ELSEVIER

Nuclear Instruments and Methods in Physics Research A 412 (1998) 440–453

NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH Section A

Performance evaluation of a hit finding algorithm for the ICARUS detector

F. Arneodo[a], P. Benetti[b], A. Bettini[c], A.Borio[d], E. Calligarich[b], C. Carpanese[c], D. Cavalli[e], F. Cavanna[a], P. Cennini[e], S. Centro[c], A. Cesana[d], C. Chen[f], Y.B. Chen[f], D. Cline[g], I. De Mitri[a], R. Dolfini[b], A. Ferrari[d], A. Gigli Berzolari[b], K.L. He[f], X.P. Huang[f], F.Lu[f], Z.H. Li[f], J.M. Ma[f], G. Mannocchi[h], F. Mauri[b], L. Mazzone[b], C. Montanari[b], M. Nicoletto[c], S. Otwinowski[g], S. Parlati[a], D. Pascoli[c], A. Pepato[c], L. Periale[i], G. Piano Mortari[a], A. Piazzoli[b], P. Picchi[h], F. Pietropaolo[c,*], A. Rappoldi[b], G.L. Raselli[b], S.Resconi[d], J.P. Revol[e], M. Rossella[b], C. Rossi[a], C. Rubbia[f], P. Sala[d], D. Scannicchio[b], F. Sergiampietri[j], S. Suzuki[i], M. Terrani[d], P. Torre[b], S. Ventura[c], M. Verdecchia[a], C. Vignoli[b], H. Wang[g], J. Woo[g], G.F. Xu[f], Z.Q. Xu[f], C. Zhang[f], Q.J. Zhang[f], S.C. Zheng[f]

[a] *Dipartimento di Fisica e INFN (LNGS), Università dell'Aquila, via Vetoio, I-67010 L'Aquila, Italy*
[b] *Dipartimento di Fisica e INFN, Università di Pavia, via Bassi 6, I-27100 Pavia, Italy*
[c] *Dipartimento di Fisica e INFN, Università di Padova, via Marzolo 8, I-35131 Padova, Italy*
[d] *Dipartimento di Fisica e INFN, Università di Milano, via Celoria 16, I-20133 Milano, Italy*
[e] *CERN, CH-1211 Geneva 23, Switzerland*
[f] *Institute of High Energy Physics, Beijing, People's Republic of China*
[g] *Department of Physics, UCLA, Los Angeles, CA 90024, USA*
[h] *Lab. Naz. di Frascati dell'INFN, via E. Fermi 40, I-00044 Frascati (ROMA), Italy*
[i] *ICGF del CNR, Corso Fiume 4, I-10133 Torino, Italy*
[j] *INFN Pisa, via Livornese 1291, I-56010 San Piero a Grado (PI), Italy*

Received 18 December 1997; accepted 11 February 1998

Abstract

A procedure to filter the ICARUS digitized data is presented. The final multi-kiloton detector will provide raw data at a rate of the order of 100 Gbytes/s and therefore a crucial element of the acquisition chain will be the on-line data reduction system. A filter has been developed to process the signals and to tag the meaningful data. The hit detection efficiency has been evaluated on a sample of signals obtained from cosmic rays events in the 3 t ICARUS prototype. Its robustness has been tested also on simulated events with different signal to noise ratio. The hit finder hardware implementation is also presented. © 1998 Elsevier Science B.V. All rights reserved.

* Corresponding author.

PII: S0168-9002(98)00347-7

1. Introduction

The ICARUS project [1] aims to build a five thousand ton liquid Argon time projection chamber (LAr-TPC) to be operated in the Gran Sasso Laboratory to search for rare events such as neutrino interactions or proton decay. A detailed description of the detector operating principles can be found in Ref. [2].

Ionizing events, taking place in the volume, will produce free electrons that will drift parallel to the electric field, inducing a current on the electrodes. The read-out is based on wire planes placed at the end of the drift volume. A three dimension view of an ionizing event can be obtained by correlation of different plane signals and drift time.

The ICARUS collaboration, after many studies on small volume prototypes, at present is working on a 600 t detector module [3]. This module will be the first of a series in order to reach an active mass of several kt. It has four readout chambers; each chamber has three wire planes with wires oriented along different directions (stereo angle of 60°), the wire pitch is 3 mm and the the spacing between successive planes is also 3 mm; the total number of wires is about 46.000. The maximum drift length is 1.50 m giving a maximum drift time of 1.0 ms for an electrical field of 500 V/cm.

The readout system is structured as a multichannel wave form recorder that stores the charge information collected by each sense wire during the drift of the electrons. Each wire is equipped with a current integrating amplifier followed by a 10 bit ADC that samples the signal every 400 ns. The converted data are stored in a digital memory. Given the resolution, both in space and time, the size of a complete drift volume image results in about 200 Mbytes. The detector is continuously sensitive, giving a total throughput of nearly 200 Gbytes/s, but according to the event rates expected under the Gran Sasso Laboratory (see Table 1), the effective data rate will not exceed few Mbytes/s.

The collected data can be highly reduced by recognizing Regions Of Interest (ROI) of the signal, by means of a dedicated hit finding unit, pipelined in the data acquisition path. Only data inside ROI's will be saved, obtaining a high efficiency zero skipping. Contiguous ROI's can be correlated to enhance detector self triggering capability.

Table 1
Expected events rate

Event type	Rate	Total throughput
Proton decay	~ 0	~ 0
Atmospheric Neutrinos	1/day	~ 0
Solar Neutrinos	2/day	~ 0
Cern Neutrinos	4/day	0.1 KBytes/s
Muons	100/h	~ 1 Kbytes/s
Low engery background	1000/s	~ 1 MBytes/s

The developed hit finding algorithm can be easily implemented in hardware. It has been extensively tested on real and simulated events to evaluate its efficiency and robustness as a function of the signal shape with different S/N ratio.

2. The hit finding algorithm

The 3D imaging of ionizing events in the active volume of the ICARUS LAr-TPC is built exploiting the free electrons produced along the tracks and drifting toward the read-out planes. Each wire of the read-out planes senses a segment of track by reading the charge induced on the wire itself by the electrons drifting a. Drifting electrons can be read by several wire planes as the read-out is non destructive. The electrons are collected on the last wire plane (collection plane). The different orientation of the wires on the successive read-out plane provides different 2D views of the event; the arrival time of the electrons on the wire plane gives the space coordinate in common to all the 2D views. The combinations of at least two 2D views allows a three dimensional reconstruction of the event.

