

# Signal Processing and Machine Learning Algorithms for Precise Timing with PICOSEC Micromegas Detectors

**A. Kallitsopoulou<sup>a,b,1,2</sup> I. Maniatis<sup>a,b,3</sup> I. Manthos<sup>c</sup> T. Papaevangelou<sup>d,e</sup> L. Sohl<sup>d,e,4</sup> A. Tsiamis<sup>a,b</sup> S.E. Tzamaras<sup>a,b</sup>**

<sup>a</sup>*Department of Physics, Aristotle University of Thessaloniki, University Campus, GR-54124, Thessaloniki, Greece*

<sup>b</sup>*Centre for Interdisciplinary Research and Innovation (CIRI-AUTH), GR-57001, Thessaloniki, Greece*

<sup>c</sup>*Institute of Experimental Physics, University of Hamburg, Luruper Chaussee 149, 22761, Hamburg, Germany*

<sup>d</sup>*Université Paris-Saclay, F-91191, Gif-sur-Yvette, France*

<sup>e</sup>*Commissariat à l'énergie atomique et aux énergies alternatives (CEA) - IRFU, F-91191, Gif-sur-Yvette, France*

E-mail: [alexandra.kallitsopoulou@cea.fr](mailto:alexandra.kallitsopoulou@cea.fr)

**ABSTRACT:** High particle rates in current and future experiments make pile-up phenomena a critical issue for extracting useful information. In this context, timing can be important as the 4<sup>th</sup> dimension parameter for triggering or event reconstruction. The PICOSEC-Micromegas detector has been shown to offer precise timing of the order of tens of ps. In this work, novel signal processing algorithms are being developed and evaluated to demonstrate the technology's ability for online precise timing. We propose, an algorithm based on Artificial Neural Networks (ANN). This algorithm uses a model to train the ANN. The performance of the different algorithms is evaluated using experimental data, resulting in a timing resolution of  $18.3 \pm 0.6$  ps, comparable to the standard analysis based on the Constant Fraction Discrimination technique. Additionally, an alternative algorithm using the charge of the pulse exceeding a threshold as a parameter to correct for systematic effects is reported.

**KEYWORDS:** Gaseous detectors, Micromegas, Timing detectors, Analytical Signal Processing, Statistical Methods, Modeling, Artificial Neural Networks

<sup>1</sup>Corresponding author.

<sup>2</sup>Now at CEA-IRFU Université Paris-Saclay, Gif-sur-Yvette, France.

<sup>3</sup>Now at Department of Particle Physics and Astronomy, Weizmann Institute of Science, Rehovot, Israel.

<sup>4</sup>Now at TUV NORD EnSys GmbH & Co. KG.

---

## Contents

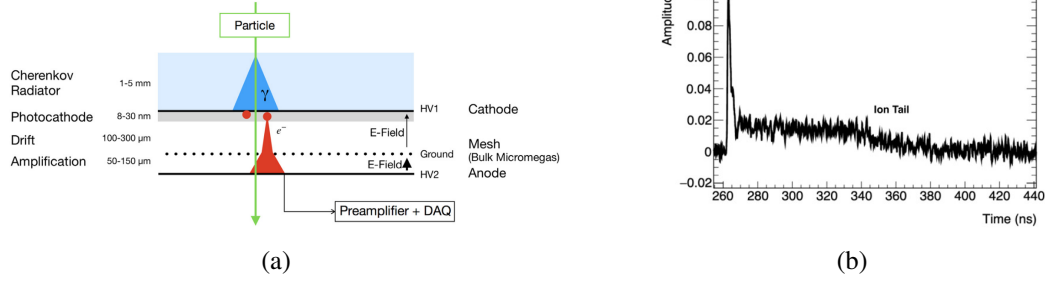
<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Alternative Timing Technique with Constant Threshold</b>	<b>3</b>
<b>3</b>	<b>An Emulation Model for PICOSEC-MM Pulses</b>	<b>5</b>
3.1	Synchronizing and Summing Single Photo-electron Waveforms	7
<b>4</b>	<b>Validation of the Emulation Model</b>	<b>9</b>
4.1	Comparison of the Emulation Model with Experimental Data (EXP-set)	10
4.2	Evaluating the Timing Properties of the Emulation Model	11
<b>5</b>	<b>Using the Emulation Model to Train an Artificial Neural Network</b>	<b>12</b>
5.1	Validation of the Artificial Neural Network's Performance	12
5.2	Validation of the ANN's Unbiased Performance	14
<b>6</b>	<b>Conclusions</b>	<b>15</b>

---

## 1 Introduction

The PICOSEC detector, shown in Fig. 1a, is based on the Micromegas technology [1](hereafter called PICOSEC-MM), and includes a Cherenkov radiator and a photocathode. This design allows to an incoming relativistic charged particle to create Cherenkov photons inside the radiator, which convert to prompt photoelectrons in the photocathode. Additionally, the drift region has been reduced to the order of 200  $\mu\text{m}$ , compared to a typical value of 3-5 mm to decrease the probability of direct ionization in the active volume, whilst also allowing for pre-amplification avalanche in the drift region due to the increased electric field.

The data used for this study were collected at the Saclay Laser-matter Interaction Center (IRAMIS/SLIC, CEA) with the LYDIL laser [2, 3]. The measurements were performed using an optimized PICOSEC-MM prototype with 119  $\mu\text{m}$  drift gap, operated at a high drift field and filled with  $Ne : CF_4 : C_2H_6$  (80:10:10) gas mixture. The electric field for these tests was set to 44 kV/cm for the drift and 21 kV/cm for the amplification detector's volumes. The photocathode consisted of a thin film (18 nm) of CsI deposited on a 10 nm Aluminum layer, serving as the cathode. The laser beam was split, with one part directed at the PICOSEC-MM detector and the other illuminating a high-precision photodiode, with  $\sim 3$  ps timing resolution, serving as a timing reference. The light beam that reached the detector was properly attenuated to control the number of photons hitting the PICOSEC photocathode, allowing us to study the detector's response to either a single-photoelectron or few, prompt photoelectrons [4]. The Signal Arrival Time (SAT) of the PICOSEC-MM was measured relative to the photodiode's signal on an event-by-event basis



**Figure 1:** (a) A graphical representation of a PICOSEC MicroMegas detector [2]. (b) The typical pulse of a PICOSEC-MM Detector is a two-component signal. It comprises a fast e-peak and an extended ion tail of the slow-moving ions.

(hereafter called reference signal). Previous studies have shown that this optimized PICOSEC-MM detector has the capability of timing the arrival of a single photon with an accuracy better than 50 ps [5].

In this report, novel timing methods are being developed and evaluated using data sets collected with the optimized PICOSEC-MM detector. The arrival time of a PICOSEC-MM signal (Fig. 1b) is typically extracted by analyzing the fully digitized electron peak waveform (e-peak), i.e. by fitting a logistic function to the leading edge of the e-peak. The SAT was then defined at the 20% of the e-peak maximum relative to the time-reference device signal (hereafter, this procedure is called full offline analysis, and the arrival time of the time-reference device is called reference time). Finally, the detector’s timing resolution was defined as the RMS of the SAT distribution.

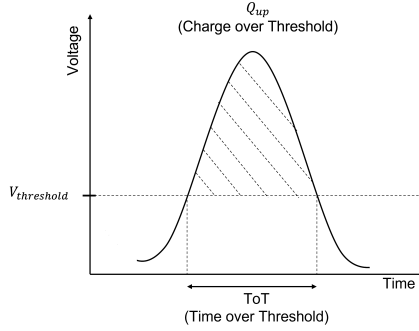
Two types of data sets were used: the SPE-set, comprising waveforms of single photoelectrons, and the EXP-set, composed of waveforms of multiple photoelectrons. All waveforms were digitized by a fast (20 GSamples/s), high-bandwidth oscilloscope (2.5 GHz) after passing through a CIVIDEC current-sensitive preamplifier. Section 2 first introduces an alternative timing method, based on the time-over-threshold technique. This algorithm uses the integral of the pulse above multiple thresholds (i.e. the equivalent of the charge) to correct for systematics. Building on the insights gained from these analytical signal processing approaches, the SPE-set is then used to emulate PICOSEC-MM signals with multiple photoelectrons, as described in Section 3. After evaluating the performance of this model in Section 4, it is used to train an Artificial Neural Network (ANN) for timing the arrival of PICOSEC-MM waveforms, as detailed in Section 5. In all methods discussed, the performance of the timing algorithms is assessed using the EXP-set waveforms.

