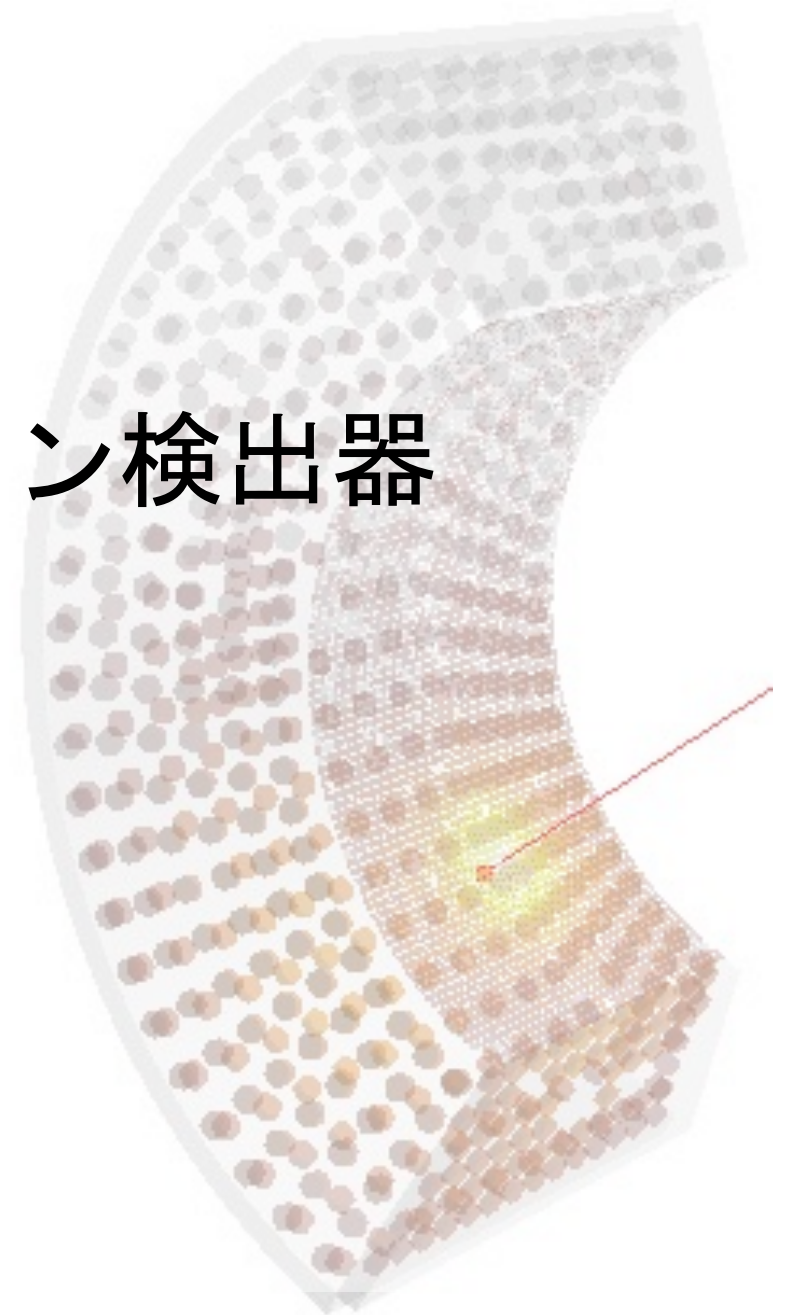


機械学習を用いたMEG IIキセノン検出器 におけるパイルアップ解析

日本物理学会第81回年次大会

榑原 滯, 他MEG IIコラボレーション

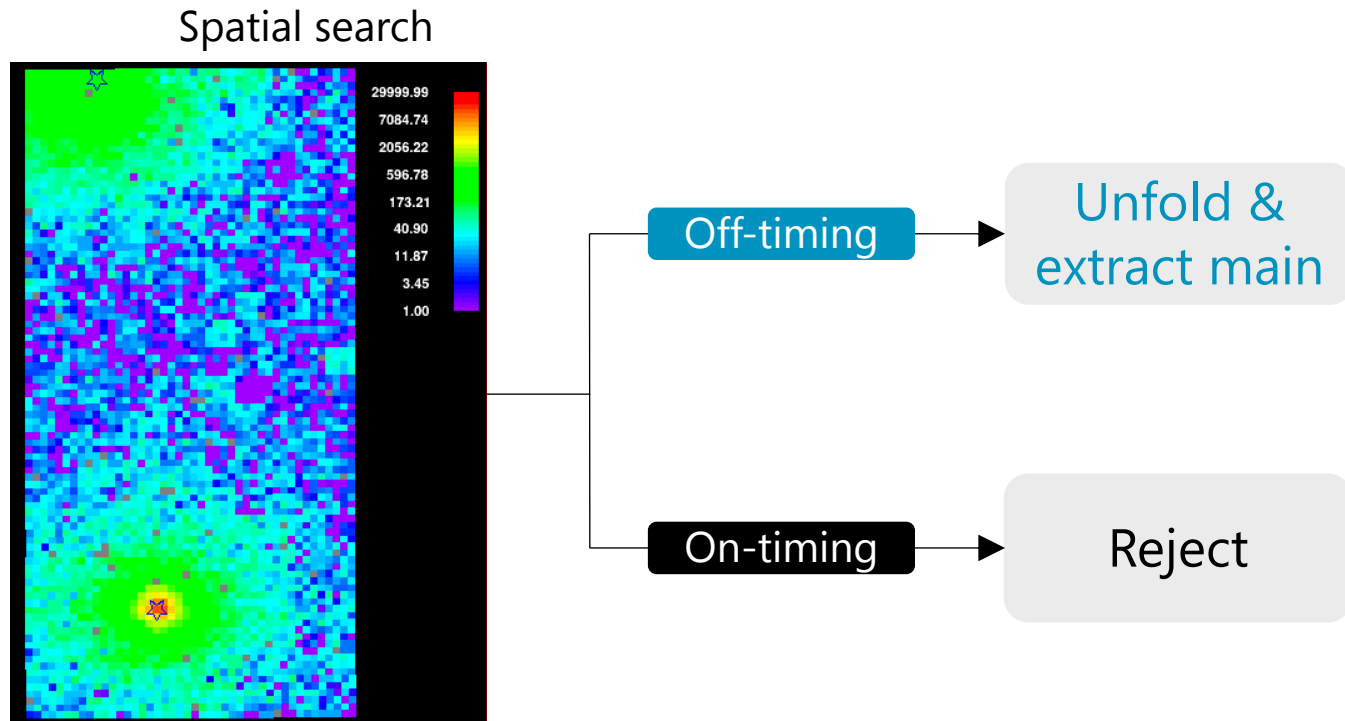
2026-09-17



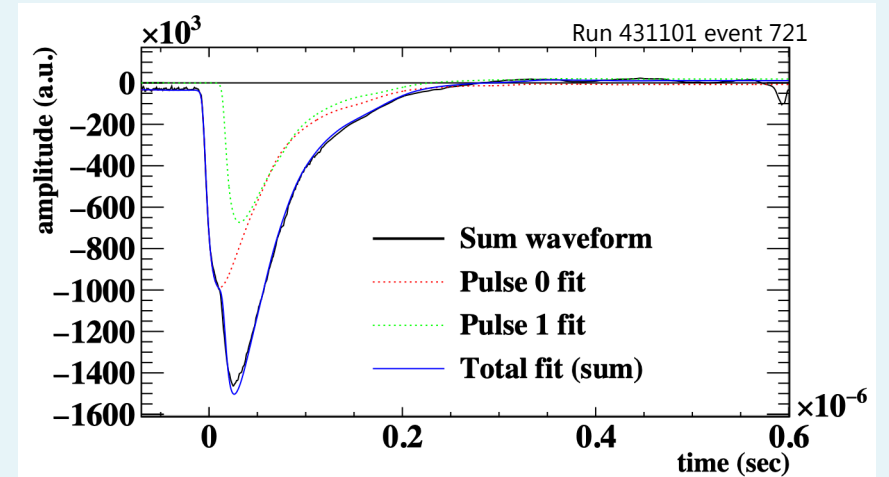
Multi-gamma events

- Many pileups observed due to high-rate muon decay ($3 - 5 \times 10^7 \mu / \text{sec}$)
- Current analysis consists of **spatial search** and **temporal search**

