Cluster counting algorithms for particle identification at future colliders

Brunella D'Anzi[1,2]**, Gianluigi Chiarello**[3]**, Alessandro Corvaglia**[3]**,
Nicola De Filippis**[2,5]**, Walaa Elmetenawee**[1,2]**, Francesco De Santis**[3,4]**,
Edoardo Gorini**[3]**, Francesco Grancagnolo**[3]**, Marcello Maggi**[2]**,
Alessandro Miccoli**[3]**, Marco Panareo**[3]**, Margherita Primavera**[3]**,
Andrea Ventura**[3,4]**, Shuiting Xin**[6,7]**, Fangyi Guo**[6,7]**, and Shuaiyi Liu**[6,7]

[1] Dipartimento di Fisica, Università di Bari Aldo Moro, Via E. Orabona n.4, I-70126 Bari,Italy
[2] Istituto Nazionale di Fisica Nucleare Sezione di Bari, Via E. Orabona n.4, I-70126 Bari, Italy
[3] Istituto Nazionale di Fisica Nucleare Sezione di Lecce, Via Arnesano, 73100 Lecce, Italy
[4] Dipartimento di Matematica e Fisica, Università del Salento, Via Arnesano, 73100 Lecce, Italy
[5] Dipartimento di Fisica, Politecnico di Bari, Via Edoardo Orabona, 4, Bari BA, 70126, Italy
[6] Chinese Academy of Sciences, 19A Yuquan Road, Shijing District, Beijing, 100049, China
[7] Institute of High Energy Physics, 19B, Yuquan Road, Shijing District, Beijing, 100049, China

E-mail: `brunella.danzi@ba.infn.it`

Abstract. Recognition of electron peaks and primary ionization clusters in real data-driven waveform signals is the main goal of research for the usage of the cluster counting technique in particle identification at future colliders. The state-of-the-art open-source algorithms fail to find the cluster distribution Poisson behavior even in low-noise conditions. In this work, we present cutting-edge algorithms and their performance to search for electron peaks and identify ionization clusters in experimental data using the latest available computing tools and physics knowledge.

1. Introduction

The large statistical fluctuations in the ionization energy loss high energy physics process by charged particles in gaseous detectors implies that many measurements are needed along the particle track to get a precise mean, and this represents a limit to the particle separation capabilities that should be overcome in the design of future colliders. The cluster counting (CC) technique (dN/dx) represents a valid alternative that takes advantage of the Poisson nature of the primary ionization process and offers a more statistically robust method to infer mass information.

It consists in singling out, in ever recorded detector signal, the electron peak structures related to the arrival of the electrons belonging to a single primary ionization act (cluster) on the anode wire. Simulation studies by using Garfield++ and Geant4 prove that the cluster counting allows to reach a resolution two times better than the traditional dE/dx method over a wide momentum range in the use-case of a helium-based drift chamber [1]. To validate the simulations results, two beam tests at CERN/H8 facility, using a muon beam ranging from 40 GeV/c to 180 GeV/c on a setup made of 1 cm, 1.5, and 2 cm size drift tubes, equipped with Mo,
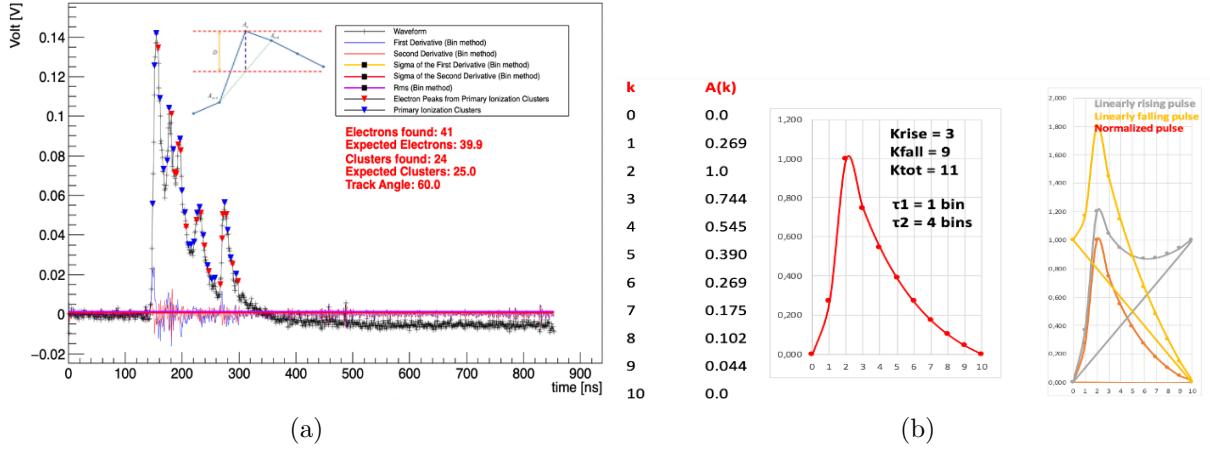
k	A(k)
0	0.0
1	0.269
2	1.0
3	0.744
4	0.545
5	0.390
6	0.269
7	0.175
8	0.102
9	0.044
10	0.0

