Study of CMS iRPC waveforms parameters and classification of signals using Machine Learning

Dalmo da Silva Dalto*, Helio Nogima, Mehar Ali Shah,

On behalf of the CMS Collaboration

*Rio de Janeiro State University – UERJ, (Brazil)
dalmo.da.silva.dalto@cern.ch

Motivation and Objectives

Motivation

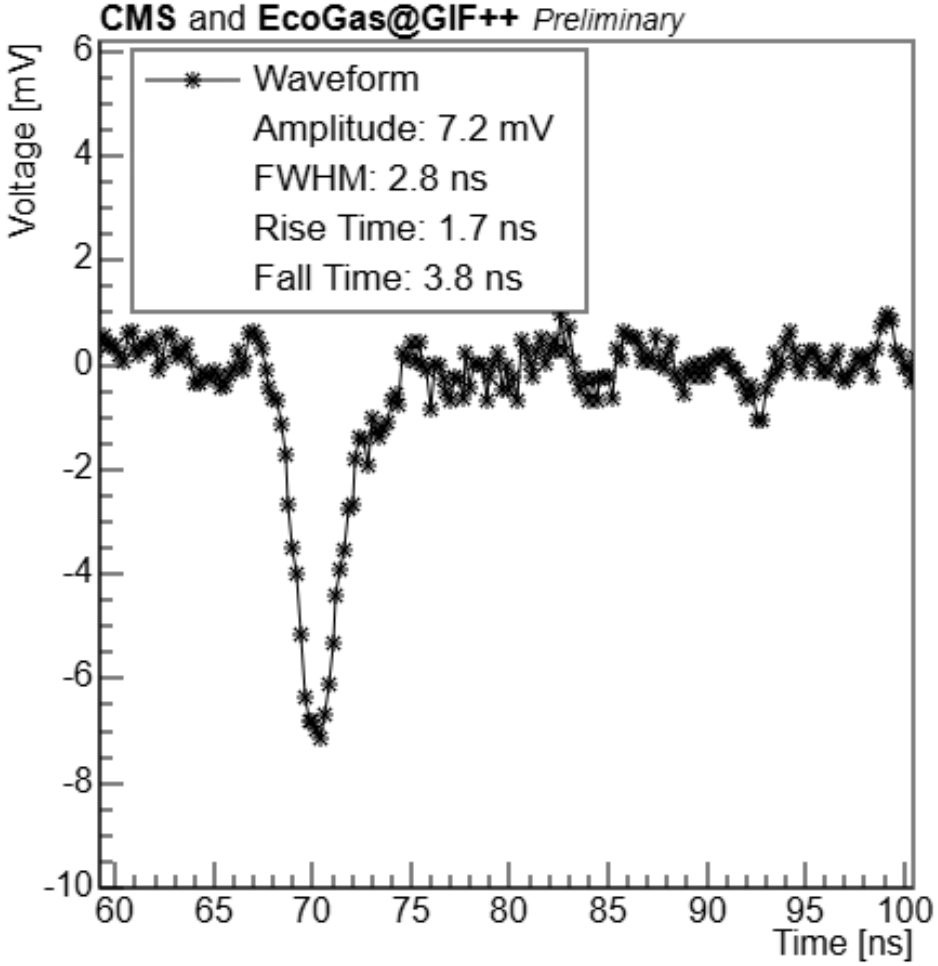
- RPC signal waveforms encode avalanche development dynamics, but the link between signal shape and underlying physics remains underexplored.
- Environmental regulations require low Global Warming Potential (GWP). standard $C_2H_2F_4$-based mixture (GWP ≈ 1485).
- Waveform analysis is time-intensive and lacks reproducibility — data-driven methods can improve consistency.

Objectives

- **Explore** correlations between signal shapes and avalanche dynamics
  → *Link waveform features to physical processes inside the detector*
- **Characterize** eco-friendly gas mixtures as low-GWP alternatives
  → *Evaluate HFO/CO_2-based mixtures against the standard $C_2H_2F_4$ blend*
- **Develop** algorithms to extract waveform parameters
  → *Automate charge, amplitude, rise time, and Full Width at Half Maximum (FWHM), measurements*
- **Implement** Machine Learning (ML) models for signal discrimination based on signal shape.

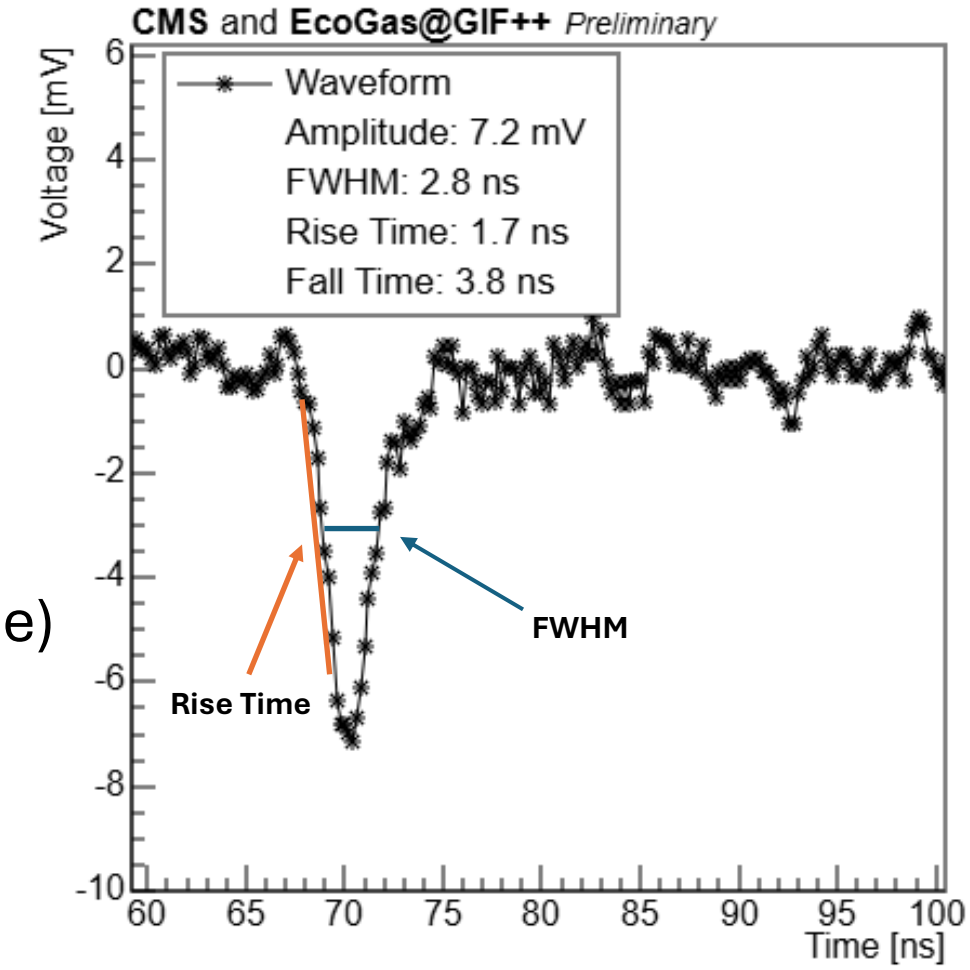
Summary

- Waveform Analysis
  - Charge
  - Rise Time
  - FWHM
- Cluster Size
- Time Resolution
- Machine Learning
  - Principal Component Analysis (PCA)
  - Hierarchical Clustering Analysis (HCA)



Waveform Analysis – Introduction

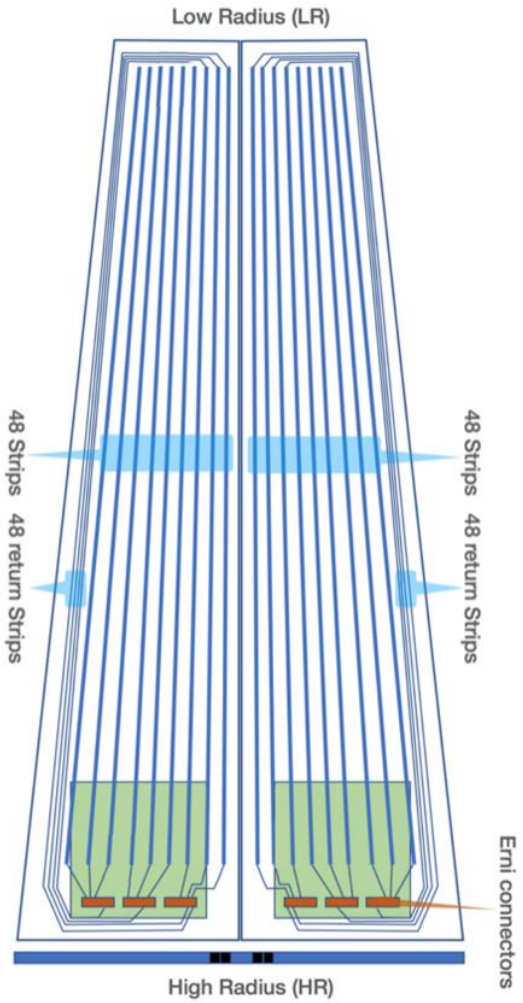
- Size Parameters
  - Charge ($Q = \int \frac{V(t)dt}{\Omega}; \Omega = 50$ ohm)
  - Amplitude
- Temporal Parameters
  - Rise Time (from 10% to 90% of its peak amplitude)
  - FWHM



Improved Resistive Plate Chambers (iRPC)

LHC Muon Beam Pipe: 

- Double gap (1.4 mm)
- Bakelite Resistive Plates (Resisitivity $0.9 - 3 \times 10^{10}$ Ωcm)
- Strips and return strips
- Pitch Strip 0.5–1 cm (Depending on the Chamber)
- Intrinsic timing resolutin (0.5ns).



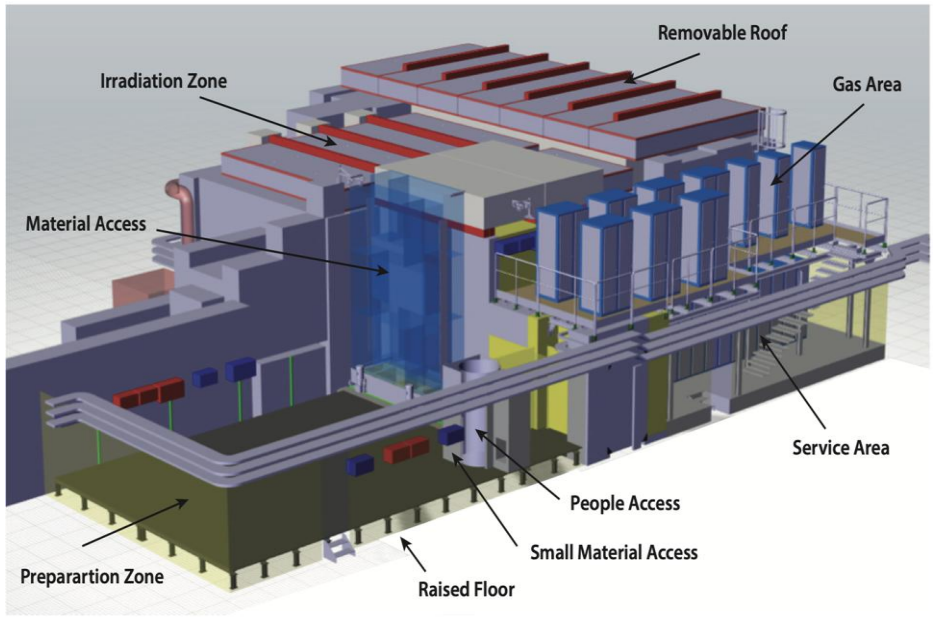
Strip PCB allows the return line concept

FEB positioned external and @ HR (lowest dose region)

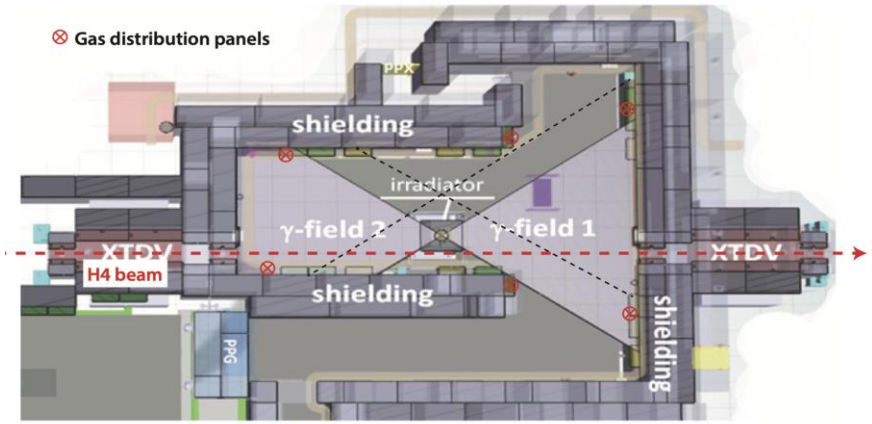
Published in: *Nucl.Instrum.Meth.A* 1083 (2026) 171125

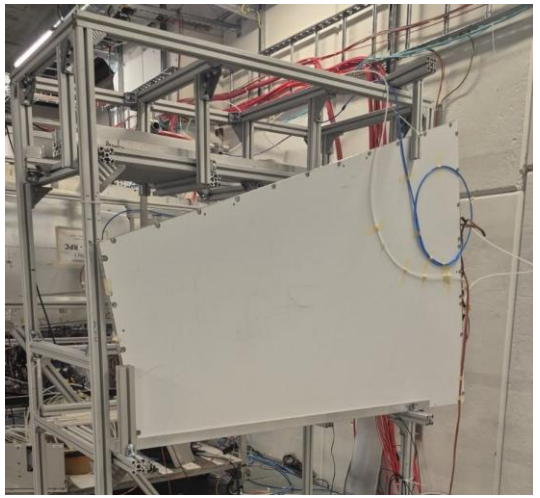
CERN Gamma Radiation Facility (GIF++)

