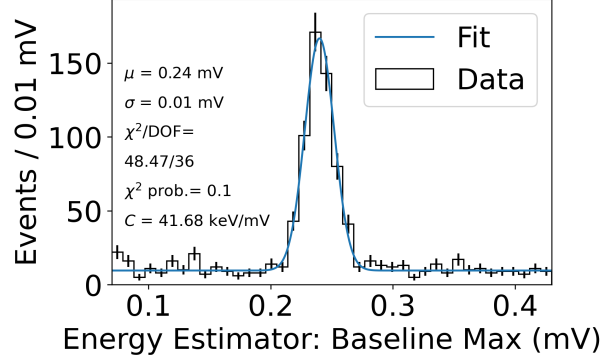

Figure 3: Result of max-min estimation with baseline.

This is not much better, with a very similar chi squared probability, given that the chi-squared is just slightly smaller at 48.5. However, it does show slightly better results, as the deviations from the flat background away from the peak are not as prevalent. We consider the more interesting case of integral estimation, as below:
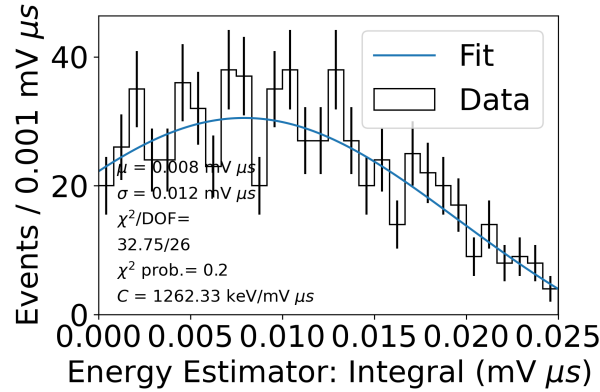

Figure 4: Result of integral estimation.

where some of the values have been cut off due to them being negative. This is an important detail, since the energy content cannot be negative, but due to the nature of this specific estimate, we have negative values as some signals have much more negative voltage noise than positive. Note also that we chose less bins as the data became too noisy otherwise.

Observing the method, we see that the scale of this estimator is many orders lower than that of the max-min estimators, since the integral done is over a very small microsecond time interval. This is reflected in the different units of the result.

Given this, the Gaussian seems to fit pretty well with the result, even having a higher chi squared probability of around 0.2. However, the standard deviation is very high, which will be amplified even more when multiplied by the calibration factor. Hence, although it is a better fit, the resolution will most likely be much worse than before.

On that note, it may be useful to instead consider integration after removing a baseline average from the signal, which gives us the following spectrum:
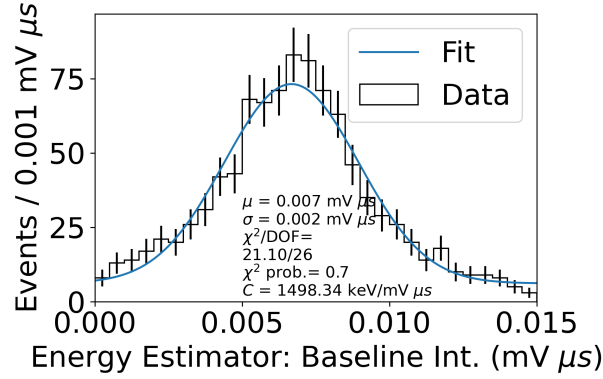

Figure 5: Result of integral estimation with baseline.

Note that there is much less noise, the peak is much more prominent and shifted into the positive spectrum, while the resolution is also higher as the deviation is lower. Although there is still a very large calibration factor, this is not as significant considering the chi squared probability is 0.7, much higher than before, as well as a chi squared value only a few integers below the degrees of freedom. This suggests slight overfitting, but still means that the fit is strong.