In the present ICARUS configuration there are three read-out planes with wires directions at 60° in one plane with respect to the next. When drifting electrons cross the first induction plane the signal is as in Fig. 1a (first induction signal); when they cross the second induction plane the shape of the signal is a triangular pulse lasting few microseconds (Fig. 1b, second induction signal); when the

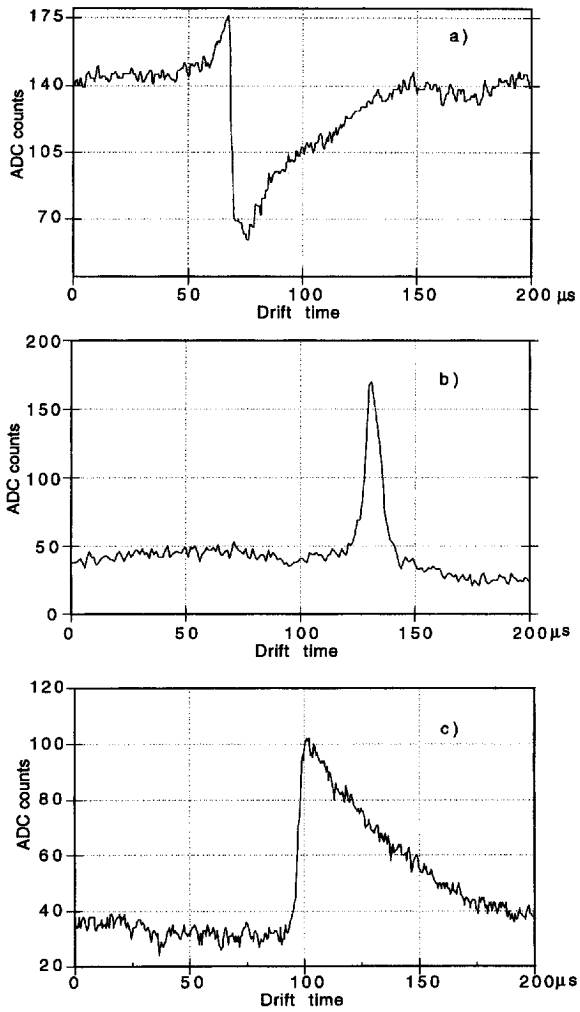

Fig. 1. (a) First induction signal (no grid in front), (b) second induction signal, (c) collection signal.

electrons reach the last plane they are collected on the wires and give a step signal proportional to the collected charge (Fig. 1c; collection signal).

A ROI has been defined as a time window around a rising edge (hit). Only data around the detected peaks are then stored for a later analysis. The relevant information carried by signals are: peak time that gives the track position inside the volume; pulse height which is proportional to the track ionization; rise time, that depends both on track angle and on the distribution of the electron cloud.

To identify a hit a time-sliding window strategy has been adopted. Inside the window, of fixed size, the positive differences, between every two consecutive signal samples, are summed and stored in an accumulator register while negative and zero differences are counted in two separate counters.

During this search another time window, whose size can be stretched, mirrors the first one keeping memory of the same sum and counter values. When the fixed size window overlaps a rising edge, the positive sum increases quickly its value while the negative and zero counters keep low. A hit is found as soon as the positive sum is above a given threshold (*rising sum_threshold*) and the two counter values are below two other thresholds (*rising zero_count* and *rising negative_count*). Once the hit has been recognized, the second window fixes its lower bound and continues to accumulate the positive differences, stretching its size step by step, until it covers the whole rising edge. The end of the rising edge is detected when, in the fixed window, three conditions are met: the positive sum is below a given threshold (*falling sum_threshold*) and at least one of the two counters is above its corresponding threshold (*falling zero_count* or *falling negative_count*). The size of the stretched window is then restored to its original value and the hit finding procedure restarts.

The end of the rising edge gives the time information for saving a suitable number of samples around the signal peak position. The accumulated differences inside the stretched sliding window, that now on will be called filtered output, mirror the signal rising edge as shown in Fig. 2. Drift time is given in 400 ns sampling time units (clock counts).

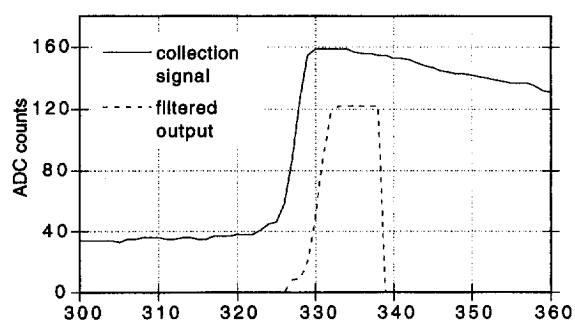

Fig. 2. Collection signal and filtered output.

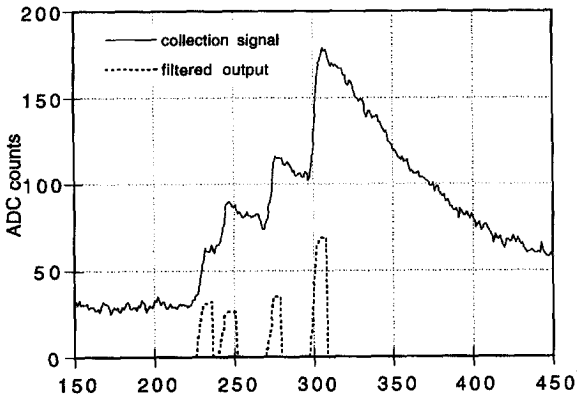

Fig. 3. An example of multiple hits detection.

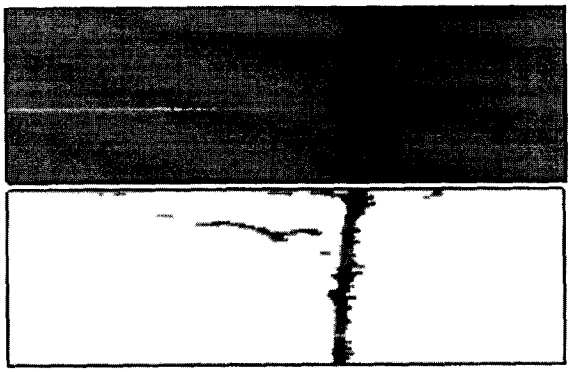

Fig. 4. A track image of a crossing muon as seen on the collection plane. A pair from gamma conversion is also visible to the left of the muon. The horizontal axis is the drift time (the full scale corresponds to 200 μs or 30 cm); the vertical axis is the wire numbering (full scale corresponds to 192 wires or 40 cm). The top view are the raw data; the bottom view is the image filtered with the alogrithm described in this paper.