Published in: *PoS* TIPP2014 (2014) 102



Published in: *PoS* TIPP2014 (2014) 102



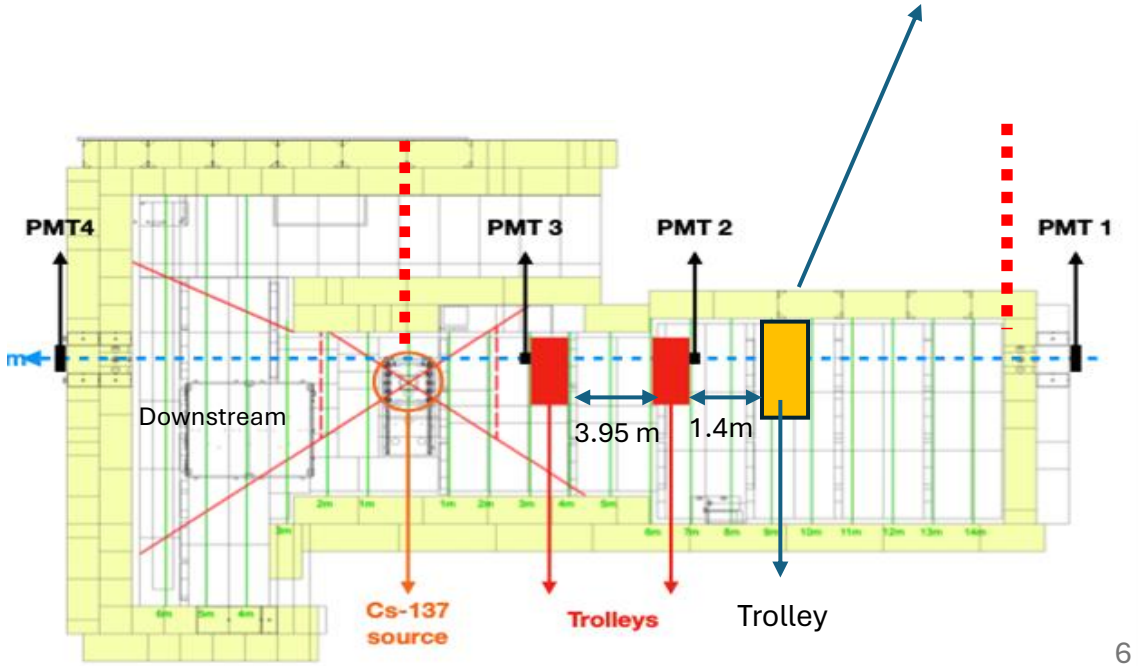


November 2025 – Chamber 185 RE3-1

100 m^2 bunker with a 11.5 TBq ^{137}Cs source (Jan 2024), providing two symmetric fields with a variable attenuator system

SPS H4 muon beam of 150 GeV/c, with ~ 10^4 muons/spill every 400 ms

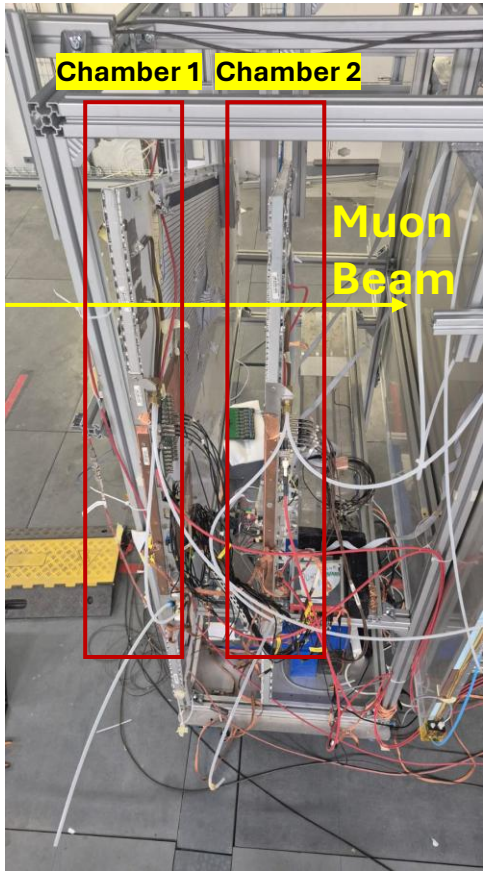
Published in: *Nucl.Instrum.Meth.A* 866 (2017) 91-103

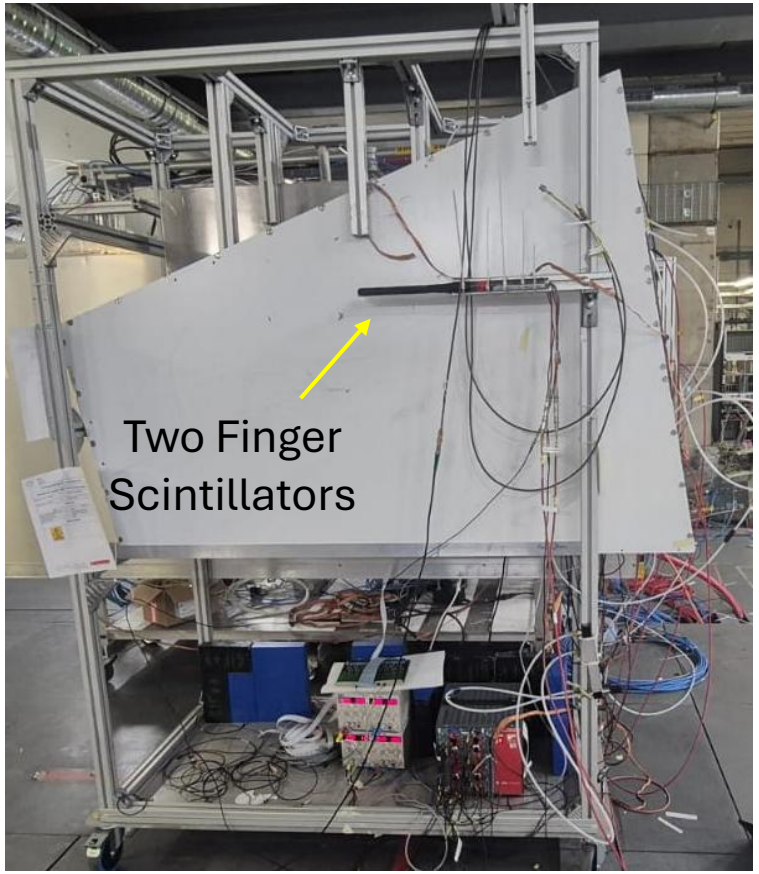


Published in: *Nucl.Instrum.Meth.A* 1052 (2023) 168271

Experimental Setup

- The two chambers were separated by a distance of approximately 20 cm.
- ERNI connector with 16 strips and 16 returns.
- Strips from the middle partition connected to the ERNI connector.
- Digitizer connected to 8 strips from one chamber and the other 8 strips connected to the other chamber.
- Trigger area of 2x22 cm^2 composed of two finger scintillators positioned in parallel.
- Gas line flowing in series for the chambers.





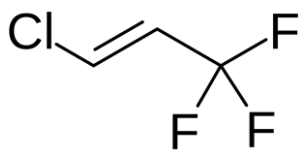
July 2026 – two Endcap chamber of Ring 3, Station 1 (RE3/1) type chambers

Gas Mixture

- GWP is a measure of how much heat a greenhouse gas traps in the atmosphere compared to carbon dioxide over a specific time period.
- HFO-1233zd is the candidate to replace SF6, the highest GWP.

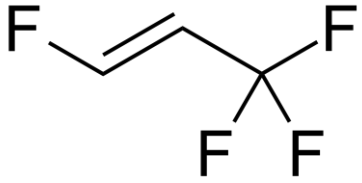
	HFO-1234ze	CO2	i-C4-H10	HFO-1233zd	R134a	SF6
Eco2_zd_0.5 [%]	35	59.5	5	0.5	0	0
Eco2_zd_2 [%]	20	73	5	2	0	0
Eco2 [%]	35	60	4	0	0	1
STD [%]	0	0	4.5	0	95.2	0.3
GWP	6	1	3	1	1430	23500

Published in: *Nucl.Instrum.Meth.A* 1066 (2024) 169580



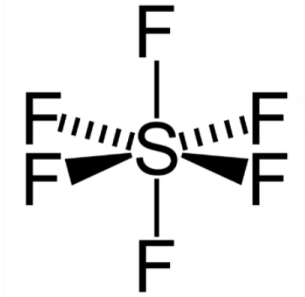
Chlorotrifluoropropene

HFO-1233zd



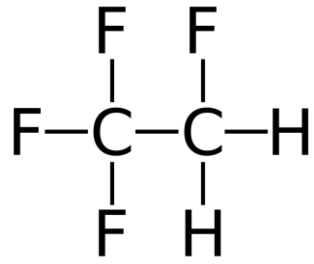
Tetrafluoropropene

HFO-1234ze



Sulfur Hexafluoride

SF_6



Tetrafluoroethane

R134a

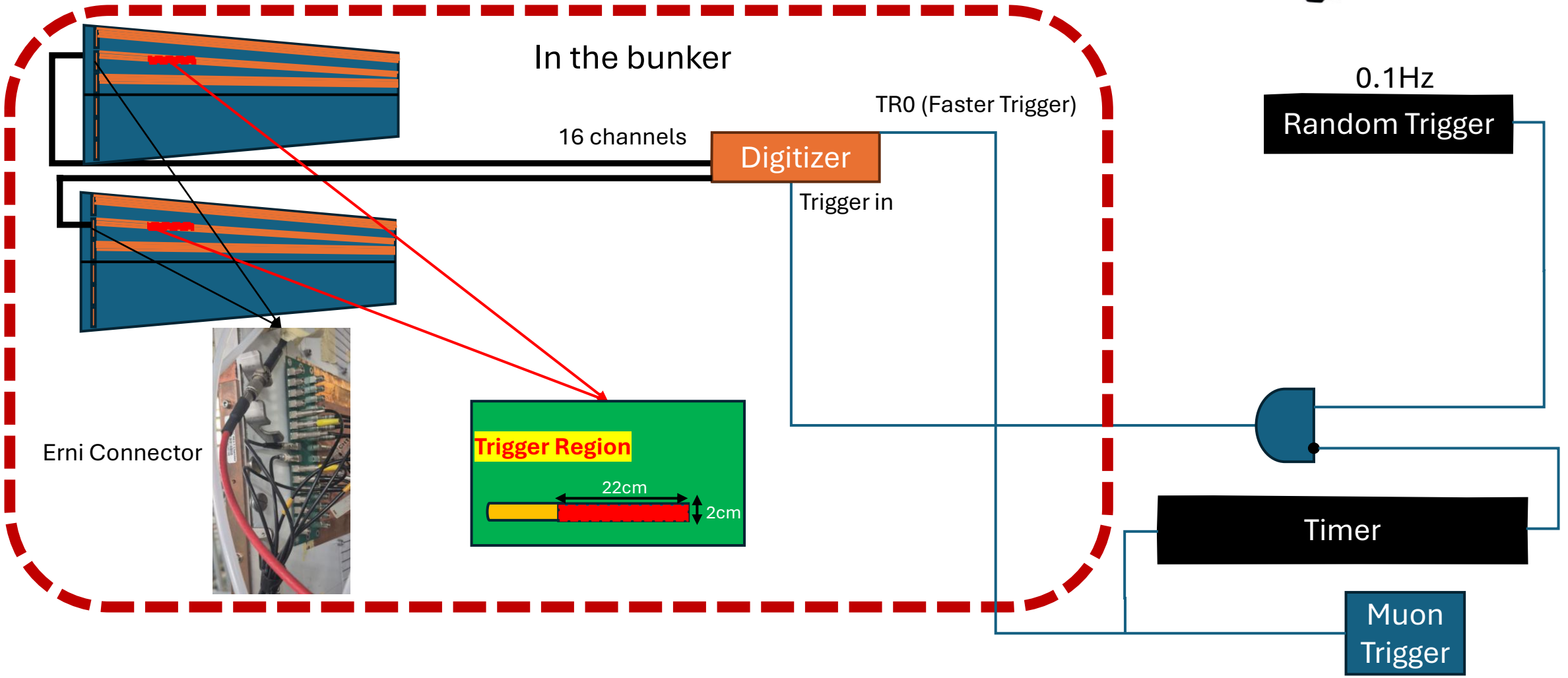
DAQ Diagram Test Beam July 2026

- Digitizer CAEN DT5742 with 16 channels, a bandwidth (BW) of 500 MHz, and 12-bit.
- Data acquired at 5GS/s sampling of two Endcap chambers of Ring 3, Station 1 (RE3/1 type chambers).
- ERNI connector fixed in the middle partition, distributing the 8 channels to each chamber.
- Beam area crossing 22x2 cm^2 (region between the high radius and the middle of the chamber).

CAEN DT5742
12 bit
5 GS/s
500 MHz
16 ch

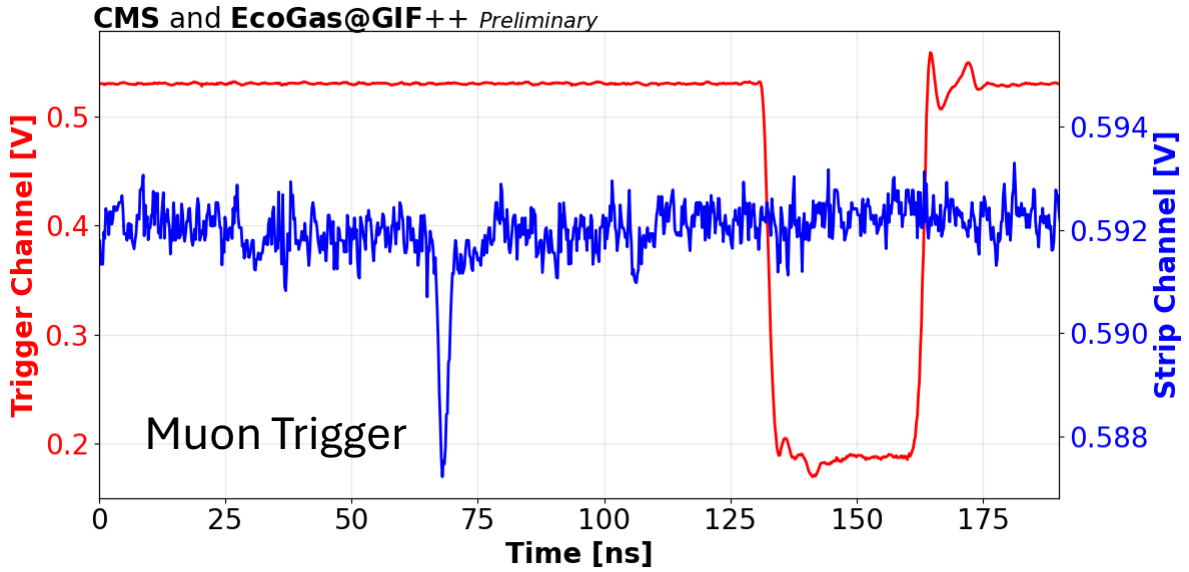
Bandwidth $= \frac{0.35}{\text{rise time}}$
Fastest Rise time $\therefore \sim 0.7ns$