It should be noted that the ANN-based timing technique performs optimally when the PICOSEC signal is relatively unaffected by time-walk. Due to the thin gap design, the prototype used in this study has a minimal time-walk effect, typically of only a few picoseconds [6]. However, in cases where the signals experience more significant time-walk dependencies, the technique is no longer a good approximation, potentially degrading the timing resolution.

## 2 Alternative Timing Technique with Constant Threshold

In our effort to propose an efficient timing technique, for scalable electronic devices with the PICOSEC-MM and reduce the cost of the experimental set-up, we evaluated the Constant Fraction method [7] on extracting the timing resolution of the PICOSEC-MM detector, by using the minimum information required. The goal of the developed signal processing algorithm is to retain the timing results achieved by the Constant Threshold method that takes into account the full waveform. In this approach, instead of using the time over threshold measurements, we use the integrals of the e-peak when exceeding certain thresholds. Hereafter such integrals will be referred to as *Charge over Threshold* and symbolized as  $Q_{up}$ . This technique, like the Time over Threshold, suffers from systematics, introduced by the strong dependence of SAT on the pulse height. We prove that we can extract the experimental information that can to correct this effect, achieving precise timing accuracy [8, 9].

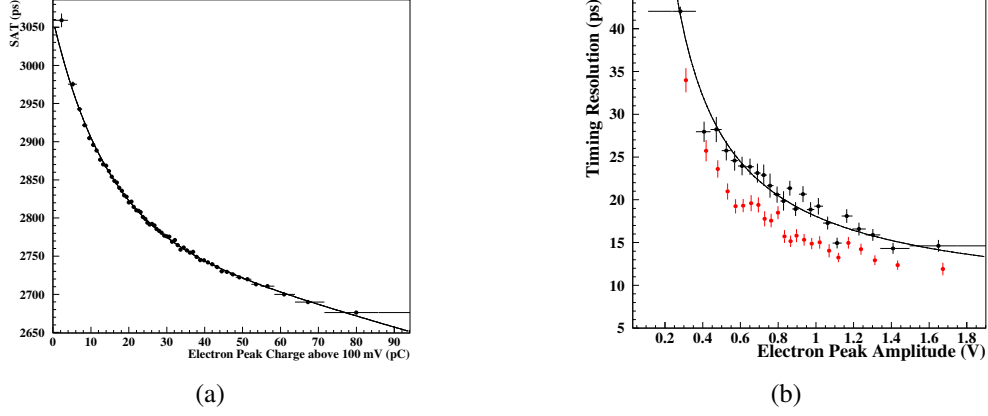
This technique requires thresholds adjusted to the amplitude range of the available experimental data. For the EXP-set, 4 thresholds were selected: 100 mV, 200 mV, 400 mV and 600 mV. The two crossing points of the threshold with the rising and falling edges of the waveform are used to integrate this region, where available, and calculate each  $Q_{up}$  value of the corresponding pulse. To make available the information of the e-peak amplitude, read-out electronics to digitize the waveform are required. However, on a large-scale experimental set-up, this will imply a significant number of channels, which can skyrocket the cost of the experimental set-up. Introducing pulse integrators such as the NINO-chip [10], or other new generation front-end electronics, in addition to the discriminator device, can achieve the time resolution of the Constant Fraction Discrimination Technique, at a significantly lower cost.



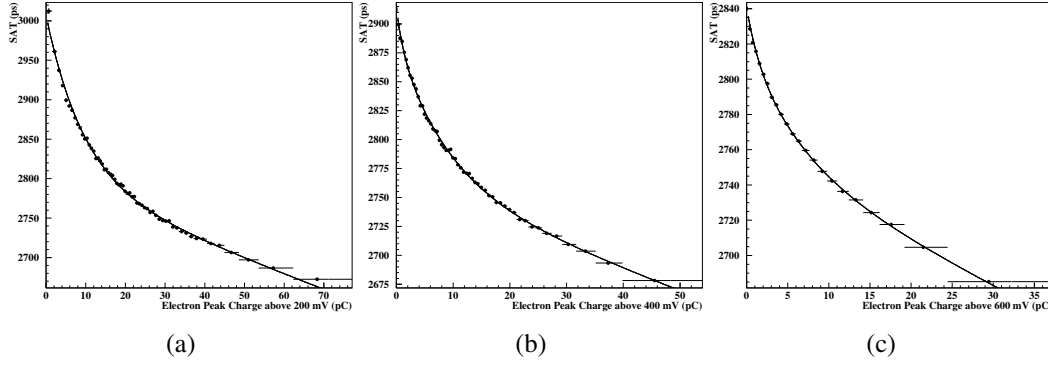
**Figure 2:** Definition of the charge above relevant threshold,  $Q_{up}$ .

The behaviour of mean SAT, can be evaluated as a function of e-peak charge above the applied threshold (100 mV), Fig.3a, parameterising the time walk effect, using a power law plus a constant term fit. Correcting the SAT for time walk using this fit results is presented with black points on Fig.3b. The algorithm that provides results with the best time resolution uses the lowest threshold for defining the SAT, and the parameterization as a function of  $Q_{up}$  of the highest available threshold for time walk correction, on an event-by-event basis [8]. This signal-processing algorithm minimizes the experimental information to a few numbers provided by the pulse integrators.

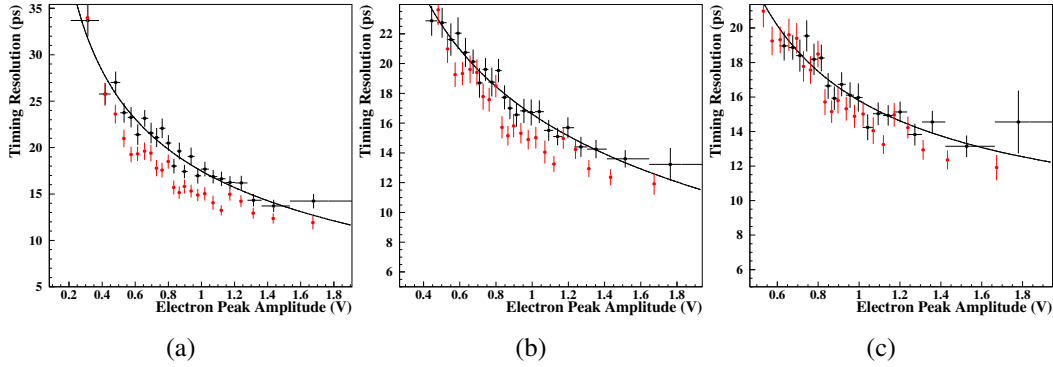
The event-by-event time walk correction using the calibration curve of the higher available



**Figure 3:** (a) SAT as a function of Charge above Threshold for the lowest threshold at 100 mV, fitted with the corresponding power law curve. (b) Resolution as a function of e-peak amplitude after time walk correction, using both Constant Fraction Discrimination (red points) and multi-charge over threshold timing techniques (black points), for the same amplitude threshold.

