Istituto Nazionale di Fisica Nucleare

POLITECNICO MILANO 1863

RadLab POLITECNICO MILANO 1863

Embedded Artificial Intelligence for Position Sensitivity in Thick Scintillators

Giacomo Ticchi[1,2], Luca Buonanno[1,2], Davide Di Vita[1,2], Fabio Canclini[1,2], Marco Carminati[1,2], Franco Camera[2,3], and Carlo Fiorini[1,2]

[1]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, Italy

[2]Istituto Nazionale di Fisica Nucleare, Sezione di Milano, Milano, Italy

[3]Università degli Studi di Milano Statale, Dipartimento di Fisica, Milano, Italy

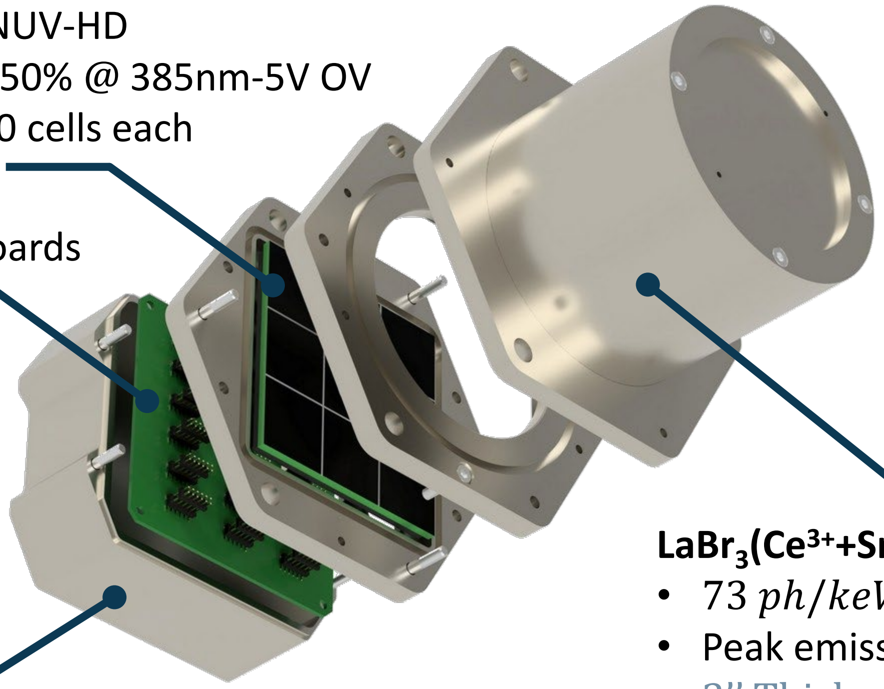
VCI VIENNA CONFERENCE ON INSTRUMENTATION

2022 Vienna Conference on Instrumentation

GAMMA Module

The GAMMA Module is a 144 channels gamma-ray spectrometer composed as in figure.
It can perform spectroscopy measurements from 20 keV to 30 MeV showing a state-of-the-art energy resolution of 2.6% at 662 keV and 1% at 8.9 MeV.

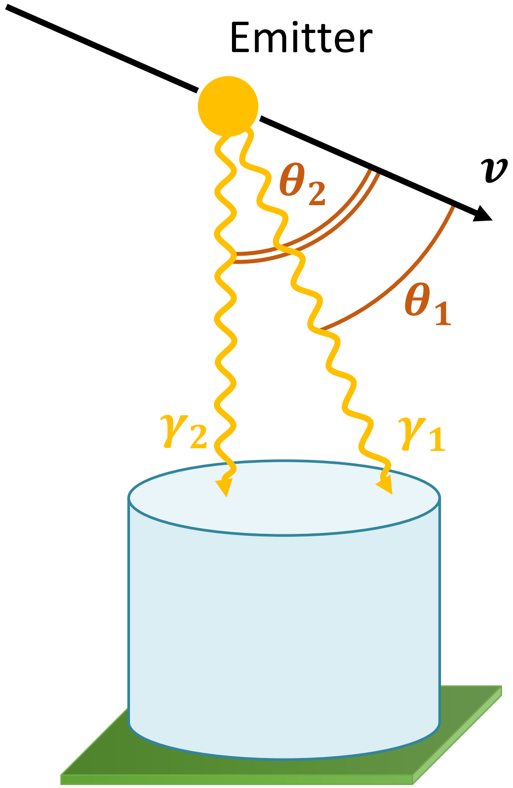
For more information about the GAMMA Module please visit presentation #45 by Davide Di Vita.



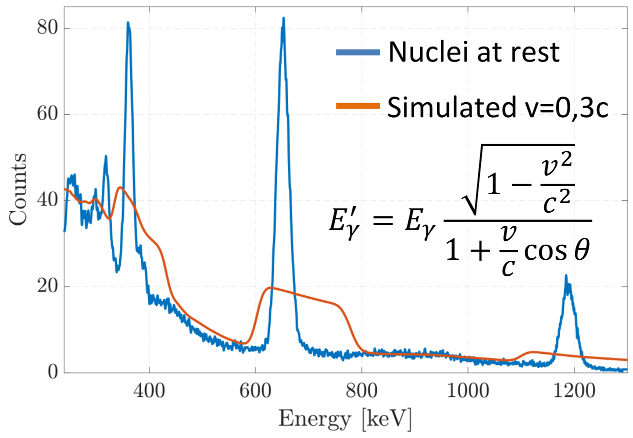
RadLab POLITECNICO MILANO 1863

Project Aim: Compensate Doppler Effect

Gamma-rays emitted by sources moving at $v \approx c$ suffer from Doppler Effect that degrades the energy resolution in the final spectrum because of peaks broadening.