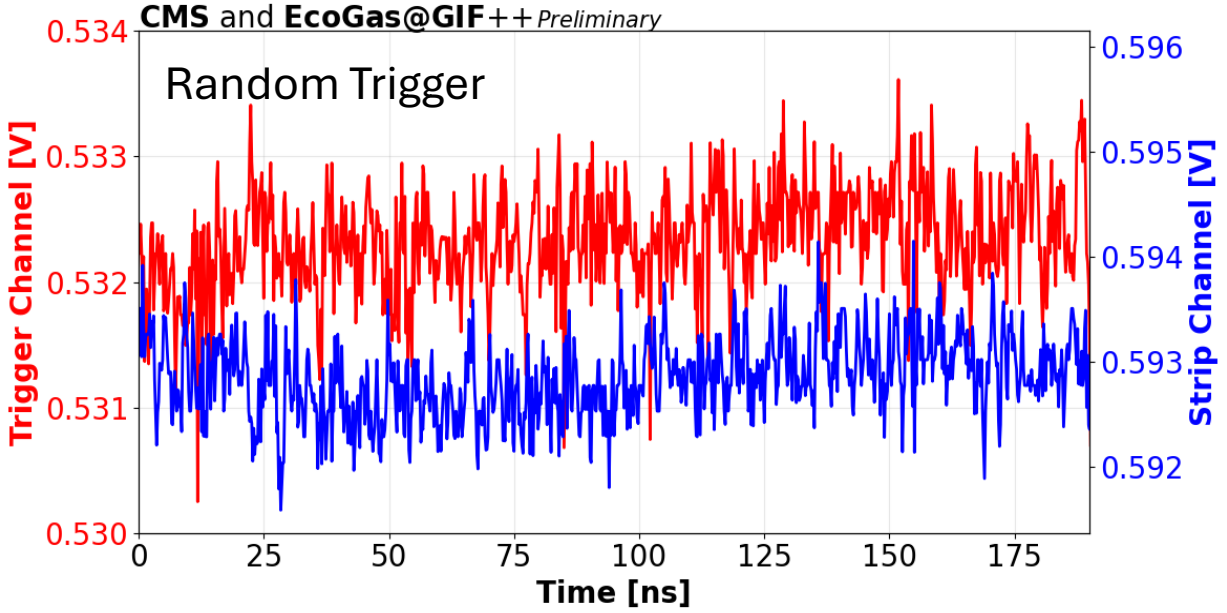




Offline Selection between Muon Trigger and Random Trigger

- Muon trigger digitalization
- Offline selection flag distinguishes random trigger from muon trigger events
- Pedestal measurement

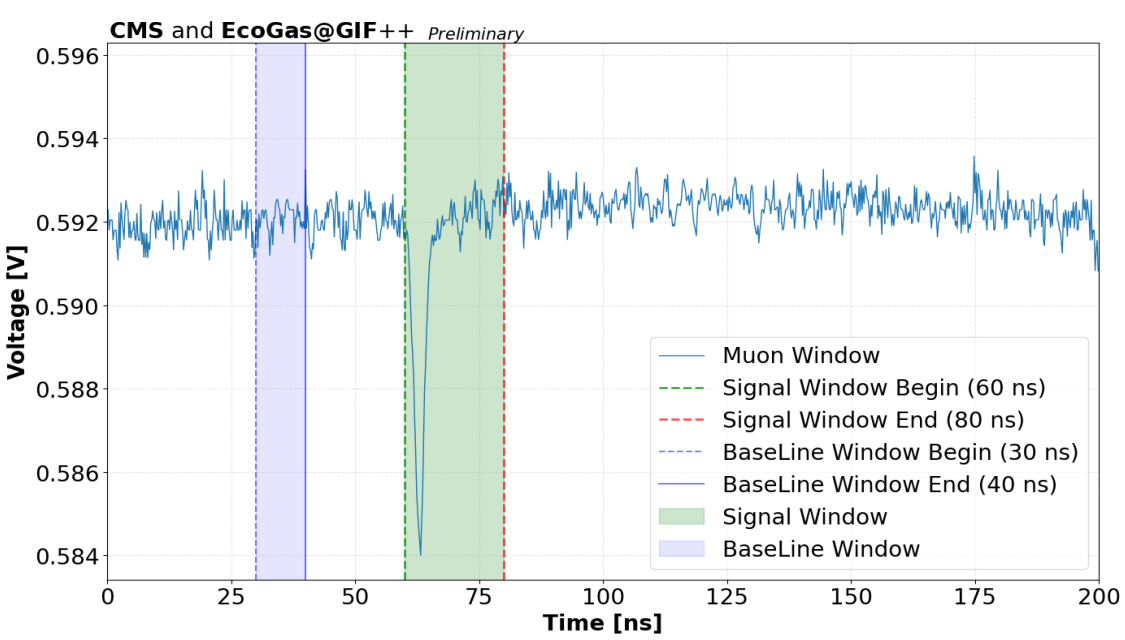


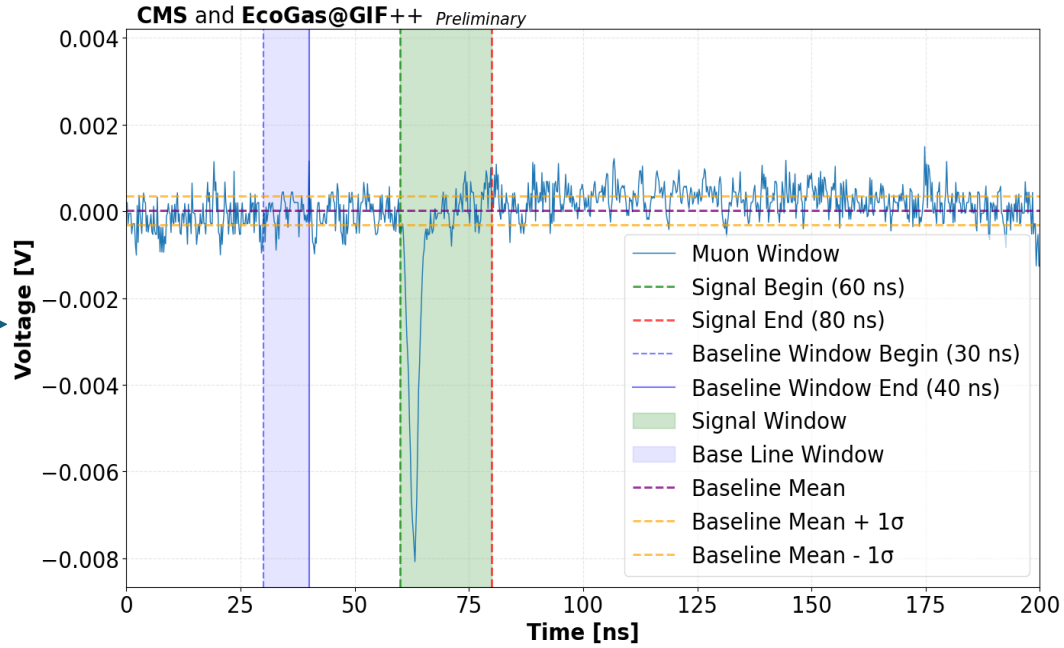


The First Step

- Conversion from Analog-to-Digital Converter (ADC) data to electric tension units.
- Baseline correction per waveform using its own noise window fluctuation V_{mean}.

Base Line Correction

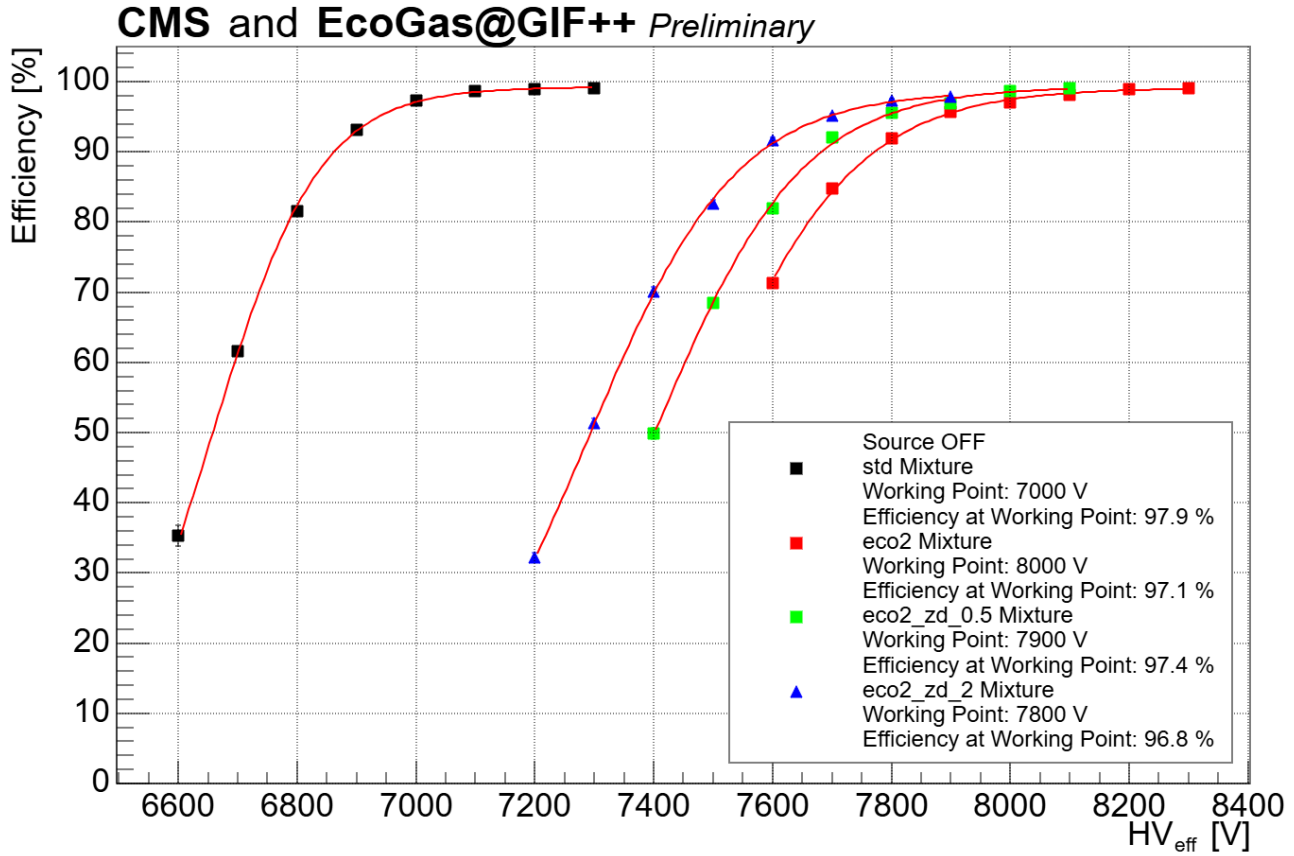




The dashed horizontal lines are the 1 σ limits of the background noise. The number of sigmas is used as a threshold for the signals in the muon window.

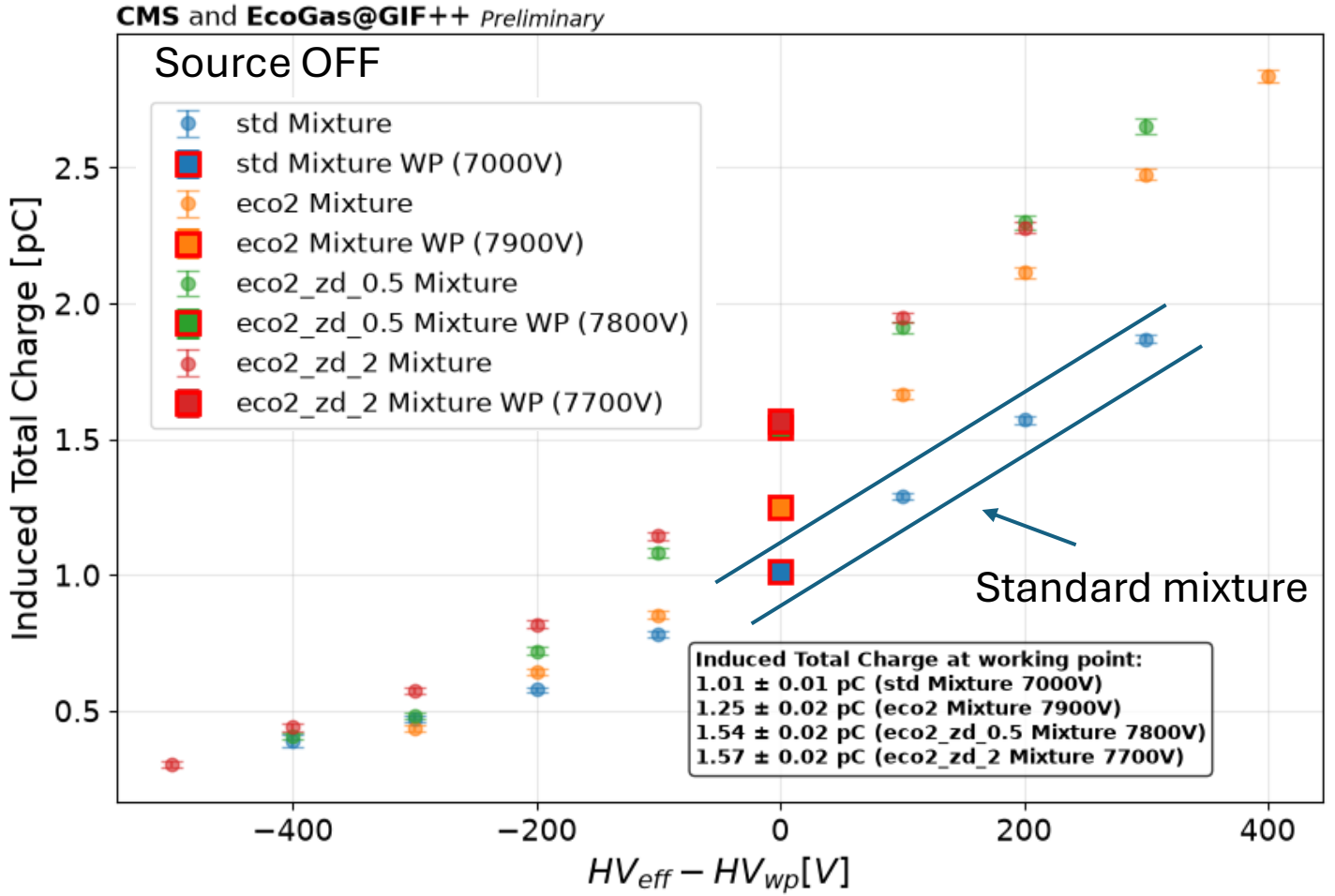
Efficiency Curve

- The effective voltage (HV_{eff}) is the applied voltage corrected for pressure and temperature.
- The efficiency is greater than 97% at the working point (WP).
- 5 sigmas threshold (≈ 1.8 mV)
- V_{wp}=V_{knee}+150V
- Eco mixtures need a higher electric field than standard mixture.