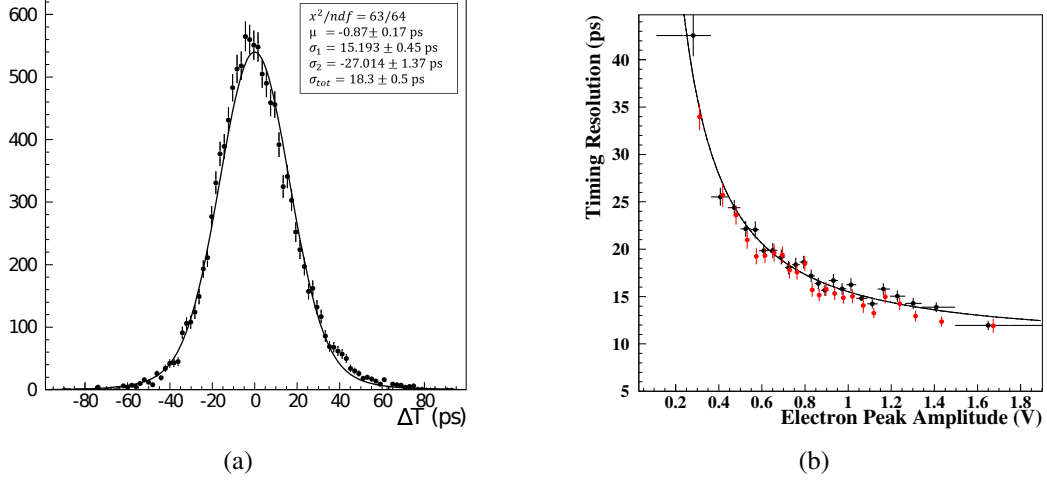


**Figure 4:** SAT as a function of Charge above Threshold for the three other applied thresholds, fitted with the corresponding power law curve. Threshold at: (a) 200 mV, (b) 400 mV and (c) 600 mV.



**Figure 5:** Time resolution as a function of e-peak amplitude after time walk correction, using both Constant Fraction Discrimination (red) and multi-Charge over threshold timing techniques (black), for threshold at (a) 200 mV, (b) 400 mV, (c) 600 mV.

crossing threshold requires calibration measurements before any analysis ( Fig.4 ). The analysis of the EXP-set, using the  $Q_{up}$  technique, is shown in Fig.6a, resulting in a timing resolution performance of the detector, as the one presented in Fig. 6b, with black points. This is similar to the one achieved with the CFD Technique, which is expressed with red points in the same Figure and signifies the successful performance of the timing algorithm. After correction for systematic effects, SAT distribution is symmetric, with an RMS of  $18.3 \pm 0.5$  ps. The same results were achieved with the full offline analysis with CFD reported also in [8].



**Figure 6:** (a) SAT distribution of EXP-set after time walk corrections, using the multi-Charge over threshold Technique, resulting in a global timing resolution of  $18.3 \pm 0.5$  ps. (b) Resolution as a function of e-peak amplitude after time walk correction, using Constant Fraction Discrimination (red points) and multi-charge-over-threshold timing techniques (black points).

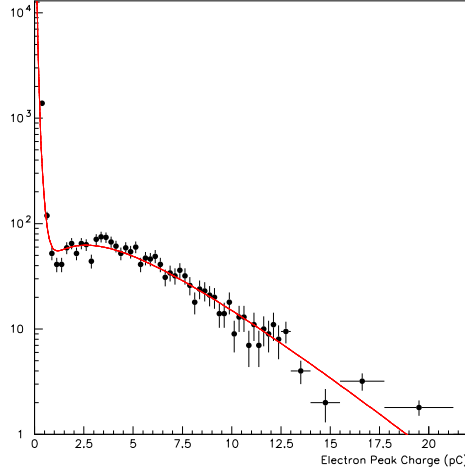
### 3 An Emulation Model for PICOSEC-MM Pulses

While analytical signal processing algorithms provide valuable benchmarks for timing extraction, their complexity can limit applicability in real-time environments. To address this, previous studies, such as [9, 11], have shown that ANN techniques can be used for online estimation of the arrival time of PICOSEC-MM waveforms with accuracy similar to the results of the full offline analysis. However, these studies were limited by the size of the available ANN training set and therefore relied on k-fold cross-validation methods for analysis [12]. In this work, a method for simulating large sets of waveforms is developed to train the ANN and provide more conclusive results.

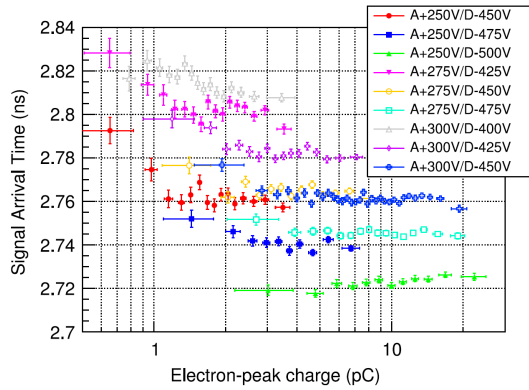
The fully digitized waveforms of the SPE-set are used to emulate waveforms of the EXP-set, on the response of the PICOSEC-MM to a few pes. As it has been previously shown (e.g. [2, 3]), the e-peak charge distribution of the single-pe waveform is well described by the Gamma probability distribution function (pdf) of Eq.3.1:

$$P_1(Q; \bar{Q}, \theta) = \frac{(1 + \theta)^{(1+\theta)}}{\bar{Q} \Gamma(1 + \theta)} \left( \frac{Q}{\bar{Q}} \right)^\theta \exp \left[ - (1 + \theta) \frac{Q}{\bar{Q}} \right] \quad (3.1)$$

where  $\bar{Q}$  is the mean value of the distribution and  $\alpha = \theta + 1$  is the shape parameter. Fig.7 shows the e-peak charge distribution of the SPE-set while the solid line is the graphical representation of Eq.3.1 with  $\bar{Q} = 0.126$  pC and  $\theta = 1.72$  (corresponding to an RMS of 0.077 pC).

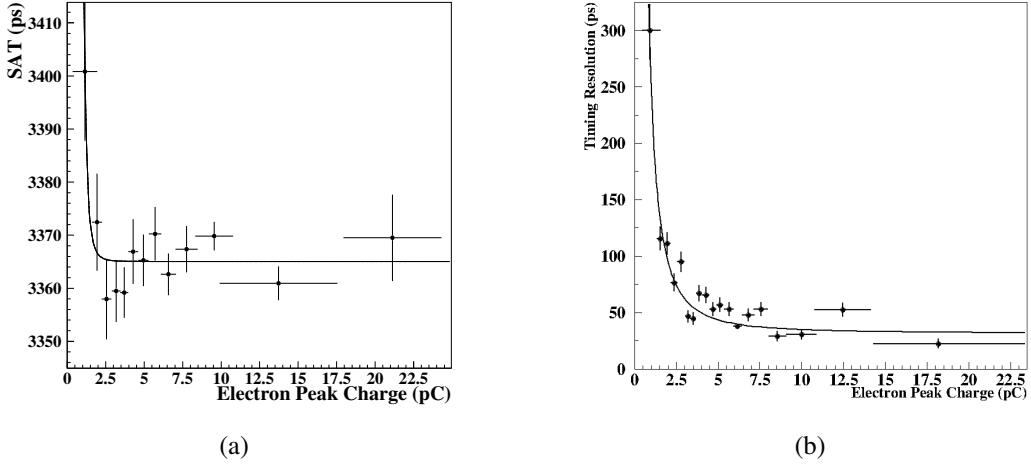


**Figure 7:** Single Photo-electron Charge Distribution, of the SPE-set. The red solid line corresponds to the fitting curve, the convolution of an exponential for the falling edge of the noise peak, and a Polya distribution for the rest of the data.