This process is made flexible through the parametrisation of all the thresholds and the window size. Their values can be tuned depending on signal characteristics and signal to noise ratio. Furthermore it allows to resolve signals of different tracks that hit the same wire closely in time. Fig. 3 shows a four tracks collection signal and the related filtered outputs.

The procedure is applied to detect both the induction and the collection signals. In the following sections the performance on both will be presented.

3. Efficiency measurements on the 3 t detector signals

As a part of the R & D ICARUS program, a 3 t prototype detector has been built in the early nineties [4,5] and had been in operation for several years. It was equipped with two read out wire chambers of two coordinates each and a wire pitch of 2 mm. The length of induction and collection wires was 2.4 and 0.9 m, respectively.

In Fig. 4 the upper part shows the image of a crossing muon track raw data recorded from the collection plane. The lower part shows the filtered image after analysis. The vertical axis corresponds to 192 wires and the horizontal axis spans a drift time of 200 μs. The gray level is proportional to signal amplitude.

The efficiency of the algorithm has been evaluated both on collection and induction data extracted from cosmic ray muon events collected in 1994 and 1995.

Some 120 collection signals with a single hit per wire and 80 induction signals have been selected from muon crossing and muon stopping events. In both cases the signals have eight bit resolution and sampling time of 400 ns.

3.1. Performance of the algorithm on collection signals

The amplitude and the rise time distribution of the collection signals, with 10% accuracy, are given in Figs. 5 and 6. The rms. noise is 2.3 (in ADC counts) and the mean noise spectral density is shown in Fig. 7.

It is important to stress, before discussing the efficiency, that although the described algorithm is suitable for exhaustive analysis, as it could be inferred from Fig. 3, the actual purpose of this procedure is not the extraction of the relevant information (pulse height, peak timing, and rising edge) in compressed form, but rather the efficient recognition of every ROI. It means that the algorithm parameters must be chosen in order not to loose significant data that will be later off line analyzed.

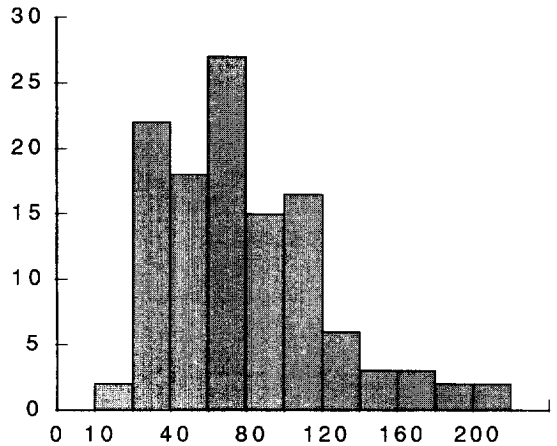

Fig. 5. Amplitude distribution of collection signals (amplitude in ADC counts).

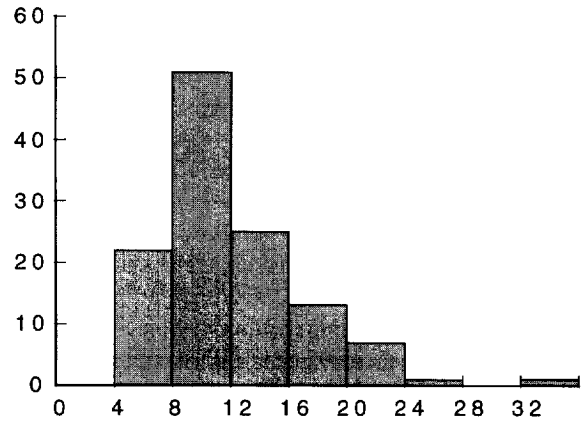

Fig. 6. Rise time distribution of collection signals (time in clock counts).

As the hit finding is based on local calculation, it is largely affected by high frequency disturbances. To reduce them a non recursive median filter [6] has been applied to the raw data. The output of the median filter is the sequence of median values obtained on a fixed size time window running over the entire raw data sample. The size of the filter is the length of the time window. Its most important characteristic is the capability to preserve sharp edges of a signal, removing efficiently impulsive noise. The performance of the median filter as a function of its size, is shown in Fig. 8. The rms noise values are plotted versus the filter size in Fig. 9.

To evaluate the hit detection efficiency a few a priori decisions have been taken. The median filter size has been set to 9 as a reasonable trade off between amplitude preservation and noise reduction. The amplitude threshold, *rising sum_threshold*, for the minimum detectable signal, has been set to 6 as a trade off between the noise level and the minimum amplitude. It is worth to point out that in our sample a minimum ionizing particle corresponds to 24 counts as set by the overall DAQ gain.

Given the median filter effect it is easy to observe that in the rising edges there are no negative slopes. So the *rising negative_count* is set to 0. For events with a worse S/N ratio, it could be necessary to

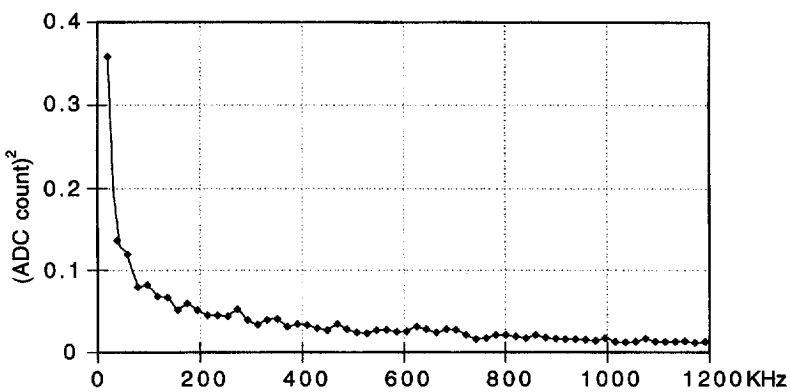

Fig. 7. Mean noise density of collection signals.