(a) (b)

Figure 1: (1a) Example of a real waveform collected from 1 cm size drift tube (90/10 He/iC$_4$H$_{10}$, muon momentum 165 GeV, Track Angle 60°, sampling rate 1.2 GSa/s, sense wire diameter 20 μm). The blue and red arrows represent the cluster and electron peaks found with the DERIV algorithm. First, second derivative and the associated errors are drawn as indicated from the legenda.The expected number of electrons and clusters is computed by multiplying the number of cluster/cm for a MIP [5], the drift tube size, the inverse of the track angle and a relativistic rise correction factor [6]. (1b): Sketch of a electron template pulse used to scan the waveforms by the RTA algorithm.

Al, and W sense wires in the 10–40 μm diameter range, have been performed by collecting data with the gas mixtures He/iC$_4$H$_{10}$ 90/10, 80/20, and 85/15 by making a scan in high voltage, sampling rate and complementary angle between the wire direction and ionizing tracks (15° to 60°) with a total of more than 800k events and a 2×10^5 gas gain [2, 3].

To apply the cluster counting technique successfully, the pulses associated with electrons from different clusters should have a small probability of overlapping in time, but the time distance between electrons coming from the same cluster should be short enough to prevent over-counting. The trade-off for an efficient CC, in a helium-based drift chamber case, requires the usage of a high front-end electronics bandwidth and a high sampling rate, respectively of at least 1 GHz and 1 GSa/s at 12 bits, entirely based on a DRS system [4], which has been crucial in our measurements. Moreover, retrieving the correct electron and clusters time arrival information requires advanced techniques able to spot peaks above the noise level in discretized signals.

In this work, we show the data analysis results concerning the ascertainment of cluster counting technique Poisson nature, the establishment of the most reliable electrons clustering algorithms, and the identification of limiting effects for a fully efficient cluster counting, like the space charge density around the sense wire and the dependency on attachment and recombination effects.

2. Electron Peak Finding Algorithms

Two electron peak finding methods have been used to retrieve the results shown in section 3. Before applying them, a preprocessing phase for data collected during beam tests is needed to set the waveform baseline to zero. For this purpose, we define and subtract to each waveform a root mean square (r.m.s.) error performing an average over its first 30 ns in which only noise effects are present.