Current pileup analysis flow



Sum waveform of the PMTs

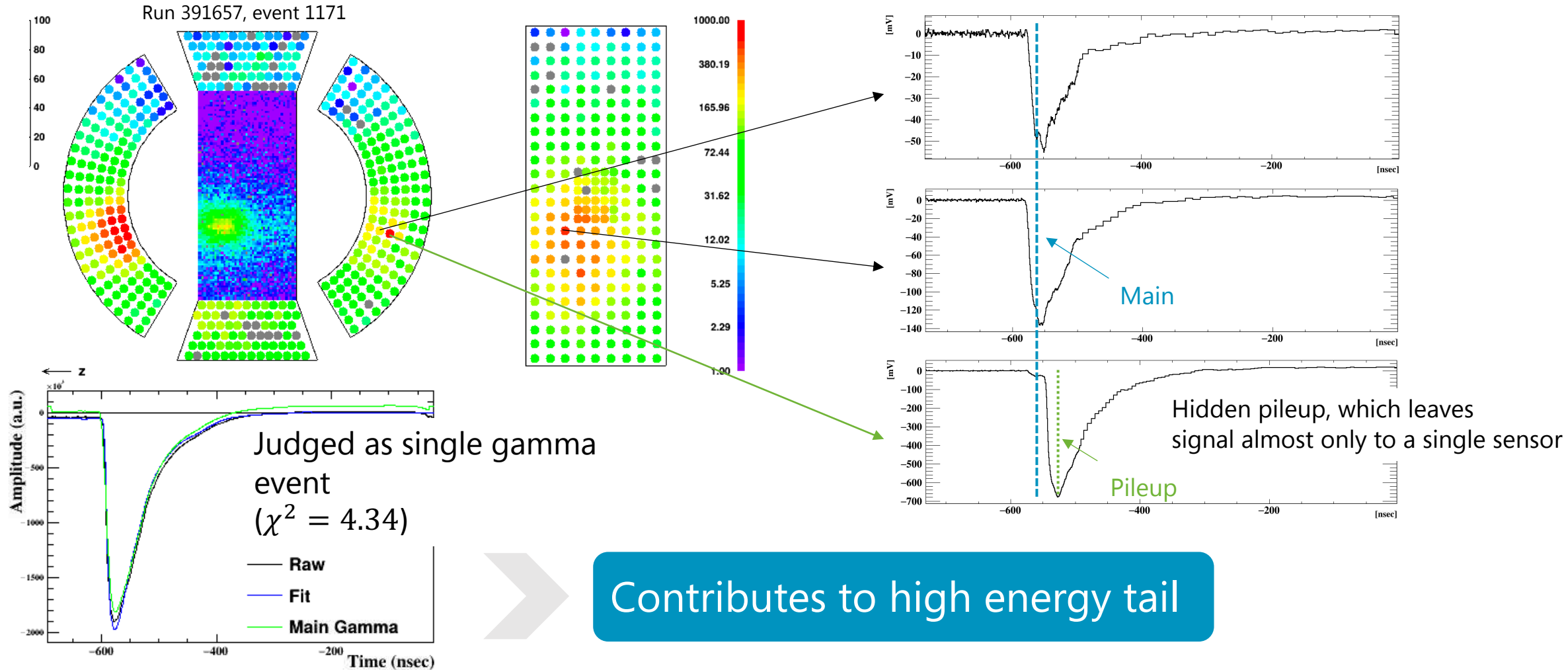


Fit initialization source:

- Differential waveform
- Trigger waveform
- Spatial search result

Limitation of current analysis

Some hidden pileups are found in event judged as “no pileup”

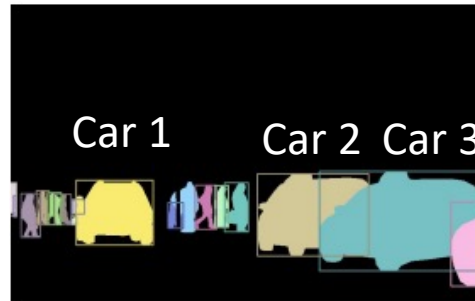
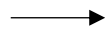


New Approach: Machine Learning Approach

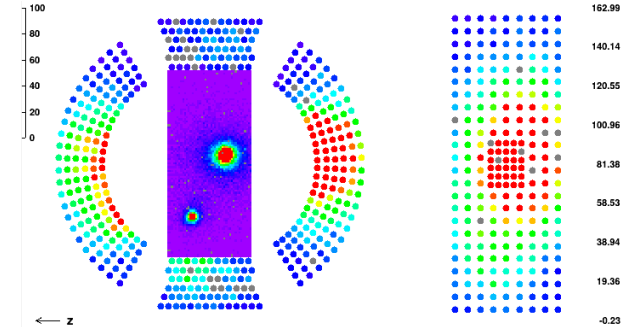
ML can effectively exploit spatial and temporal sensor correlations comprehensively

