MATHEMATICAL CHALLENGES IN HIGH-THROUGHPUT MICROCALORIMETER SPECTROSCOPY

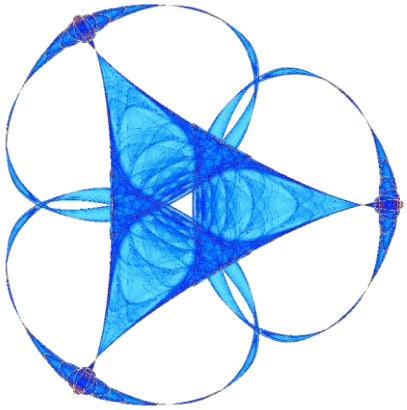
By

Vincent Morissette-Thomas, Alice Nadeau, Louis-Xavier Proulx, Yun Wei, Heng Zhu, and Jielin Zhu

Mentor: Bradley K. Alpert

IMA Preprint Series #2435

(September 2014)



INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS
UNIVERSITY OF MINNESOTA
400 Lind Hall
207 Church Street S.E.
Minneapolis, Minnesota 55455-0436
Phone: 612-624-6066 Fax: 612-626-7370
URL: http://www.ima.umn.edu

Mathematical Challenges in High-Throughput Microcalorimeter Spectroscopy

Mentor:
Bradley K. Alpert, National Institute of Standards and Technology

Participants:
Vincent Morissette-Thomas, Université de Sherbrooke
Alice Nadeau, University of Minnesota
Louis-Xavier Proulx, Université de Montréal
Yun Wei, University of Michigan
Heng Zhu, University of Calgary
Jielin Zhu, University of British Columbia

August 20, 2014

1 Introduction

In recent years, microcalorimeter sensor systems (Figure 1) have been developed at NIST, NASA, and elsewhere to measure the energy of single photons in every part of the electromagnetic spectrum, from microwaves to gamma rays. These microcalorimeters have demonstrated relative energy resolution, depending on the energy band, of better than 3×10^{-4}, providing dramatic new capabilities for scientific and forensic investigations. They rely on superconducting transition-edge sensor (TES) thermometers and derive their exquisite energy resolution from the low thermal noise at typical operating temperatures near $0.1\mathrm{K}$. They also function in exceptionally broad energy bands compared to other sensor technologies. At present, the principal limitation of this technology is its relatively low throughput, due to two causes: (1) limited collection area, which is being remedied through development of large sensor arrays; and (2) nonlinearity of detector response to photons arriving in rapid succession. Both introduce mathematical challenges, due to variations in sensor dynamics, nonstationarity of noise when detector response nears saturation, crosstalk between nearby or multiplexed sensors, and algorithm-dependent noise of multiplexing. Although there are certain inherent limitations on calibration data, this environment is extremely data-rich and we exploit data to attack two of these mathematical challenges.

1.1 Overview and mathematical model

A TES microcalorimeter is an exquisitely precise (low-noise) thermometer. Its response is nearly linear, for a limited photon energy range but pile-up of photon arrivals exacerbates nonlinearity. Microcalorimeter detectors are maintained to a temperature near absolute zero. As a photon hits one of the microcalorimeters in the array (Figure 2), the temperature of the microcalorimeter increases and consequently increases the resistance of the circuit. This increase of the resistance produces a drop in the TES current, which we correspondingly call a pulse. Figure 3 shows a series of pulses from the data stream. The height of each individual pulse enables the computation of the energy of each photon and the accuracy of the computed energy relies heavily on an accurate measure of each pulse height. The sensitivity of a microcalorimeter depends on the transition between its normal and superconducting phases. Figure 4 shows the transition phase for the Mo-Cu bilayer film. The temperature of the microcalorimeter is held around $96\mathrm{mK}$ for which a small change in temperature will lead to an significant change in the resistance.

Figure 1: Stage that accommodates up to 256 microcalorimeter detectors (on top) for the soft x-ray band and the associated multiplexing and readout electronics (sides). During operation, the pictured stage is nested inside a superconducting magnetic and radiation shield.

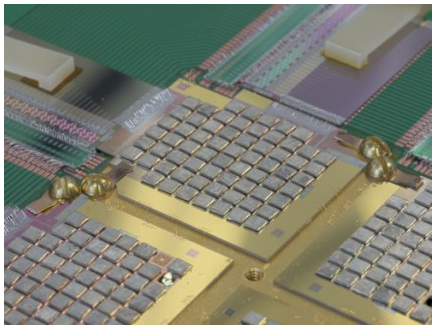


Figure 2: Close-up of an array of microcalorimeter detectors.

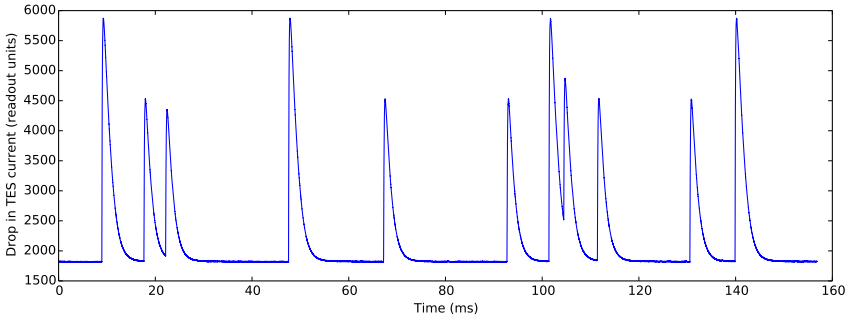


Figure 3: Stream of pulses representing a drop in the TES current.