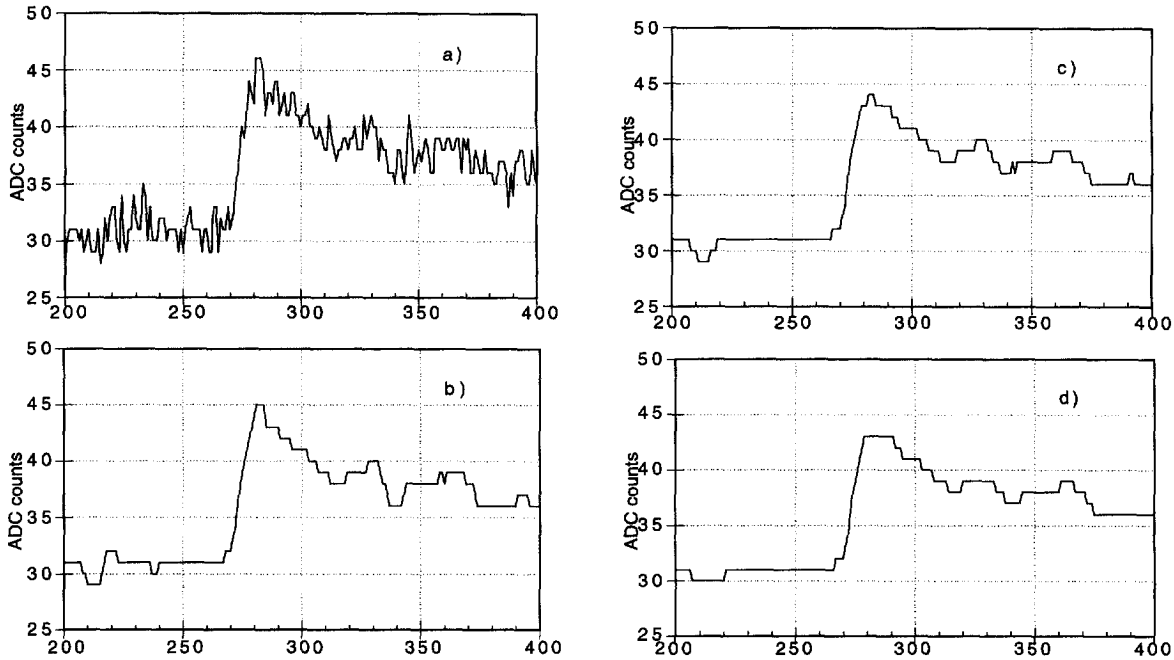

Fig. 8. Effects of the median filter size on the collection signal shape: (a) no filter, (b) 7 points, (c) 9 points, (d) 11 points.

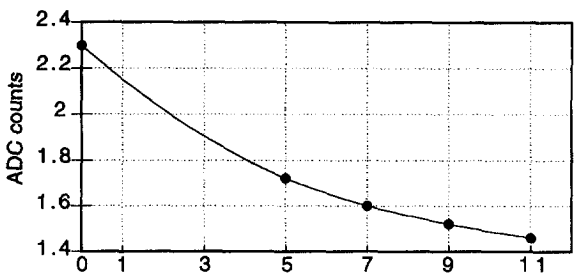

Fig. 9. Rms. noise value as a function of filter size (0 means no filter applied).

accept some negative differences. For the same reason the *rising zero_count* is not controlled to make the decision. Therefore a signal has been found if, inside the window, the sum of positive differences is over 6 counts and there are no negative differences while the zero differences count can be of any value.

To detect the peak, the end of the rising edge of the signal has to be identified. The requirement is that the sum of positive differences has to be less than the one on the rising edge to avoid, through some digital histerysis, multiple detection of the same signal. The *falling sum_threshold* is then set to 4. To take into account both flatness at the peak and negative slopes the *falling zero_count* is set to 3, while *falling negative_count* is set to 0.

Therefore the peak of the detected signal is reached when the sum of positive differences is less than 4 and either zero differences count is over 3 or there is at least 1 negative difference.

The peak detection efficiency is then evaluated as a function of only the size of the searching window. To check how many signals of the sample are detected and how many are noise generated, the position of the peak is compared to the one obtained from the off line analysis within a given tolerance. The tolerance is set to 10 counts. Fig. 10a and b report the efficiency and the false detection's versus the filter window size.

The detection efficiency is 100% for a searching window size larger than 4 counts while the false detection does not exceed 2.5%.

We recall that this performance is evaluated on real data and that the efficiencies are verified event

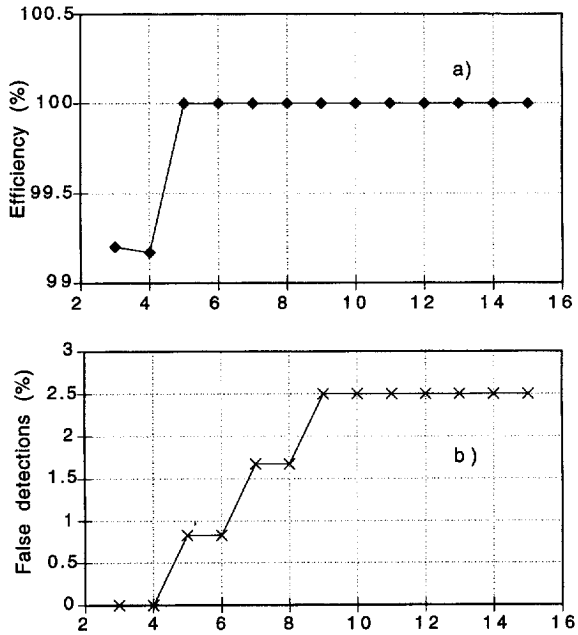

Fig. 10. (a) Detection efficiency as a function of the window size, (b) percentage of false detection as a function of the window size.

by event by visual inspection. The same efficiencies are found on Monte Carlo simulations of these events.

This algorithm can be applied with similar parameters to the "first induction" signal since the shape of it is close to that of the collection signal (but of reverse polarity). As a consequence the extraction efficiency and false detection's are much the same as that found for the collection signals.

3.2. Performance of the algorithm on induction signals

The same hit finding strategy is applied although the signal shapes are quite different (see Fig. 1). In Figs. 11 and 12 the amplitude and the width distributions of the induction signals are shown. The width has been measured at 10% of the amplitude value. Due to the large low frequency noise the accuracy is within 15%. The rms. noise (in ADC counts) is 4.1 and the mean spectral density is shown in Fig. 13.

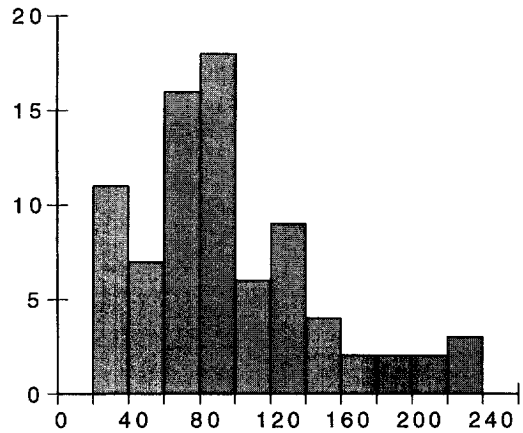

Fig. 11. Amplitude distribution of induction signals (amplitude in ADC counts).