Last but not least out of the integration methods, the ranged integral provides the following spectrum estimate:
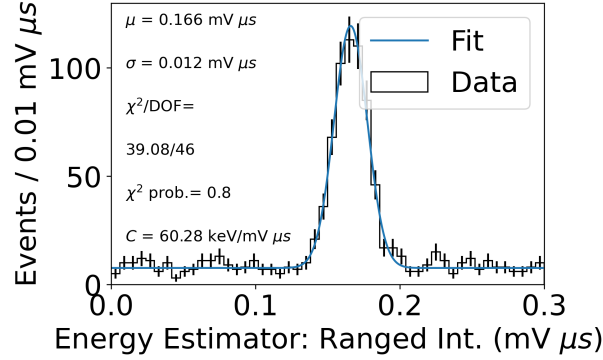

Figure 6: Result of integral estimate in $100\mu s$ range of pulse onset.

This surpasses the other two in precision, having a chi squared probability of 0.8 and a chi squared value relatively closer to the system's degrees of freedom. An added benefit comes from the standard deviation being smaller percentage wise, as well as the calibration factor, so the scaling will be much smaller. As a result, the actual resolution is much higher, suggesting that this is the most powerful out of the three integration methods in energy estimation.

Finally, we consider an alternative method we proposed by fitting the pulse to its ideal shape, which gives us the spectrum as below:
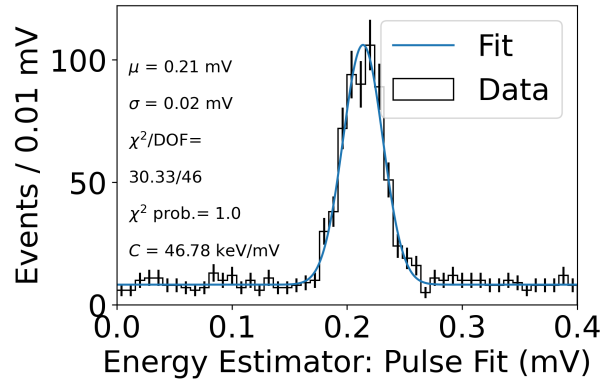

Figure 7: Result of pulse fitting estimation.

This has a chi squared probability of 1.0, the highest out of these. Although it is much more overfit than the others, it has a relatively small σ and C, meaning its resolution must be much higher. This accuracy can be observed in the

histogram, where the flat background of the Gaussian fits the surroundings of the peak very well, while the peak itself under the error bars also is fit very well.

We summarize the key elements of each estimation method in a table as below:

Energy estimator	Calibration factor (C)	Energy resolution $(\sigma \cdot C \text{ / keV})$	χ^2 probability
Max-min	41.54 keV/mV	0.69	0.1
Max-min with baseline	41.68 keV/mV	0.50	0.1
Integral	1262.33 keV/(mVμs)	14.78	0.2
Integral with baseline	1498.34 keV/(mVμs)	3.42	0.7
Integral in range	60.28 keV/(mVμs)	0.70	0.8
Pulse fitting	46.78 keV/mV	0.81	1.0

Out of these, pulse fitting provides the best fit to the calibration data's energy spectrum, while maintaining a relatively low deviation/resolution (high resolution in the qualitative sense) in relation to its other well-fitting competitors from the integral estimators.

We now proceed to use this energy estimator in the spectral analysis of an unknown signal of photons incident on our detector, with 1000 samples of pulses contained within it as well. To get an idea of the signal, just to make sure it is of the same type, we consider these pulses on a graph as below:
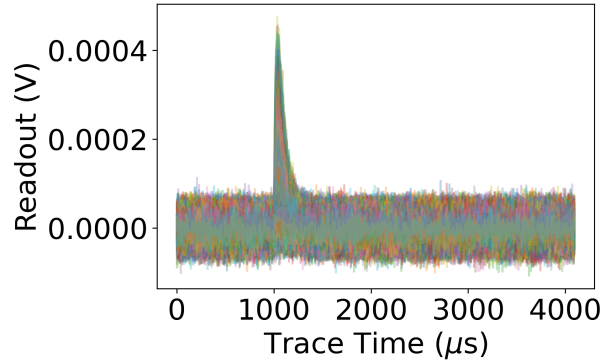

Figure 8: All 1000 pulses in signal, superposed on top of each other.

which evidently has the same shape, and is in the same unit range as our calibration. Hence, using pulse fitting as our energy estimation method is justified here. Note that here, we will perform the fitting while applying the calibration factor, so the resulting histogram will already be in the energy domain.