Irwin and Hilton's [1] ODE model describes the evolution of the temperature, T, of the TES and the electrical current, I, through the TES over time, t:

$$C\frac{dT}{dt} = -P_{bath} + P_J + \delta(t - t_0) \cdot E \tag{1}$$

$$L\frac{dI}{dt} = V - I \cdot R_L - I \cdot R(T, I). \tag{2}$$

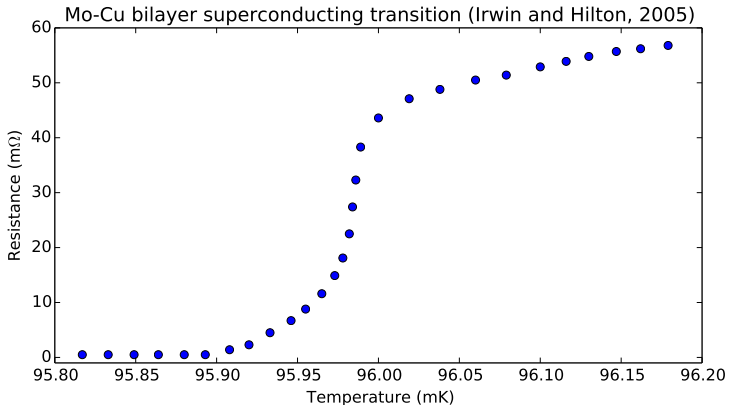


Figure 4: The transition of a superconducting film (Mo-Cu bilayer) from the normal to the superconducting state near 96mK.

Here C is the heat capacity (of both the TES and any absorber), P_{bath} is the power flowing from the TES to the heat bath, P_J is the Joule power dissipation, E is the energy of the photon impulse, L is the inductance, V is the Thevenin-equivalent bias voltage and $R(T, I)$ is the electrical resistance of the TES, which is generally a function of both temperature and current [1].

These coupled ODEs have three main nonlinear terms which are given by the expressions:

$$P_{bath} = k \cdot (T^n - T_{bath}^n) \tag{3}$$

$$P_J = I^2 R(T, I) \tag{4}$$

$$R(T, I) = \frac{R_N}{2}\left[1 + \tanh\left(\frac{T - T_c + (I/A)^{2/3}}{2\ln(2)T_w}\right)\right] \tag{5}$$

where T_w and A are obtained from logarithmic temperature sensitivity α_I and logarithmic current sensitivity β_I of R at $T = T_0,\ I = I_0$.

Although we don't explicitly use these equations in the following analysis, we do use the idea that the response of the system is nonlinear. This nonlinearity becomes extremely important when considering the analysis of a pulse followed closely by another, as the relationship is not a simple superposition. In the following we attempt to deal with this nonlinear pulse problem as well as address the problem of noisy signals.

1.2 Two mathematical challenges

In order to obtain the best signal-to-noise ratio, current practices assume a linear device response, isolated pulses, and temporal alignment of pulses. Unfortunately, these three assumptions are violated, often significantly, in real life experiments. In order to truly obtain the best signal-to-noise ratio, one must account for device nonlinearity and pulse pile-up.

A recorded pulse stream as in Figure 3 is often divided into records containing only one pulse triggered at the same place in the record. Figure 5 shows two examples of records with pile-up: one where a pile-up occurs after the triggered pulse and one in which a previous pulse tail is present in the triggered pulse. The general question for this project is how can one deal with pulse pile-ups? Pile-up of two very close successive pulses are very hard to deal with and are often removed from the dataset before analysis. If we restrict the data record to only single, isolated pulses, significant loss of data occurs when the incoming photon rate is high. If, instead, we keep the piled-up pulses (but with an understood higher variance in their calculated energy) we should be able to extract some information to use. Our analysis of noise approximation describes a probabilistic approach for extracting such information from noisy datasets and is presented in the third section. Our analysis of the pile-up problem is presented in the following section.

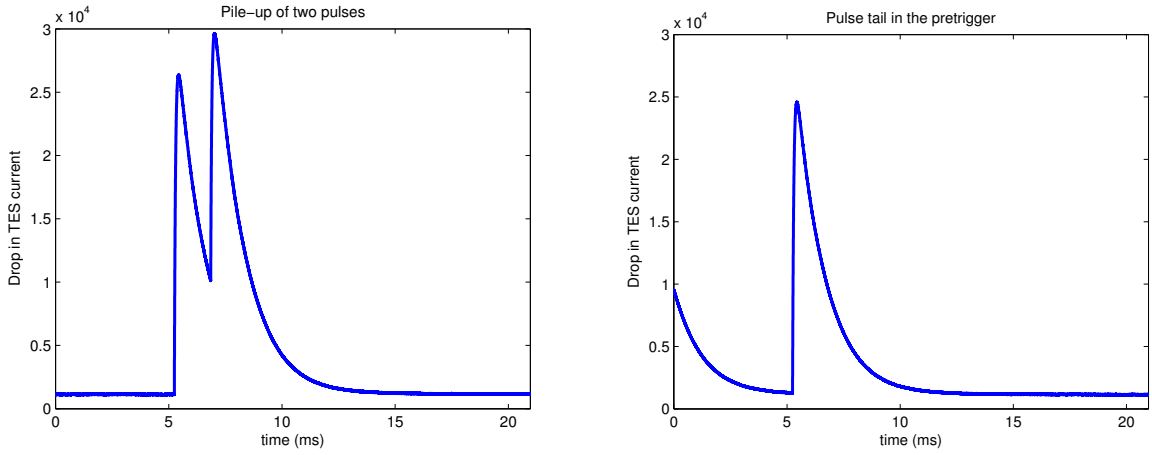


Figure 5: Record with two consecutive pulses (left) and record with pulse tail in triggered pulse (right).

2 Pile-up problem

The ultimate goal of any experiment using microcalorimeters is to compute the total energy of a photon using the current drop (pulse) of the TES. Since the energy of incoming photons is directly related to the pulse height captured by the sensor, one must be able to compute accurately the height of the pulse. This information is difficult to compute accurately in the case of a high frequency photon stream which contains mostly piled-up pulses since the relationship is nonlinear. Ideally, one can find a relation between the pulse heights and some pretrigger value which might not be linear and for which we can characterize the contribution of each previous pulse tail to the successive pulses. Below we detail our attempt to find such a characterization.

2.1 Effects of filtering

Our work was initially done using records of the pulse stream. This approach was restrictive since for a given record, the pretrigger captures only part of the information of the pulse that came before the pulse in the record. By recordizing the data stream, we eliminated the possibility of obtaining information about the prepulse height and the time interval between two successive pulses for analysis.