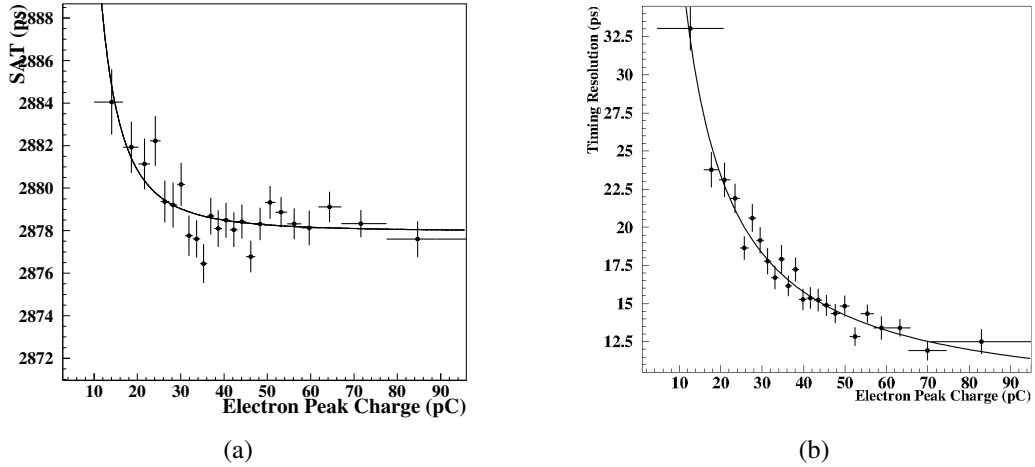


**Figure 8:** Dependence of SAT on e-peak charge, for different voltage settings [3].

Operating the detector in such a high gain (i.e. 44 kV/cm at the drift and over 21 kV/cm at the amplification stage, generated by 525/275 V ), while creating the two data-sets, leads to a very small dependence of the pulses arrival time on their size, meaning that there is almost no time walk effect, as can be seen in Fig.8 . Using the full offline analysis algorithm, we can extract the timing performance of the two data sets as a function of the e-peak charge. Fig.9a, shows the SAT as a function of the e-peak charge, confirming the small effect of the time-walk on the single photoelectron data-set, and it's corresponding resolution, as in Fig.9b. Respectively, Fig.10 represents the SAT of the EXP-set and its corresponding resolution, as a function of the e-peak charge.



**Figure 9:** (a) Dependence of the SAT as a function of e-peak charge, for the SPE-set pulses. The solid line represents a power law fit to the experimental data. (b) Timing resolution as a function of the e-peak charge of the SPE-set. The solid line represents a double exponential fit to the experimental data.



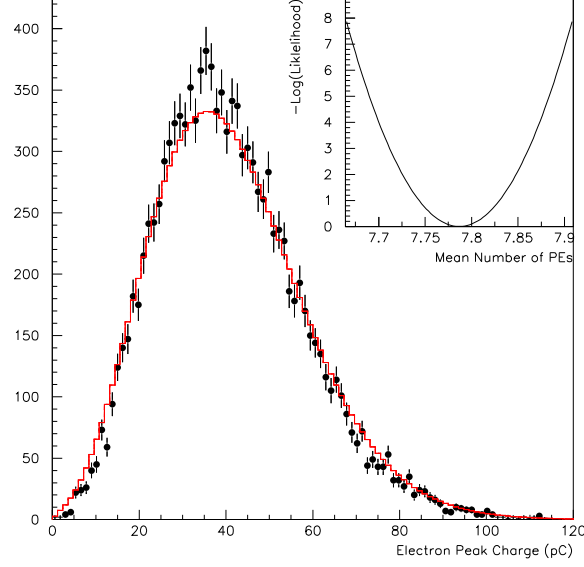
**Figure 10:** (a) Dependence of the SAT as a function of e-peak charge, for the EXP-set pulses. (b) Timing resolution as a function of e-peak charge for the EXP-set. The solid line corresponds to a power law fit to the data points, both for the SAT and Timing Resolution data points.

### 3.1 Synchronizing and Summing Single Photo-electron Waveforms

In the case that the PICOSEC-MM signal is produced by  $N_{pe}$  photoelectrons the charge distribution of the corresponding e-peak is described by the convolution of  $N_{pe}$  identical pdfs, following the functional form of Eq.3.1. This convolution results in the multi-pe pdf of Eq.3.2:

$$P_N(Q; \bar{Q}, \theta, N_{pe}) = \frac{(1 + \theta)^{N_{pe}(1+\theta)}}{\bar{Q} \Gamma(N_{pe}(1 + \theta))} \left( \frac{Q}{\bar{Q}} \right)^{N_{pe}(\theta+1)-1} \exp \left[ - (1 + \theta) \frac{Q}{\bar{Q}} \right] \quad (3.2)$$





**Figure 11:** (a) E-peak Charge Distribution of EXP-set. With red solid line is the prediction of the deconvolution algorithm, as a fit to the experimental data (black points). In the inset plot the Log-Likelihood minimization for the estimation of the number of photoelectrons.

As in [13], we assume that the EXP-set comprises PICOSEC-MM waveforms corresponding to  $N_{pe}$  photoelectrons, where  $N_{pe}$  follows a Poisson distribution with a mean value equal to  $\mu_{pe}$ . Then, the e-peak charge distribution of the EXP-set waveforms should follow the following pdf:

$$G(Q; \mu_{pe}, \bar{Q}, \theta) = \sum_{N_{pe}=0}^{\infty} \frac{\mu_{pe}^{N_{pe}} \cdot e^{-\mu_{pe}}}{N_{pe}!} \cdot P_N(Q; \bar{Q}, \theta, N_{pe}) \quad (3.3)$$

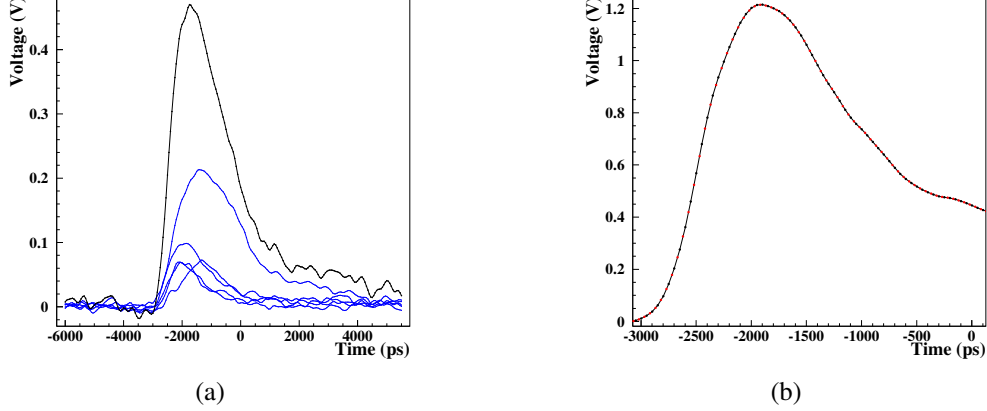
where  $P_N(Q; \bar{Q}, \theta, N_{pe})$  is given by Eq.3.2.

The e-peak charge distribution of the EXP-set waveforms is represented with the solid red line of Fig.11. The mean value of the number of photoelectrons, that have produced these waveforms, has been estimated by fitting Eq.3.3 to the above data points. In this fit, the  $\bar{Q}$  and  $\theta$  parameter values have been fixed to 0.126 pC and 1.72 respectively, which describe well the single-pe charge distribution (Fig.7). The fit results to  $\hat{\mu}_{pe} = 7.8 \pm 0.1$  pes. A graphical representation of Eq.3.3 with  $\hat{\mu}_{pe} = 7.8$  is shown, as the solid red curve, in Fig.11.