- Pileup separation is similar to object detection task

Object detection



Pileup detection

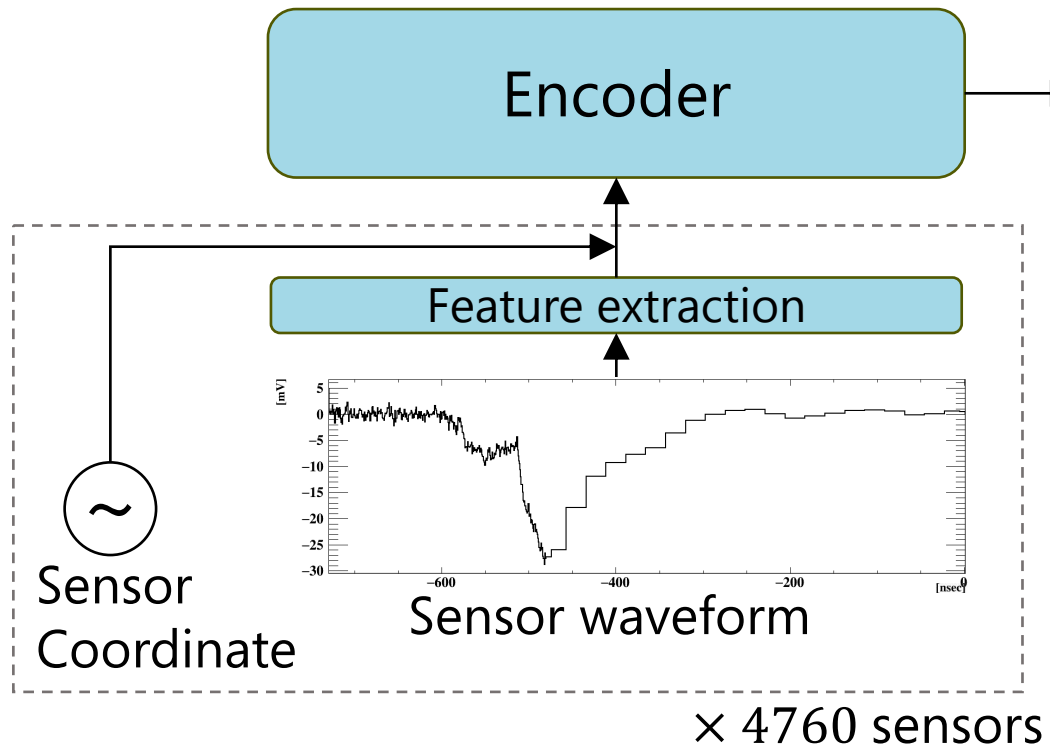


Photon 1,
Photon 2

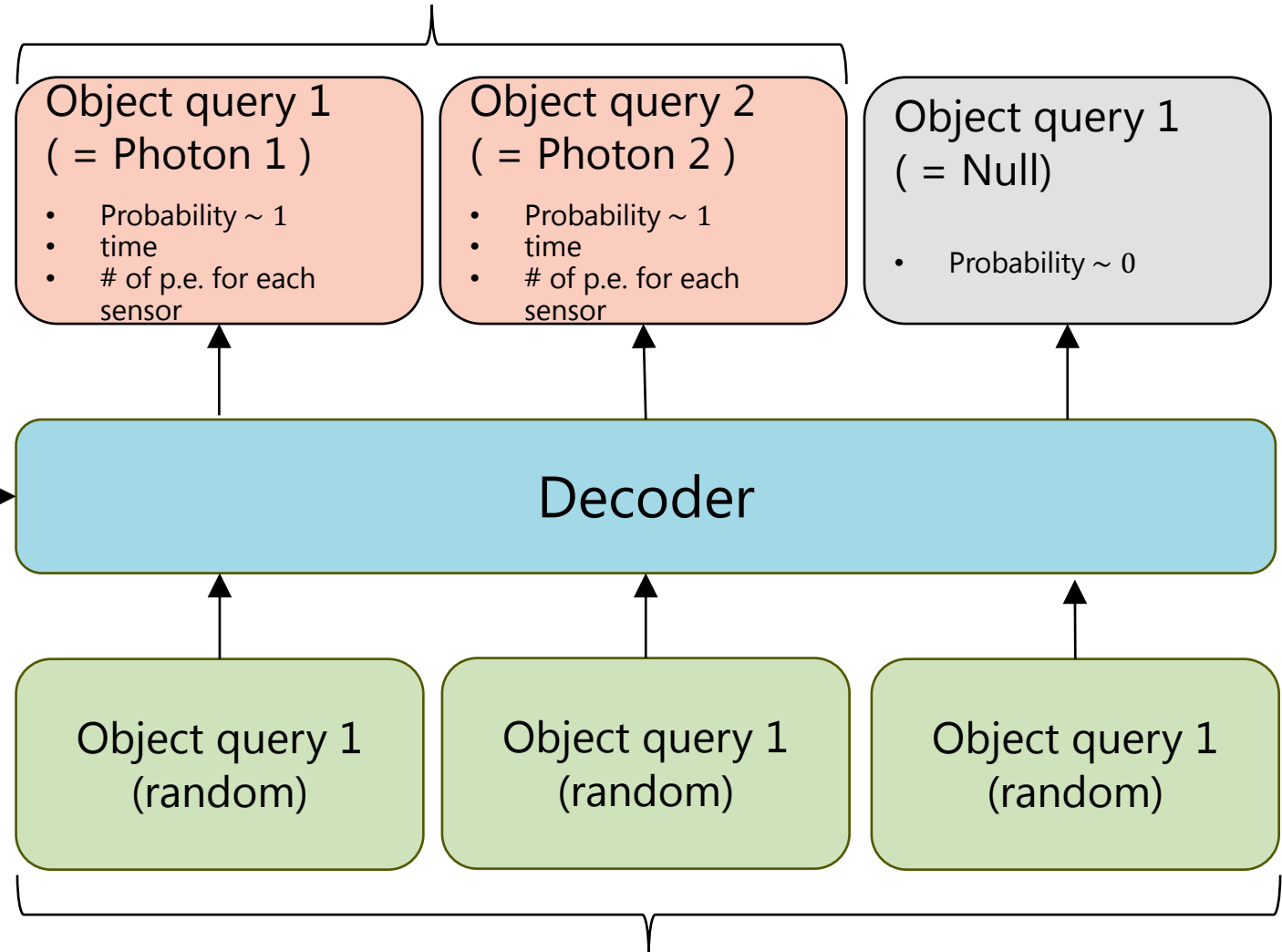
- Challenge: simulation vs data discrepancy
 - Simulation is not perfect → Good performance against simulation is not enough
 - Despite its strong performance on MC validation samples, the model trained on MC data exhibited a significant performance drop when evaluated on real data

Model architecture

- Model developed based on [Object Detection Transformer](#)
- Input: Sensor waveform
- Output: Variable number of photons and their information



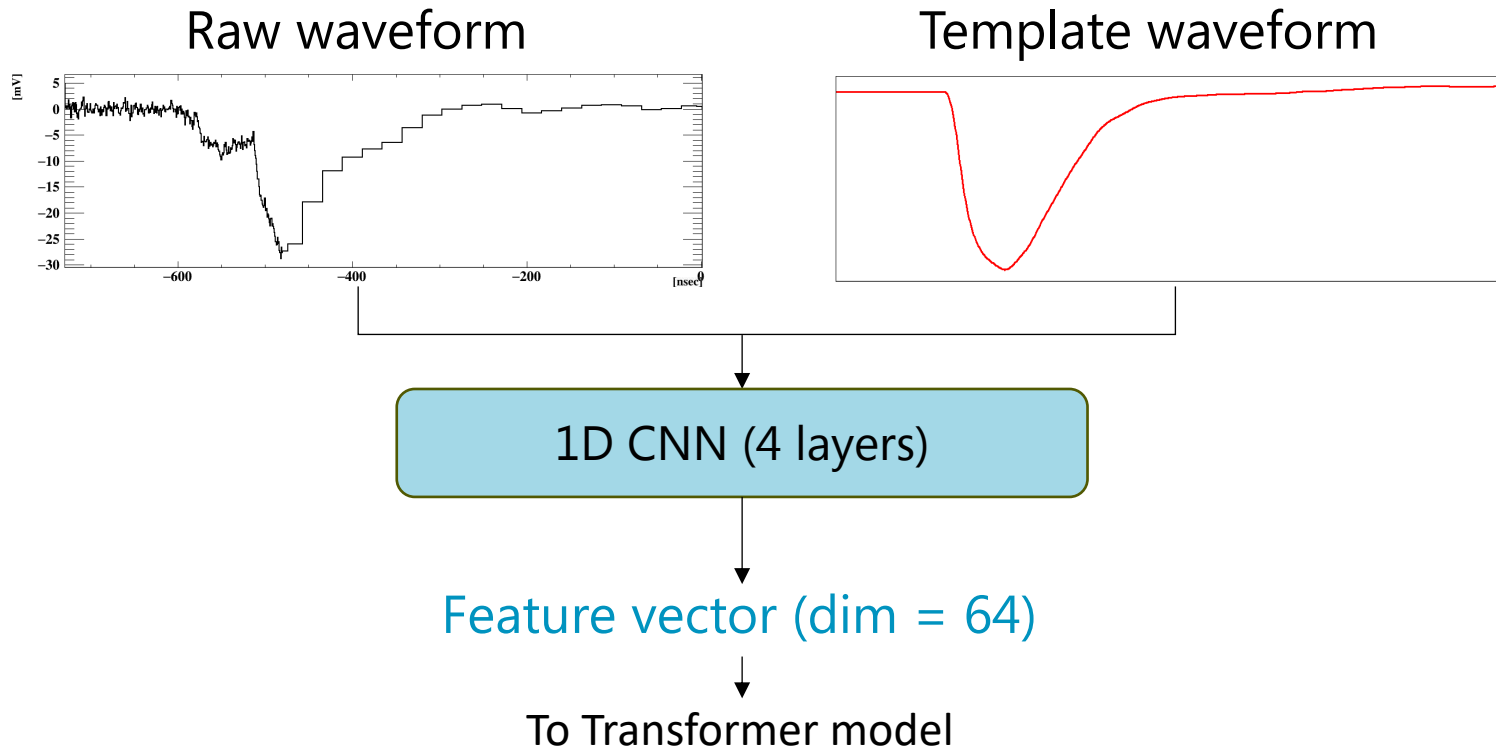
Variable number of prediction photons



Fixed number of object queries
(assumed maximum photons)

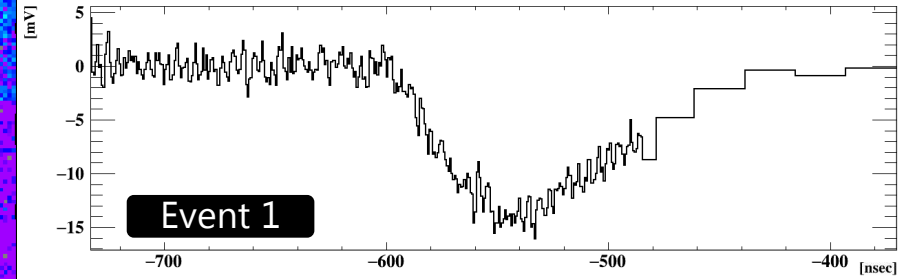
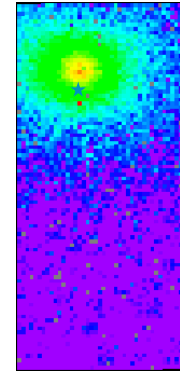
Feature extraction by 1D CNN

- 1D CNN extracts a feature vector from the raw 512-bin waveform
- The waveform shape for a single gamma event varies sensor-by-sensor
 - This was not modeled in the simulation, causing part of the data-simulation discrepancy
 - The model corrects for this by also feeding a template waveform into the CNN

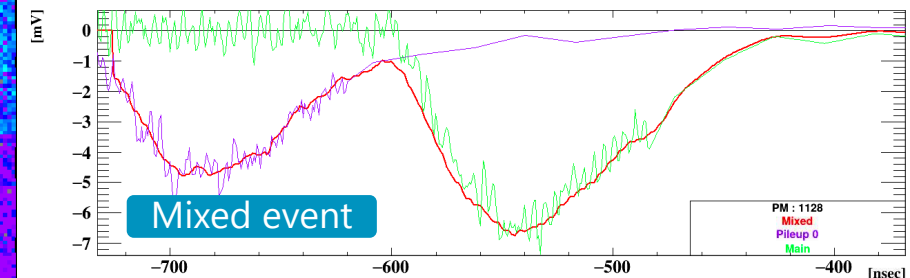
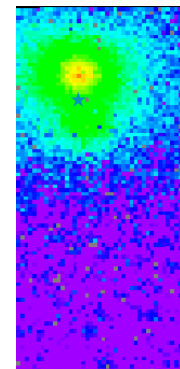
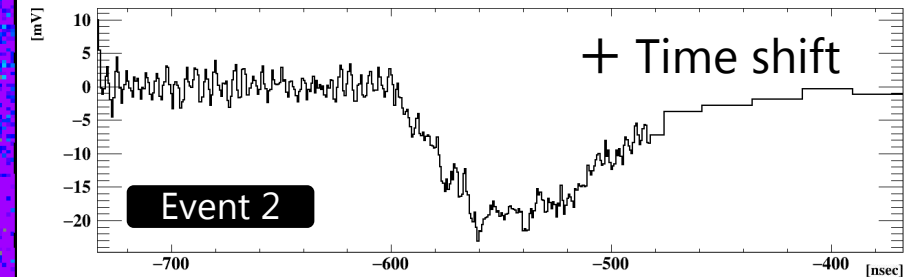
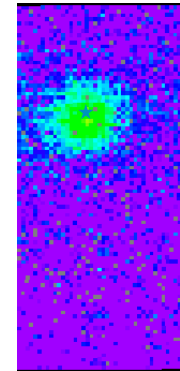