Figure 6 shows the information that we extract from the pulse stream. Each pulse and prepulse pair satisfying a criterion are triggered and their heights, pretrigger height, and time interval are stored. The flow chart in Figure 7 describes the algorithm we use for this filtering process.

Two filters were used during the project. Figure 8 shows three different filters and the corresponding histograms of the pulse heights spectrum; we used the second and third. For the sake of comparison, we have included the first row, corresponding to an optimal filter (see [2]). The histogram for the optimal filter presents three spikes that corresponds to K_{α_1}, K_{α_2} and K_{β} emission lines.

The second row corresponds to the filter using an average of all pulses of the stream that do not have a baseline value for the entire record before the pulse. The histogram for this filter seems to be in good agreement with the histogram of the optimal filter but the K_{α_1} and K_{α_2} emission lines have merged. Finally, the last row shows the filter using the average pulse with no prepulse adjusted to the noise power spectrum. This filter broadens the histogram of the pulse heights and a lot of information seems to be lost with this filter.

2.2 Possible correlations

This section presents some relationships between the pulse height and different pre-trigger values. In an attempt to obtain stable estimates of the pre trigger values, the prepulse is filtered and its height is used

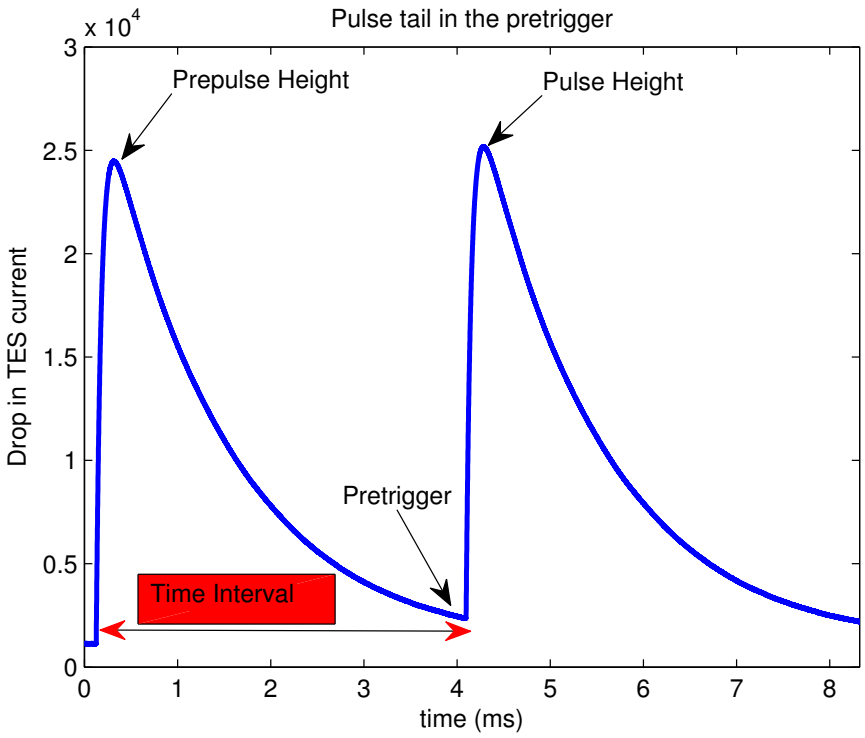


Figure 6: Computing the pretrigger for a given pulse.

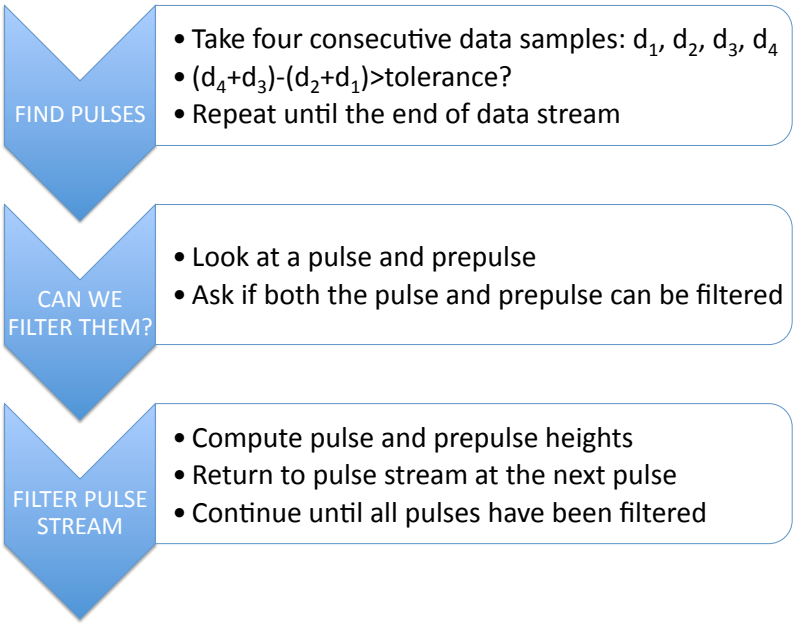


Figure 7: Flow chart for filtering the pulse stream.