Induced Total Charge

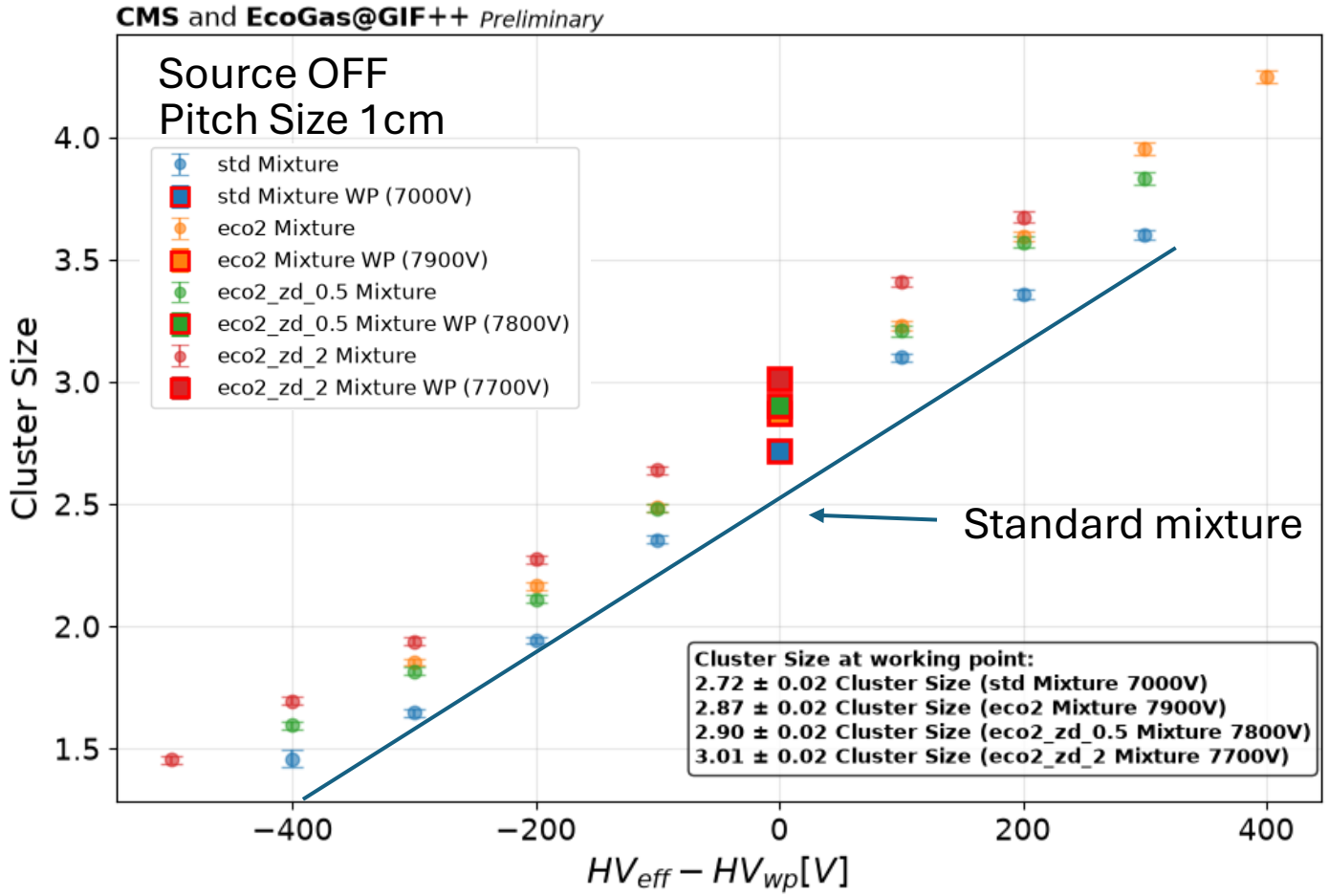
- Induced total charge (sum of charges from 8 strips, subtracting the pedestal charge)
- Average of the induced total charge distribution
- The charge increases with the voltage.
- Standard gas mixture presents a charge smaller than eco2 and eco_zd blends.



Cluster Size

Cluster Size

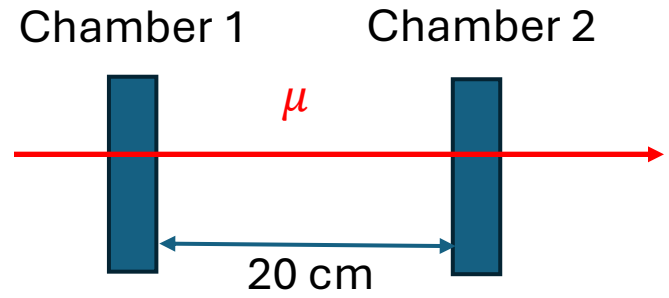
- Cluster Size (Number of adjacent strips hit)
- Average of the cluster size distribution
- The cluster size becomes larger as the voltage increases.
- The standard gas mixture presents a smaller cluster size than the eco2 and eco_zd blends.



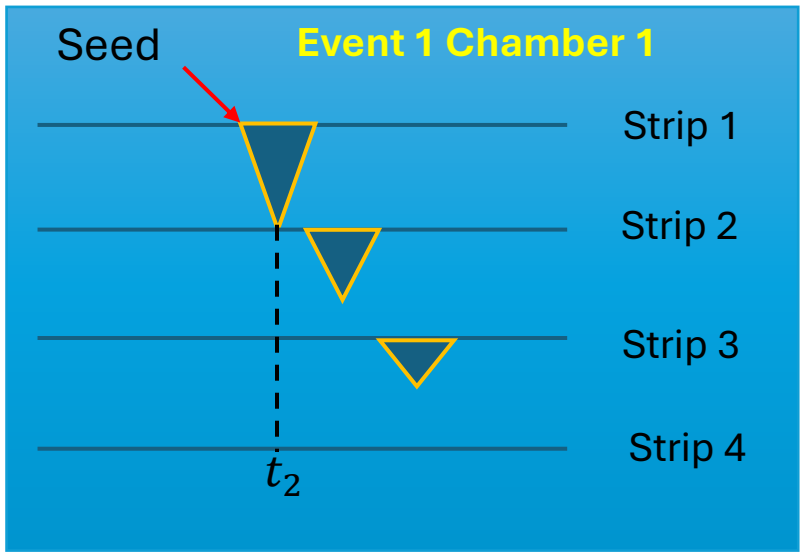
Time Resolution

Seed Strip Selection

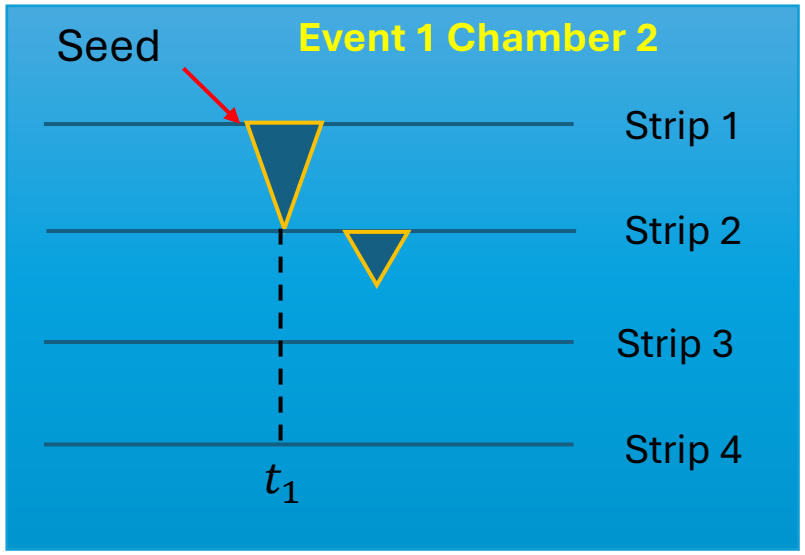
- **Seed Strip:** The strip with the highest charge.



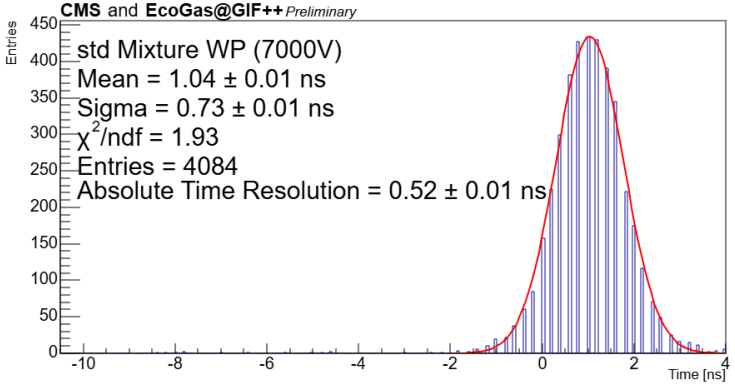
1 Cluster with size 3



1 Cluster with size 2



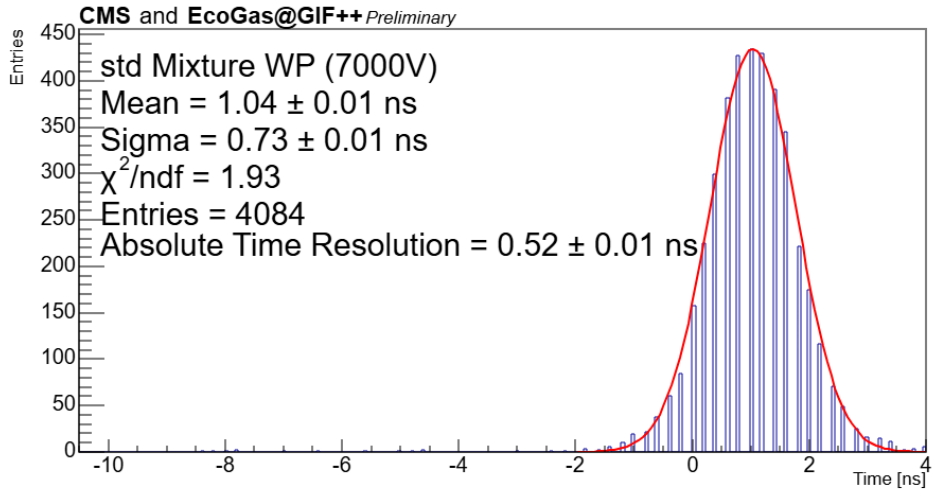
$$\Delta t = t_2 - t_1$$

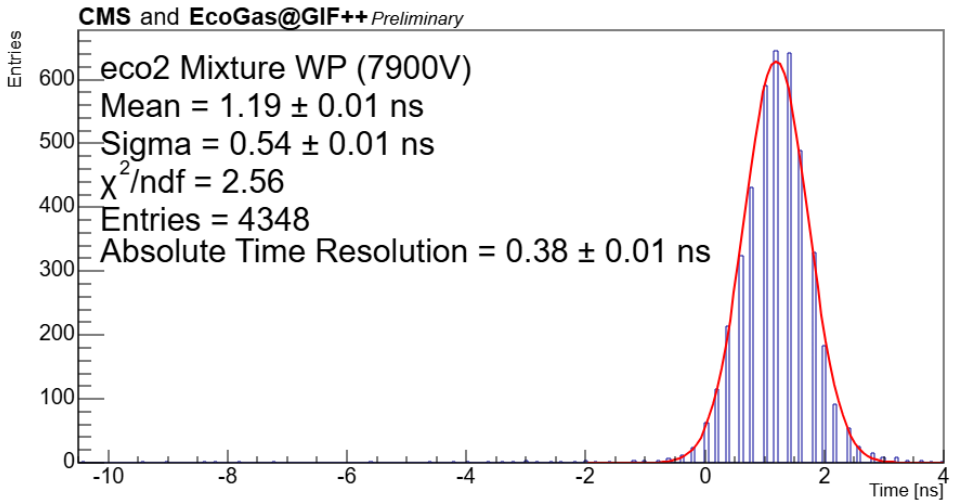


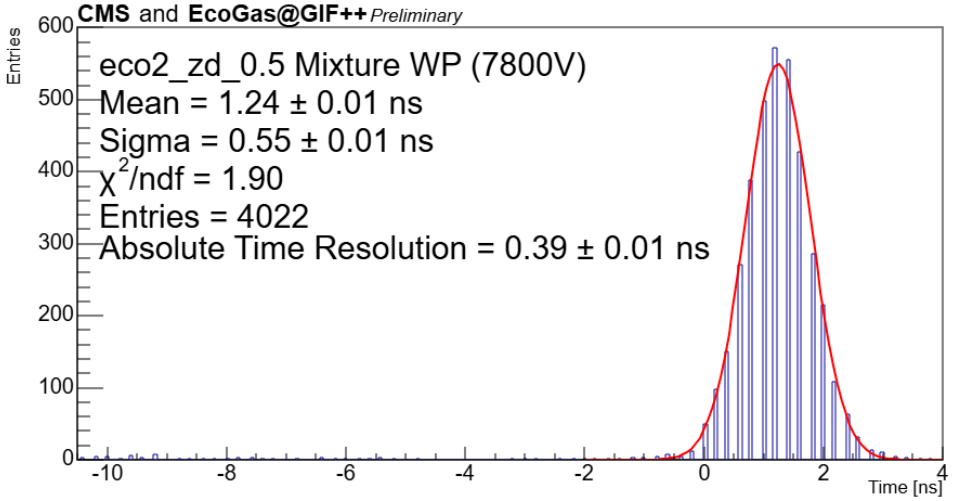
Time Resolution

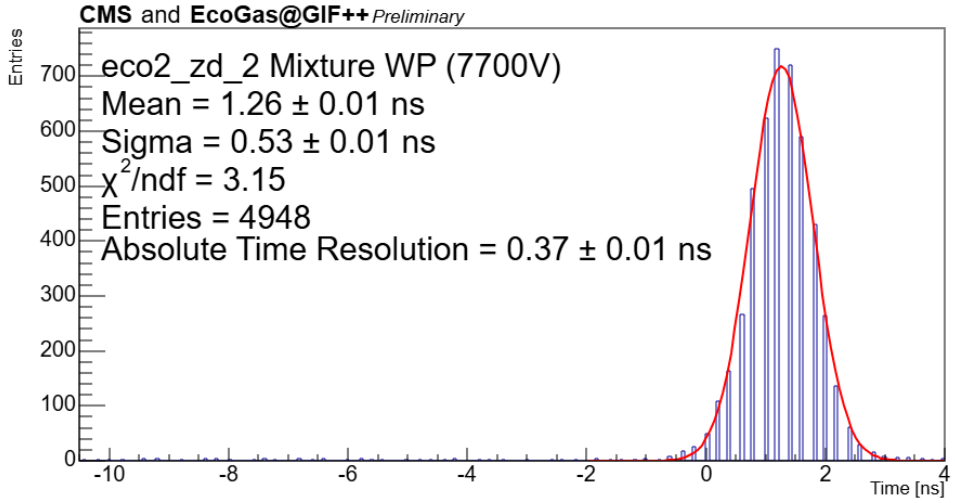
- Time difference between two chambers at a distance of 20 cm.
- The absolute time resolution is defined as the sigma from the Gaussian fit of the time difference distribution divided for the $\sqrt{2}$.
- Selection by the seed strip (the strip with the highest charge in the event).
- The eco mixtures present an improvement in time resolution