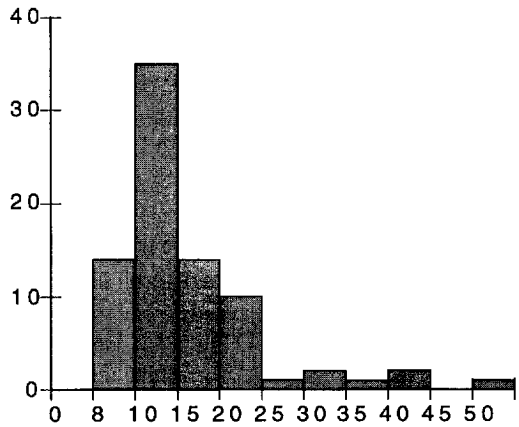

Fig. 12. Width distribution (in clock counts) of induction signals.

The effect of the median filter as a function of its size is shown in Fig. 14 while the rms. noise value is plotted versus the filter size in Fig. 15.

The detection efficiency has been evaluated applying a median filter of 5 points as a reasonable trade off between noise reduction and amplitude preservation. The algorithm parameters have been set at the same values used for the collection signals, apart from the *rising and falling sum_thresholds* set to 12 and 10, respectively. Fig. 16a and b report the efficiency and the false detection versus the filter window size.

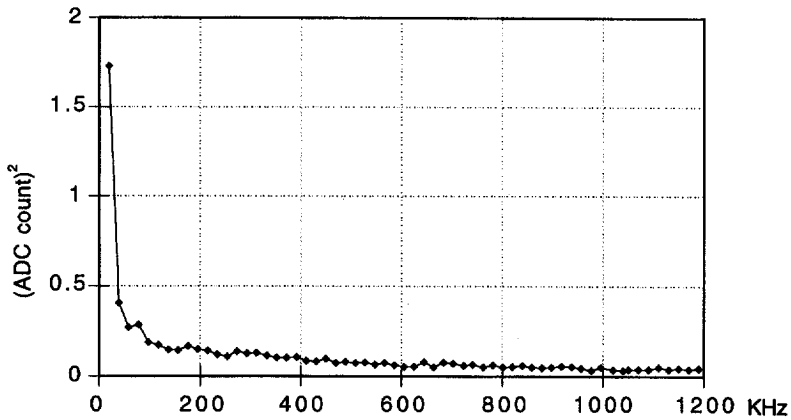

Fig. 13. Mean noise density of induction signal.

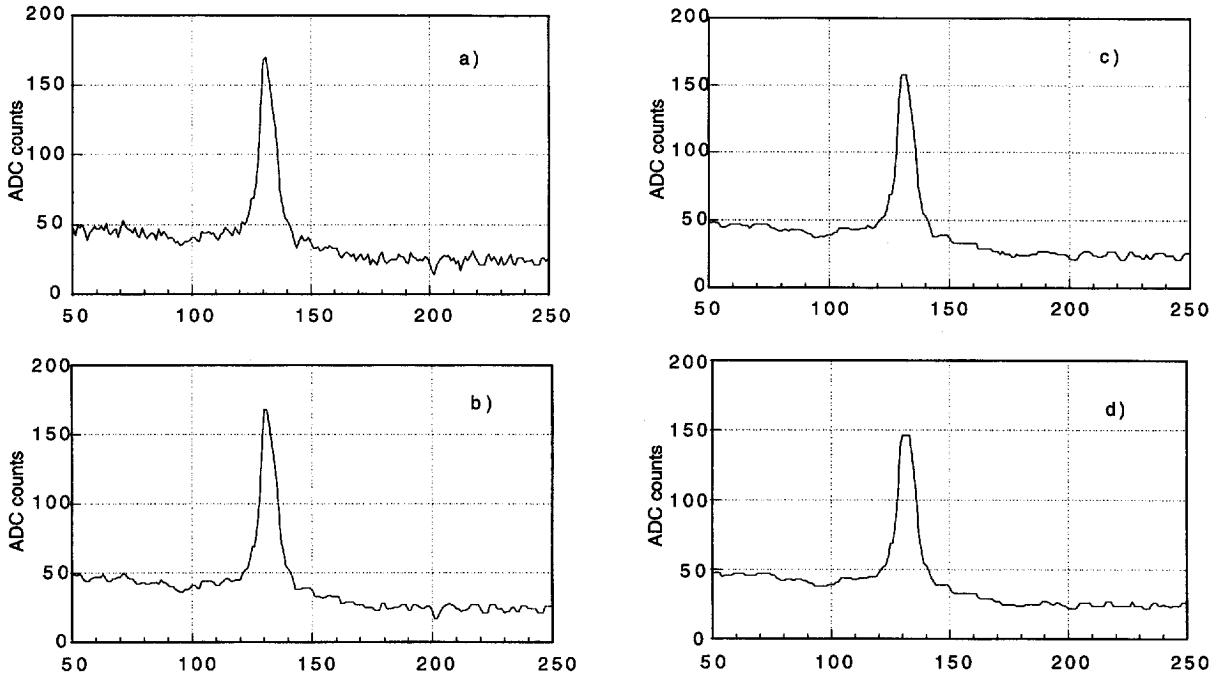

Fig. 14. Effects of the median filter on the induction signal shape; (a) no filter, (b) 3 points filter, (c) 5 points filter, (d) 7 points filter.

The optimum search window is around 6 and this is easily understood keeping in mind the induction signal shape that has a mean width of about 15 counts.

It has been extensively verified that the algorithm performance does not depend on a fine tuning of the parameters both for collection and induction.

We can conclude that a maximum detection efficiency (100%), with a reasonable number of false detection (within 2%), can be obtained in a rather wide interval of the selection parameters values. It

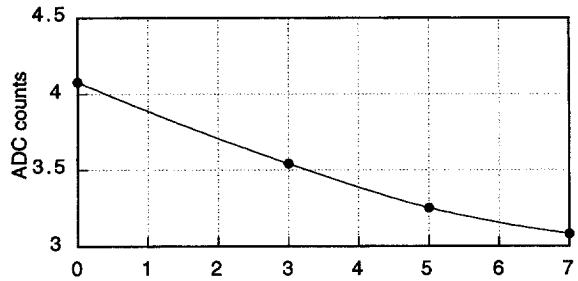

Fig. 15. Rms. noise value versus filter size (0 means no filter applied).