Supposing that PICOSEC-MM (including the front-end electronics) is a linear device, the detector response to N-photoelectrons should be emulated by the sum of N single-pe waveforms. For the simulation of waveforms like those comprising the EXP-set, we implement the following algorithm:

A multiplicity, N of single-photoelectron (single-pe) signals is chosen according to a Poisson distribution with a mean value of 7.8. N waveforms are randomly selected among those of the SPE-set. Each selected waveform is then time-shifted, using a common set of time coordinates for all the SPE-set, where zero corresponds to the reference time. To accurately estimate the relative position of each single-pe signal, a  $3^{rd}$  degree polynomial interpolation is applied between the digitization points of the leading edge. This interpolation ensures that the pulses are synchronized,

allowing for the summation of these signals into an equivalent multi-pe pulse. Synchronizing the reference time for all pulses by setting them to zero, creates a consistent clock for triggering and digitization. However, this synchronization removes the natural time jitter. To replicate this, we artificially introduce a Gaussian shift of  $\pm 50$  ps (with a standard deviation of 25 ps) to each summed pulse. This can be seen in an example of a simulated pulse in Fig. 12a. An example application of the algorithm, using  $N = 5$  single-pe waveforms, is shown in Fig. 12b.



**Figure 12:** (a) Example of the simulated pulse of the PICOSEC-MM response to 5 photoelectrons (black), by summing randomly selected 5 waveforms of the SPE-set (blue). (b) Example of a simulated multi-pe pulse shifted within  $\pm 50$  ps. Black points are the original digitizations of the waveform, while the red points are the shifted ones. The line connects the black digitization points.

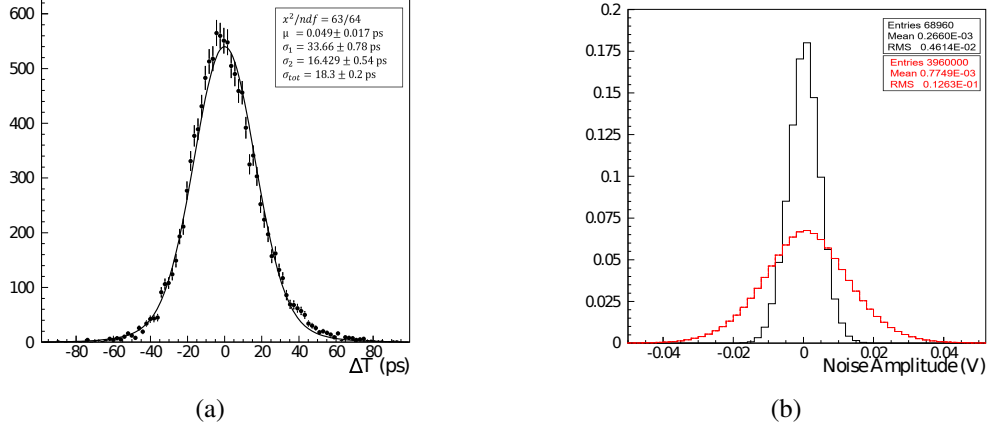
Using the bootstrap technique [14] and the Poisson Distribution as a metric, we can generate random samples of single-pes to create simulated multi-pe pulses. However, this method has the disadvantage of introducing a correlation between the generated pulses since the final pulses share similar single-pe waveforms.

#### 4 Validation of the Emulation Model

The data set of emulated multi-pes waveforms is expected to be identical to the response of the PICOSEC-MM detector on the laser beam (EXP-set). and share the same timing properties as the EXP-set. EXP-set data have been fully analyzed offline, following the algorithm described in [15]. After applying time walk corrections, parameterized as a function of the SAT to the e-peak charge, results in a timing resolution of  $18.3 \pm 0.2$  ps.

The full offline analysis of PICOSEC-MM waveforms, requires the calculation and subtraction of the baseline offset, while the baseline RMS is also calculated on an event-by-event basis [16]. For the SPE-set, the baseline RMS distribution is presented with black in Fig. 13b. Thus, during the summing process, the generated pulses present an increasing RMS baseline noise (red-colored distribution in Fig. 13b), analogous with the square root of the number of photoelectrons.

The simulated pulses suffer from extra artificial noise, caused by the summing method of the single-pes pulses. For the simulation, data of the SPE-set were used with e-peak charge greater than 1 pC, or amplitude of 30 mV, to reduce the noise contribution. Unfortunately, this selection

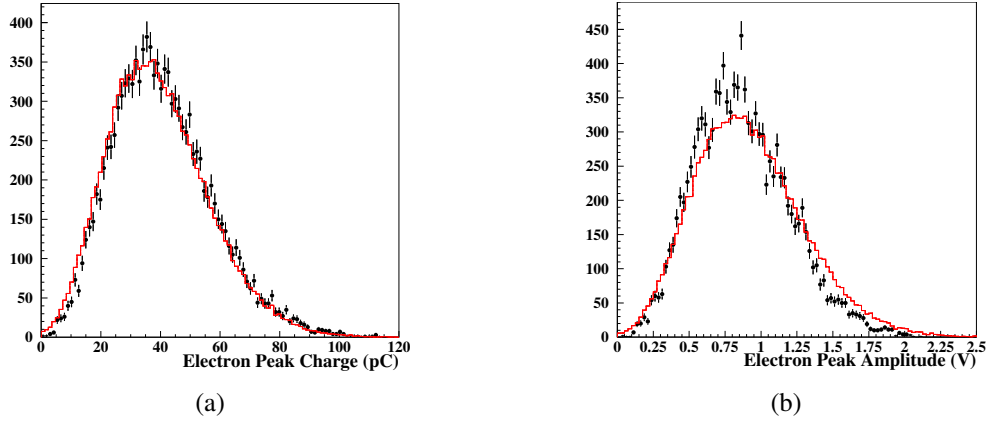


**Figure 13:** (a) SAT distribution of the EXP-set using the CFD Technique, resulting to a global timing resolution of  $18.3 \pm 0.2$  ps after time walk corrections and fitted with the weighted sum of two Gaussian distributions. (b) Distribution of baseline RMS noise on single-pes(black) and emulated multi-photoelectrons(red).

cut restricts the number of available events, hence the missing part of the spectrum needs to be replaced. This replacement is realized by scaling down larger pulses, in accordance with the Polya charge distribution.

#### 4.1 Comparison of the Emulation Model with Experimental Data (EXP-set)

As a cross-reference, to validate the simulated results, we compare them with the EXP-set. The distributions of e-peak charge and the peak amplitude are presented in Fig. 14a, where black points are the simulated events and red are the experimental ones.



**Figure 14:** (a) E-peak Amplitude and (b) E-peak Charge Distributions for both the simulated-set (black) and the EXP-set (red) of waveforms.

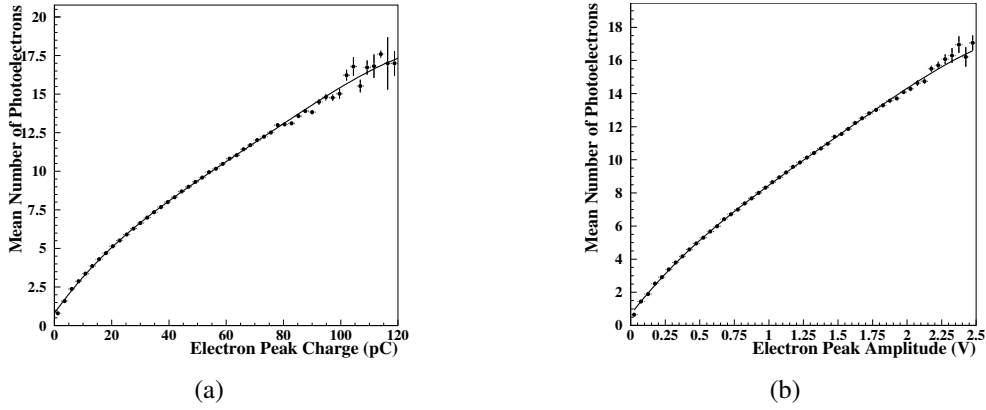
Fig. 14b also provides evidence that EXP and simulated sets differ slightly regarding the e-peak amplitude distributions. Defining the maximum value for the amplitude of a signal requires the localization of the point with the highest value. It is obvious, that this measurement is biased from

noise contribution. As shown (Fig. 13b), simulated signals suffer from extra noise and as a result, the peak amplitude distribution is shifted to higher amplitude values. This constitutes a limitation of the model, which has to be considered in the full offline analysis of the simulation set. On the contrary, the e-peak charge is a more reliable parameter for the simulation, as it neutralizes noise effects and produces a similar distribution to the experimental data.