$$\text{Absolute Time Resolution} = \frac{\sigma}{\sqrt{2}}$$



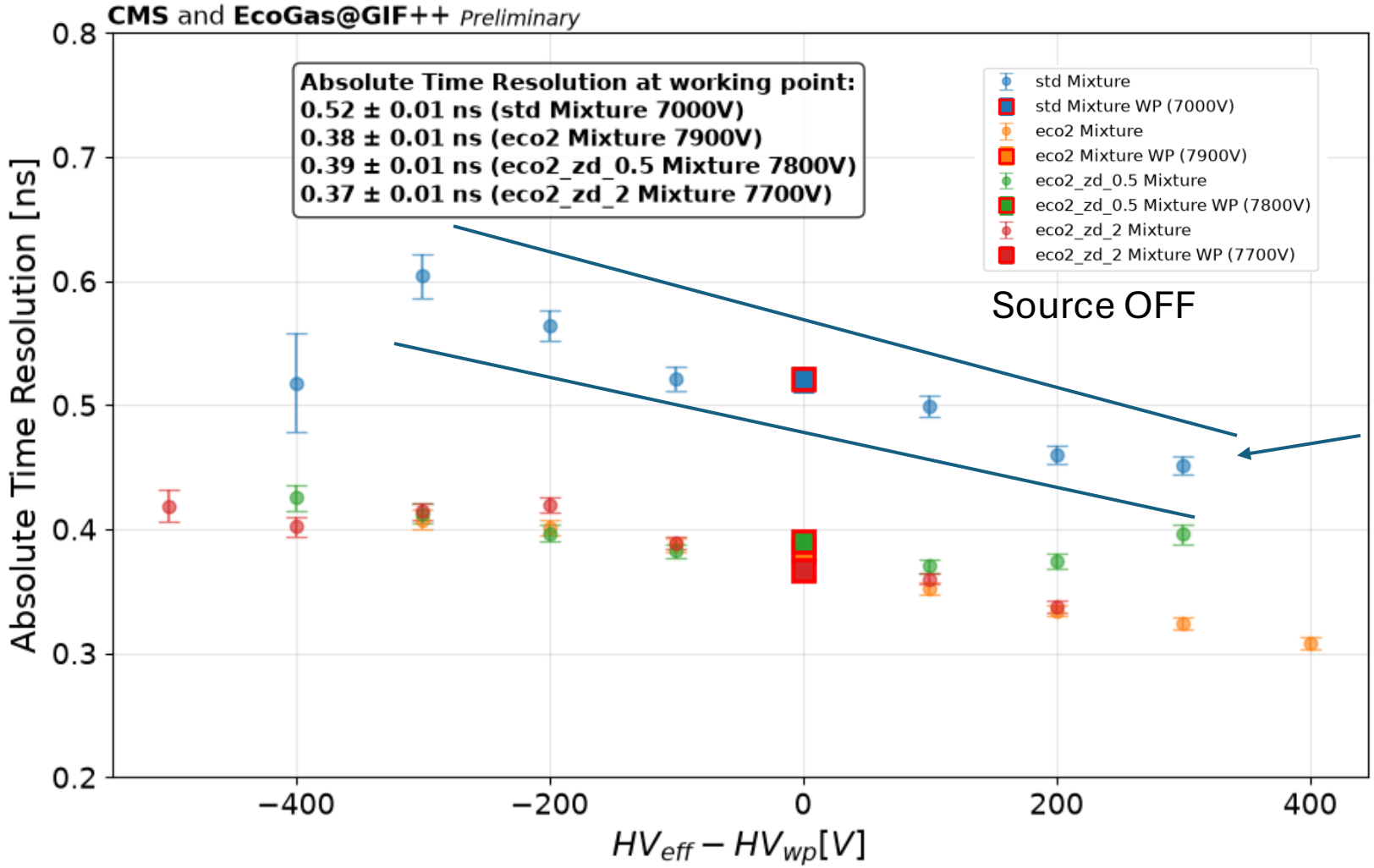






Time Resolution

- Time difference between two chambers at a distance of 20 cm.
- The time resolution continues to improve with the increase in voltage.
- The eco mixtures present an improvement in time resolution.

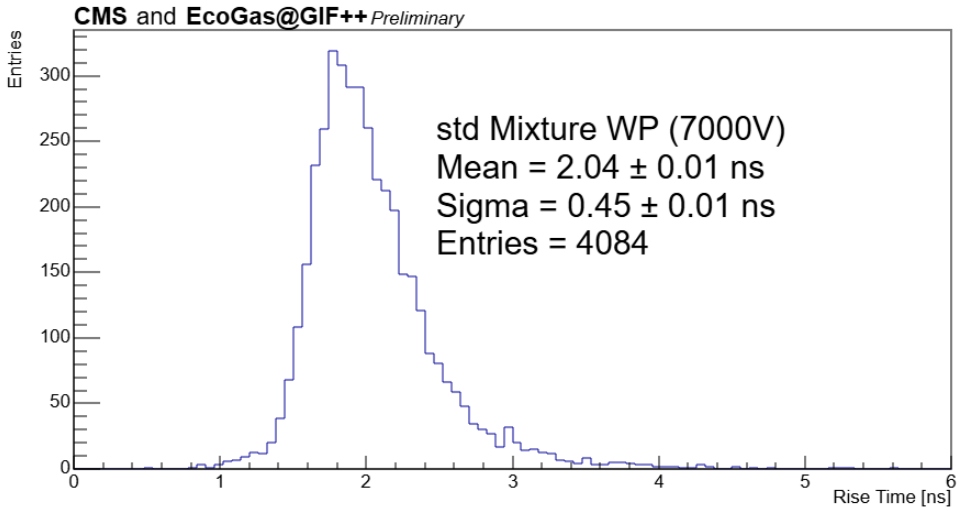


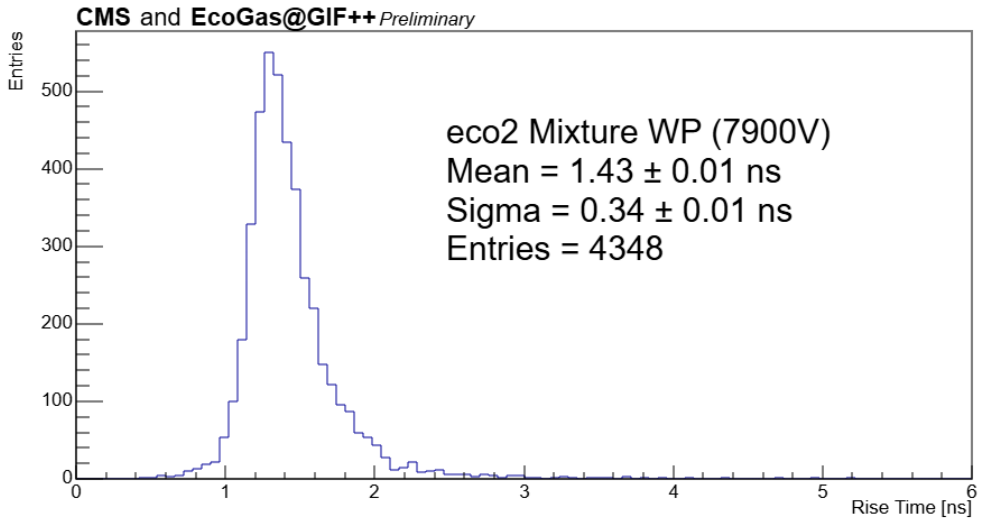
Standard mixture

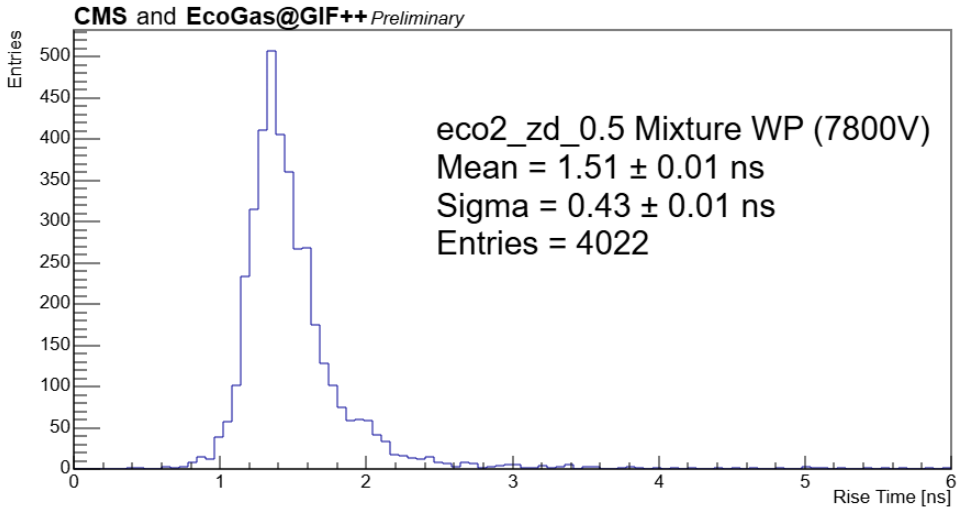
Time Parameters

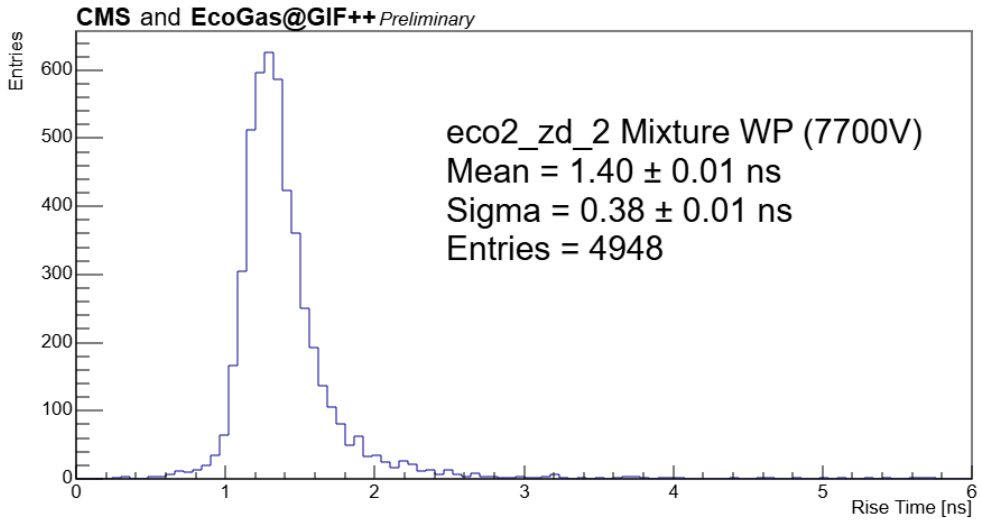
Rise Time

- Selection by the seed strip (the strip with the highest charge in the event).
- The standard mixture shows a slower rise time than the standard mixture.