Preparation of training samples

- Made from real data to avoid simulation-data discrepancy
- Mix events with high probability of being a single gamma
 - Single gamma selection ... By setting a stricter cut on fit χ^2 value than the usual threshold
- Enhancing difficult samples
 - Converted in deep region
 - Small energy
 - On-time with main
 - Close to main



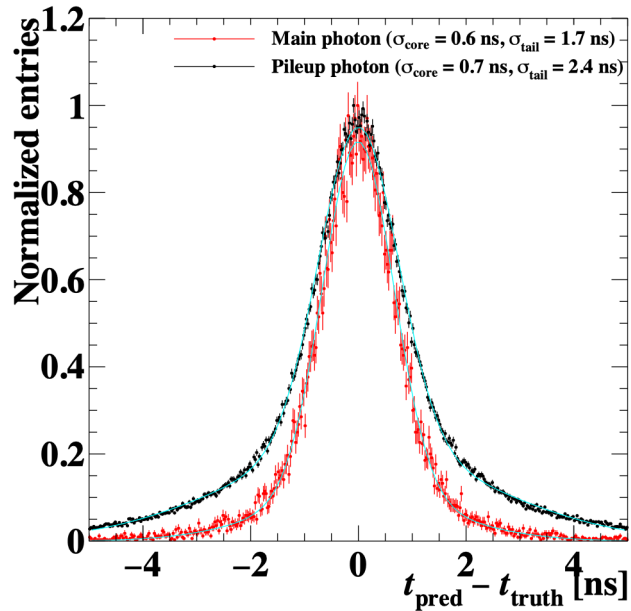
+



Mixed waveforms are scaled to the typical calibration factors

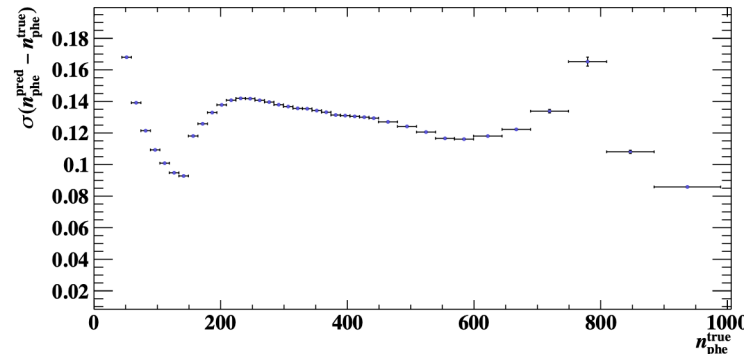
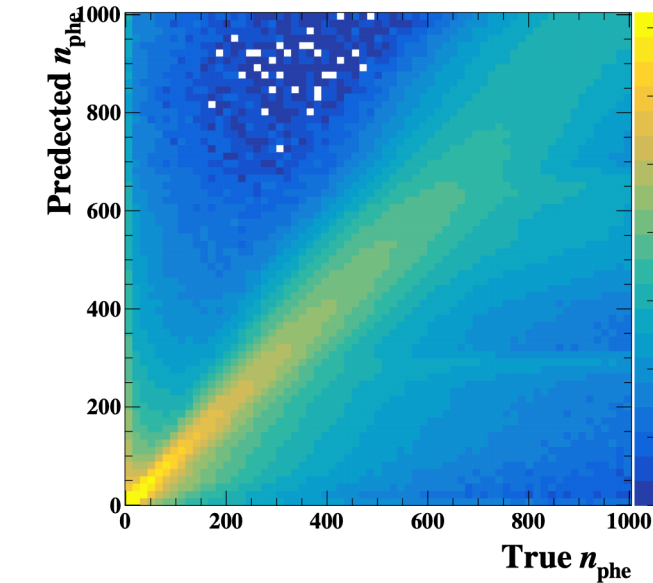
Performance on artificial pileups

Time



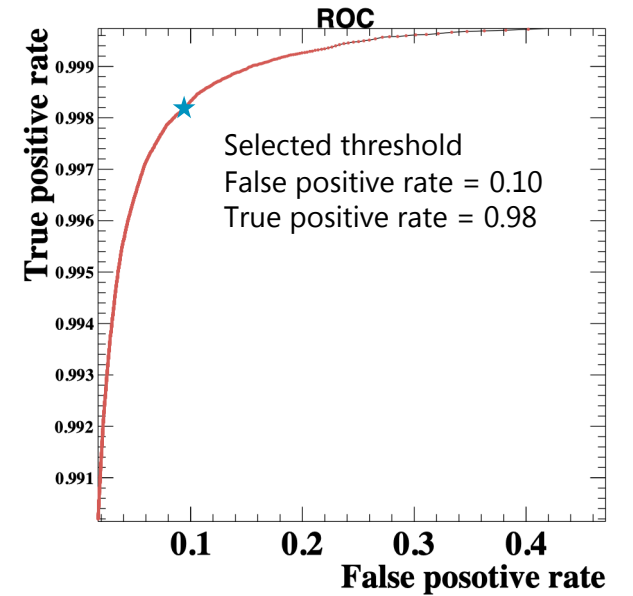
Below 1 ns core
time resolution

of photoelectrons



10~15 % n_{phe} resolution

Photon probability



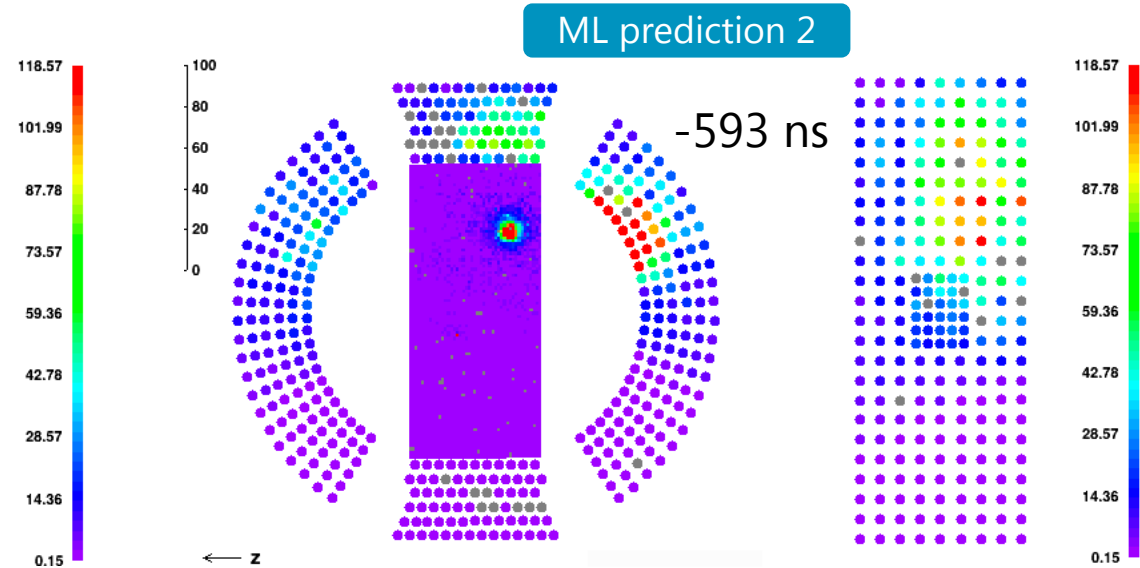
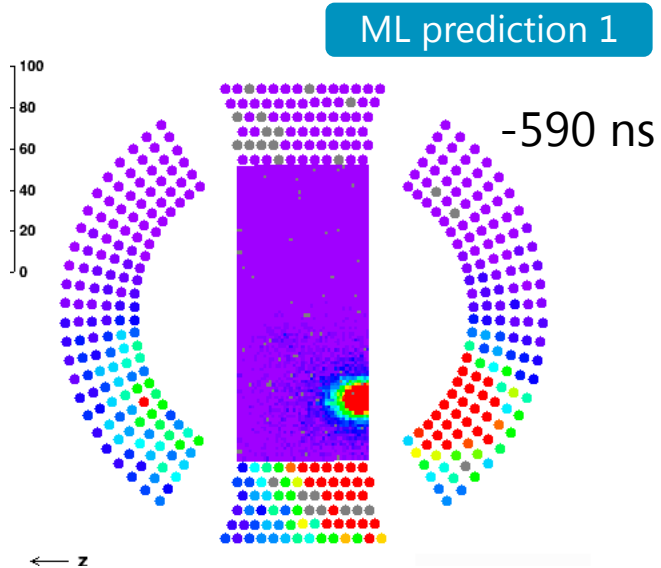
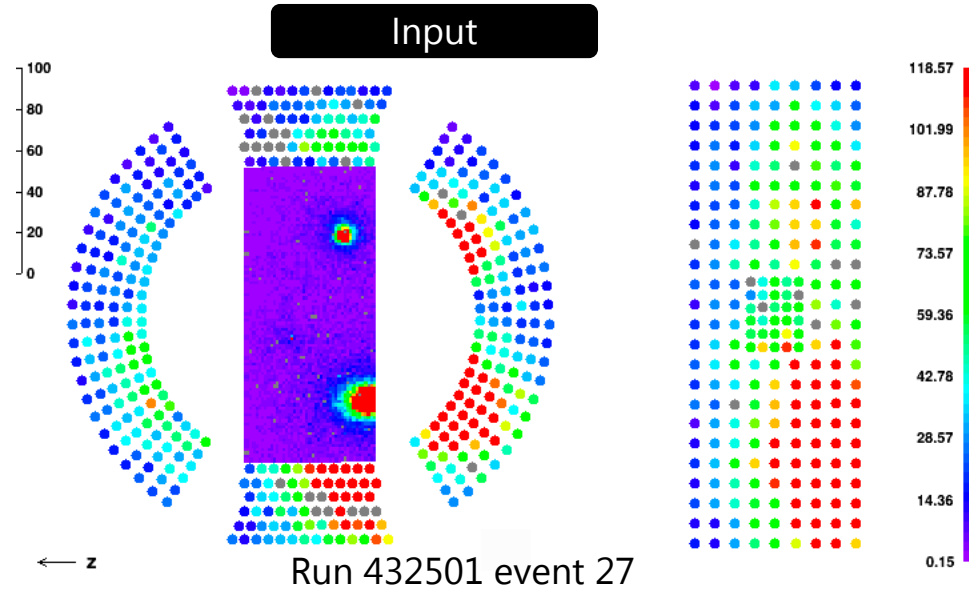
Prediction