The first technique used is a second derivative-based (DERIV) algorithm. It computes the first and second derivatives as the ratio between amplitude averages over two consecutive bins
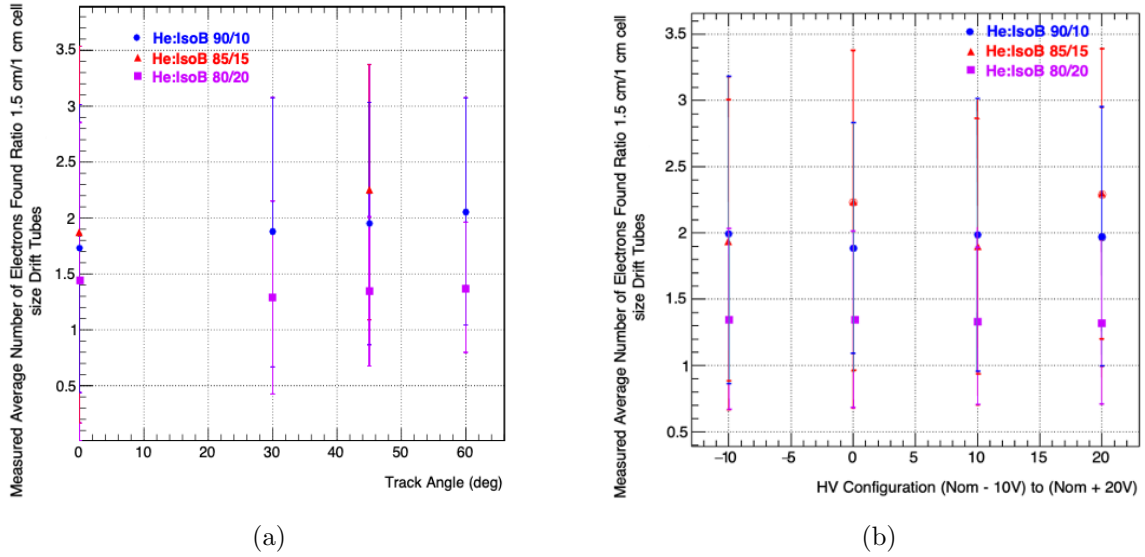
Figure 2: Ratio plots between the average number of electrons found by the DERIV algorithm for 1 cm and 1.5 cm cell drift tubes as a function of the angle (2a) complementary to the one formed between the sense wire and the muon track passed in the detector and the drift tube high voltage configuration sets for different helium-based gas mixtures (2b).

and two times their bin size depending on the sampling rate. Subsequently, it requires that, at the peak candidate position, the derivatives are less than a scanned quantity proportional to the waveform r.m.s. (ϵ) and they increase (decrease) before (after) the peak candidate position of a ϵ-based small quantity. R.m.s.-based lower limit boundaries on the amplitude at peak candidate positions and on the voltage difference between the peak candidates amplitude and the previous (next) bin ones are imposed as well (see the example in Figure 1a).

A second methodology, the running template algorithm (RTA) is under test. It defines an electron pulse template having a raising (K_{rise}) and falling (K_{fall}) exponential over a fixed number of bins (K_{tot}) and digitized (A(k)) to mimic the data behavior (see the example in Figure 1b). The algorithm scans the waveform and runs over the K_{tot} bins window by comparing the electron pulse shape to the normalized data, builds an agreement assessment quantity (χ^2), and defines a cut on it. RTA subtracts the found peak from the signal spectrum and iterates the search until no new peaks are found. The electron pulse is fine-tuned for 1.2 GSa/s setting and its sampling rate dependence is under implementation.

In both cases, the ratios of measured electron peaks found for different size drift tubes as a function of scan angle, high voltage, and sampling rate are compatible with the expectations within error bars (see Figure 2a, and Figure 2b). Those plots are more reliable than the efficiency ones since the expected values for the number of electrons (clusters) strongly depend on assumptions for the number of clusters/cm [5] for a minimum ionizing particle (MIP) for a given drift cell-size and the relativistic rise correction factor [6]. Finally, the expected Landau distribution behavior for the number of electron peaks found has been obtained (see Figure 3a).