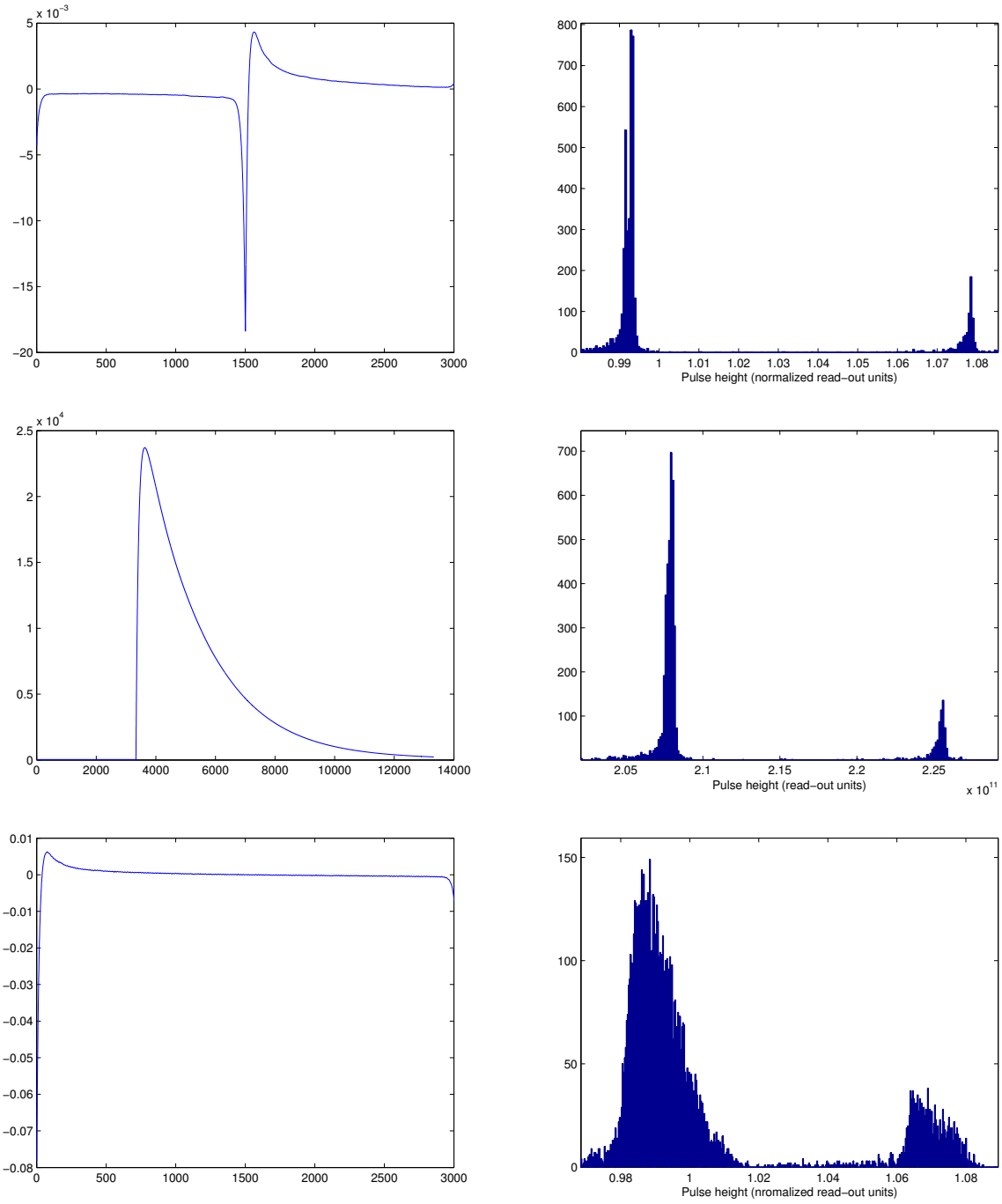


Figure 8: Filters and histograms of the resulting filtered pulse heights: optimal filter (top), average pulse filter (middle), no prepulse filter adjusted for noise (bottom).

to scale the average pulse at the corresponding time since arrival. Figure 9 shows the difference between computed and averaged trigger values which is mostly zero with some noise.

Figure 10 gives the relation between the prepulse height and the pulse height. This relation has very interesting patterns although it appears there is no correlation between the two.

Figure 11 shows the relation between the pulse height and the time between the current and previous

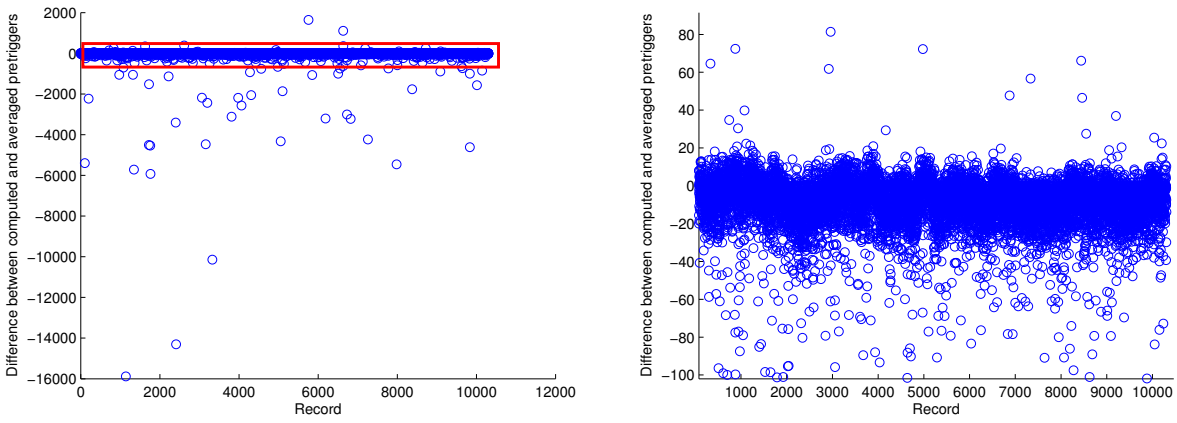


Figure 9: Difference between the computed and averaged trigger values.

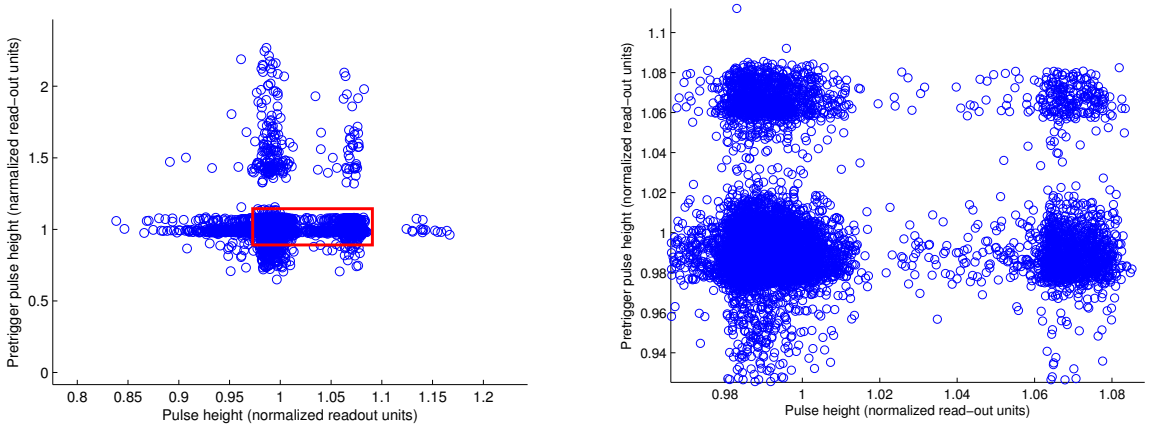


Figure 10: Relation between the prepulse height and the pulse height.