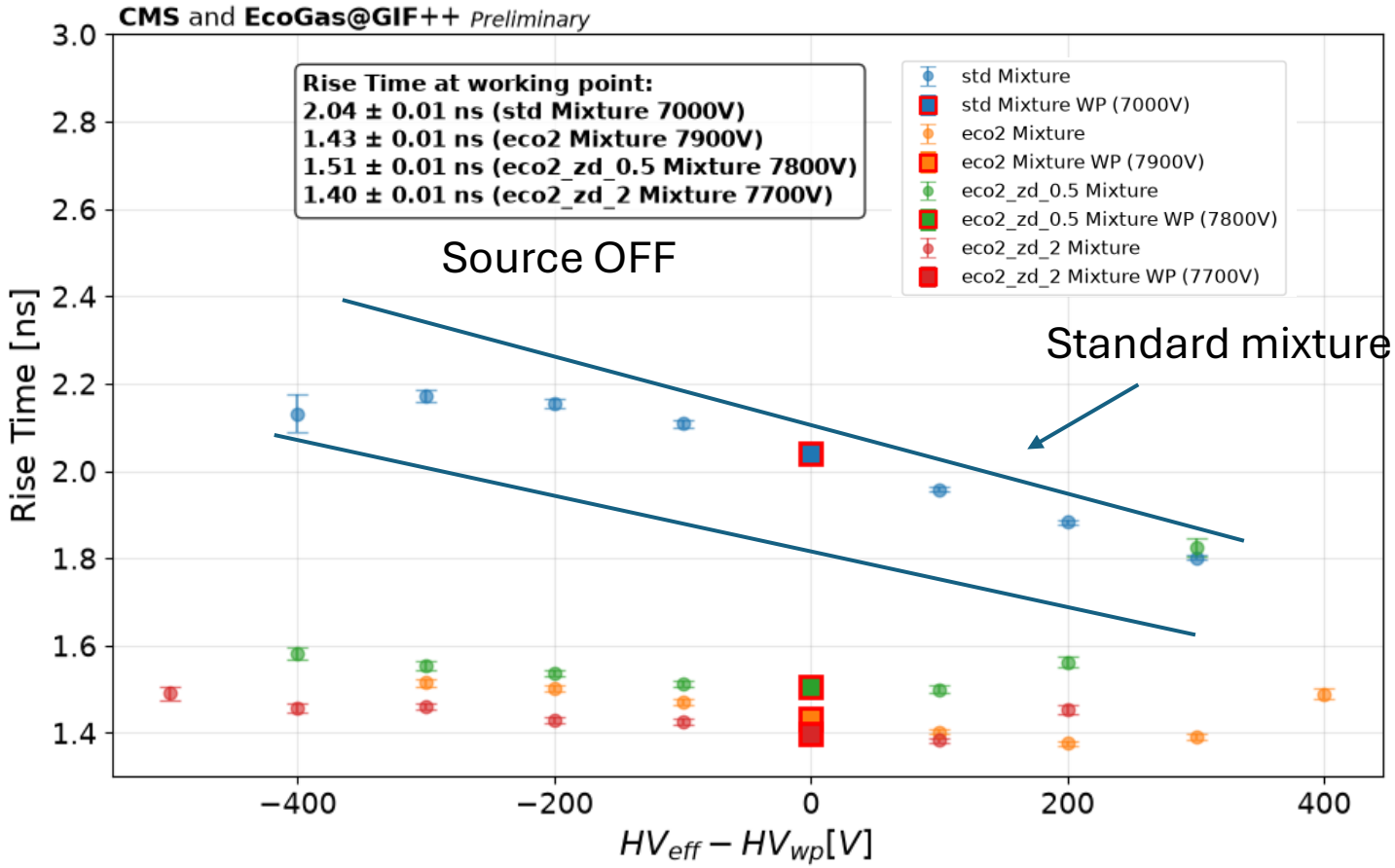






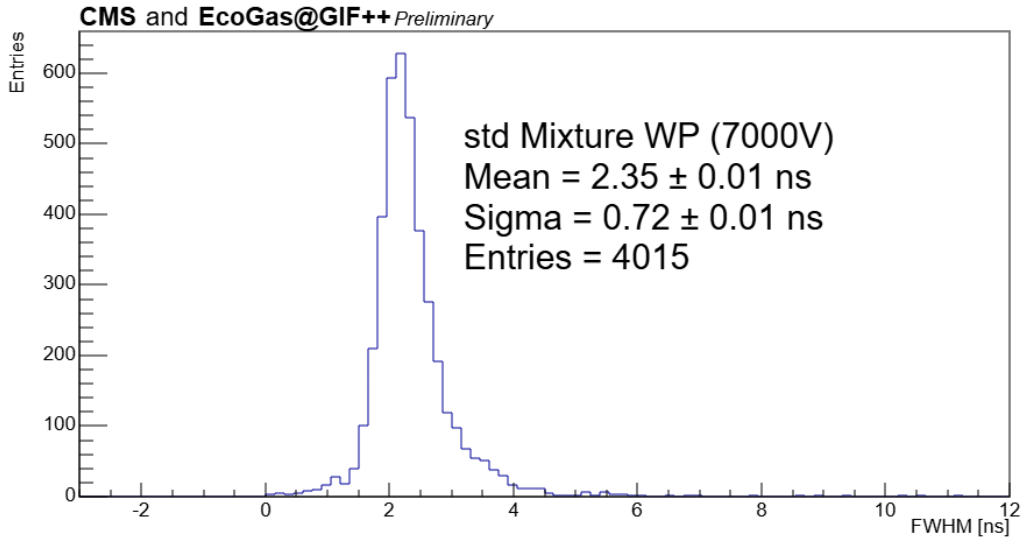
Rise Time

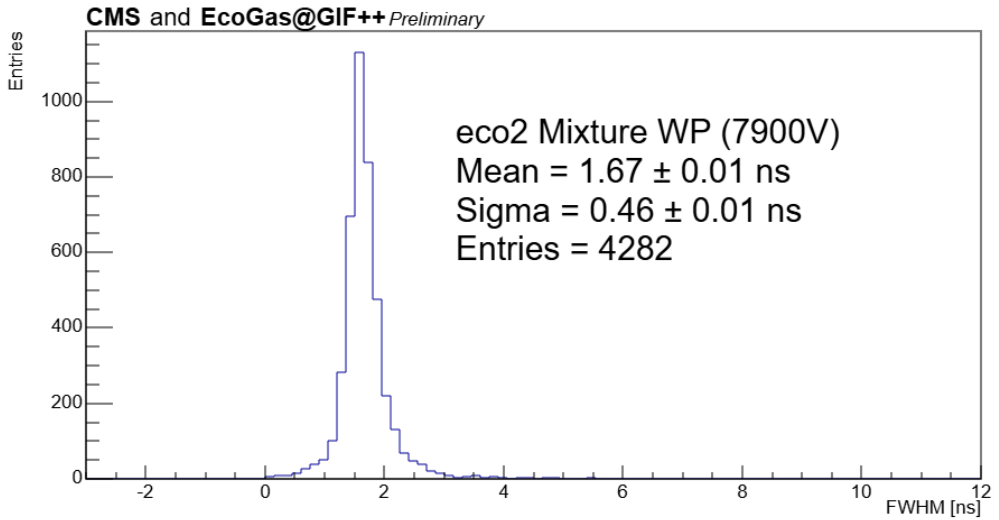
- The rise time becomes faster as the voltage increases.
- At the working point, the standard mixture presents a rise time greater than the eco mixture.
- The eco mixture starts to increase the rise time after the working point.

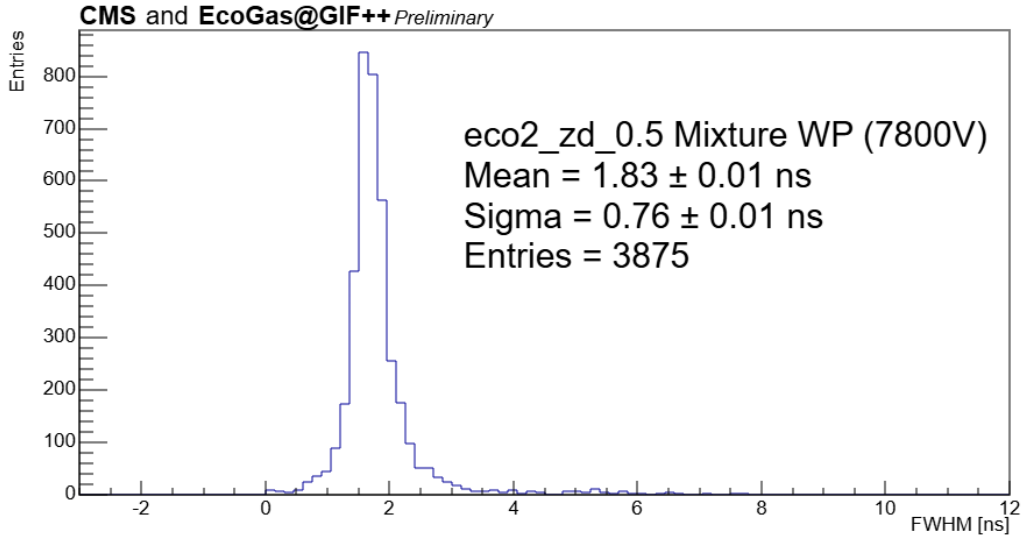


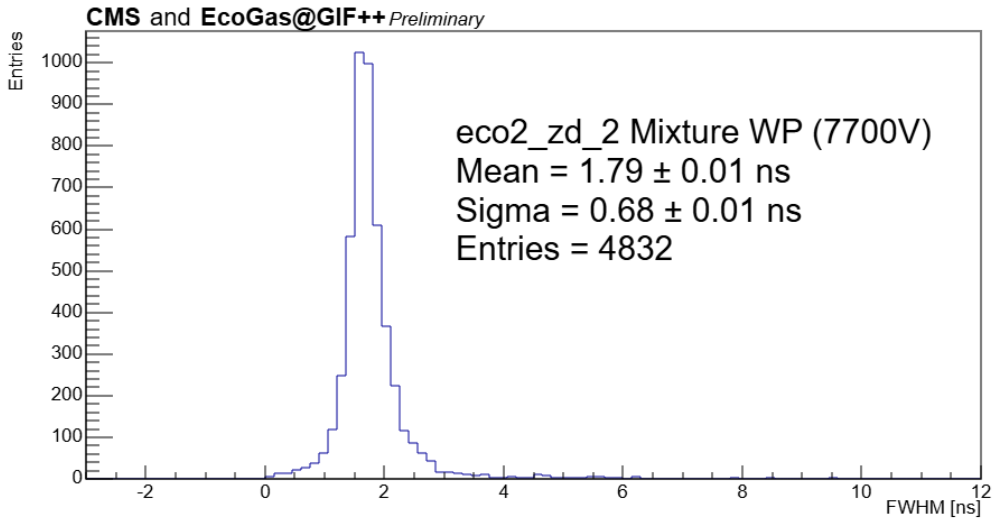
FWHM

- Seed strip selection (strip with the highest charge in the event).
- The standard mixture presents a larger width than the eco mixture.



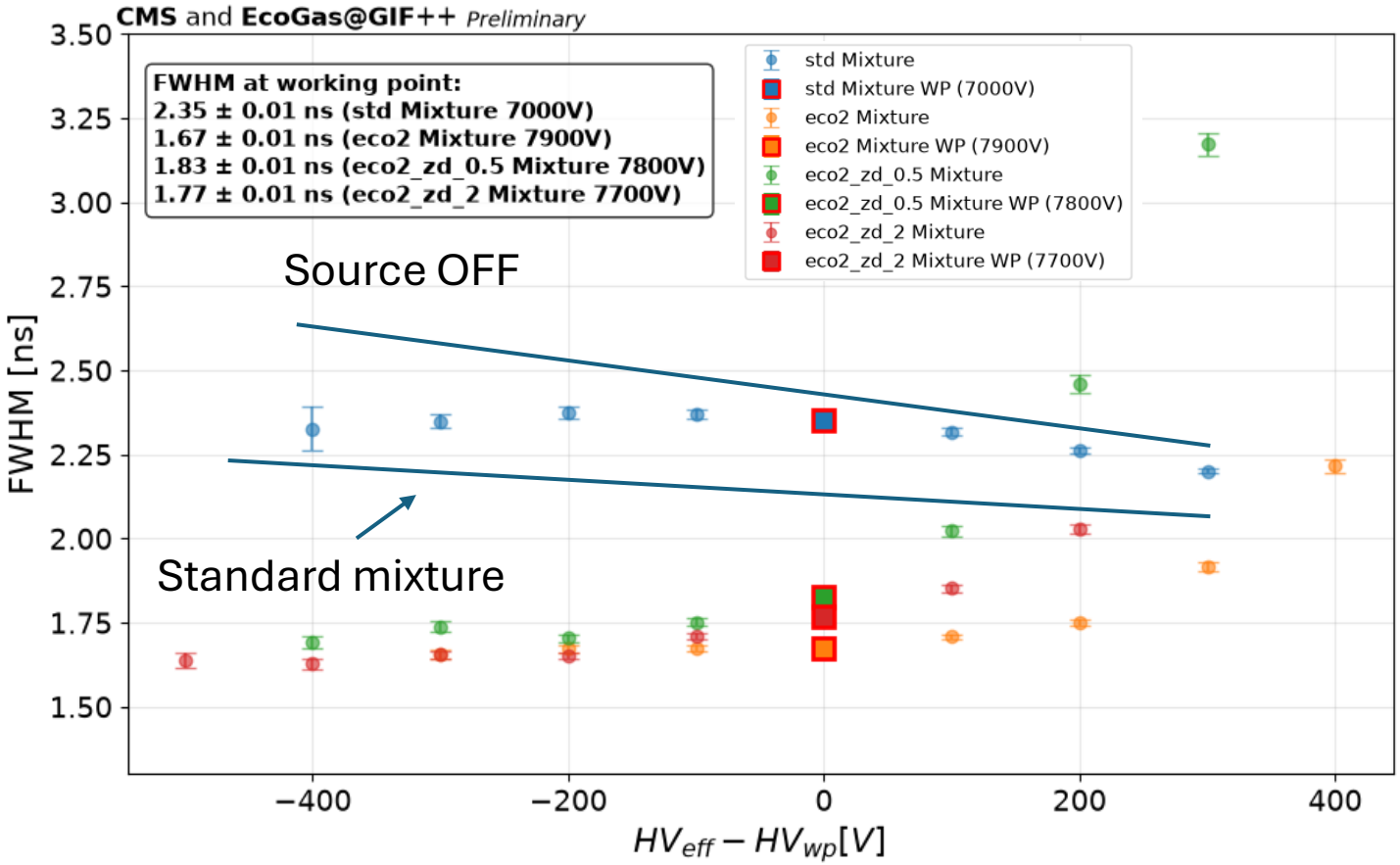






FWHW

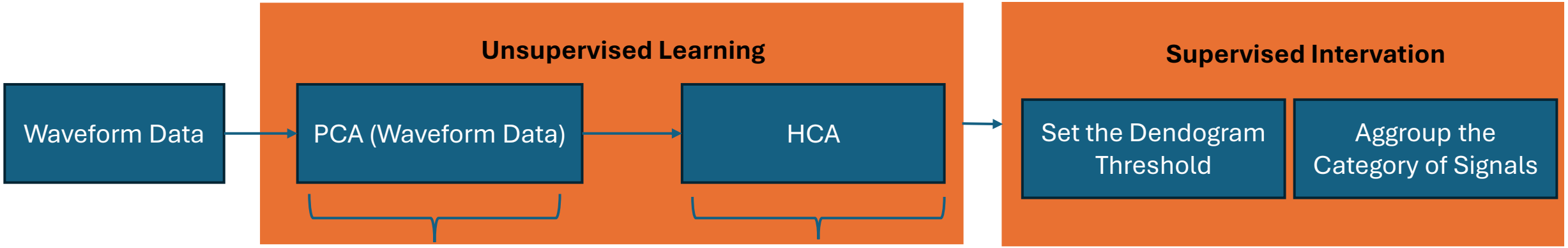
- The FWHM for the standard mixture becomes narrower with an increase in voltage
- The eco mixtures increase the FWHM as the voltage rises.
- At the working point, the standard mixture presents an FWHM larger than the eco mixture.



Machine Learning Analysis

Hierarchical Clustering Analysis with Principal Component Analysis

- Hierarchical Cluster Analysis (HCA) is an unsupervised machine learning method that groups similar data points into a tree of clusters.
- Principal Component Analysis (PCA) is used to reduce the dimensionality of the data so that the information is preserved and concentrated into the components.
- **Each gas mixture will be analyzed individually at the Working Point (WP).**
- **The analysis is performed in the strip with more efficiency.**
- **From this point on, the term "cluster" refers to a grouping of classified population.**

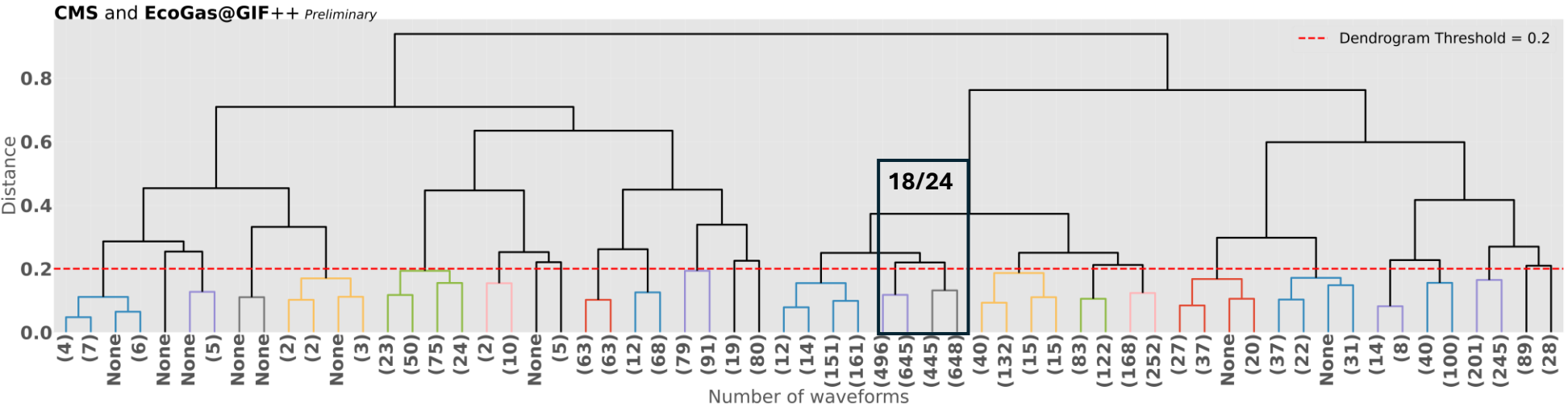


$$Waveform \rightarrow [\,V_1, V_2, V_3, V_4, \ldots V_n] \xrightarrow{\text{PCA}} PCA(Waveform) = [\,c_1, c_2, c_3, c_4, \ldots c_i]$$

$$Similarity = \sqrt{\sum_{i=0}^{components} (PCA(wf_1[i]) - PCA(wf[i]))^2}$$

Machine Learning Analysis (HCA)