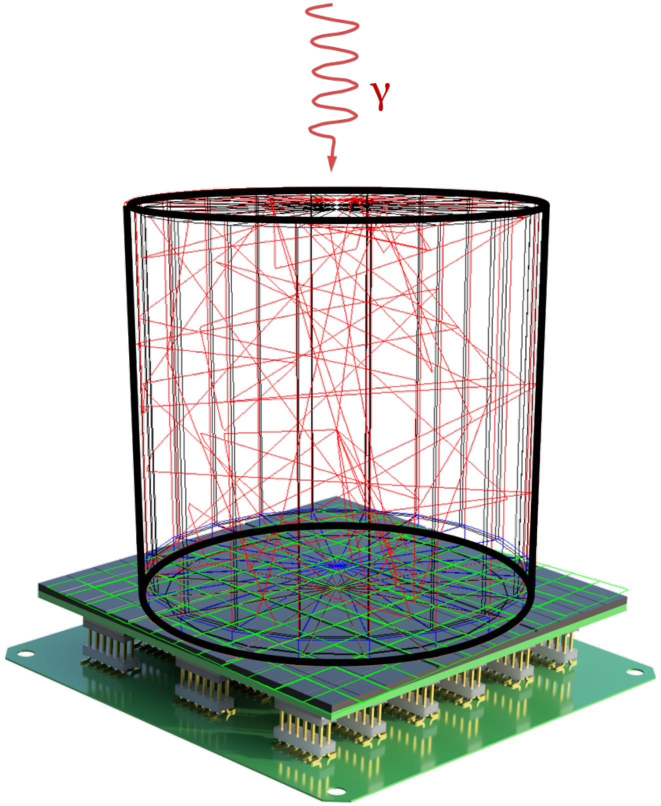
Peaks Broadening

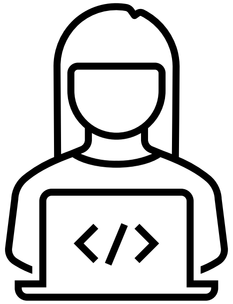


The aim of this work is the reconstruction of the gamma-ray interaction position on the GAMMA Module. In this way the θ angle can be retrieved leading to a Doppler Effect compensation.

Interaction Position Reconstruction: How?

There are no analytical models that explain the complex light distribution on the SiPM matrix after the gamma-ray interaction in a thick crystal. Due to different problems such as scintillator thickness and internal reflections, the interaction position reconstruction is a sophisticated problem that can be solved with Artificial Intelligence.