## 4.2 Evaluating the Timing Properties of the Emulation Model

The evaluation of the timing properties of the model and their comparison with the EXP-set using the same offline analysis method, as performed in [8], will ensure the validity of the model.

It is essential, for the analysis of the emulated pulses, to quantify the correlation between e-peak size and the number of photoelectrons. For this, the distribution of e-peak charge is divided by the distribution of e-peak charge, weighed with the number of photoelectrons. The result is plotted as a function of e-peak charge in Fig. 15a, while the errors are calculated from the e-peak charge distribution weighted with the squared number of photoelectrons. The same results can be obtained for an analysis using the e-peak amplitude as a parameter (Fig. 15b). Each plot is parameterized using a 4<sup>th</sup> degree polynomial.



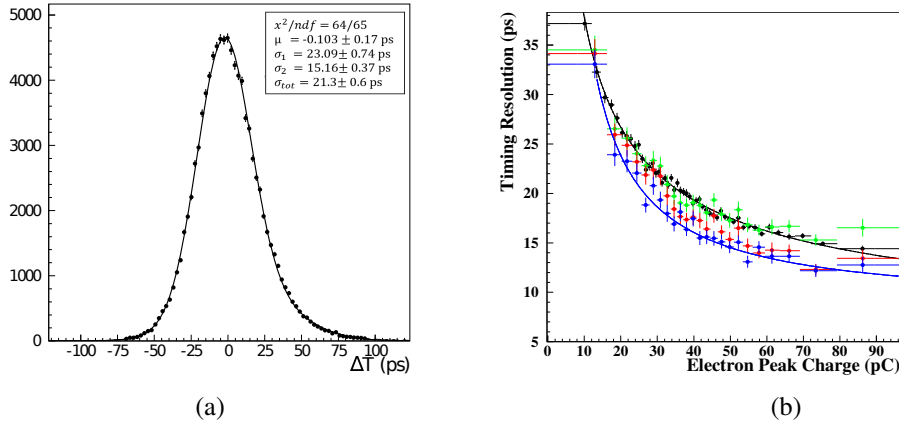
**Figure 15:** Correlation plot for the number of photoelectrons estimation relevant to e-peak size. (a) E-peak Charge distribution and (b) E-peak amplitude distribution. In both plots, the black solid line represents a 4<sup>th</sup> degree polynomial fit to the data points.

The full offline timing analysis of the simulated pulses, using the calibration curves for time walk correction, results in a timing resolution of  $21.3 \pm 0.6$  ps, as it can be seen in Fig. 16a. The timing resolution of the model is presented with black points in Fig. 16b. The timing precision of the model is lagging by 3 ps compared to EXP-data, shown as blue points in Fig. 16b. The deteriorated resolution of the simulated data is attributed to two main sources: the baseline RMS noise and the synchronization limitations from using a common time reference. The sum of  $\bar{N} = 7.8$  single-photoelectron pulses accumulates additional baseline RMS noise, while the timing accuracy of the reference device restricts resolution based on the number of photoelectrons,  $N$ . Though noise accumulation constrains the method for developing new pulses, these two error components can be added to the EXP-data, making both sets subject to the same noise.

To account for additional noise in the EXP-set, each data point is shifted event by event based on the number of photoelectrons, which provide signal information. The extra RMS noise, represented as  $\sigma_1 = \sqrt{N_{pe}} \cdot \sigma_{1pe}$ , where  $\sigma_{1pe} = 0.0046$  (estimated in the Fig.13b), is added by shifting waveforms through a Gaussian with mean equal to zero and  $\sigma = \sigma_1$ . This adjustment shifts the EXP-data points toward the simulated ones, as seen with the red points in Fig.16b. Additionally, the photodiode's precision contributes further error based on the number of photoelectrons,  $N_{pe} \cdot \sigma_{phd}$ , which is combined quadratically with the RMS noise, ultimately shifting the red points to become the green ones, as depicted in the same figure. The total effect results in a variance of :

$$V[Data] = V[Data_{noise}] + V[phd] = N_{pe} \cdot \sigma_{1pe}^2 + N_{pe}^2 \cdot \sigma_{phd}^2 \quad (4.1)$$

where the error of photodiode's precision ( $\sigma_{phd}$ ) is approximately 3 ps[4].



**Figure 16:** (a) SAT distribution of the model, after time walk corrections, resulting to a timing resolution of  $21.3 \pm 0.6$  ps, fitted with the weighted sum of two Gaussian distributions. (b) Resolution as a function of e-peak charge of: Simulated data (black), EXP-set (blue), EXP-set shifted due to extra RMS baseline noise according to the number of pes (red) and EXP-set additionally shifted due to photodiode's precision as a function of the number of pes (green). Lines correspond to fit with power law.

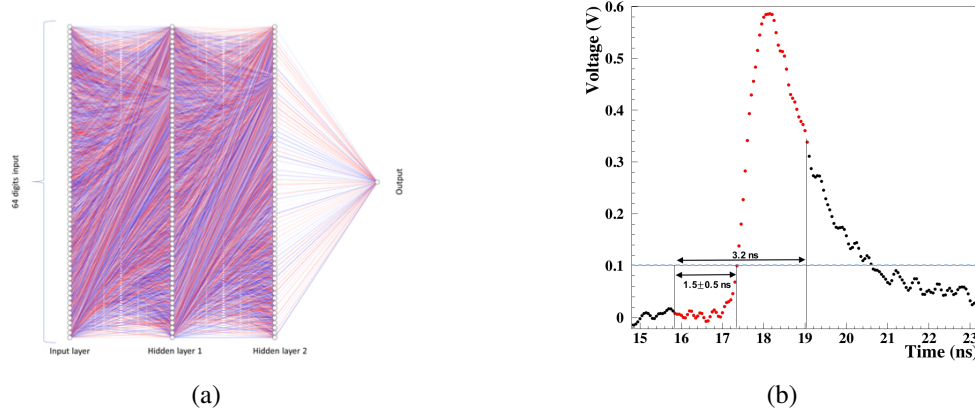
## 5 Using the Emulation Model to Train an Artificial Neural Network

Estimating timing resolution using Artificial Neural Networks can be highly beneficial, enabling real-time or near-real-time timing analysis. Additionally, it can easily be combined with other data analysis techniques, such as machine learning algorithms for signal classification or data filtering. It can become an important tool to enhance overall timing resolution estimation.

### 5.1 Validation of the Artificial Neural Network's Performance

In the present study, the Neural Network implementation was based on a Feed-Forward NN architecture [17]. In this context, the ANN consists of an input layer, two hidden layers with 64 neurons each, and an output layer. This model has been found to give adequate results and can be visualized

in Fig.17a. For all nodes, *ReLU* activation function [18] was used apart from the output layer, an essential part of regression tasks. In parallel, the mean squared error was used as a loss function.



**Figure 17:** (a) Artificial Neural Network architecture comprising an input layer, two hidden layers, and one output layer. The number of input and hidden nodes is defined by the number of selected digitizations of the e-peak waveform. (b) Typical PICOSEC-MM digitized waveform. Red points denote the digitized information presented to the ANN, in this case -the 64 digits around the trigger point- correspond to 3.2 ns time interval.