Truth		1	2	≥ 3
	1	95.2 %	4.8 %	0.0 %
	2	1.7 %	94.1 %	4.2 %
	≥ 3	0.0 %	1.1 %	98.9 %

Number of total photons

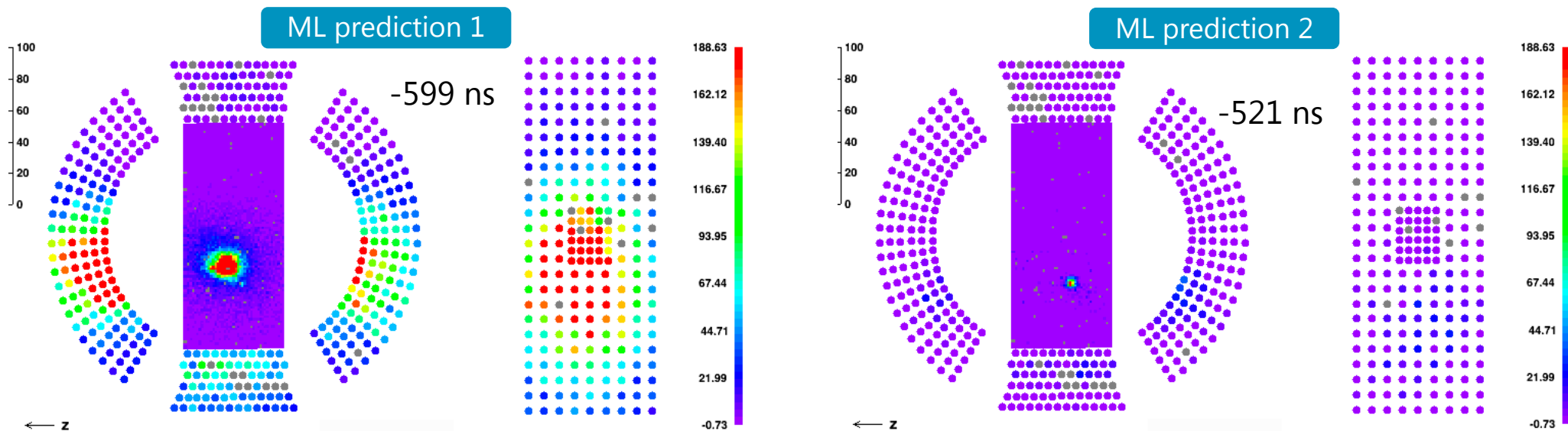
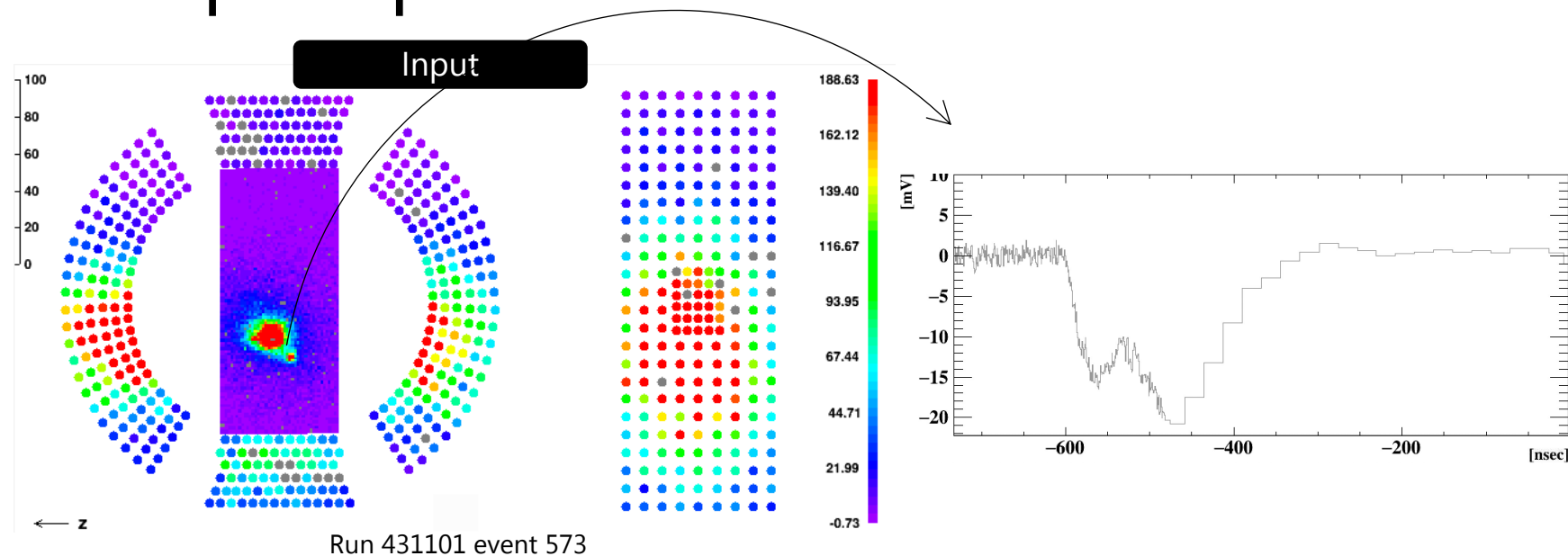
Performance on real pileups