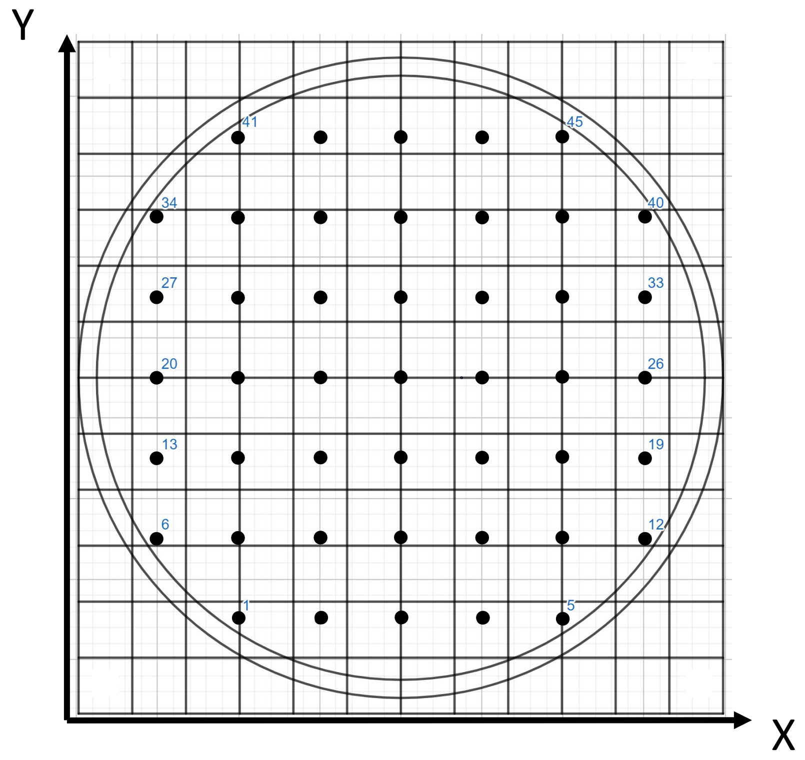


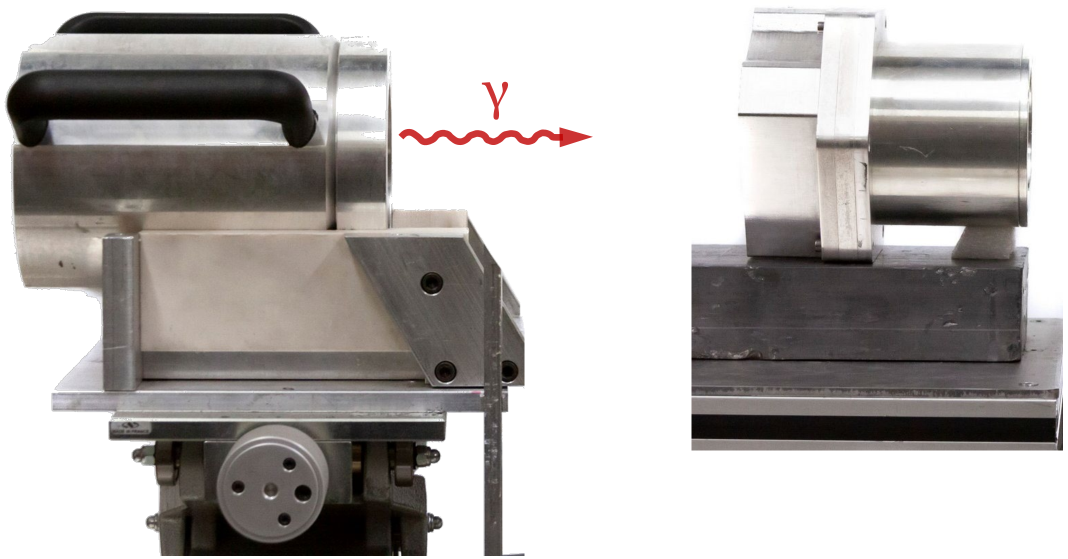
Artificial Intelligence Solutions:

- Neural Network (NN)
- Decision Tree (DT)

Supervised Learning Approach

AI algorithms have been trained with a Supervised Learning approach: a 45 points grid has been defined on the circular scintillator surface and each of them is a <u>class</u>.





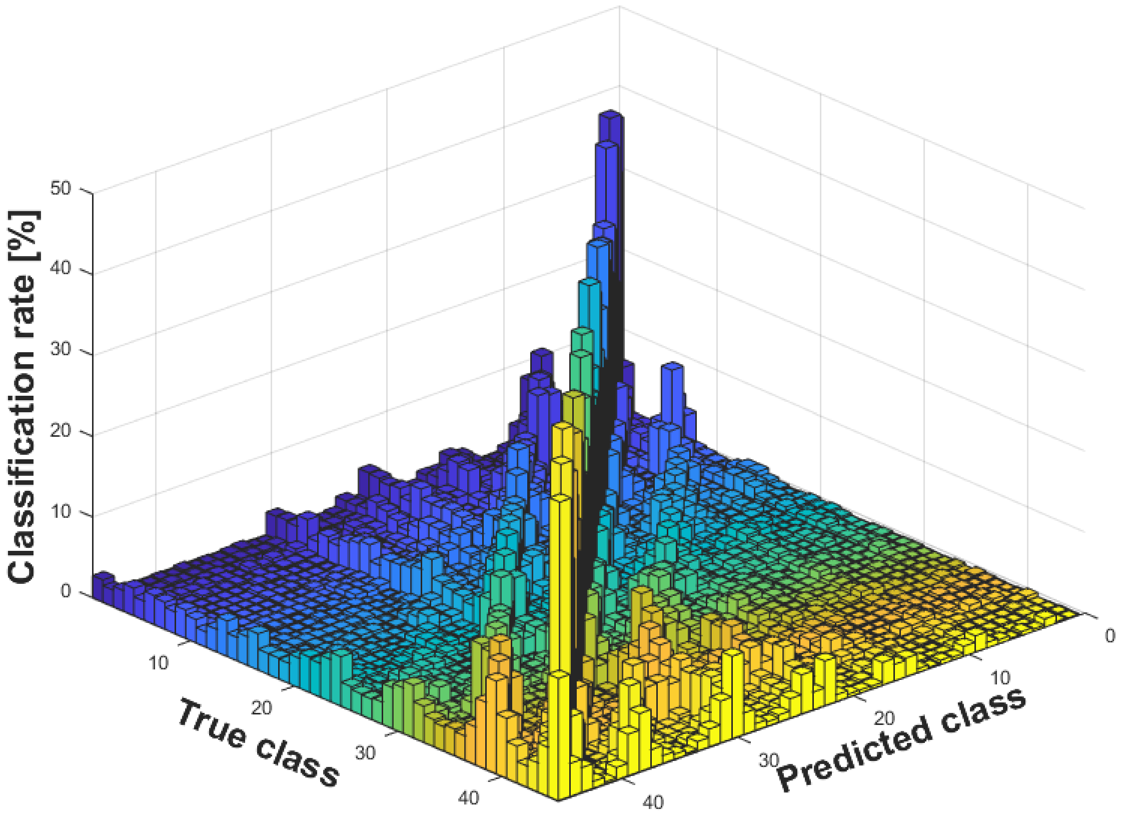
Collimated ^{137}Cs
Photon Beam (1 mm)

The training dataset is acquired moving a collimated ^{137}Cs source that radiates every single class one at a time.