pulse. It is possible to see two horizontal lines which are seen as the K_α and K_β emission lines in the histograms. After zooming to the red square region, we see a nonlinear increase for the first 5ms. The increase can be explained by the choice of the filter, here shown in the last row of Figure 8. This filter is negative for a small interval at the left side of the x axis which means that if there is a large pretrigger value, the filter will under estimate the actual pulse height. Since the relation in that time interval does not seem to be linear, we suggest further investigation in this direction.

3 Noise approximation problem

The current procedure to address the problem of pulse pile-up is to eliminate any record that contains more than one pulse. This eliminates many records that still contain information, albeit with higher uncertainty in the calculated photon energy. A natural question would be: How can we keep those records and still have a good estimation of the photon energy spectrum? The idea here is for an unknown photon energy distribution D, we observe $y_1, \ldots, y_n$ and $\sigma_1, \ldots, \sigma_n$ where $y_i = x_i + \epsilon_i$ with $x_i \sim D$ and $\epsilon_i \sim N(0, \sigma_i^2)$. Thus the instrumental broadening σ_i depends on sample number i. Our goal is to obtain the best estimate of the distribution D. In our analysis we've considered two approaches; one tries to unbroaden the signals

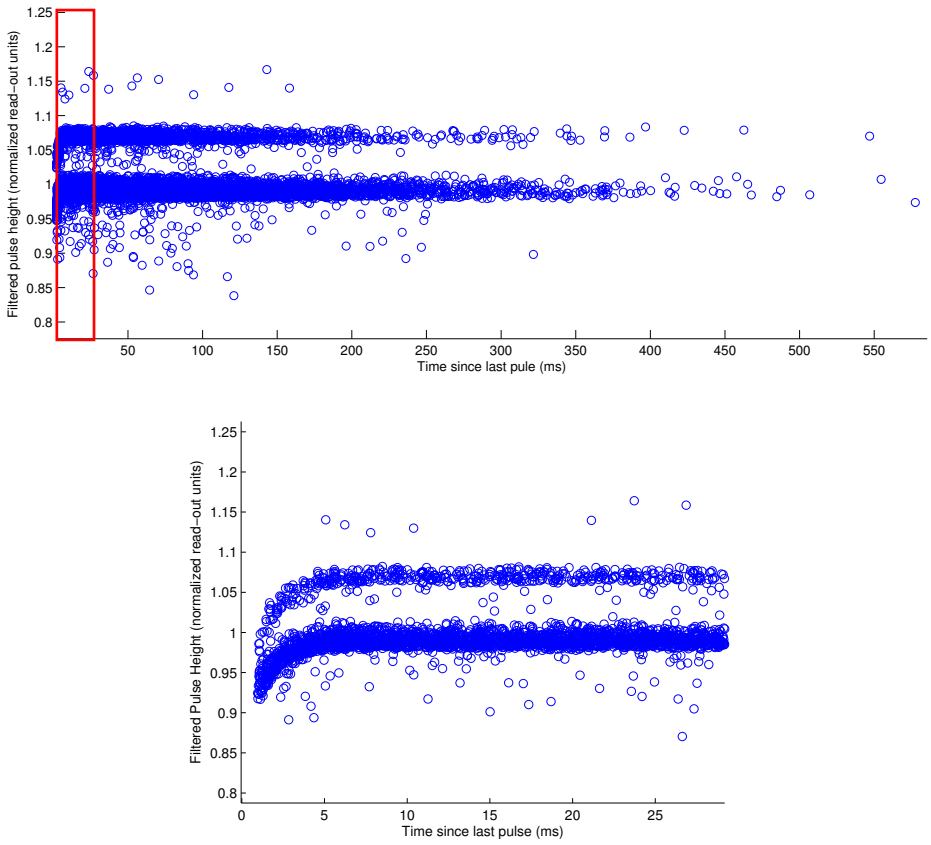


Figure 11: Relation between the pulse height and the time between the current and previous pulse.

and the other attempts to combine high and low variance data in order to obtain the most information about the distribution. In the following we describe both ideas in detail and provide simulations to test our hypotheses.

3.1 Unbroaden the signal

We would like to unbroaden the signals because if we can 'get rid of' the noise, ϵ_i, the real distribution, D, will be easy to estimate. Random variables such as Y, X and ε can be identified by their characteristic functions which is the inverse Fourier transform of the probability density function (PDF). For example, the characteristic function of Y is:

$$\varphi_Y(t) = E[e^{itY}] = \int_{\mathbb{R}} e^{ity} dF_Y(y) \tag{6}$$

where $F_Y(y)$ is the cumulative distribution function of Y. Then we can use a property of the characteristic functions which states that the sum of two independent random variables X_1 and X_2 can be written as

$$\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t) \cdot \varphi_{X_2}(t)$$

In our case, we would obtain

$$\varphi_{Y_i}(t) = \varphi_{X_i}(t) \cdot \varphi_{\varepsilon_i}(t) \Rightarrow \varphi_{Y_i}(t) \cdot \varphi_{\varepsilon_i}^{-1}(t) = \varphi_{X_i}(t)$$

Unfortunately, the samples y_i are not continuous and their probability density function will be a set of Dirac delta functions. The integral form of the characteristic function (6) is no longer proper, but it is of the Riemann-Stieltjes kind. This means, for a real-valued functions $f(x)$ and $g(x)$ where $x \in \mathbb{R}$, the Riemann-Stieltjes integral is

$$\int_a^b f(x)\, dg(x).$$

It is defined to be the limit of the approximating sum

$$S(P, f, g) = \sum_{i=0}^{n-1} f(c_i)(g(x_{i+1}) - g(x_i))$$

as the norm of the partition of the interval $[a, b]$

$$P = \{a = x_0 < x_1 < \cdots < x_n = b\}$$

approaches zero. Here, c_i is in the i-th subinterval $[x_i, x_{i+1}]$.