Two coincident photons
found in conventional



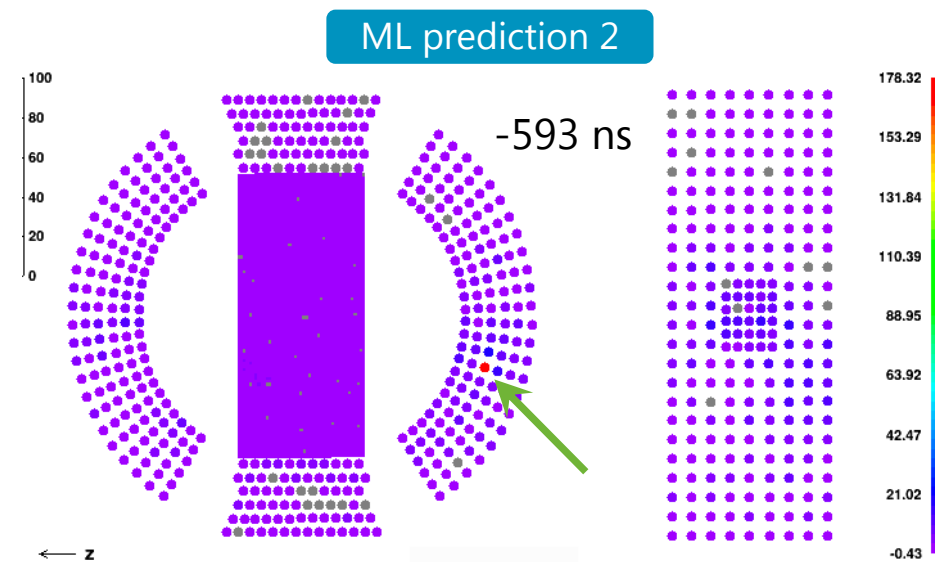
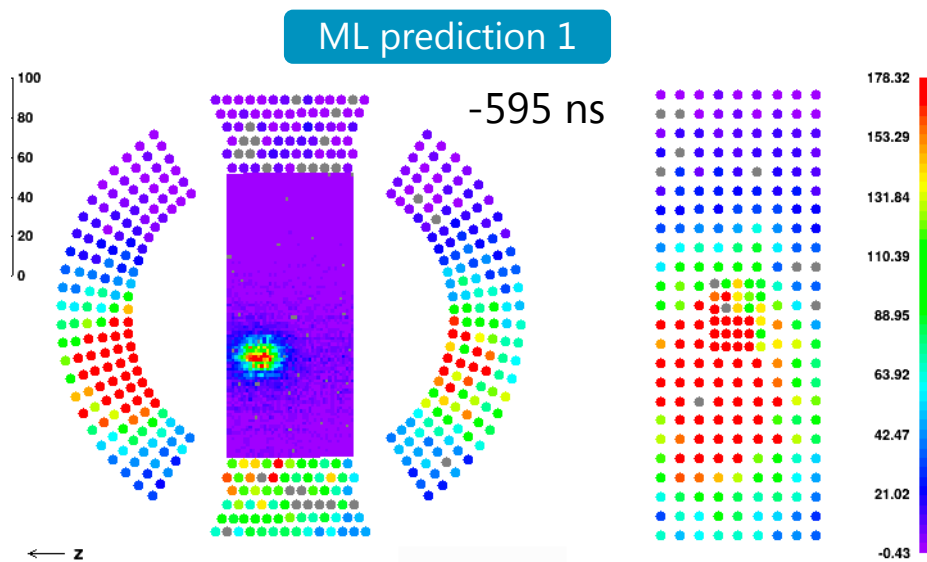
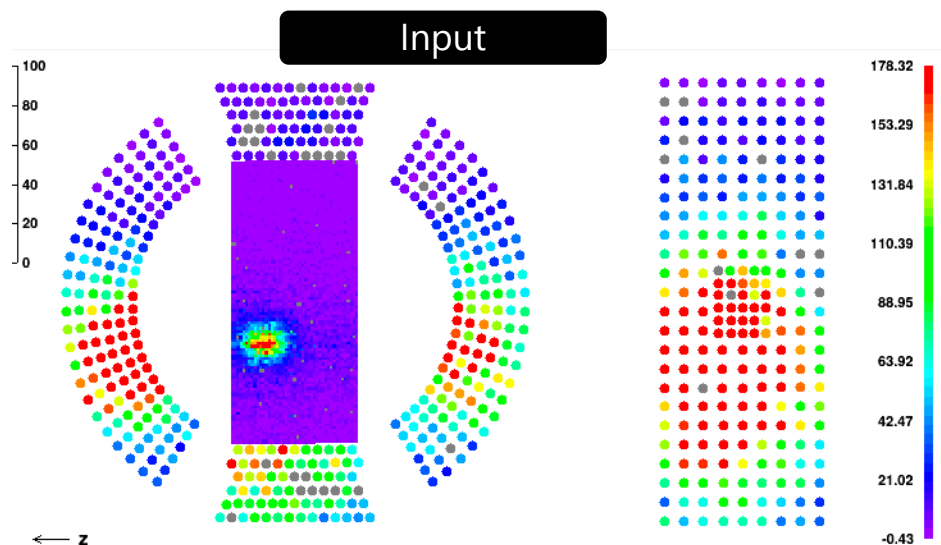
Performance on real pileup

"no pileup"
in conventional



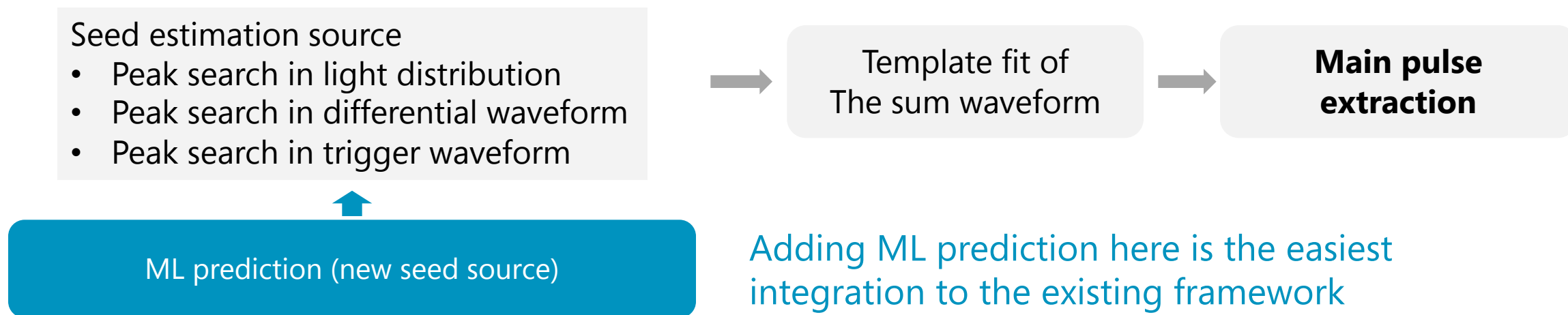
Performance on real pileup

“no pileup”
in conventional
(same event as in page 3)



Preliminary integration: ML outputs as fitting seed

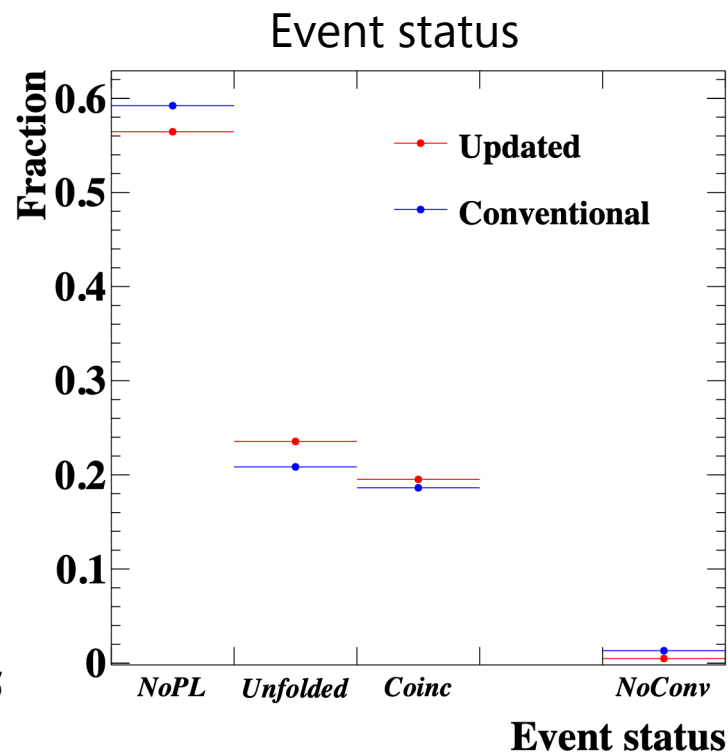
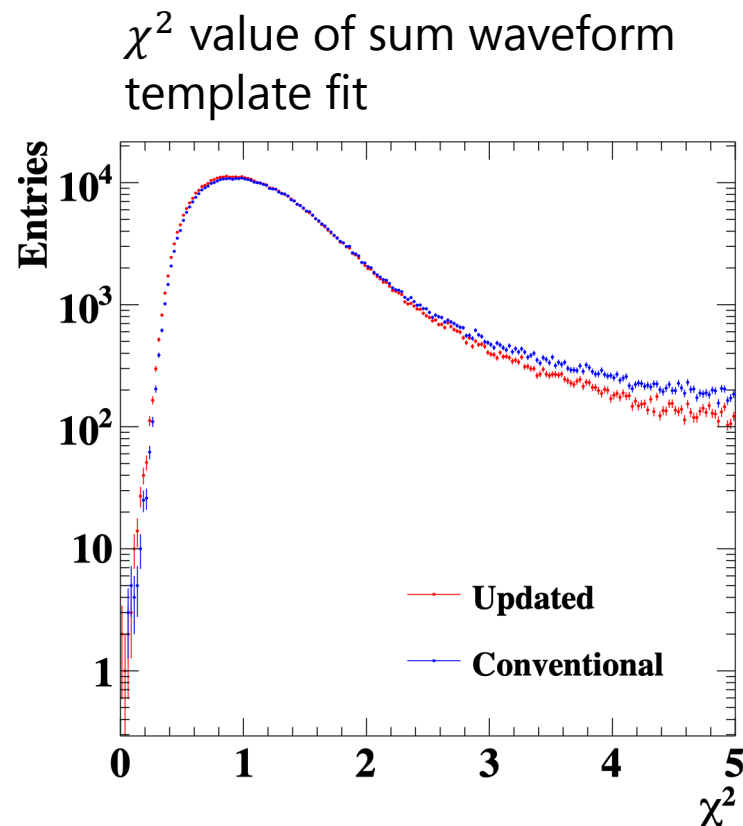
- Current pileup analysis: Template fit of the sum waveform
- Number of fitted pulses and their amplitude & time are estimated before fitting ("seeds")