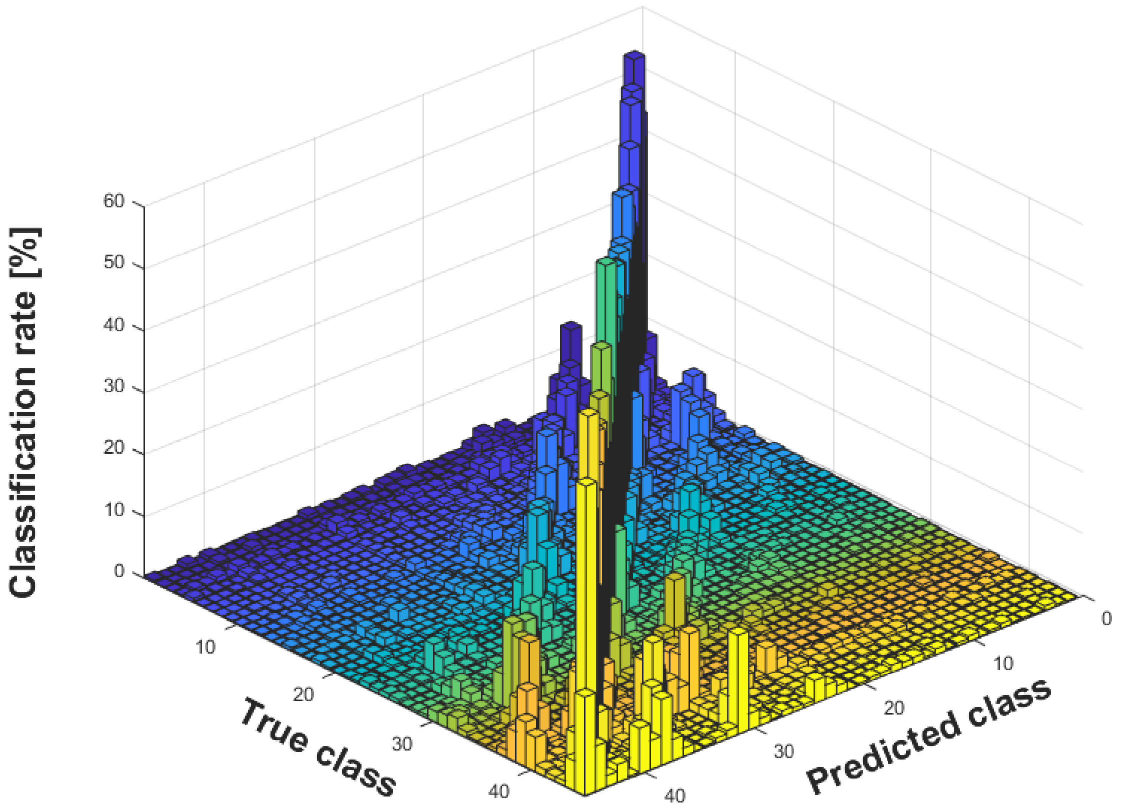
Confusion Matrix Comparison

Both Decision Tree and Neural Network have been tested and results are shown.