In the present study, two training scenarios were performed:

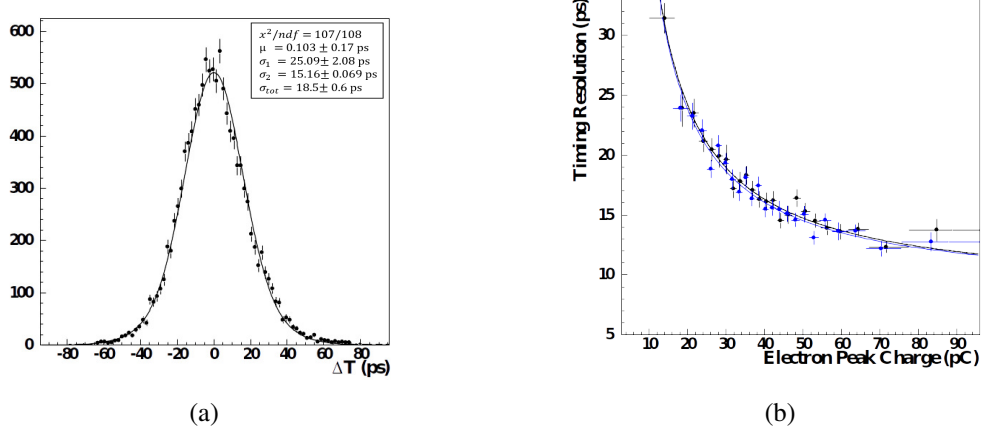
- The predictive capabilities of the ANN were first studied with experimental data using the k-fold cross-validation technique, which was already mentioned in the introduction of Section 3 [12].
- After the model implementation, a large data-set of pulses is available. The training procedure using the model is analogous to the previous case, without the folding. The model was later on tested using the total data set of the experimental set.

The training set used in each case was split into training and validation sets, with a ratio of 0.2. The model was trained for 10000 epochs. At each iteration, the model is evaluated on an unseen validation set, and its weights are updated only if the cost function shows improvement on this validation set.

Waveforms fed in the ANN were like the one seen in Fig.17b. A 100 mV threshold trigger is chosen, giving a selection criterion to the waveforms and providing a timestamp. This timestamp is used to determine the starting time point which is 1.5 ns before the crossing point with the threshold. A shift of the starting time with a Gaussian with  $\sigma=500$  ps, for the ANN to be unbiased of the selected starting point is applied. Subsequently, this will define the total time window of 3.2 ns (64-bit digitization with a sampling rate of 50 ps) given to the input layer of the ANN (red-colored points). Various input lengths were tested, and it was found that the optimum performance was achieved when the input layer contained mostly the leading edge. Moreover, during the training period, the reference time was provided to the network as the target value.

Using the method of training with k-fold cross-validation, only with experimental data (EXP-set, with 7.8 pes), the ANN results in a timing resolution of  $18.5 \pm 0.6$  ps, shown in Fig.18a. In

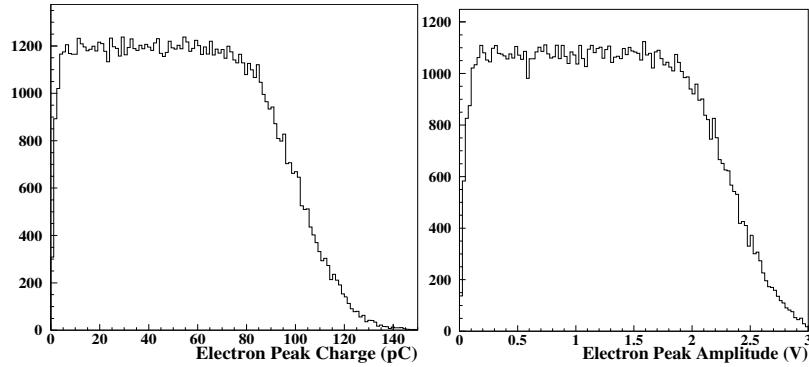
Fig.18b, the timing resolution using the ANN is compared with the results of the full offline analysis using the CFD technique. The similarity confirms the validity of the ANN timing method performance.



**Figure 18:** (a)The PICOSEC-MM Signal Arrival Time, evaluated by the Neural Network, fitted with a Gaussian distribution with RMS that determines the global timing resolution at  $18.5 \pm 0.6$  ps. (b) Timing Resolution as a function of e-peak charge. The ANN results (blue points) are compared with those of the CFD Technique (black points).

## 5.2 Validation of the ANN's Unbiased Performance

The second training procedure approach uses as a training set the simulated pulses. Aiming for a training session unbiased of the e-peak size, the model was developed with uniform charge and amplitude distributions<sup>1</sup>, as shown in Fig.19. Again, the ANN reaches the same timing accuracy as the full signal processing analysis of  $18.5 \pm 0.6$  ps.



**Figure 19:** Training of the ANN with uniform distribution of e-peak amplitude (right), and e-peak charge (left).

Since there is strong evidence that ANN works properly, a variety of tests performed to prove its unbiased and consistent operation. It has been shown that the injection of noise to input training

<sup>1</sup>The uniform region of distribution represents the vast majority of the EXP-set signals

data can act as a regularization technique, preventing the ANN from overfitting. For this reason, a 15 mV Gaussian noise was added to the training phase, resulting in better performance of the neural network algorithm and leaving unaffected the timing resolution. Additionally, aiming to understand whether or not the ANN performs with respect to the signal processing algorithm, a significant amount of noise was added to the testing data-set (EXP-set). Consequently, on an event-by-event basis, the EXP-set was shifted within a Gaussian with mean zero and RMS 30 mV. The full offline analysis of the shifted EXP-set, results in a resolution of  $23.6 \pm 0.5$  ps. For the same data, used as a testing set, the ANN results to  $24.7 \pm 0.6$  ps (Fig.20a).

The next evaluation test uses only the simulated data produced by the model. The simulated data set produced with uniform charge distribution was used as a training set, then the simulated data set was used as a testing set according to the Polya charge distribution of the EXP-set. In this way, the ANN results in a timing resolution of  $21.3 \pm 0.6$  ps, similar to the one achieved with the full offline analysis (Fig.20c).

Another test is to provide fewer input points of each waveform on the ANN training session, i.e. 1.25 ns instead of 1.5 ns (with the same 50 ps spacing). Additionally, the jitter of the timestamp is changed to 100 ps. This procedure results in a timing resolution of  $18.4 \pm 0.6$  ps, similar to the CFD analysis result (Fig.20b). The last test utilizes of different sampling rate with a digitization of 200 ps instead of 50 ps, resulting in 16 digits as the input layer of the ANN. This demands also a different training set, with the same digitization before the ANN train (i.e the multi-pes waveforms will have a wider digitization interval). Performing the full offline analysis on this sample results in a timing precision of  $19.01 \pm 0.5$  ps. Respectively, the ANN achieves  $19.02 \pm 0.5$  ps (Fig.20d).