We can express the cumulative distribution function $F_Y(t)$ with the empirical cumulative distribution function (ECDF). The ECDF of Y is

$$\hat{F}_Y(t) = \frac{1}{n} \sum_{1}^{n} \mathbb{1}\{y_i \leq t\}$$

where $\mathbb{1}$ is the indicator function. From this estimate, we can estimate $\varphi_{Y_i}(t)$ by Riemann-Stieljes integration as

$$\hat{\varphi}_{Y_i}(t) = \sum_{i=0}^{n-1} e^{ity} (\hat{F_Y}(y_{i+1}) - \hat{F_Y}(y_i)). \tag{7}$$

Since the noise is a Gaussian, we don't need to approximate it with the equation above. Its closed form is

$$\varphi_{\varepsilon_i}(t) = e^{\frac{-1}{2} t^2 \sigma_i^2}$$

We can see that a problem will occur when we process $\varphi_{Y_i}(t) \cdot \varphi_{\varepsilon_i}^{-1}(t)$ because the inverse of the characteristic function of ε_i goes to infinity as t increases. One solution is to limit the domain of $\hat{\varphi}_{Y_i}(t)$ with $t \in [a, b]$ such that we still have a good estimation of $\hat{\varphi}_X(t)$.

To check this approach, we simulated a complex distribution. The simulated data is a combination of Gaussian curves with random $\mu \in [0, 1]$ and small standard deviations ($\sigma \in [10^{-3}, 0.5 \times 10^{-2}]$). Intuitively, the noisy data broadens the pure signal and the larger the noise, the larger the broadening effect is. Here, we disrupted the sample with only one type of noise to simplify the analysis and apply the unbroadening process mentioned above. We can see the PDFs plotted in Figure 12.

Notice that the PDF of the approximation of D fits quite well (we still need to quantify what "quite well" means; further information below), but the frequency range is very limited, for example the high peaks fit worse than the lower peaks. This happens because we are restraining the estimation of the PDF by cutting the characteristic function at t. This type of cut creates artifacts, seen at the bottom of Figure 12, which is far from ideal. Additionally, these results consider only one type of noise. What happens if the noise changes for every record like the original assumptions? And how can we deal with the large frequencies? Perhaps the unbroadening solution is not quite right.

3.2 Combine data with different variance

Suppose we can collect different types of data in an experiment: some with low uncertainty collected at high costs and others with high uncertainty but collected at lower costs. We would like to try to combine both types in order to improve the information that we could get from only collecting one kind.

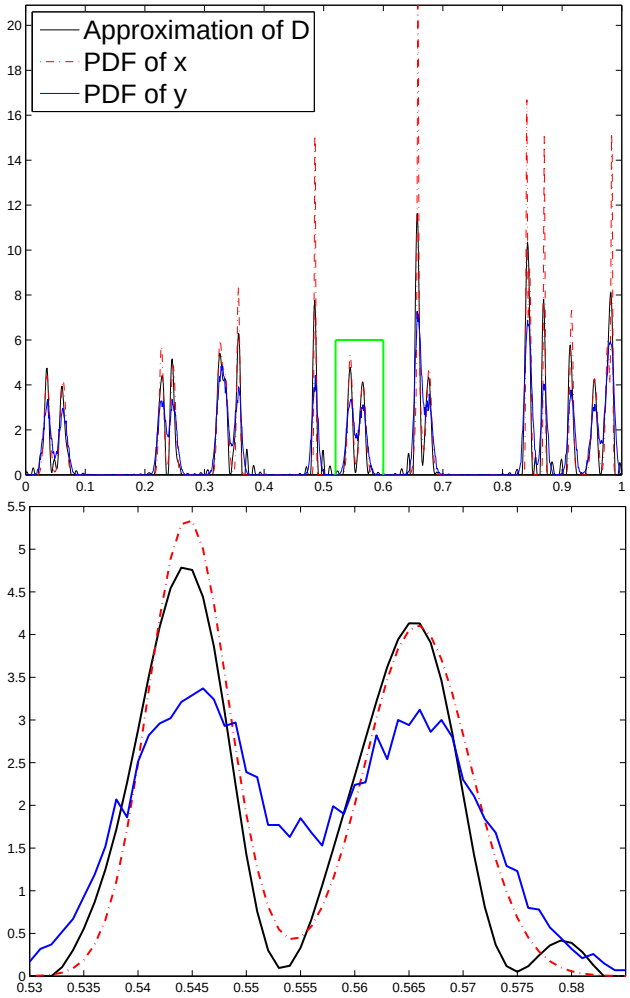


Figure 12: Top: Probability density function for noisy sample (blue), real baseline sample (dashed red) and the approximation of the real density (black) with $t \in [-700, 700]$. Bottom: Expanded region from above plot.

In our analysis we chose to have only two kind of noises (for simplicity): one with a smaller variance equal to the minimum variance of the Gaussians before combination ($\sigma^2 = 10^{-6}$) and one with a larger variance equal to the maximum variance of the Gaussians before combination ($\sigma^2 = 2.5 \times 10^{-5}$). We also 'collect' the samples in different proportions. The lower disrupted sample (Y_1) is $1/9$ of the total samples and the proportion of the higher disrupted sample (Y_2) is $8/9$. This procedure was made to fit what we would expect from the observed records. For each disrupted sample and for the baseline we computed their characteristic functions with equation (7), see Figure 13. Visually, the noisiest sample looses accuracy as the frequency increases. To measure how well the noisy samples actually fit the baseline sample, we propose to use the L^2 norm between Y_1, Y_2 and X which can be written as

$$L^2 = \|\varphi_{Y_i}(t) - \varphi_X(t)\|$$

for $i \in \{1, 2\}$. We could try to combine them as they appear in Figure 13, but the signal may be lost at higher frequencies. We can amplify the signal of the noisiest sample and then combine them with a convex combination. The amplification process can be thought as an "unbroadening" process with the following transformation

$$\hat{\varphi}_{\tilde{Y}_2}(t) = \hat{\varphi}_{Y_2}(t) \cdot e^{(\frac{1}{2}t^2(\sigma_2^2 - \sigma_1^2))}$$