Decision Tree



Neural Network

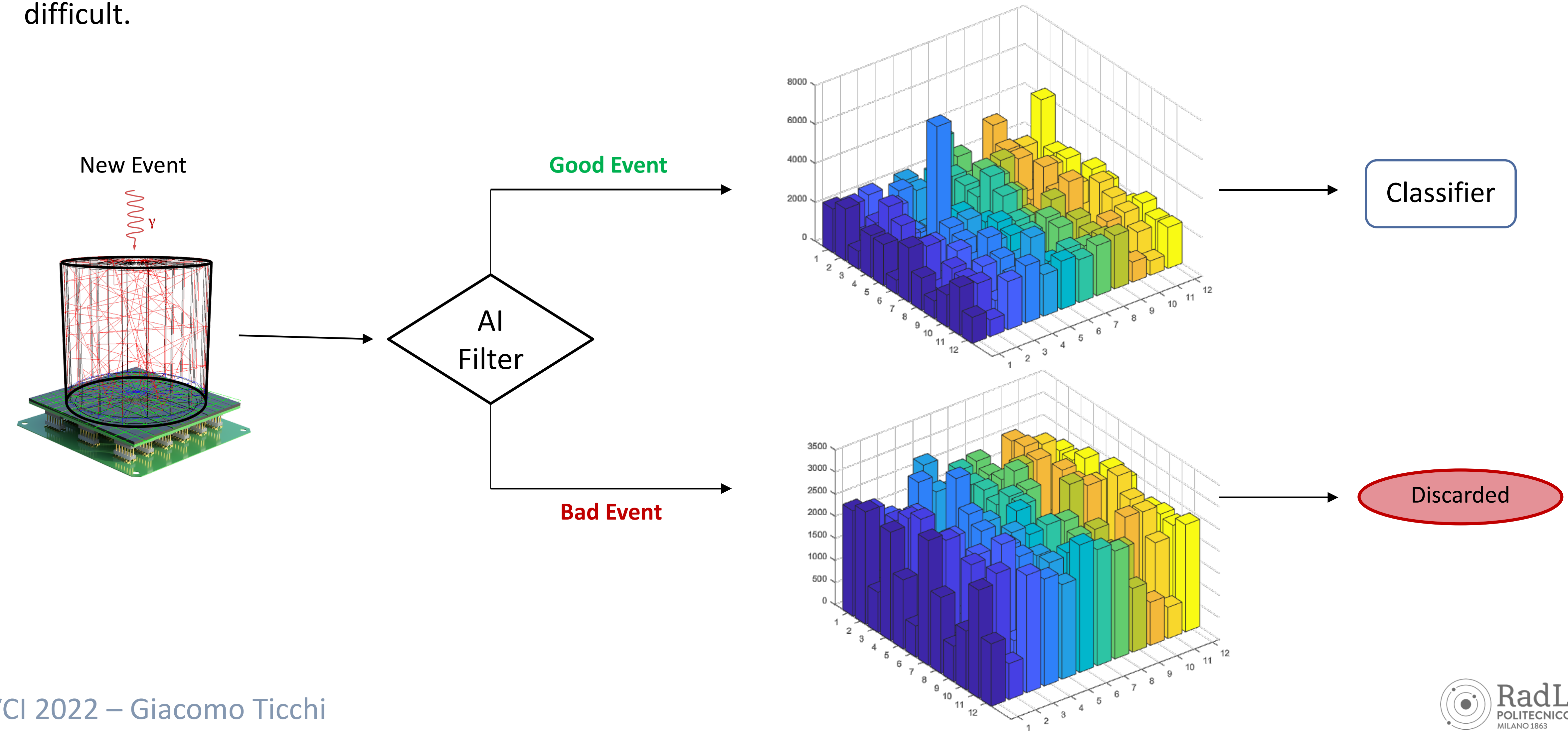


Mean Error	1.52 cm	1.1 cm
RMS Error	2.13 cm_{rms}	1.75 cm_{rms}

RadLab POLITECNICO MILANO 1863

Artificial Intelligence Filter

An AI Filter has been developed to improve the spatial resolution. The Filter is trained to recognize and discard events that generates a complex light distribution on the SiPMs making the position reconstruction really difficult.