- Clustering the standard mixture data, the dimensionality is reduced to 165 components.
- Setting a dendrogram threshold of 0.2, the data is split into 26 clusters.
- Looking inside each cluster, three categories of waveforms were observed.
  - Avalanche with one-peak waveform
  - Not active strips
  - Anomalous waveforms



	Cluster	Number of Waveform	Percentage
0	Cluster 0	172	3.44%
1	Cluster 1	202	4.04%
2	Cluster 2	170	3.40%
3	Cluster 3	446	8.92%
4	Cluster 4	140	2.80%
5	Cluster 5	91	1.82%
6	Cluster 6	8	0.16%
7	Cluster 7	80	1.60%
8	Cluster 8	12	0.24%
9	Cluster 9	338	6.76%
10	Cluster 10	6	0.12%
11	Cluster 11	80	1.60%
12	Cluster 12	2	0.04%
13	Cluster 13	85	1.70%
14	Cluster 14	18	0.36%
15	Cluster 15	420	8.40%
16	Cluster 16	126	2.52%
17	Cluster 17	1	0.02%
18	Cluster 18	1093	21.86%
19	Cluster 19	5	0.10%
20	Cluster 20	28	0.56%
21	Cluster 21	22	0.44%
22	Cluster 22	19	0.38%
23	Cluster 23	1	0.02%
24	Cluster 24	1141	22.82%
25	Cluster 25	205	4.10%
26	Cluster 26	89	1.78%
27	ALL	5000	100.00%

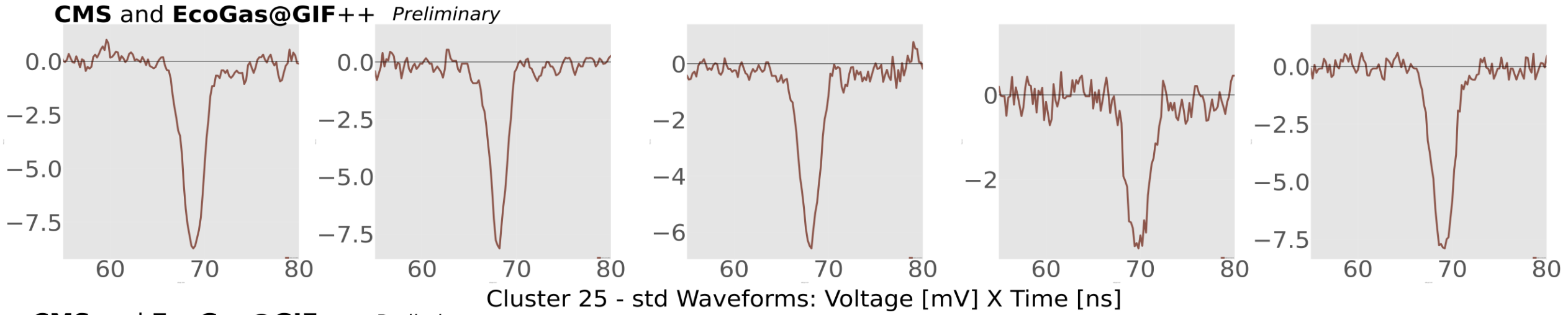
Summary HCA Analysis

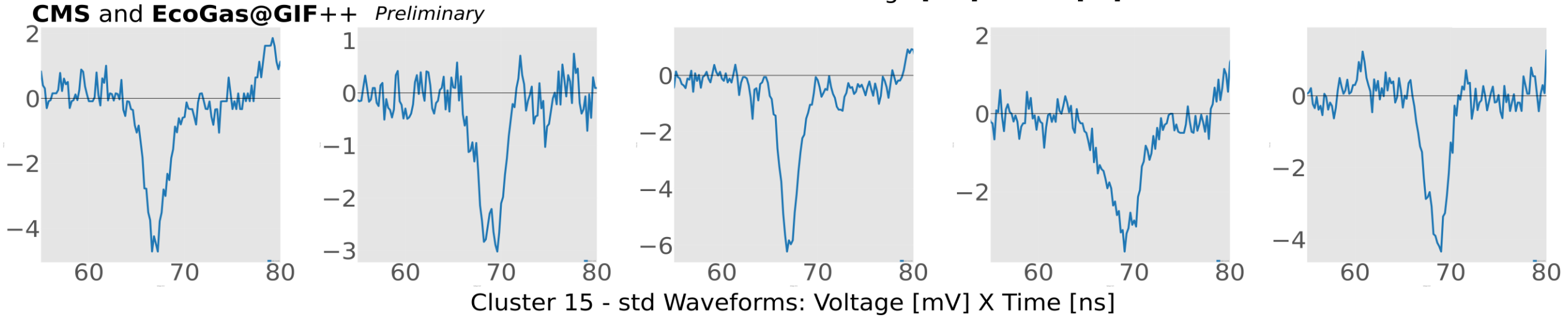
- The analysis presents more than ≈50% activated strips (analysis extends to just one strip).
- Anomalous waveforms independent of the mixture.
- Candidates for streamers (double peak waveforms) present only in eco mixtures at the working point, as well as long-end waveforms.
- The number of clusters is greater for the eco mixture.
- The model was able to split the cluster into 5 different categories of waveforms.
- Eco mixtures present 2 additional categories compared to the standard mixture.

Waveforms Cluster Category	std	eco2	eco2_zd_0.5	eco2_zd_2
One Peak Waveforms Population (A) [%]	53.7	37	28	41
Not Activated Strips Waveforms Population (B) [%]	44.7	46	51	45
Anomalous Waveforms Population (C) [%]	1.6	1.3	2	2.5
Double Peak Waveforms Population (D) [%]	0	0.02	3	1
Long-end Waveforms Population (E) [%]	0	15.68	16	10.5
Number of Clusters	26	46	66	62

Category A (one peak waveform)

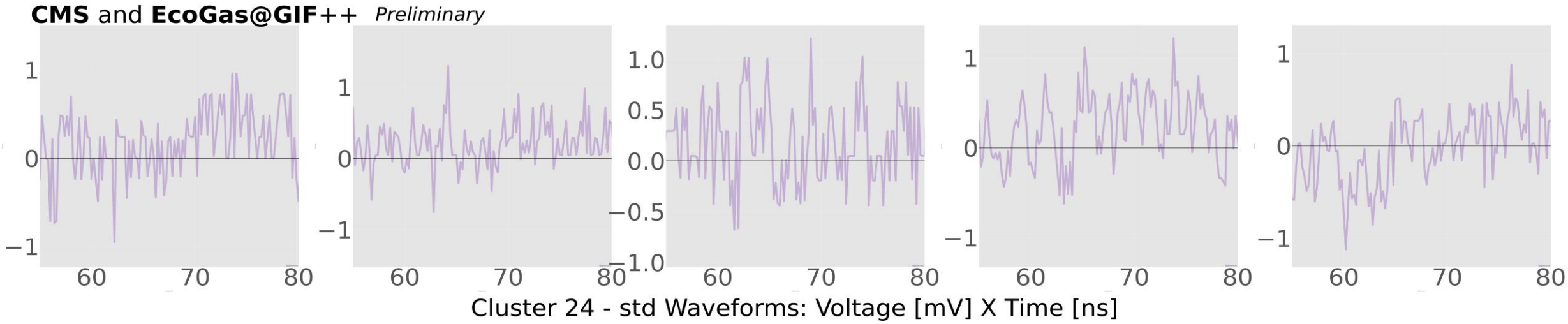
- Standard mixture waveforms
- Clusters from the same category.

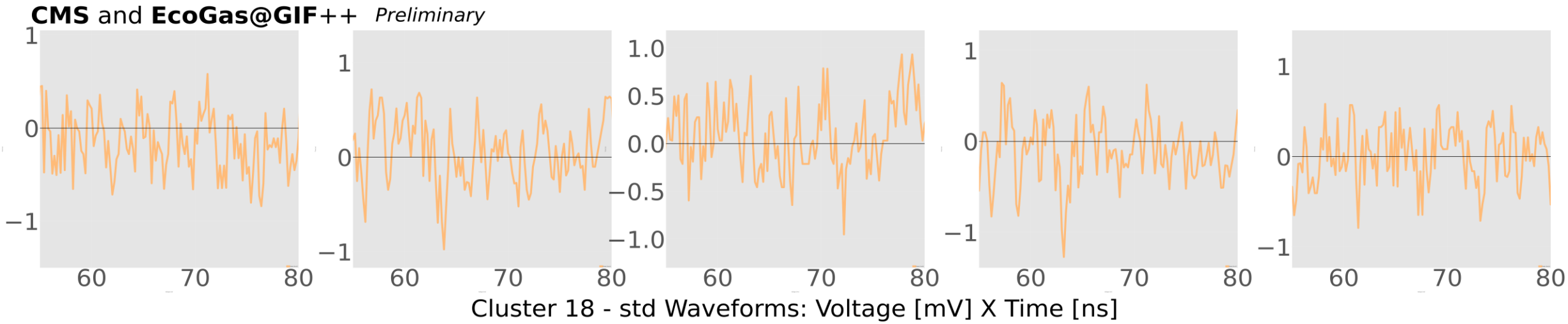




Category B (Not Active Strips)

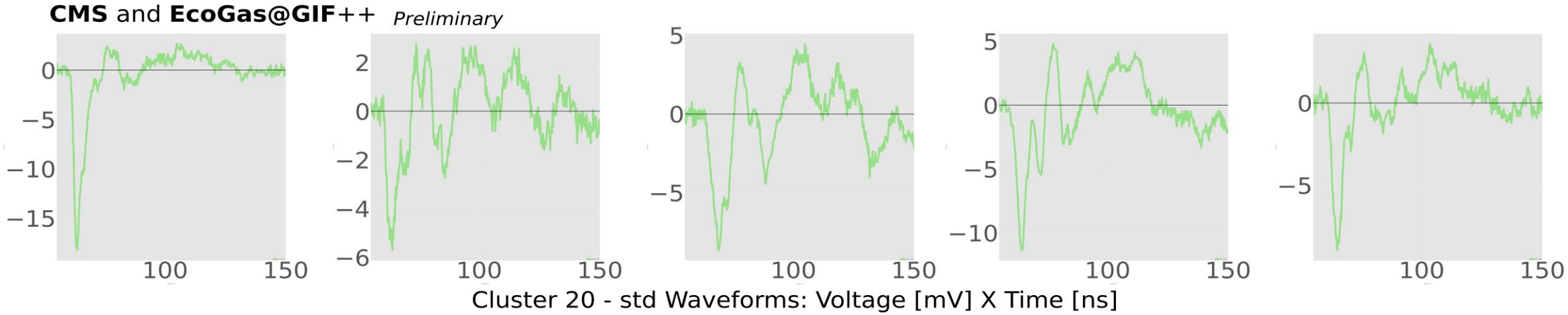
- Standard mixture waveforms
- Clusters from the same category.

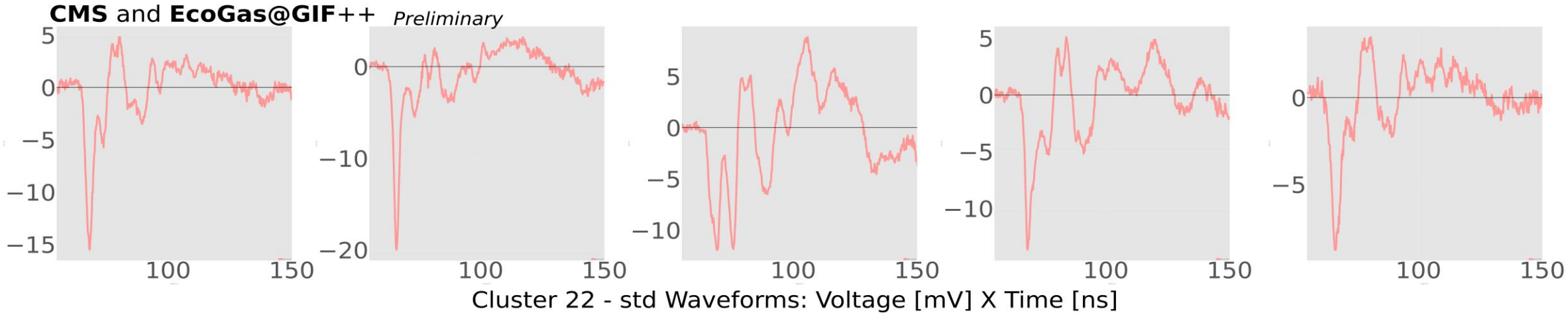




Category C (Anomalous waveforms)

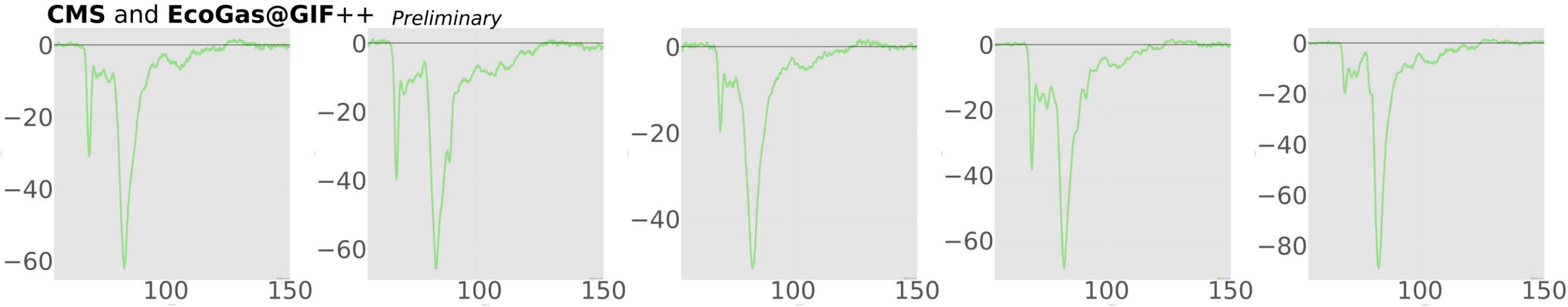
- Standard mixture waveforms
- Clusters from the same category.