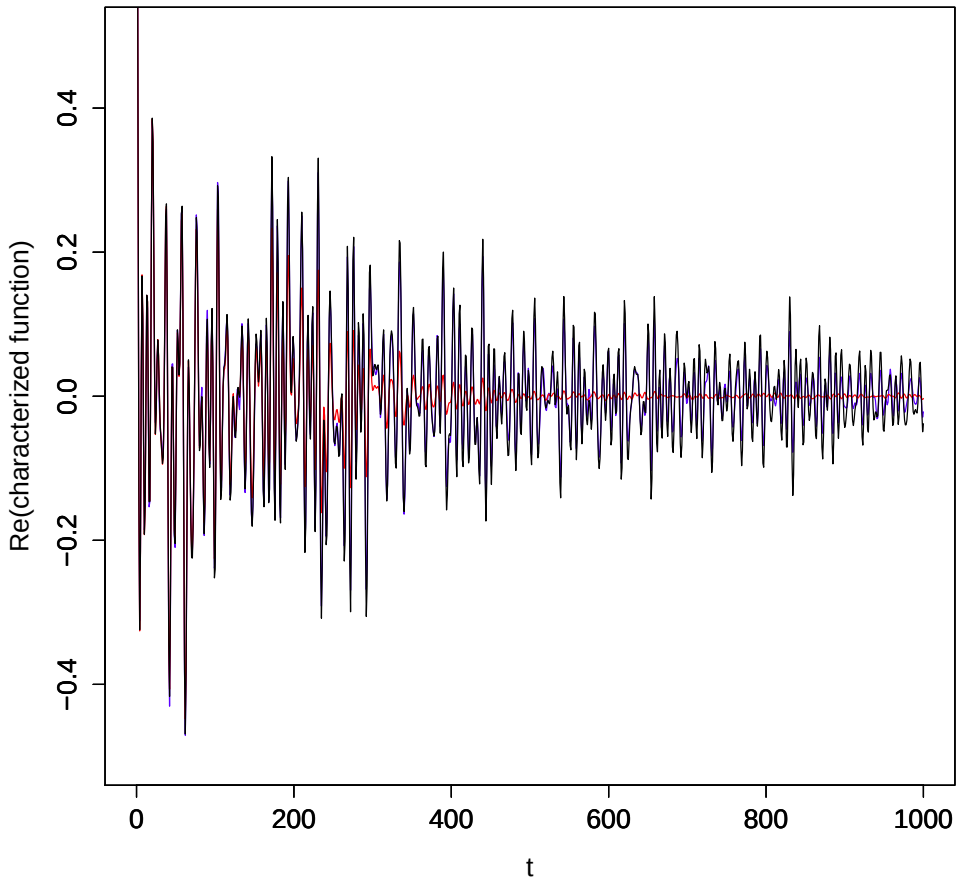


Figure 13: Real part of the characteristic functions for sample with small noise (blue), large noise (red) and baseline sample (black) with $t \in [0, 1000]$.

The value of $\alpha(t)$ in the convex combination $\alpha(t) \cdot \hat{\varphi}_{\tilde{Y}_2}(t) + (1 - \alpha(t)) \cdot \varphi_{Y_1}(t)$ is not clear, but it most certainly is a function of t. We computed the best values of α and formed the best characteristic function with the convex combination. The algorithm tests for which α the accumulation of the L^2 norm from 0 to t is lowest for every point, t, of the real part of the characteristic function. We obtain the graph of the different accumulated L^2 norms in Figure 14. We can see that the combination improves the signal and that the L^2 norm is lower. Unfortunately, the best α's are not easily characterized. This phenomenon is seen in Figure 15. Notice that the choice of α for a particular t changes quite rapidly offer the range of t, but that for small t α tends to be around 1 and for large t α tends to be around 0.

Instead of 'cherry-picking' the optimal values of α for every frequency, we could instead introduce a monotonic constraint. Then the problem becomes a quadratic programming problem and can be solved by an interior-point-convex algorithm. We then find that the optimal $\alpha(t)$ is a step function as shown in Figure 16. Comparing Figure 14 and Figure 17 shows that the monotonic $\alpha(t)$ keeps the majority of the improvement of the L^2 norm.

The problems of dealing with different noise, processing them and combining them is not an easy task. We have a better understanding of how the simulated data behave, but unfortunately, the real "observe" data are not quite as friendly as the simulated samples. Further knowledge is needed to advance and extended research needs to be done to resolve this complex problem. One suggestion is to investigate whether or not $\alpha(t)$ is always a step function, regardless of the distribution of X and samples of Y. If so,

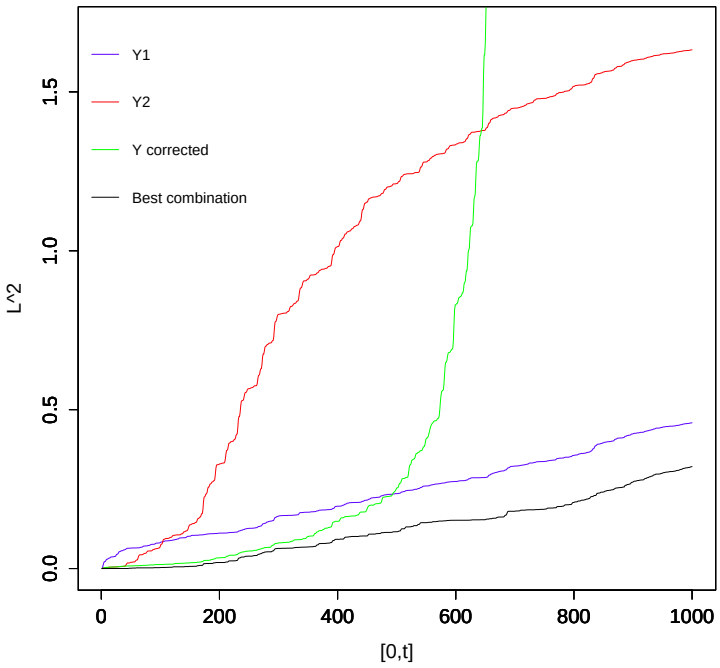


Figure 14: Accumulated L^2 norm of the characteristic function for small noise sample (blue), larger noise sample (red), corrected sample (green) and convex combination sample (black).

can we find the relationship between α and t and generalize the convex combination to samples with n different noise distributions? Ultimately, we would hope to be able to use the convex combination the to characteristic function to get a robust and accurate estimation of the distribution, D.