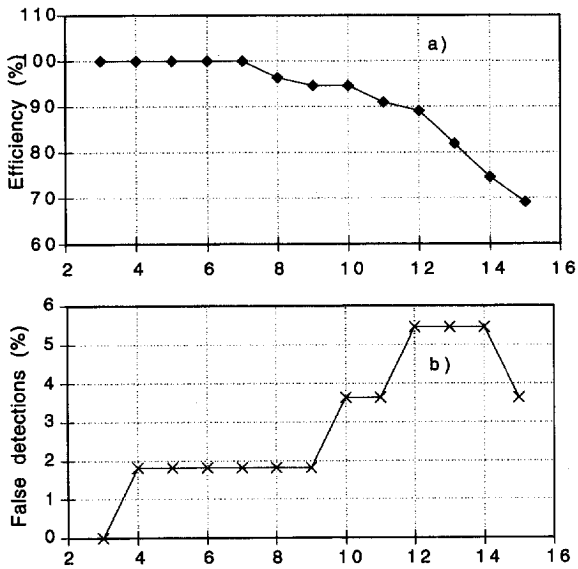

Fig. 16. (a) Detection efficiency as a function of the window size, (b) percentage of false detection's as a function of the window size.

must be noted that in both cases the S/N ratio is around 5.

4. Efficiency versus S/N ratio

Simulations of the ICARUS 600 t detector indicate that the signal to noise ratio should be of the order of 10, because of the increased input capacitance of the front-end electronics (due to longer wires and cables). Hence it is important to evaluate the algorithm performance for decreasing signal to noise ratio.

For this purpose a sample of 100 events with amplitude and noise parametrized have been generated. Amplitude is variable between 10 and 37 in 10 steps of 3 counts. The noise is the superimposition of a $1/f$ component of fixed value and a white component whose rms. value can be varied from 1 to 8 counts; a cut of 500 KHz has been provided to take into account the front end amplifier bandwidth.

Fig. 17 shows the mean spectral density for a white noise of 2 counts while Table 2 gives the total noise versus the white noise.

The shape of the simulated signals is given in Fig. 18. The rise and decay time of the collection signal are 8 and 90 counts respectively while the width of the induction signal (at 10% of its maximum) is 12 counts.

4.1. Performance of the algorithm on collection signals

Due to the rather sharp rising edge, the search window has been set to 6, while the other parameters are the same as for the experimental data, with the exception of the *rising* and *falling sum_threshold* that are increased for increasing noise. It is important to define the minimum detectable signal. A rather demanding condition has been defined to detect a signal: the efficiency has to be 100% for three contiguous thresholds and the false detection for the highest threshold, must not be greater than 1%.

Fig. 19 gives the amplitude of the minimum detectable signal and the related false detection versus the total noise.

A 100% efficiency is obtained with a S/N ratio of 5 as shown by the slope of the curve in Fig. 19a.

4.2. Performance of the algorithm on induction signals

As the signal has been chosen rather short to simulated severe conditions, the median filter size have been set to 3. For the same reason the search window is set to 5. The *rising negative_count* is 0 for noise less than 5 counts, 1 for larger noise, *falling*

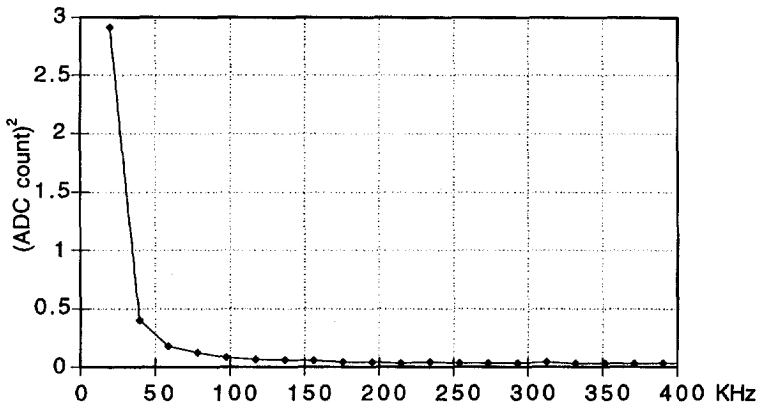

Fig. 17. Mean noise density for white noise value of 2 counts.

Table 2
Total noise versus white noise.

White noise	Total noise
1	2.2
2	3
3	3.5
4	4.1
5	4.7
6	5.6
7	6.2
8	7.2

negative_count is set to 1, while the zero counts parameters do not care. The *rising sum_threshold* increases for increasing noise.

To detect a signal the conditions are more relaxed: the efficiency has to be 100% for three contiguous thresholds and the false detection for the highest threshold, must not be greater than 20%. Fig. 20 gives the amplitude of the minimum detectable signal and the false detection's versus the total noise.

From the Fig. 20a the minimum S/N ratio required is around 5. However the false detection is much higher than for the collection.

As expected, the hit finder appears to be less effective for induction events, whose shape has been chosen close the one of impulsive noise, than for collection events.

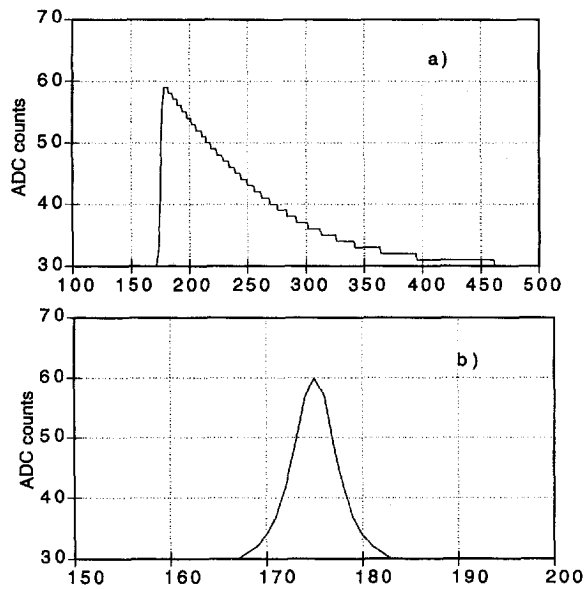

Fig. 18. Simulated signals shape; (a) collection signal, (b) induction

5. Hardware implementation

As the selection of the ROI's has to be done in real time, a specific hardware signal analyzer, named Daedalus, has been designed in VLSI CMOS technology [7]. To preserve the signal shape and to have, at the same time, a compact hardware structure, data acquisition is performed

