Development of Data Processing and Analysis Pipeline for the RICOCHET Experiment

J. Colas1 · J. Billard1 · S. Ferriol1 · J. Gascon1 · T. Salagnac1 · for the RICOCHET collaboration

Received: 22 October 2021 / Accepted: 18 October 2022 / Published online: 21 November 2022 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Achieving a percentage-level precision measurement of the coherent elastic neutrino nucleus scattering (CEνNS) spectrum requires a robust data processing pipeline which can be characterised with great precision. To fulfil this goal, we present hereafter a new Python-based data processing pipeline specifically designed for temporal data analysis and pulse amplitude estimation. This pipeline features a data generator allowing to accurately simulate the expected data stream from the RICOCHET experiment at the Institut Laue Langevin nuclear reactor, including both background and CEνNS signals. This data generator is pivotal to fully understand and characterise the data processing overall efficiency, its reconstruction biases, and to properly optimise its configuration parameters. We show that thanks to this optimised data processing pipeline, the CryoCube detector array will be able to achieve a 70 eV energy threshold combined with electronic/nuclear recoil discrimination down to ∼100 eV, hence fulfilling the RICOCHET targeted performance.

Keywords RICOCHET · CENNS · Signal processing · Optimal filter

1 Introduction

The RICOCHET experiment is a reactor neutrino observatory which will host two complementary detectors inside a shielded cryostat installed 8.8 m away from the 58 MW ILL research nuclear reactor [1]. One of these detectors, the CryoCube presented in [2] is an array of 27 high-purity Ge sub-detector. Each 40 g sub-element will be instrumented with a NTD-Ge thermal (neutron transmutation-doped germanium) sensor and multiple aluminium electrodes for heat and charge measurements, respectively. Overall, the CryoCube will provide at most 150 data channels. Each
detector channel will be connected to a high electron mobility transistor (HEMT) low-noise cold pre-amplifier described in [3].

The data taking strategy is to perform continuous stream acquisition at a sampling rate between 10 kHz and up to 100 kHz, as required for an efficient anti-coincident muon veto tagging.

The data acquisition is planned to be running for months continuously such that we expect $O(100)$ terabyte of raw data for one operating year at ILL.

The processing will be done using analysis window with a duration between 0.5 and 2 s. The typical value of 1 s is well suited to contain completely the heat pulse without including multiple pulses in the same window. The heat pulses are characterised by the sum of three exponentials, each describing a time response of the detector thermal system. They are much slower ($O(10)$ μs) than the ionisation pulses which can be described by a Heaviside function because of the very fast ($O(10)$ ns) charge collection to the electrodes.

2 Simulating the CryoCube

2.1 Heat and Ionisation Pulses

We expect to measure two types of events: electronic recoils (ER) and nuclear recoils (NR) which differ from their ionisation yield $Q = E_{\text{ion}} / E_{\text{recoil}}$. We use the quenching model given by Lindhard for NR [4] and $Q = 1$ for ER. The NR quenching coefficient goes from 0.1 at 100 eVnr to 0.3 around a few keVnr. A third type of event may occur in our detector, which is the so-called heat-only (HO) event. These events form a low energy excess background, from a yet unidentified source [5], and are characterised by a null quenching factor $Q = 0$. When the electrodes are biased to a nonzero electric potential we have to consider the Neganov–Trofimov–Luke (NTL) effect which corresponds to an additional heat contribution produced by the drift of the charges in the Ge crystal under the electric field. With this additional contribution, the energy relations can be written as [6]

\[
E_{\text{ion}} = E_{\text{recoil}} \cdot Q \quad E_{\text{heat}} = E_{\text{recoil}} \cdot \left(1 + Q eV / \epsilon \right)
\]

The simulation of an ER, NR and HO event of 10 keV recoil energy is shown in Fig. 1 with $\epsilon = 3$ eV (Ge), $V = 2$ V and $\epsilon$ the elementary charge. The NTL effect is clearly visible since the amplitude of the heat pulse increases with the amplitude of the ionisation signals. The noise added to these traces corresponds to the noise expected for the future CryoCube electronic and is described in the following subsection.