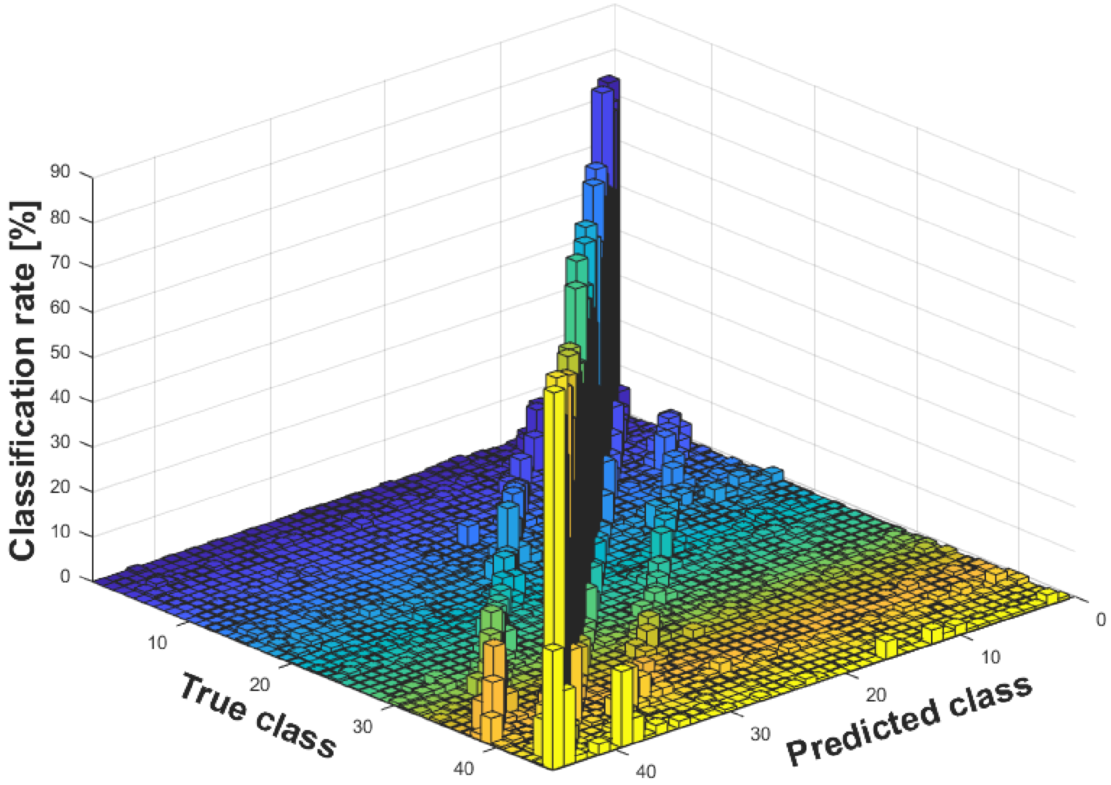


RadLab POLITECNICO MILANO 1863

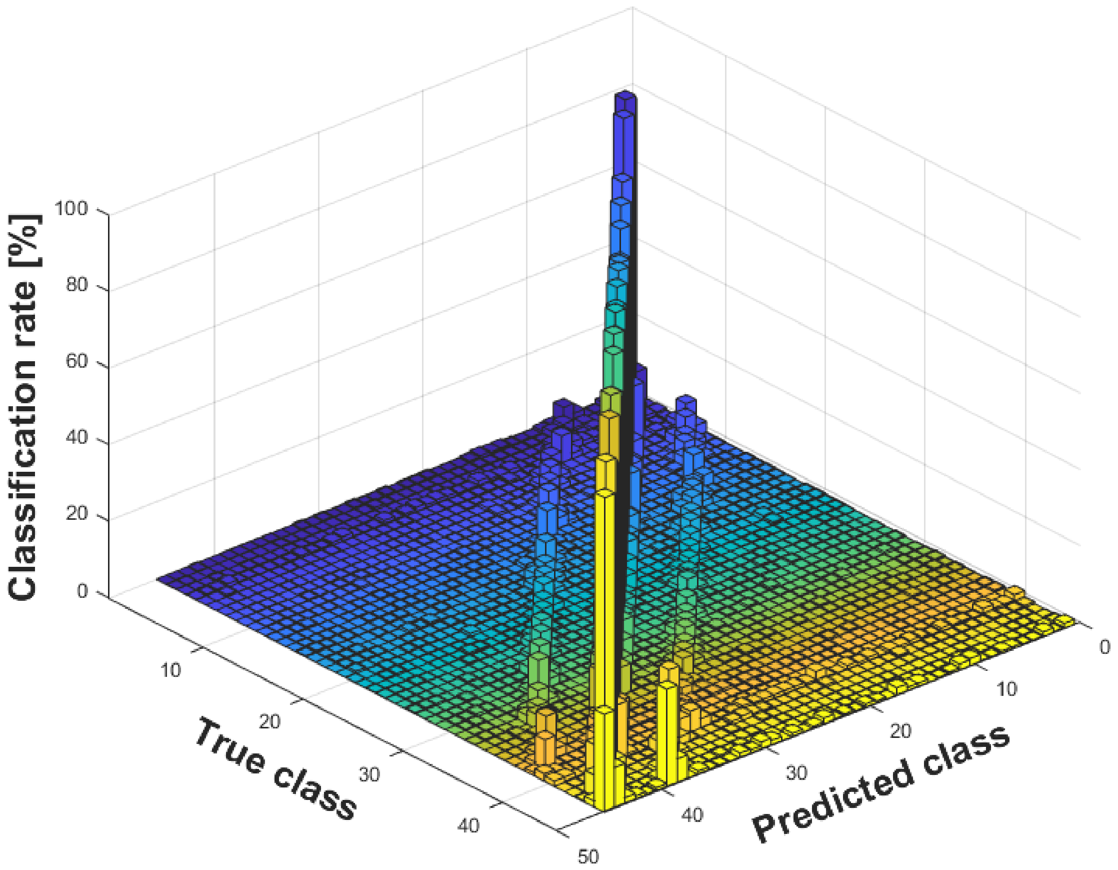
Confusion Matrix Comparison with AI Filter

Thanks to the Filter the spatial resolution has been greatly improved.

Decision Tree



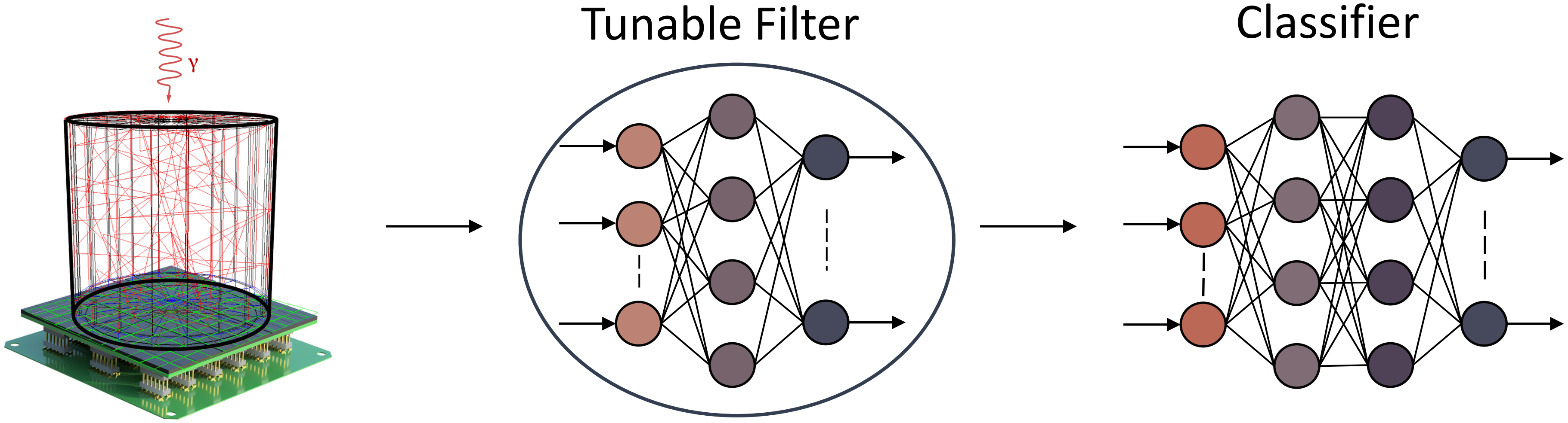
Neural Network



Mean Error	0.58 cm	0.53 cm
RMS Error	$1.29\ cm_{rms}$	$1.21\ cm_{rms}$

Discarded Events Percentage VS Spatial Resolution

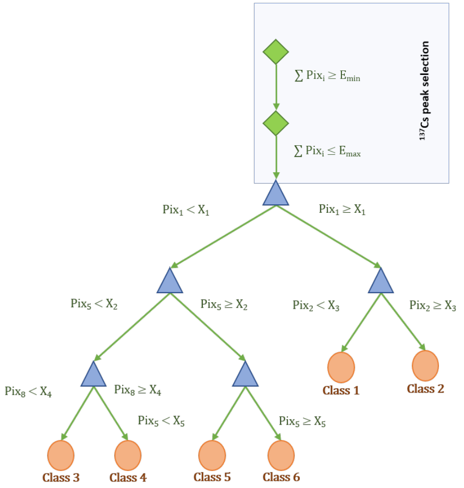
The AI Filter introduces a trade-off between discarded events and spatial resolution. The more the filter is selective, the more events will be discarded reaching a better spatial resolution. However, in the training phase the filter can be tuned to accept specific spatial errors leading to a lower percentage of discarded events.