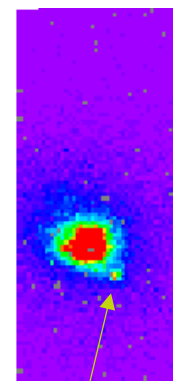
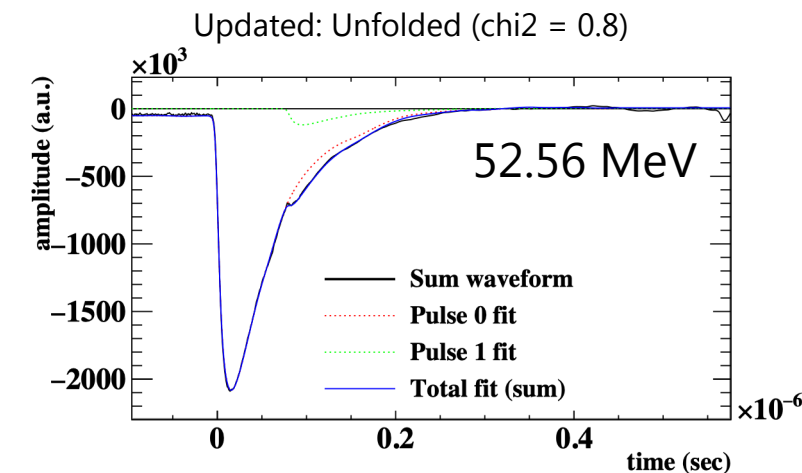
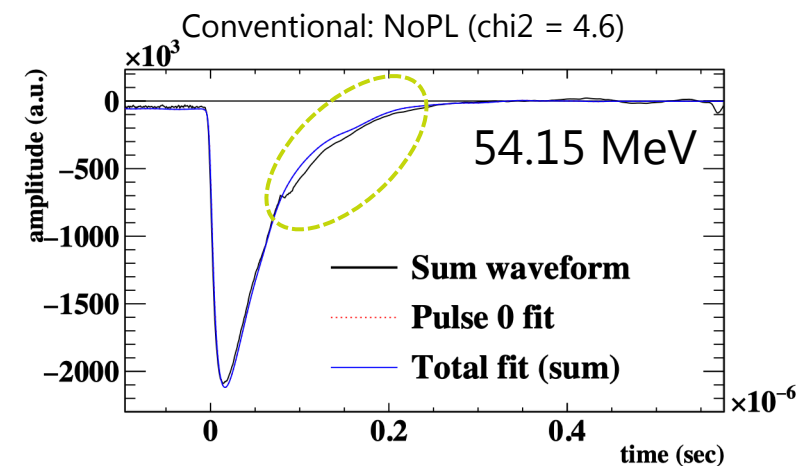
Expected effect

- Newly found pileup by ML improves fitting
- Accurate seeding by ML gives better convergence (even for the existing seeds)

Performance with the ML seed



- 65 % reduction of events which did not converge
- 18 % reduction of events with $\chi^2 > 2$



Newly found pileup by ML

Summary & Prospect

Summary

- Gamma-ray pileup occur frequently in MEG II LXe detector
- Newly developed a Machine Learning model for pileup analysis
- Its performance seems promising
 - Found pileups that conventional method could not find
 - Improved fit quality by using as the fitting seed in conventional sum waveform fitting

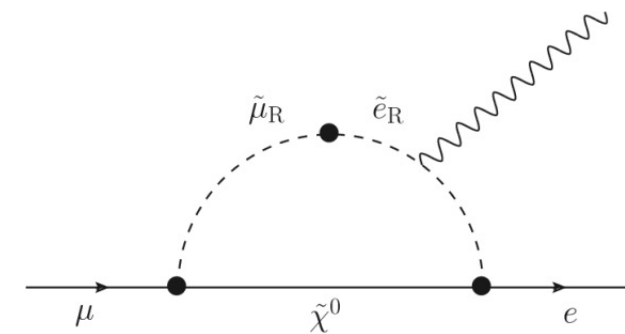
Prospect

- Improve model performance
 - Tune hyper-parameters
 - Better consider event type fraction in the training model (maybe retry with MC)
- Establish validation method against real data
- Reduce computational time (currently consuming 10-20 sec / event)
- Develop an analysis algorithm that can fully utilize ML output

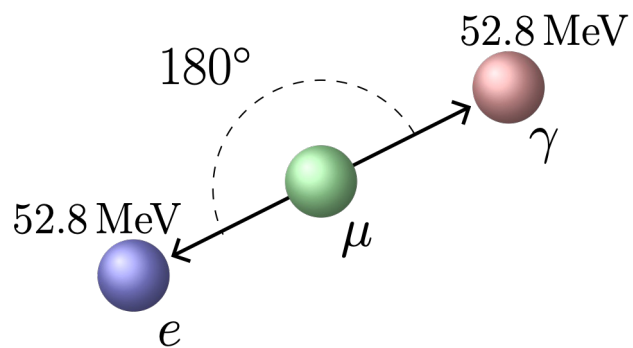
Back up

$\mu \rightarrow e\gamma$ Search

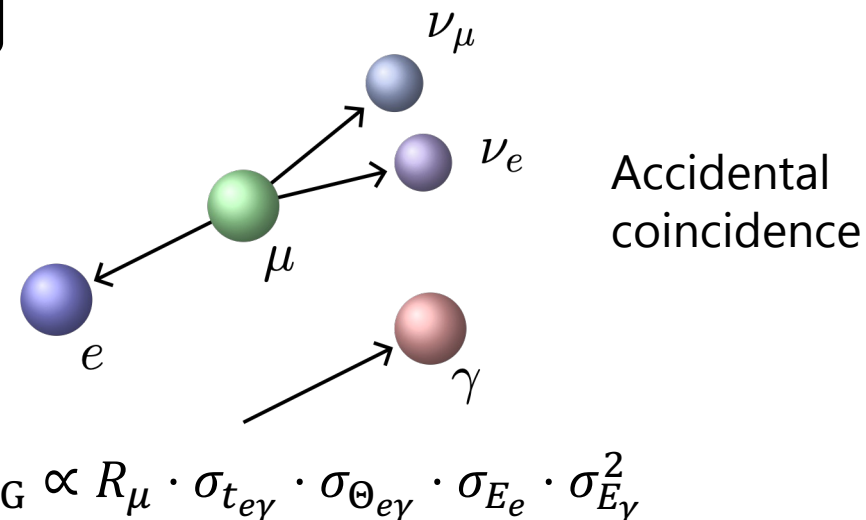
- $\mu \rightarrow e\gamma$: One of the Charged Lepton Flavor Violation (CLFV)
 - Strongly suppressed in SM + neutrino oscillation ($Br(\mu \rightarrow e\gamma) \sim 10^{-54}$)
 - Measurable branching ratio predicted by many new physics (e.g., SUSY-seesaw, SUSY-GUT, ...)
- Current upper limit: Set by the MEG II experiment @ PSI
 - $Br(\mu \rightarrow e\gamma) < 1.5 \times 10^{-13}$ (90 % C.L.) with 2020-2021 data