3. Clusterization Strategy and Results

The primary ionization cluster identification (CLUSTER algorithm) consists in associating properly the electron peaks recognized by the algorithms described in section 2, and defining a
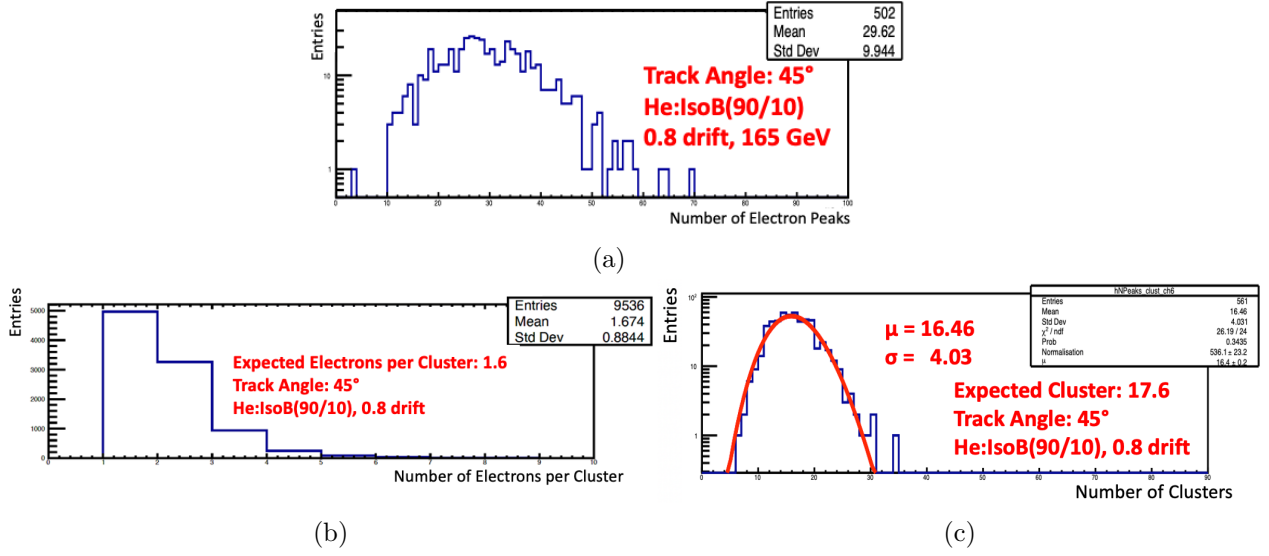
(a)

(b)

(c)

Figure 3: Examples of number of electrons (3a), electron per cluster (3b) number of clusters (3c) distributions found by the DERIV algorithm for 1 cm cell size tubes.
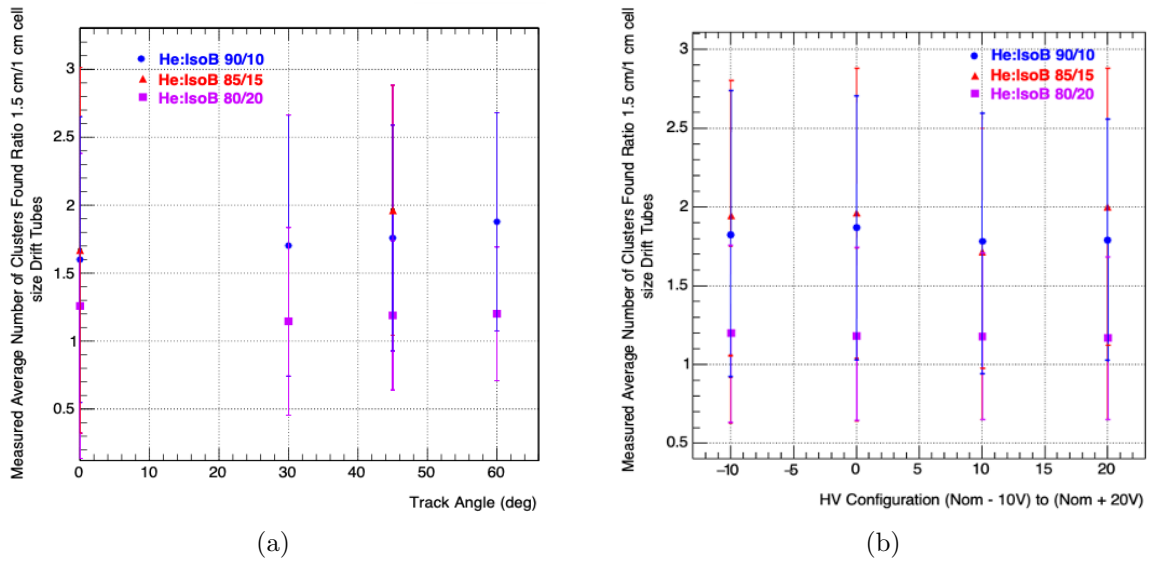