Acknowledgements

All participants would like to thank IMA and PIMS for organizing this workshop and UBC for hosting this event. They are infinitely grateful to their mentor, Bradley Alpert, who shared his time, knowledge, advice, and laughs throughout these 10 days.

References

[1] Irwin, K.D. and Hilton, G.C., *Transition-Edge Sensors*, Cryogenic Particle Detection : Topics in Applied Physics. **99**, pp 63-150, Springer (2005)

[2] Alpert B.K. et al., *Note: Operation of gamma-ray microcalorimeters at elevated count rates using filters with constraints*, Rev. Sci. Instrum. **84**, 056107 (2013)

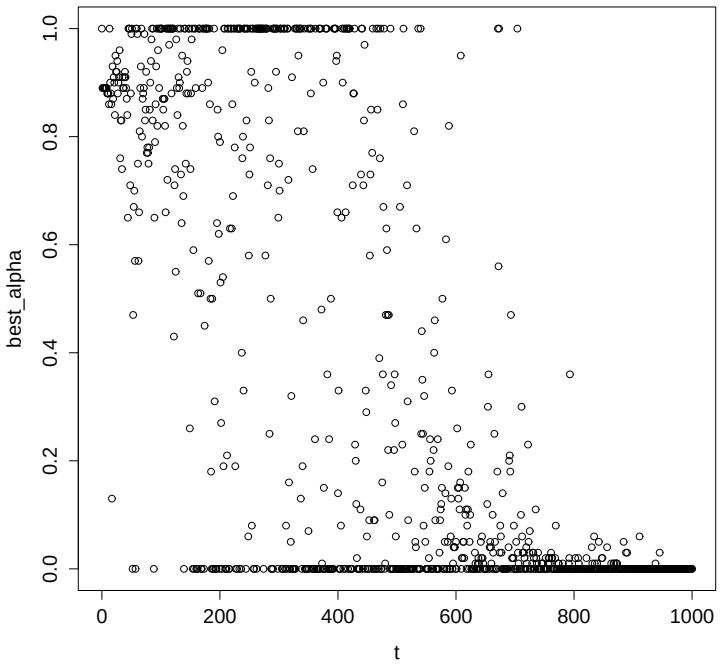


Figure 15: Scatterplot of the best α's from the convex combination

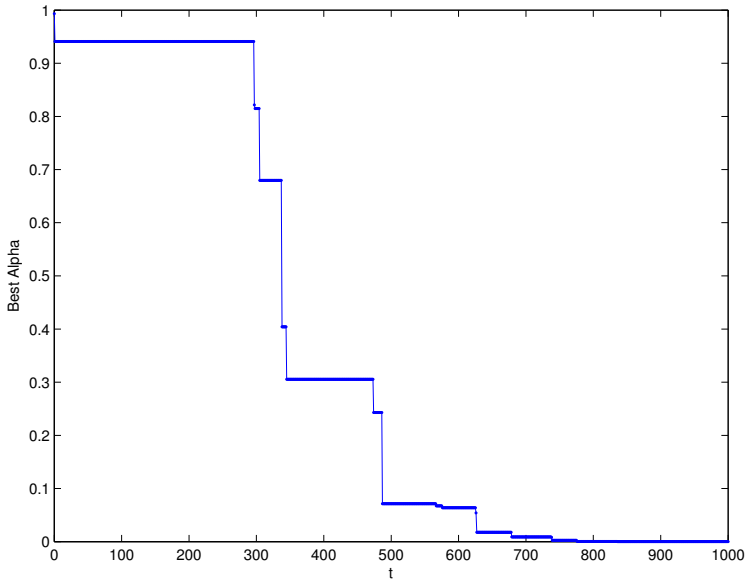


Figure 16: Optimal $\alpha(t)$ for the convex combination with a monotonic constraint.

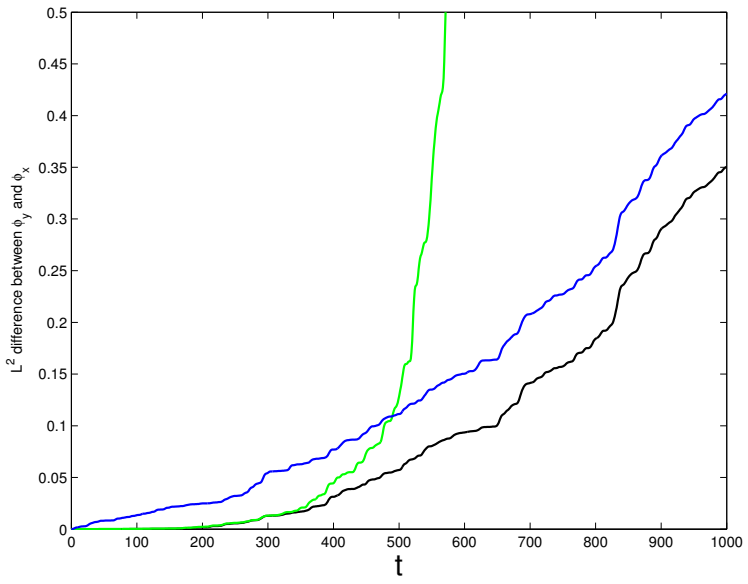


Figure 17: Cummulative L^2 norm of the characteristic function for small noise sample (blue), corrected sample (green), and convex combination (black) with monotonic $\alpha(t)$.