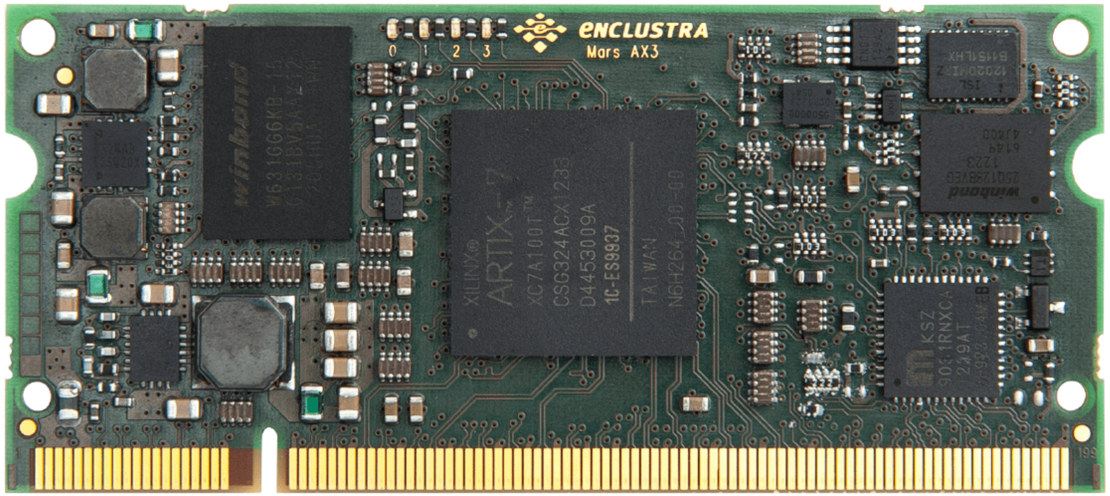


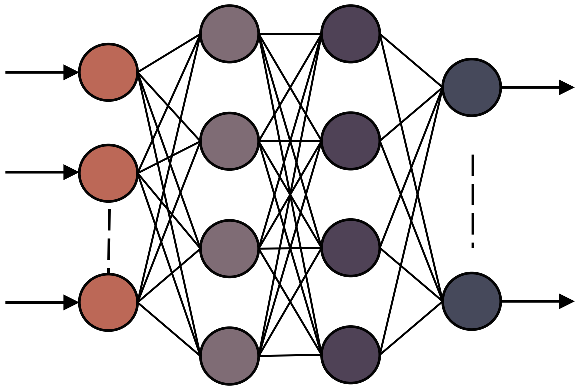
Tolerance	0 cm	1 cm	1.5 cm	No Filter
Discarded Events	**62 %**	**33 %**	**16 %**	0 %
Mean Error	0.53 cm	0.75 cm	1 cm	**1.1 cm**

Real Time 2D Position Reconstruction

Both the Neural Network and the Decision Tree have been implemented on FPGA (Artix-7 xc7a100t) thanks to Vivado High Level Synthesis. Latencies and hardware resources needed are summarized in the Table below.







	Latency	BRAM	DSP	FF	LUT
Decision Tree	10 ns ÷ 440 ns	9 %	~ 0 %	~ 0 %	~ 0 %
Neural Network	**35.5 μs**	73 %	**91 %**	41 %	63 %

RadLab POLITECNICO MILANO 1863

Real Time 2D Position Reconstruction

The Neural Network shows a big latency due to the high number of matrix-vector multiplications that are made by 32 bit Floating Point numbers.

$$\begin{bmatrix} Float & \cdots & Float \\ \vdots & \ddots & \vdots \\ Float & \cdots & Float \end{bmatrix} \begin{bmatrix} Float \\ \vdots \\ Float \end{bmatrix}$$

Up to $30 \cdot 10^3$ operations

Quantization Aware Training (QAT)

To minimize latency and computational complexity the NN has been also trained with the QAT technique which converts weights and biases from Float (32 bit) to integer numbers (8 bit).

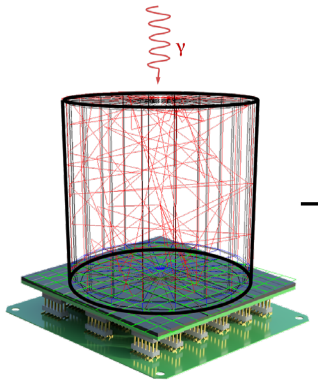
$$\begin{bmatrix} int8 & \cdots & int8 \\ \vdots & \ddots & \vdots \\ int8 & \cdots & int8 \end{bmatrix} \begin{bmatrix} int8 \\ \vdots \\ int8 \end{bmatrix}$$

Up to $30 \cdot 10^3$ operations

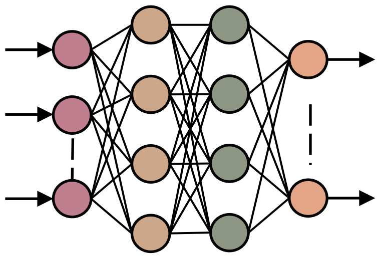
	Latency	BRAM	DSP	FF	LUT
Neural Network	**35.5 μs**	73 %	**91 %**	41 %	63 %
QAT Neural Network	**25.1 μs**	55 %	**54 %**	9 %	23 %

Real Time 2D Position Reconstruction with Filter

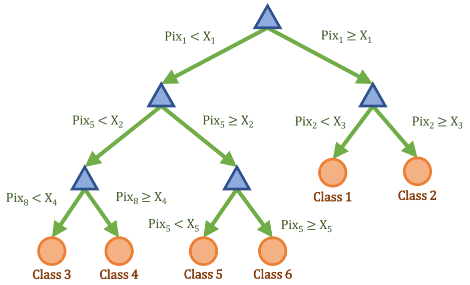
Implementing the AI Filter in the FPGA (Artix-7 xc7a100t) keeping low the lantency and the hardware resources is a challenging task. The best way is to use a NN as AI Filter and a Decision Tree as Classifier.