(a)

(b)

Figure 4: Ratio plots between the average number of clusters found by the DERIV algorithm for 1 cm and 1.5 cm cell drift tubes as a function of the angle (4a) complementary to the one formed between the sense wire and the muon track passed in the detector and the drift tube high voltage configuration sets for different helium-based gas mixtures (4b).

time position and amplitude for the cluster. Firstly, it merges the electron peaks in consecutive bins into a single one to reduce the fake electron rate.

Secondly, contiguous electron peaks that are compatible with the electrons' diffusion time are associated with the same ionization cluster. For them, a counter for electrons per cluster is incremented. The CLUSTER algorithm takes into account that the electrons' diffusion time has a dependence on the electron time arrival $\approx \sqrt{t_{ElectronPeak}}$ and on the gas mixture used (e.g. the drift velocity changes of $\approx 20\%$ passing from He:IsoB 90/10 to 80/20).
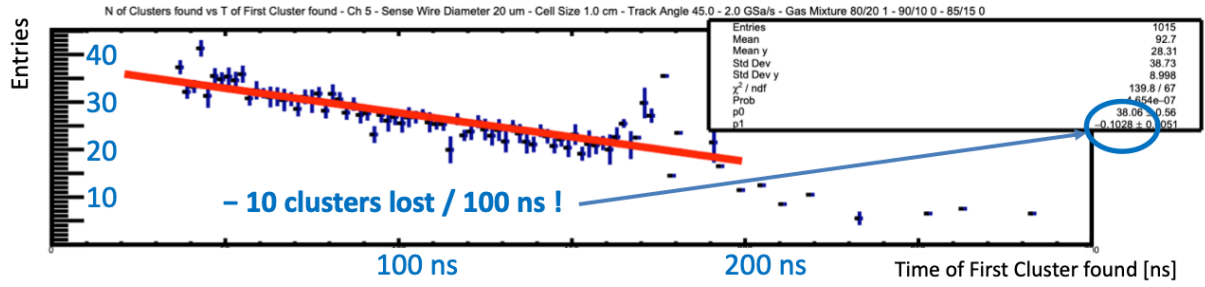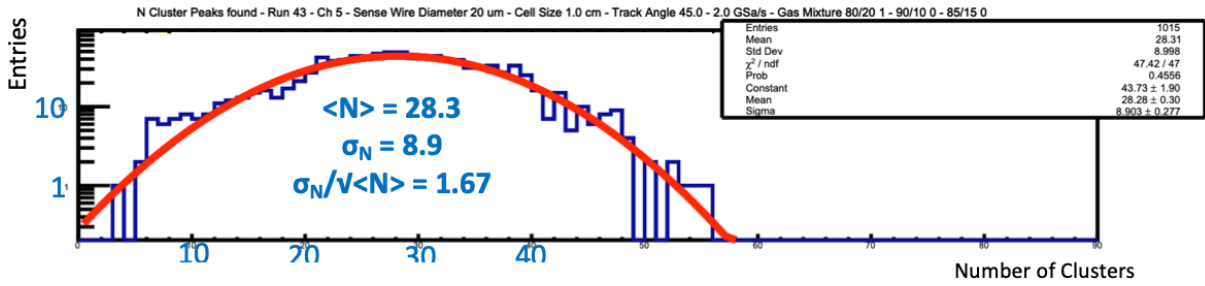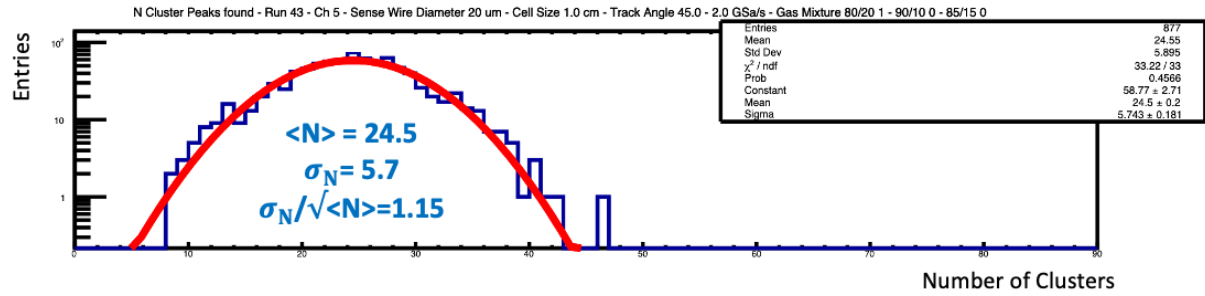
Figure 5: (5a): Plot of the average number of clusters (at a certain drift time) found by the DERIV algorithm as a function of the arrival time of the first cluster (drift time-related) for 1 cm cell-size drift tubes. A cluster loss of ≈ 10 clusters every 100 ns has been observed. (5b) Example of a no-Poissonian number of cluster distribution due to attachment and recombination effects before the correction implemented in (5c).