In hindsight, the data only is found in the energy interval from 0 to 20 keV, so we will restrict our domain of interest to be here. Moreover, due to the resolu-

tion of our energy estimator being 0.81 keV, we make our number of bins only so large as to let each bin contain a resolution interval, in which case we have chosen 20 bins to be ideal.

We find the data has a shape very close to that of an exponential $f(E) = Ae^{-\beta E} + c$ for some fitting parameters A, β, and c. After fitting it in this way, we get the following histogram, with the corresponding chi squared values and probabilities:
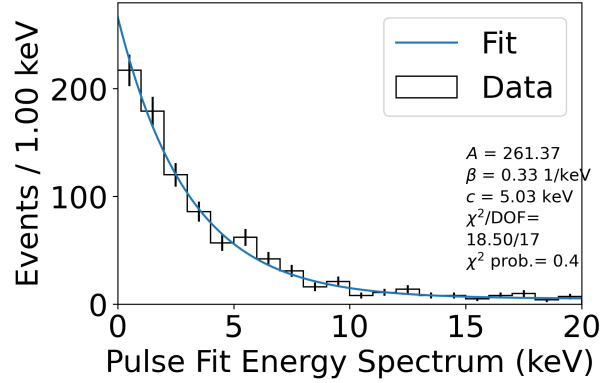

Figure 9: Pulse fit energy spectrum with the exponential fitting.

Observe that this fit is pretty good, considering $\chi^2 = 18.5$ is just 1.5 off from the actual degrees of freedom. Moreover, it is not overfit, suggesting that our data has enough detail for this fitting assumption to be justified. Given the chi squared probability of 0.4, it could do better, but it is still relatively accurate.

Hence, we have a functional form of our signal's energy spectrum, given by:

$$f(t) = Ae^{-\beta E} + c$$

with parameters $A = 261.37$, $\beta = 0.331/\text{keV}$, and $c = 5.03\text{keV}$, as a good model for the signal's approximate structure. Note that the constant term corresponds to background noise that always deposits some energy into the detector uniformly during the experiment, while the exponential term describes the signal's energy composition. This suggests that the light signal being detected has its energy distributed as a negative exponential.

4   Conclusion

In conclusion, through the comparison of several energy estimators in their chi-squared probabilities and resolutions on a calibration source, we concluded that the pulse fitting method of energy estimation had the most utility out of them, having the highest chi squared probability.

With this, we were able to successfully extract the energy spectrum of an unknown signal, and find a relatively good functional form for it as a negative exponential.

As an extension, we might consider a combination of the most useful energy estimators we considered, such as combining some form of baseline and pulse fitting in the case of more complicated ideal pulse forms for other detector types in a wide variety of experimental applications not just including photovoltaic behaviour.

Although it is beyond the scope of this report to consider possible physical interpretation of this signal, it may offer insight to find some connection with another subject. Particularly, noite that any system that obeys a negative exponential distribution is usually the result of a property or source of the system being in thermodynamic equilibrium. That is, a system in thermodynamic equilibrium obeys a Boltzmann distribution of energies.

If a large set of photons in our incident signal indeed behave in this manner, this may suggest that its source is in some form of thermodynamic equilibrium, as it emits photons in this energy distribution. Such a body that maintains thermodynamic equilibrium while emitting light resembles the behaviour of a black body, implying that this signal could possibly be the result of black body radiation incident on the detector.

This has a wide variety of possible sources (i.e., light bulb, sun, and so on), but this ability to determine a property of the signal's source shows the utility of this type of spectral analysis, especially in the way it is generalizable to more complicated systems.