2.2 Electronic Noise Generation

The noise models for both the ionisation and heat channels are extracted from [3] and can be accurately reproduced from the following three noise structures: (1)
white noise, generated using random normal distributed samples, PSD (power spectral density, in $V^2/\text{Hz}$) $\propto$ constant; (2) brown noise, generated using first-order low-pass-filtered white noise, PSD $\propto f^{-2}$; and 3) pink noise, generated using the Voss–McCartney algorithm [7], PSD $\propto f^{-1}$. These noises, characterised by their power spectral density, will be used as our basis for our noise modelling such that the expected heat and ionisation noises can be approximated by:

$$\text{PSD}_{\text{heat/ion}}(f) \sim A_{\text{heat/ion}} + B_{\text{heat/ion}}/f + C_{\text{heat/ion}}/f^2.$$  \hspace{1cm} (1)

This approximation could be improved using more different noise models or additional filters, but this is already sufficient to reproduce very well the ionisation noise and relatively well the heat noise as we can see in Fig. 2 where we compare our simulated noise with the models from [3]. The discrepancy between model and generated noise around $f = 10$ Hz shows that a high-pass filter should be used to reduce the gap between model and generated noise. But since the implementation of such filter can be tricky for bigger-than-memory signals we will develop this feature for a future work. To be able to generate hours-long signal, we had to adapt our algorithms for chunk-by-chunk data generation where the data stored in memory at a time (a chunk) is not too large and fits in RAM. Our algorithms preserve the phase continuity between each consecutive chunks.

### 3 Processing Data

The processing pipeline for the CryoCube was firstly characterised using simulated stream generated using the previously discussed approach. Using simulated realistic data is the best way to know efficiently the performances of such complex pipeline in terms of amplitude, timing resolution and trigger efficiency. The goal of our pipeline is to give an estimation of the amplitude $\hat{a}$, the location $\hat{t}_0$ and the quality of fit $\chi^2$ for each triggered pulses. The only required prior information is the analytical

![Fig. 1](image-url) Simulation of a 10 keV recoil energy event considering a Ge detector with a planar design and a 2V potential between electrodes: Left—Electronic recoil. Centre—Nuclear recoil. Right—“Heat Only” event. The pulse simulation considers a sensitivity of 1 $\mu$V/keV and 1.2 $\mu$V/keVee together with a gain of 9.8 nV/ADU and 78.5 nV/ADU for the heat and ionisation channels, respectively.
pulse shape for each channel. This knowledge gives us a starting point to evaluate the noise in the data stream which is the most important quantity for the estimation of $\hat{a}$ and $\hat{t}_0$ as we will explain in the following section.

The global view of the processing pipeline is presented in Fig. 3. Firstly the pipeline is set-up using an external configuration yaml file. Then each hour of data is sent to this pipeline and undergoes: (1) a down-sampling to reduce the sampling frequency $f_s$ (if needed for speed consideration), (2) a linear high-pass filtering to clean the low frequency end of the spectrum, (3) an estimation of the noise power spectrum, (4) a trigger event positions and (5) an amplitude and position ($\hat{a}, \hat{t}_0$) estimator for each triggered events. The output and the input of this pipeline are compressed structured files in Zarr format [8], which support multiple compression algorithm as well as simultaneous read and write.

### 3.1 Noise Estimation

To estimate the noise PSD, we use the analytical pulse shape and look at the parts of the stream which least resemble to pulses. One way to do that is to use the correlation between a pulse trace with the entire stream and look for the minimum amplitude of the obtained signal. We then take the maximum number of traces which have no data point above a particular threshold dynamically defined. Typically in our case, this first step of the noise estimator returns at least $\sim$100 noise traces per one hour long data stream from which a first PSD $J_0$ can be estimated. The second step is to take the maximum number of traces from the stream and compute the $\chi^2$ distribution such that $\chi^2_i = \sum_f \frac{|T_i(f)|^2}{J_0(f)}$ where $T_i$ is the $i$-th trace in the Fourier space normalised as a LPSD. This gives us a $\chi^2$ distribution with a peak around the number of frequency bins, corresponding to all the traces similar to the first ones considered for $J_0$. These are the good
noise traces from which the final noise PSD $J$ can be estimated with higher statistics, hence lower variance.