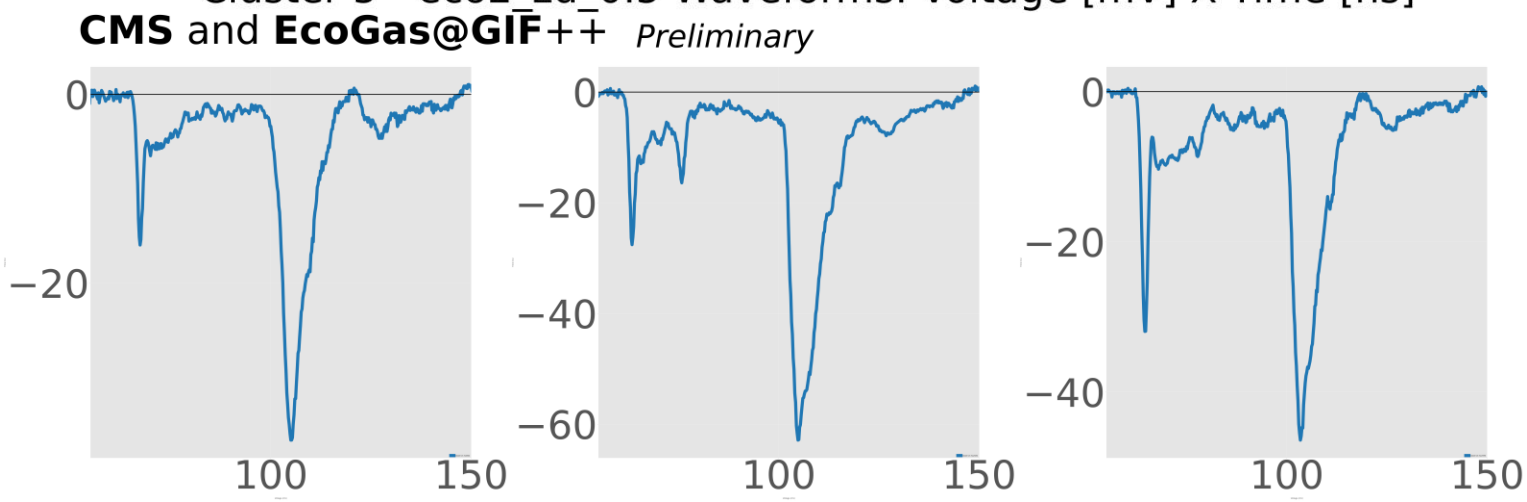




Category D (double peak waveforms)

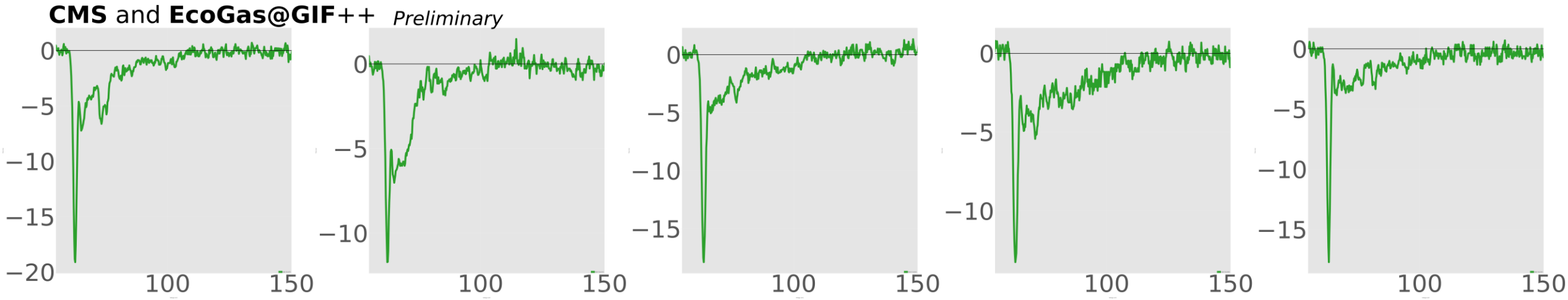
- Present only in the eco mixture waveforms at the working point.
- Candidate to streamers
- Cluster from the same category.

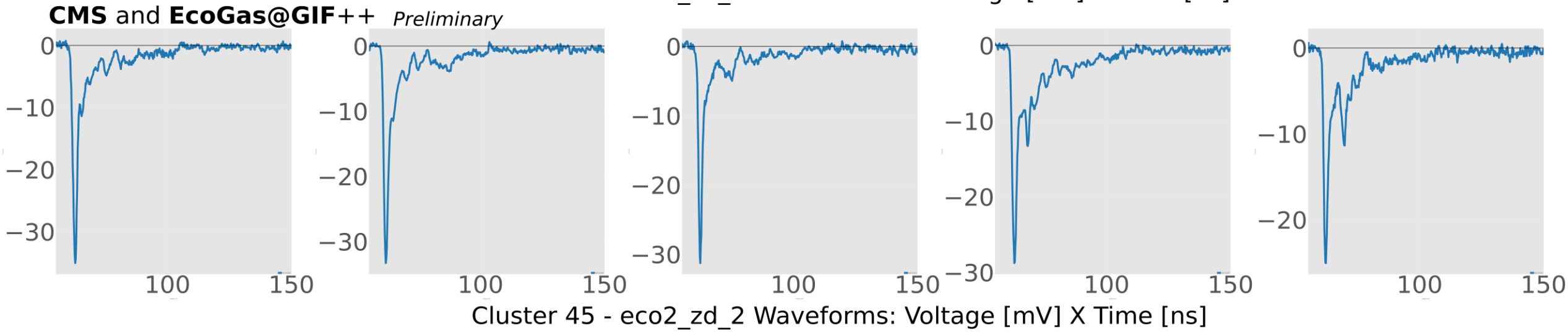




Category E (Long-End waveforms)

- Present only in the eco mixture waveforms at the working point.
- Cluster from the same category.





Conclusions on Alternative mixtures

- The algorithms developed were able to analyze the waveform parameters, giving information about how the parameters change when the electric field increases and when the gas mixture is changed.
- **Charge:** Eco mixtures show a higher average charge and larger cluster size than the standard mixture.
- **Timing:** Rise time measurements indicate that eco mixtures produce faster signals than the standard $C_2H_2F_4$-based blend.
- **Signal Width:** The width of the signal is smaller for the eco mixtures at the working point (WP).
- **Time Resolution:** The time resolution of the chamber has improved when the chamber is operating with the eco mixture.
- **GWP:** In terms of the global warming potential, the component HFO-1233zd is 23500 times lower.
- **Trade-off:** There is a trade-off between the rate capability and time resolution

Conclusions on Machine Learning Analysis

- The Hierarchical Clustering Analysis (HCA) combined with Principal Component Analysis (PCA) provided a second way to discriminate the signal using only the shape as input.

- The results are promising for classifying the signals according to their shape, making it possible to identify different shapes of signals.

Next Steps

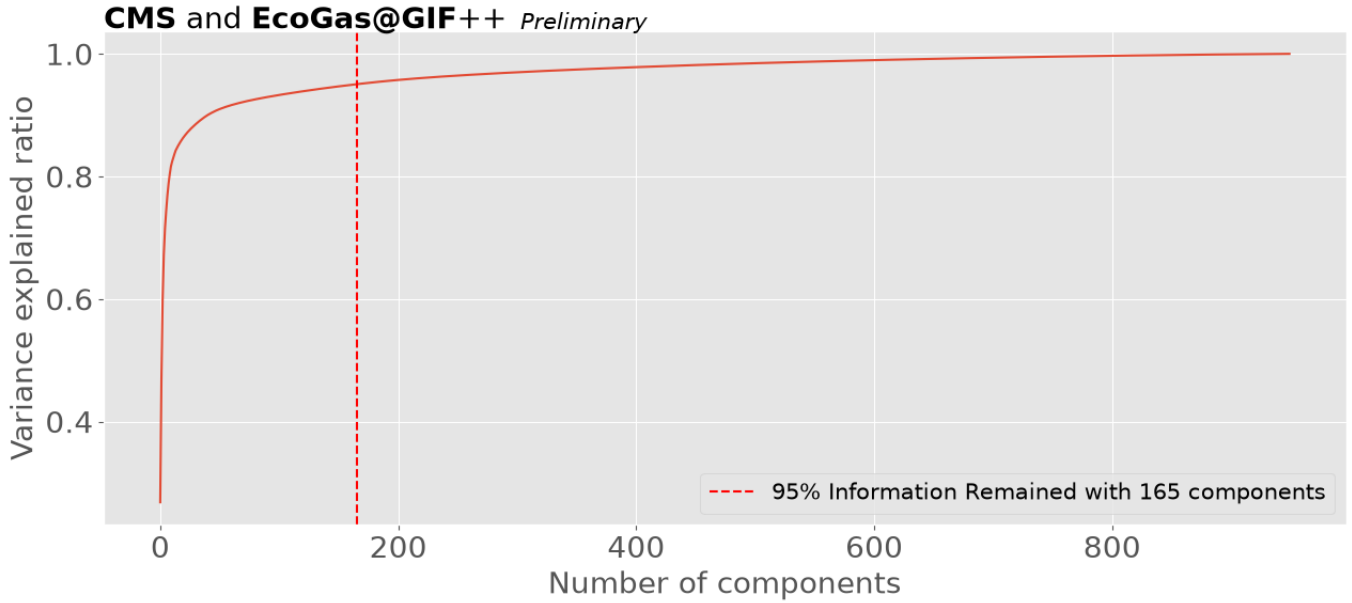
- Extend the analysis for the fall time.
- Investigate the streamer probability and FWHM broadening mechanism in eco mixtures at high voltage
- Extend analysis to additional operating voltages and gas mixture ratios
- Extend the analysis for when the chamber is being irradiated.
- Implementing seed strip selection in the Hierarchical Clustering Analysis.
- Trained a supervised learning model, such as a convolutional neural network, using discrimination from unsupervised learning.

Questions???

Charge Calculation

The plot shows an example of one waveform collected from one strip. From the shape of the waveform, it is possible to measure the **charge**, defined as $\mathrm{Q} = \int \frac{V(t)dt}{\Omega}$, the calculation is approximated to Riemann integral, approximating the area under the curve as the sum of rectangular slices over discrete time samples: $Q \approx \frac{\Delta t}{\Omega} \sum_i V_i$, where Δt is the sampling interval, $\Omega = 50\ \mathrm{Ohm}$ and V_i are the voltage sample; as well as the amplitude **amplitude**; both can be defined as size parameters.

Machine Learning Analysis (HCA)



Instead of working with hundreds or thousands of raw, often correlated features, the PCA creates a new set of uncorrelated variables called **Principal Components**. These components are ranked by how much of the original data's variance they capture. The first component captures the largest variance, the second the next-largest, and so on. To determine how many principal components to retain, we plotted the **Cumulative Explained Variance Ratio** as a function of the number of components. In the plot, the X-axis shows the number of components, and the Y-axis shows the cumulative percentage of the total dataset variance captured. The procedure here is to use the number of components that capture at least 95% of the variance; in this case, the 1024 points of voltage were reduced to 165 components.

References

1. Space resolution measurement and threshold calibration of the iRPC detector, CMS Collaboration, M. Gouzevitch et al., Nucl. Instrum. Meth. A 1083 (2026) 171125, arXiv:2511.xxxxx [hep-ex], Nov 14, 2025.
2. M.R. Jäkel, M. Capeans, I. Efthymiopoulos, A. Fabich, R. Guida, G. Maire, M. Moll, D. Pfeiffer, F. Ravotti, H. Reithler, CERN-GIF++: a new irradiation facility to test large-area particle detectors for the high-luminosity LHC program, PoS TIPP2014 (2014) 102.
3. D. Pfeiffer et al., "The radiation field in the Gamma Irradiation Facility GIF++ at CERN," Nucl. Instrum. Meth. A, vol. 866, pp. 91–103, 2017, doi: 10.1016/j.nima.2017.05.045.
4. K. Mota Amarilo et al., "RPC based tracking system at CERN GIF++ facility," Nucl. Instrum. Meth. A, vol. 1052, p. 168271, 2023, doi: 10.1016/j.nima.2023.168271.
5. G. Proto et al., "Performance of new generation of Resistive Plate Chambers operating with alternative gas mixtures," Nucl. Instrum. Meth. A, vol. 1066, p. 169580, 2024, doi: 10.1016/j.nima.2024.169580.
6. SANTONICO, R.; CARDARELLI, R. Development of resistive plate counters. Nuclear Instruments and Methods in Physics Research, v. 187, p. 377-380, 1981.
7. SANTONICO, R.; CARDARELLI, R. Development of resistive plate counters. Nuclear Instruments and Methods in Physics Research, v. 187, p. 377-380, 1981.
8. ABBRESCIA, M.; PESKOV, V.; FONTE, P. Resistive Gaseous Detectors: Designs, Performance, and Perspectives. [S.l.]: Wiley, 2018.
9. ATLAS Collaboration. Performance of the ATLAS RPC detectors with tetrafluoroethane-based gas mixtures. Journal of Instrumentation, v. 18, p. P05010, 2023.
10. CMS Collaboration. The CMS experiment at the CERN LHC. Journal of Instrumentation, v. 3, p. S08004, 2008.
11. RIGOLETTI, G.; GUIDA, R.; MANDELLI, B. Performance studies of RPC detectors operated with $C_2H_2F_4$ and CO_2 gas mixtures. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, v. 1049, p. 168088, 2023. ISSN 0168-9002
12. BLUM, W.; ROLANDI, L.; RIEGLER, W. Particle Detection with Drift Chambers. Springer, Berlin, Heidelberg, 2008.
13. JAMES, G. et al. An Introduction to Statistical Learning with Applications in R. 8. ed. Germany, Berlin Heidelberg: Springer-Verlag, 2017. 419 p.
14. CMS Collaboration. The Phase-2 Upgrade of the CMS Muon Detectors. 2017. 10.17181/CERN.5T9S.VPMI
15. LUZ, R. J. B. Development of the instrumentation and readout schemes of MARTA, an upgrade of the Pierre Auger Observatory. Tese (PhD Thesis), Lisboa, Portugal, 2020. Supervisor: Pedro Jorge dos Santos de Assis; Co-supervisor: Pedro Miguel Félix Brogueira. Jury classification: Pass with Distinction.
16. KASSAMBARA, A. Practical Guide To Cluster Analysis in R: Unsupervised Machine Learning. [S.l.]: STHDA, 2017. 187 p.
17. CERN. Garfield++. 2024: https://garfieldpp.web.cern.ch/garfieldpp.
18. CERN. Magboltz. 2024: https://magboltz.web.cern.ch/magboltz.
19. The Phase-2 Upgrade of the CMS Muon Detectors CMS Collaboration • Thomas Hebbeker (Aachen, Tech. Hochsch.) et al.
20. CMS MUON COLLABORATION. CMS iRPC FEB development and validation. Nuclear Instruments and Methods in Physics Research A, v. 1064, p. 169400, 2024.
21. RAMOS LÓPEZ, D. et al. New RPC gas mixtures for sustainable operation in the CMS experiment. Nucl. Instrum. Meth. A, v. 1069, p. 169950, 2024.
22. PROTO, G. et al. Characterization of new eco-friendly gas mixtures based on HFO for RPCs. Journal of Instrumentation, v. 16. n. 02, p. C02001, 2021.