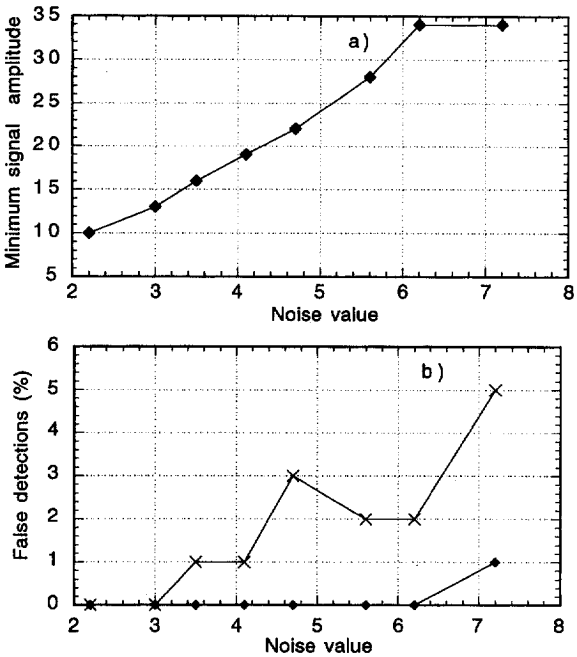

Fig. 19. (a) Minimum detectable collection signal versus total noise, (b) false detection percentage for the lower (×) and the higher (◆) threshold values versus noise.

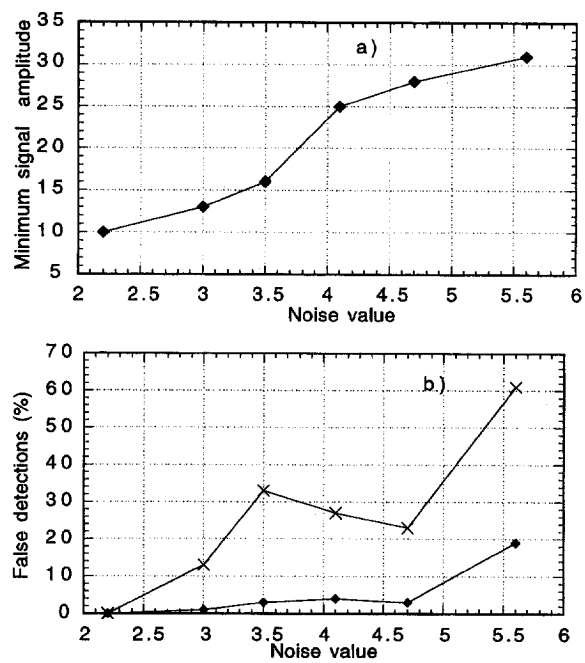

Fig. 20. (a) Minimum detectable induction signal versus total noise, (b) false detection's percentage for the lower (×) and the higher (◆) threshold values versus noise.

grouping channels in set of 16 and multiplexing then to one 40 MHz ADC in order to sample each channel every 400 ns. As shown in the block diagram of Fig. 21, the signal analyzer processes four channels, demultiplexing the 10 bits input data.

The chip architecture is a six stage pipeline; one stage is dedicated to demultiplex inputs, one to multiplex outputs, one to perform median filtering, while the hit finder requires three stages. Data move through the pipeline at a frequency of 2.5 MHz that matches the input data rate and four chips are required to process 16 channels. The chip features programmable operating parameters: median filter size (up to 15 points), window size (up to 15 counts) and all the threshold values. The median filter implementation is worth to be described in some details for its original architecture, in the last paragraph.

A few prototype chips have been produced and tested. A picture of the 4.4 mm × 4.4 mm chip is shown in Fig. 22.

They perform according the specifications and presently a complete 16 channels read out module [8,9] is under test.

5.1. Median filter

The scheme of the median filter implementation in the Daedalus chip is given in Fig. 23. Input data is loaded into the top D-register at the frequency f, typically 2.5 MHz, while the shift register formed by the full register file, up to 15 registers, is clocked at a frequency $10f$. As each register is 10 bit long, as soon as all the bit are shifted out from the top register a new data is loaded in it.

The unit that implements the median filter (Median Filter Processor) is connected to the serial outputs of each register so it can process in 10 clock cycles all the bits of the registers in parallel, starting from the MSB. The odd number of the processed registers is programmable from 3 to 15.

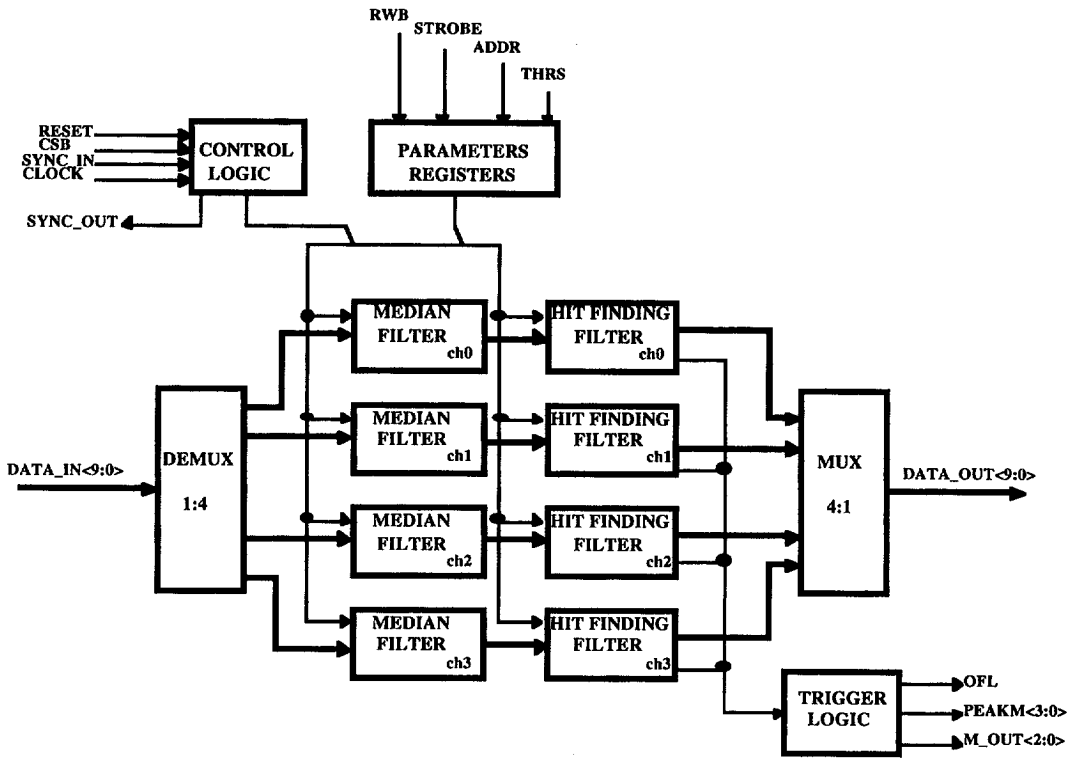

Fig. 21. Block diagram of the hit finder chip.