Finally, the position and amplitude of the clusters corresponds to those of the electrons having the highest amplitude in the cluster. An example of the CLUSTER outcome for the number of electrons per cluster distribution and the number of clusters found are shown respectively in Figure 3b and Figure 3c. The ratio plots are in good agreement with expectations as well (see Figure 4a). Moreover, results on the electron drift velocity values can be retrieved from the first cluster drift time distribution and they are compatible with the expected values (e.g. of $\approx 2.5 cm/\mu s$ for 1 cm (2 cm) cell size tubes, taking into account their maximum impact parameter of 0.5 cm (0.9 cm)). From Figure 5a we spotted the combined effect of recombination, electron attachment, and electric field suppression due to space charge affect the electron finding peak algorithm performance. Cuts on DERIV and RTA algorithms, which were optimized before including the recombination and attachment effects, need to be reformulated (see an example

of uncorrected and corrected cluster distribution in Figure 5b and Figure 5c). Also, these corrections, strongly depend on the drift length and, therefore, on the drift tube size and must be calculated for each different drift tube configuration.

4. Conclusions and Outlook
Despite the challenging environment for the search of electron peaks in real-data waveforms, promising cluster counting algorithms have been tested, and they perform well enough to be used in tracking gaseous detector prototypes for future colliders. Calibration is ongoing to take into account attachment and recombination effects in helium-based drift chambers.

5. Acknowledge

References
[1] F. Cuna *et al* Simulation of particle identification with the cluster counting technique Proceeding at LCWS2021 `arXiv:2105.07064 [physics.ins-det]`.
[2] C. Caputo *et al* Particle identification with the cluster counting technique for the IDEA drift chamber Proceeding Nucl. Instrum. Meth. A **1048** (2023) 167969 `doi:10.1016/j.nima.2022.167969`.
[3] F. Cuna *et al* Particle identification with the cluster counting technique for the IDEA drift chamber PoS ICHEP2022 **335** (2022) `doi:10.22323/1.414.0335`.
[4] S. Ritt *et al* Application of the DRS chip for fast waveform digitizing," Nucl. Instrum. Meth. A **623** (2010), 486-488 `doi:10.1016/j.nima.2010.03.045`.
[5] H. Fischle *et al* Experimental determination of ionization cluster size distributions in counting gases Nucl. Instrum. Meth. A **301** (1991) 202-214 `doi:10.1016/0168-9002(91)90460-8`.
[6] R. G. Kepler *et al* Relativistic increase of energy loss by ionization in gases. Nuovo Cim **7** 71–86 (1958) `https://doi.org/10.1007/BF02746883`.