### 3.2 Triggering Potential Events

The trigger algorithm is based on matched filtering and is close to the implementation presented in the widely used reference [9]. The idea is to apply a matched filter (matched to the pulse template) which is characterised by a transfer function $H$ such as $H(f) = \alpha \frac{S^*(f)}{J_0}$, where $S^*$ is the complex conjugate of the Fourier transform (FT) of the pulse trace and $\alpha$ is a complex normalisation constant:

$$\alpha = \sqrt{\frac{N}{2f_s}} \left( \sum \frac{|S|^2}{J} \right)^{-1} \exp \left( i 2\pi f_0 \frac{f}{f_s} \right).$$

(2)
The resulting signal \( m(t) = F^{-1}(H \ast F(x(f))) \) of the matched filtering of the data stream \( x(t) \) is then sorted by decreasing values. The first value of this sorted array is retained as trigger position \( i_0 \) and a region around this position is discarded. Typically the size of this region is the same as processing window \( N \) such that there are never two triggers in the same window. Then we repeat this first step by taking the first non-discarded value and append it to \( i_0 \) until all the data are discarded. By doing the trigger selection this way, we do not need to specify a particular threshold value and we are able to trig on the lowest pulse amplitudes, maximising our detection efficiency to the lowest event energies.

### 3.3 Estimation of \( \hat{a} \) and \( \hat{t}_0 \)

The trigger algorithm returns an array of integer \( i_0 \) representing the position in sample numbering where an event might be. We will use this information as a starting point to our fitting approach. The idea is to use the analytical pulse shape to create a 2-dimensional matrix \( M_{i,k} \) with \( i \) the index of starting \( \tau_i \) (more explanations below) and \( k \) is the window sample index between 0 and \( N-1 \). In other words, this matrix contains the traces of the pulses for each channels and for different starting time \( \tau_j \). This means that we evaluated the pulses functions on a grid of starting time \( \tau_i \) thin enough to reach sub-sample resolution. In the presented work, the sub-sampling is regularly spaced in the range \( [(i_0-p)/f_s,(i_0+p)/f_s] \) with \( L \) sub-samples (Usually: \( f_s=10 \) kHZ, \( p=1, L=101 \)). Then, to find the optimal starting time \( \hat{t}_0 = \tau_{i_{\text{var}}} \) and amplitude \( \hat{a} \) we minimise the \( \chi^2 \) defined, for one channel, as:

\[
\chi^2(\tau_i, \hat{a}) = \sum_j \frac{|D(f) - \hat{a} M(\tau_i,f))|^2}{J(f)}
\]

where \( D \) is the FT of one window returned by the trigger normalised as LPSD, like \( M \) which is the LPSD normalised FT of the template matrix \( M \). The minimisation is performed by calculating the optimal \( \hat{a} \) for each \( \tau_i \) and then we take the minimal \( \chi^2 \) value computed over the whole time range. Lastly, this simple example can be generalised to more than one channel, as done in the following, in building a global \( \chi^2 \) with its corresponding matrix of PSDs also including the cross-PSD (XPSD) between channels.

### 4 Results

#### 4.1 Noise Estimation and Trigger Efficiency

Applying the processing pipeline described above to simulated yet realistic data as presented in Sect. 2.1 allows us to estimate the performance of the future CryoCube detector with the present processing scheme. We find a 50\% efficiency at about 70 eV nuclear recoil energy, with a slightly lower threshold for ER thanks to their higher NTL heat energy boost (see Fig. 4). We also see, from Fig. 4, that the noise estimation
algorithm recovers properly the injected noise power spectral densities for all down-sampled sampling frequencies.

### 4.2 Energy and Timing Resolution

Figure 5 (left panel) shows the timing resolution $\sigma_{t_0}$ given by the RMS of the difference between the input and reconstructed time onsets of simulated events as a function of the recoil energy, for varying sampling frequencies and recoil types. This Monte Carlo estimated timing resolution can be compared with its theoretical calculation using the estimated amplitude $\hat{a}$ according to the equation

$$\sigma_{t_0}^2 = \left[ \hat{a}^2 \sum_f (2\pi f)^2 \frac{|S(f)|^2}{|J(f)|^2} \right]^{-1}$$

![Fig. 4](image1.png) Left—Trigger efficiency at $f_s = 1$ kHz for different recoil energy based on the noise prediction [3]. Right—Estimated linear noise power spectrum density for heat (red) and ionisation channel (blue) using our proposed minimal correlation algorithm for identification of noise, at multiple sampling frequencies and compared to the noise models we tried to reproduce (see Fig. 2)