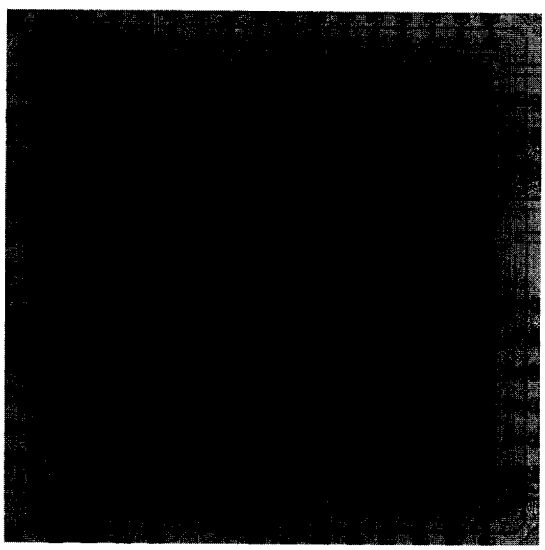

Fig. 22. Daedalus chip picture.

The Median Filter Processor starts its operation from the MSB of a odd number of registers. There are two possible cases: either all these bits are the same, zeros or ones, or they are different. In the first case no decision is taken, while in the second case the median filter value is recognized to belong to the larger subset of zeros or ones. The process then proceeds for the next bit but only inside the recognized subset and so on until the LSB is processed. The last subset contains only the median filter value(s). It is easy to understand that while the process proceeds bit by bit also the corresponding bit of the median value is found, in fact this is either the common value or the value that belongs to the larger subset. Instead of extracting step by step the smaller subset that at the end contains the median value, an equivalent procedure has been implemented. All the bits of lower weight of the data excluded from further processing are set to their

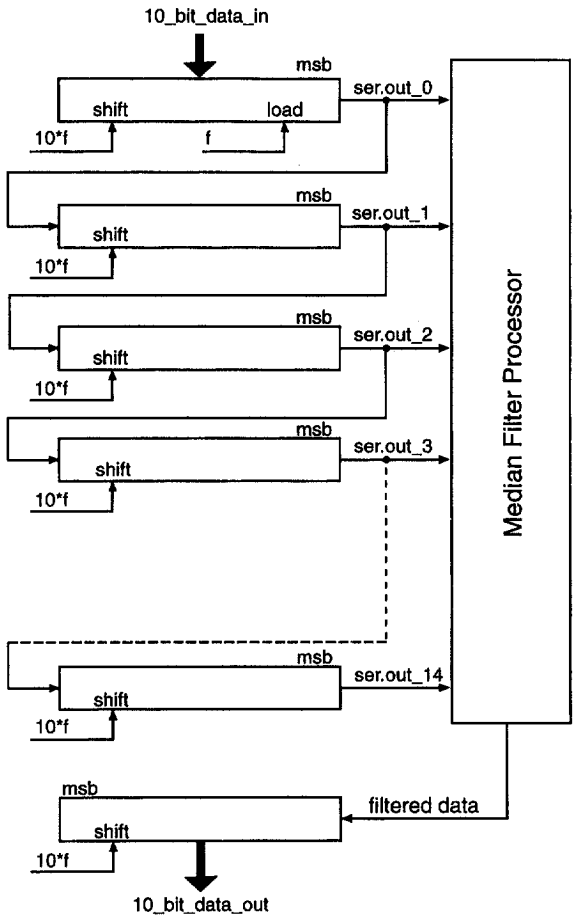

Fig. 23. Block diagram of median filter.

actual value, that is equivalent to push the excluded data to the borders, highest or lowest, of the filter window. This algorithm is easily implemented with combinatorial logic and state machines in the VLSI chip. It is relevant to note that the processing time is independent from the filter size as it is only function of the data length.

6. Conclusion

In this paper it has been proved that the hit finder algorithm for the ICARUS detector implemented in the Daedalus chip, provides a data reduction scheme that allows the storage of only few hundreds bytes per wire hit (as required for a full off-line reconstruction of the relevant physical parameters) with hit finding efficiency close to 100%.

This allows to decrease the total throughput of the ICARUS 600 t at Gran Sasso of several order of magnitudes (from hundreds of Gbyte/s to few Mbytes/s). This data flow is fully compatible with the present DAQ capability.

Recent developments of the ICARUS R & D activity suggest that the detector signals could be read in "current mode". This would provide clear advantage on the dynamic range of the data acquisition chain and the tracking ability of the detector.

The signal shape would be the discrete derivative of that in charge mode (hence the collection signal will look like the induction signal in "charge" mode and the induction signal will have bipolar shape). From the point of view of the signal extraction and the zero suppression described in this paper, this alternative makes the task easier. This is mainly due to the fact that the low frequency noise (responsible of the baseline variations) is highly reduced. This feature largely overcompensates the lower signal to noise ratio of the current mode configuration.

Simulations and tests on real signals are underway; preliminary results are very similar to those described in this paper.

References

[1] ICARUS Collaboration, ICARUS: a proposal for the Gran Sasso Laboratory, INFN/AE-85/7 Frascati, 1985.
[2] ICARUS Collaboration, ICARUS II: a second generation proton decay experiment and neutrino observatory at the Gran Sasso Laboratory, Proposal, LNGS-94/99-I & II, 1994.
[3] ICARUS Collaboration, A first 600 t ICARUS detector installed at the Gran Sasso Laboratory, Addendum to Proposal, LNGS 95/10, 1995.
[4] P. Benetti et al., Nucl. Instr. and Meth. A 332 (1993) 395.
[5] P. Cennini et al., Nucl. Instr. and Meth. A 345 (1994) 230.
[6] N.C. Gallagher Jr, G.L. Wise, IEEE Trans. Acoust. Speech and Signal Process. ASSP-29, 1981.
[7] P. Arcaro et al., DAEDALUS: a proposal for a VLSI implementation of a feature extractor for Icarus signals, ICARUS-TM-96/02, 1996.

[8] C. Carpanese, M. Lippi, Readout module for the Icarus experiment, CAEN mod.V708 technical manual, 1997.

[9] C. Carpanese et al., DAEDALUS : a hardware signal analyser for Icarus, Proc. Conf. Frontier Detectors for Frontier Physics, Isola d'Elba, 1997.