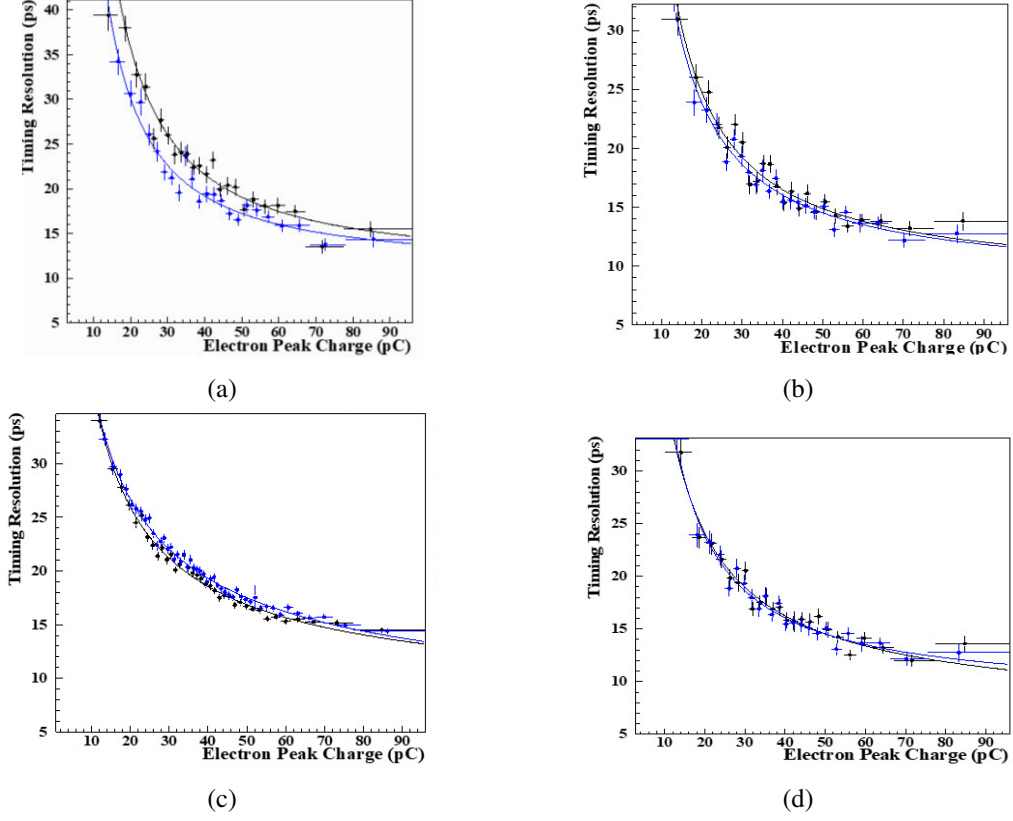
## 6 Conclusions

This study demonstrates the development of advanced signal processing techniques for online, precise timing. The EXP-data set was collected in Laser Beam Tests and consisted of PICOSEC-MM signals with an average of 7.8 photoelectrons (pes). The standard CFD timing technique was used to analyze these signals, resulting in a resolution of  $18.3 \pm 0.2$  ps, which served as a benchmark for evaluating the results of newly developed timing algorithms.

One such alternative algorithm is the Constant Threshold Discrimination timing technique, which uses the charge over threshold as the variable for time walk correction. This technique provides results in excellent agreement with the reference method and can easily be integrated into electronic devices, such as the NINO chip accompanied by ADC measurements.

Another technique proposed in this study is an Artificial Neural Network (ANN) for real-time signal processing, which holds great promise for fast event selection. To train the ANN, a model was created due to the need for a large training sample. After verifying that the model was consistent with the detector's response to many photoelectrons, it was used to train the ANN to extract timing information. The ANN passed all tests with high precision timing performance, proving that it is a reliable and unbiased tool, despite the model used as the training set consisting of waveforms that are not completely uncorrelated. The emulated data proved sufficient for the ANN to learn the signal analysis procedure for precise timing. Significantly, this first implementation was developed in a regime where the detector design ensured minimal time-walk effects—only a few picoseconds in the thin-gap prototype [6]. Under such conditions, the ANN approach is highly effective. However,





**Figure 20:** Timing resolution as a function of e-peak Charge, for the four different tests passed from the ANN(black points), in comparison with the full offline analysis using CFD Technique(blue points). (a) Adding noise of 30 mV at EXP-data-set. (b) Test of the ANN with 1.25ns timestamp. (c) Evaluation of the performance of ANN in simulated data both for train and test processes. (d) Test of the ANN with fewer digitization points (16 instead of 64) as input layer.

in detectors or conditions where time-walk contributions are more significant, the same method may no longer provide an accurate approximation, potentially limiting its timing resolution.

It should also be emphasized that the ANN technique presented here represents only the first architecture tested within this framework. Other ANN-based approaches have already been developed [19, 20] and will be reported in forthcoming publications. These complementary efforts highlight the ongoing progress in applying machine learning to PICOSEC timing, suggesting that further improvements in architecture and training strategies may unlock even higher performance.

Overall, this study highlights how both analytical signal processing algorithms and ANN-based methods can achieve precise timing resolution with PICOSEC-MM detectors. While analytical techniques provide robust and hardware-friendly solutions, ANN approaches hold great promise for flexible, real-time applications—particularly when time-walk effects are minimized.

## Acknowledgments

This research work was co-funded by the Hellenic Foundation for Research and Innovation (HFRI) under the HFRI PhD Fellowship grant (Fellowship Number: 6741). Part of the work was also supported by the Deutsche Forschungs-gemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2121 “Quantum Universe” – 390833306. The authors would like to thank Mr. T. Gustavson and the IRAMIS facility of CEA Saclay, as well as the CERN-RD51 PICOSEC Micromegas collaboration, for providing experimental data from the PICOSEC detector.

## References

- [1] Y. Giomataris, P. Rebougeard, J. Robert and G. Charpak, *Micromegas: a high-granularity position-sensitive gaseous detector for high particle-flux environments*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* (1996) .
- [2] J. Bortfeldt et al., *Picosec: Charged particle timing at sub-25 picosecond precision with a micromegas based detector*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **903** (2018) 317.
- [3] J. Bortfeldt et al., *Modeling the timing characteristics of the picosec micromegas detector*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **993** (2021) 165049.
- [4] L. Sohl, *Development of PICOSEC-Micromegas for fast timing in high rate environments*, Ph.D. thesis, Université Paris-Saclay, 12, 2020.
- [5] L. Sohl et al., *Single photoelectron time resolution studies of the picosec-micromegas detector*, *Journal of Instrumentation* **15** (2020) C04053.
- [6] A. Utrobicic et al., *Single channel picosec micromegas detector with improved time resolution*, 2024.
- [7] K. Georgakopoulou, C. Spathis, G. Bourlis, A. Tsirigotis, A. Leisos, M. Birbas et al., *A 100 ps multi-time over threshold data acquisition system for cosmic ray detection*, *Measurement Science and Technology* **29** (2018) 115001.
- [8] A. Kallitsopoulou, *Development of a simulation model and precise timing techniques for picosec-micromegas detectors*, 2021.
- [9] I. Manthos et al., *Signal processing techniques for precise timing with novel gaseous detectors*, *Journal of Physics: Conference Series* **2105** (2021) 012015.
- [10] F. Anghinolfi et al., *Nino: an ultra-fast and low-power front-end amplifier/discriminator asic designed for the multigap resistive plate chamber*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **533** (2004) 183.
- [11] A. Tsiamis, K. Kordas, I. Manthos, M. Tsopoulou and S. Tzamarias, *Timing techniques with picosecond-order accuracy for novel gaseous detectors*, *Journal of Instrumentation* **17** (2022) C09009.
- [12] F. Chollet, *Deep Learning with Python*, Manning Publications Co., USA, 1st ed. (2017).
- [13] I. Manthos et al., *Recent developments on precise timing with the picosec micromegas detector*, *Journal of Physics: Conference Series* **1498** (2020) 012014.
- [14] K. Ramachandran and C. Tsokos, *Mathematical Statistics with Applications*, Elsevier Science (2009).

- [15] S. Aune et al., *Timing performance of a multi-pad picosec-micromegas detector prototype*, [\*Nucl. Instrum. Meth. A\* \*\*993\*\* \(2021\) 165076 \[2012.00545\]](#).
- [16] Bortfeldt et al., *Precise charged particle timing with the picosec detector*, in *AIP Conference Proceedings*, p. 080009, AIP Publishing LLC, 2019.
- [17] I. Goodfellow, Y. Bengio and A. Courville, *Deep learning*, MIT press (2016).
- [18] A.F. Agarap, *Deep learning using rectified linear units (relu)*, 2019.
- [19] A. Tsiamis, *Development of innovative instrumentation of new methods for the selection and analysis of experimental data towards searching for New Physics at the High Luminosity Large Hadron Collider*, ph.d. thesis, Aristotle University of Thessaloniki, 2024. [10.26262/heal.auth.ir.359504](#).
- [20] E. Chatzianagnostou, *Study of multi-pad picosec-micromegas detector prototype*, m.sc. thesis, Aristotle University of Thessaloniki, 2022, [10.26262/heal.auth.ir.343732](#).