![Fig. 5](image2.png) Left—Timing resolution for different types of events versus the recoil energy. Right—Amplitude resolution for heat and ionisation channels versus the sampling frequency
As one can see from Fig. 5 (left panel), $\sigma_{t_0}$ improves faster than $1/E_{\text{recoil}}$ thanks to the additional fast ionisation signal. This is particularly true for ER events because of their higher ionisation yield. Our expected timing resolution suggests that a 400 Hz muon veto trigger rate, as expected with RICOCHET at ILL, is manageable. From Fig. 5 (right panel), we can see that the baseline ionisation energy resolution $\sigma_{a_0}$ slightly improves with increasing the sampling frequency $f_s$ as expected from its higher signal bandwidth. The estimated baseline resolution is compatible with our theoretical calculations: $\sim 10$ eV (heat), $\sim 20$ eVee (ionisation) [3]. As a result, this study confirmed the choice of 10 kHz as an optimal sampling frequency taking into account the quantity of raw data as well as the timing and energy resolutions [5].

4.3 Particle Identification Capabilities

Finally, the detailed simulation confirms the particle identification capabilities of the CryoCube detector, where we confirm a ER-NR discrimination threshold around 100 eV recoil energy as it is shown in Fig. 6. This threshold is defined as the intersection of the upper 1-sigma and lower 1-sigma bands from the NR and ER populations, respectively, in the quenching vs recoil energy (or equivalently the heat vs ionisation) plane. Adding the HO event population shows that they will affect our ability to observe the CE$\nu$NS signal at the lowest energies, and a mitigation of such events is therefore our highest priority.

Fig. 6 Particle identification performance obtained using the data processing pipeline with simulated data stream using the expected electronic noise and Lindhard quenching model. In addition of electronic (blue) and nuclear recoils (red), we choose to add the heat-only (black) background as well. The coloured regions in the quenching plot are defined as the ± one-$\sigma$ band of the respective event type (NR, ER, HO)
4.4 Application to Experimental Data

This processing pipeline was also used on experimental data in various situations and with several acquisition devices with great performance. This reflects the high flexibility of this Python-based processing framework which is already being used with various cryogenic detector technologies, namely with NTD-Ge [2, 10] and Kinetic Inductance Detectors [11].

5 Conclusion

We presented an efficient simulation and data processing pipeline dedicated to the CryoCube detector array. Using realistic ionisation and heat models (pulse and noise) we confirmed the expected performance of the CryoCube detector array required to observe with high precision the CEνNS signal from the ILL reactor. More detailed simulations, including the ILL background and CEνNS signal, are ongoing to estimate the future Ricochet CEνNS precision measurement and new Physics discovery potential.

Acknowledgements This work is supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program under Grant Agreement ERC-StG-CENNS 803079. We are grateful to J.-B. Filippini and H. Lattaud for their feedbacks on the use of this processing pipeline.

Data Availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

1. S. Hertel et al. Ricochet progress and status, J. Low Temp. Phys. (This Special Issue) (2021)
2. T. Salagnac et al. Optimization and performance of the CryoCube detector for the future Ricochet low-energy neutrino experiment, J. Low Temp. Phys. (This Special Issue) (2021)
3. A. Juillard, others, Low-noise HEMTs for coherent elastic neutrino scattering and low-mass dark matter cryogenic semiconductor detectors. J. Low Temp. Phys 199(3–4), 798–806 (2019)
4. J. Lindhard et al., Influence of crystal lattice on motion of energetic charged particles. Kongel. Dan. Vidensk. Selsk. Mat. Fys. Med. 34(14), 965 (1965)
5. J. Billard, Searching for Dark Matter and New Physics in the Neutrino Sector with Cryogenic detectors. Habilitation à diriger des recherches, Université Claude Bernard Lyon 1, January (2021)
6. E. Armengaud et al., Performance of the EDELWEISS-III experiment for direct dark matter searches. J. Instrument. 12(08), P08010–P08010 (2017)
7. R. Voss, J. Clarke, 1/f noise in music: music from 1/f noise, J. Acoust. Soc. Am. 63, 258–263 (1978)
8. F. Alted et al., MRC Centre for Genomics and Global Health. https://github.com/zarr-developers/zarr-python
9. S. Di Domizio, F. Orio, M. Vignati, JINST 6, P02007 (2011). https://doi.org/10.1088/1748-0221/6/02/P02007
10. H. Lattaud et al., Phonon and charge signals from IR and X excitation in the SELENDIS Ge cryogenic detector, J. Low Temp. Phys. (This Special Issue) (2021)
11. J. Colas et al., Improvement of contact-less KID design using multilayered Al/Ti material for resonator, J. Low Temp. Phys. (This Special Issue) (2021)

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.