AI Filter



Classifier



	Latency	BRAM	DSP	FF	LUT
QAT Neural Network + Decision Tree	**$3.3\ \mu s \div 20.63\ \mu s$**	37 %	37 %	7 %	19 %

Real Time 2D Position Reconstruction Summary

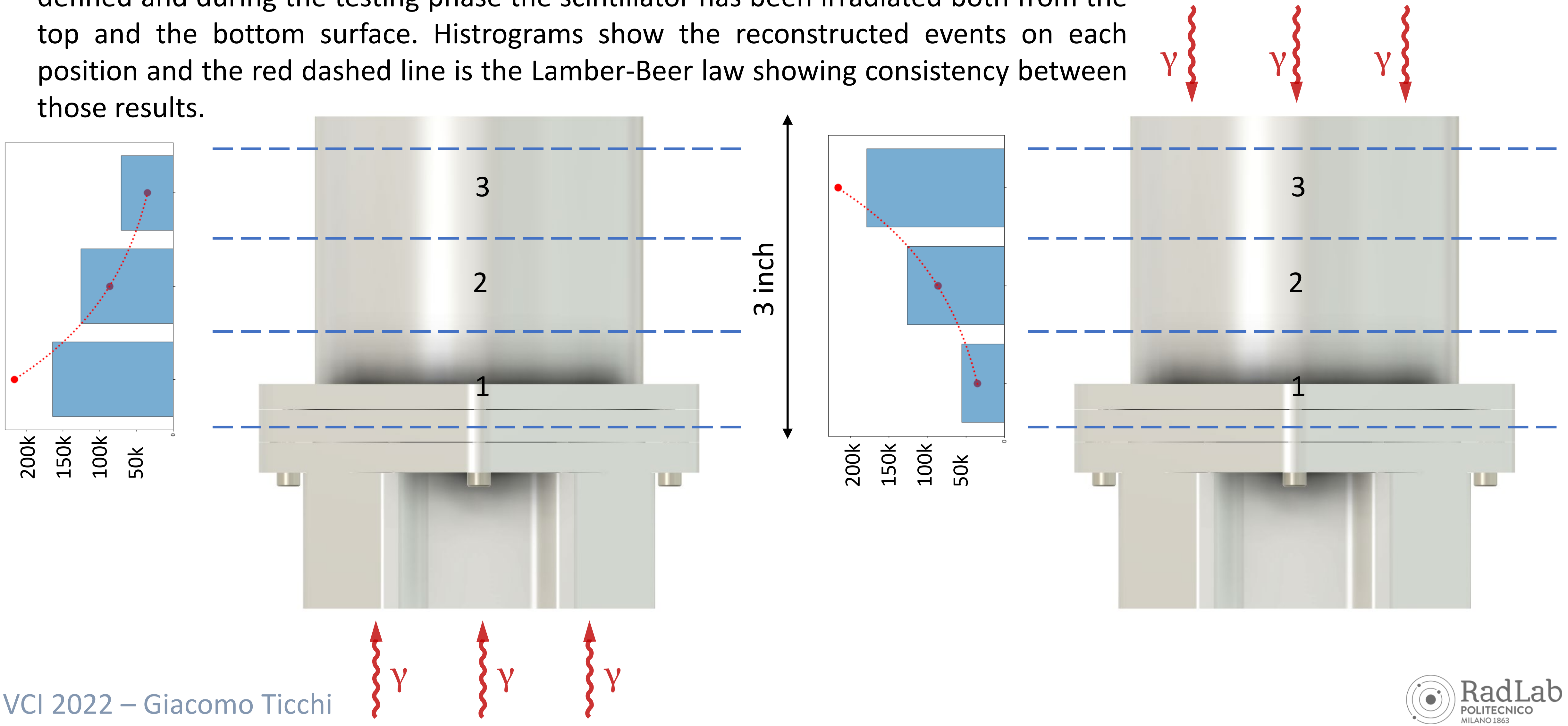
Here all the implemented solutions are summarized.

	Latency	DSP	Mean Error	Discarded Events
Decision Tree	**10 ns ÷ 440 ns**	~ 0 %	**1.52 cm**	0 %
Neural Network	**35.5 μs**	**91 %**	1.1 cm	0%
QAT Neural Network	25.1 μs	54 %	~ 1.1 cm	0%
QAT Neural Network + Decision Tree	$\mathbf{3.3\ \mu s \div 20.63\ \mu s}$	**37 %**	**0.73 cm**	**62 %**

Artix-7 xc7a100t

Depth Of Interaction

Also the Z coordinate has been reconstructed with a NN. Three classes have been defined and during the testing phase the scintillator has been irradiated both from the top and the bottom surface. Histrograms show the reconstructed events on each position and the red dashed line is the Lamber-Beer law showing consistency between those results.



INFN MILANO
Istituto Nazionale di Fisica Nucleare
Sezione di Milano

POLITECNICO MILANO 1863
DIPARTIMENTO DI ELETTRONICA INFORMAZIONE E BIOINGEGNERIA

RadLab POLITECNICO MILANO 1863

Thank you!

giacomo.ticchi@polimi.it