Signal

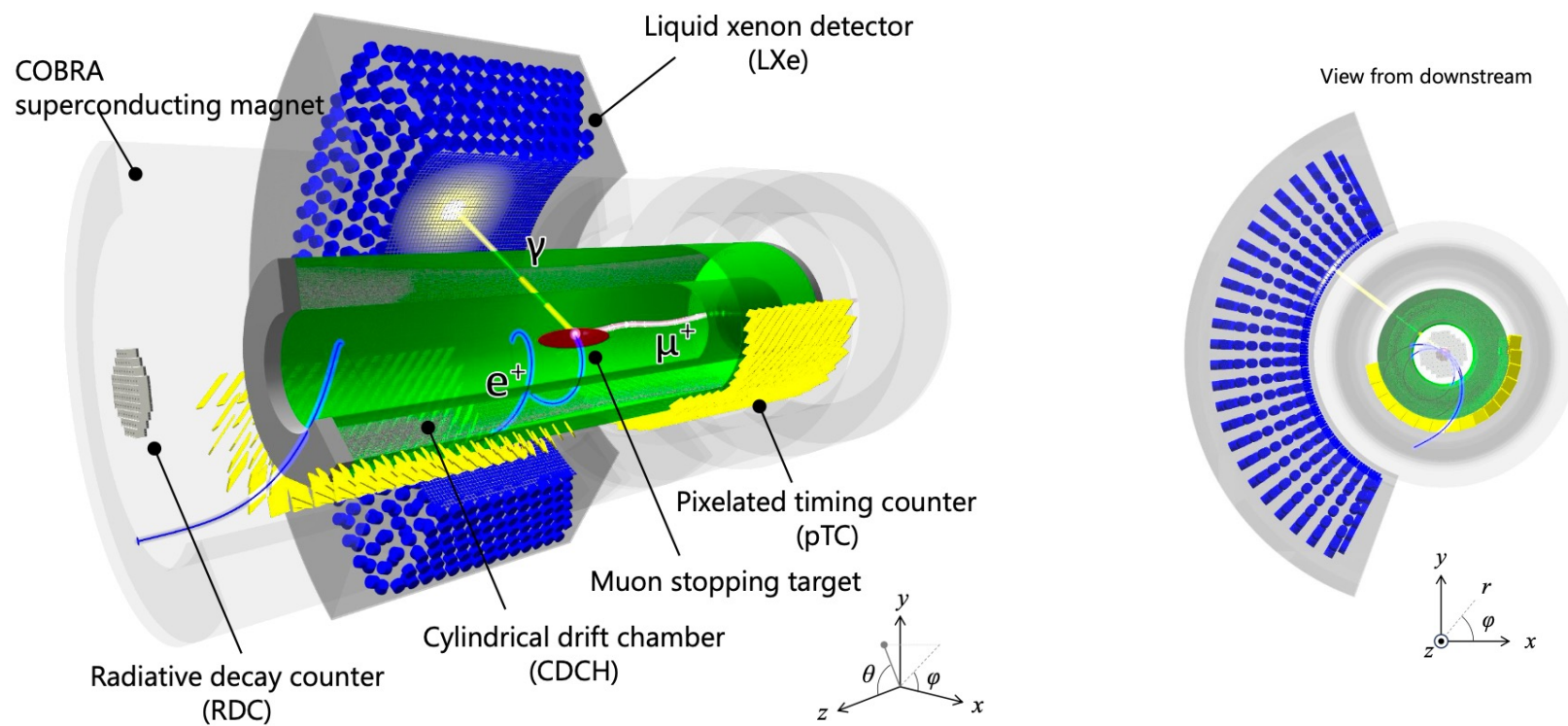


Main BG



MEG II experiment

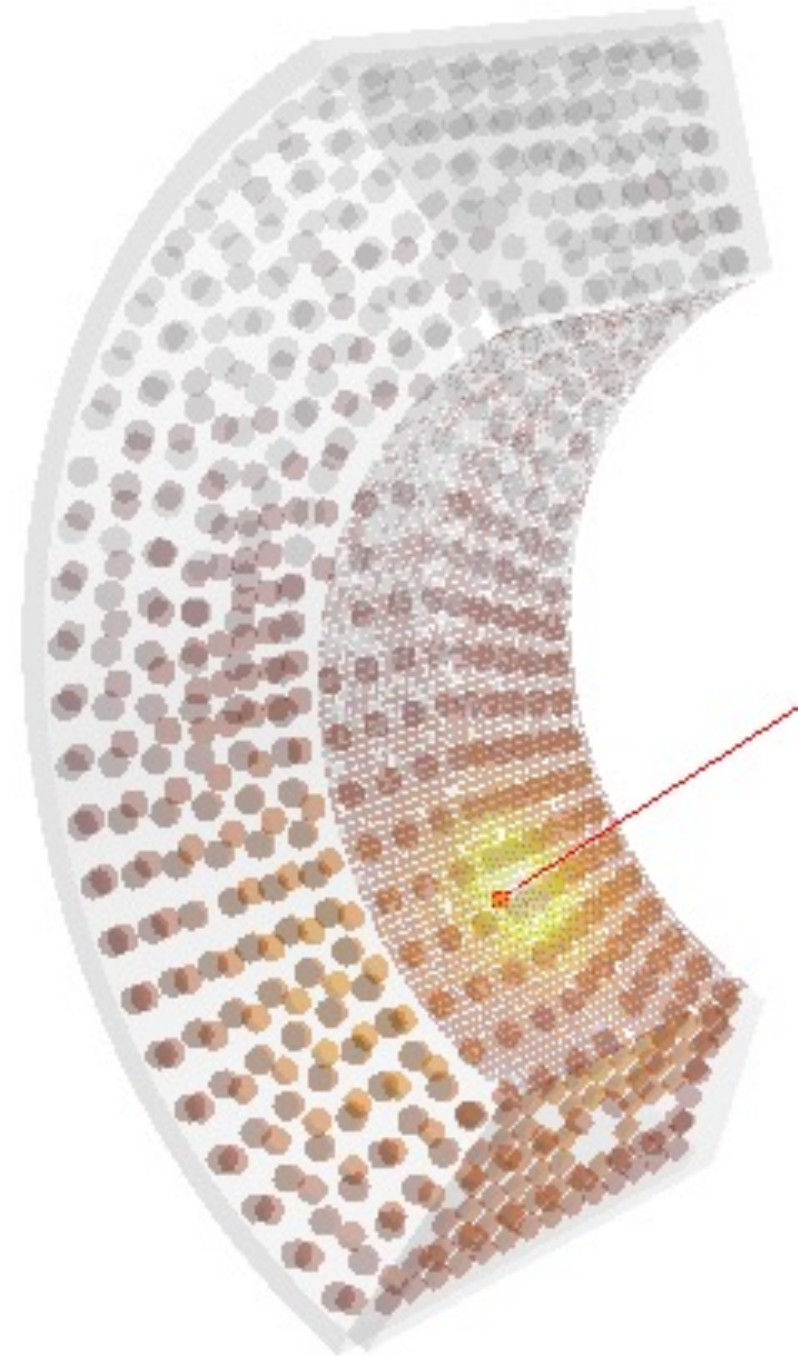
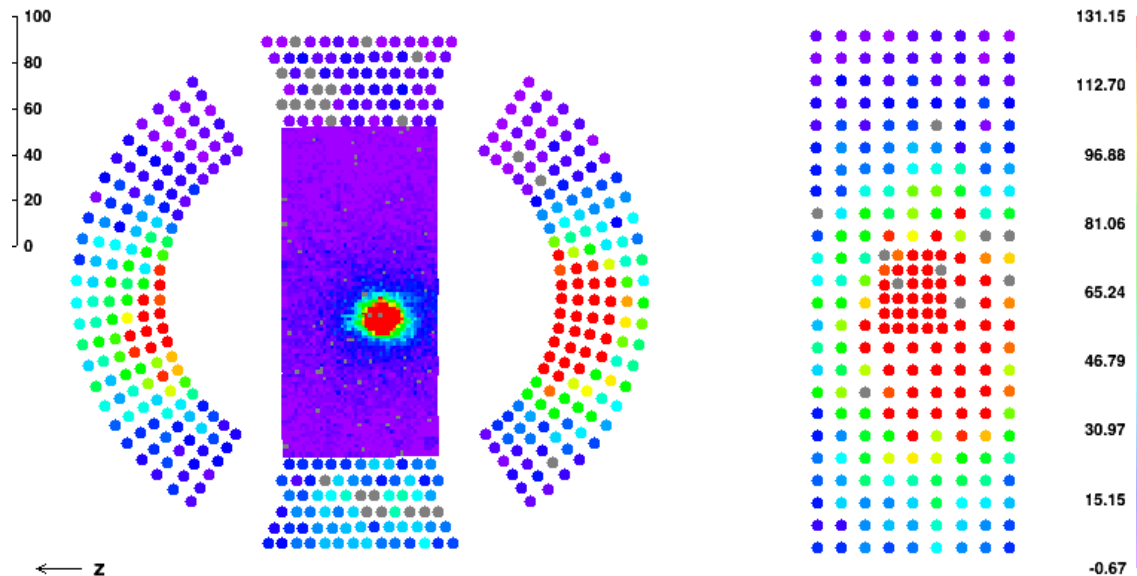
- Searches $\mu \rightarrow e\gamma$ with world's highest sensitivity
 - High intensity DC muon beam at PSI $\times$ Detectors with high resolution
- DAQ continues until the end of this month
- Target sensitivity: 6×10^{-14} (90 % C.L.)



LXe gamma-ray detector

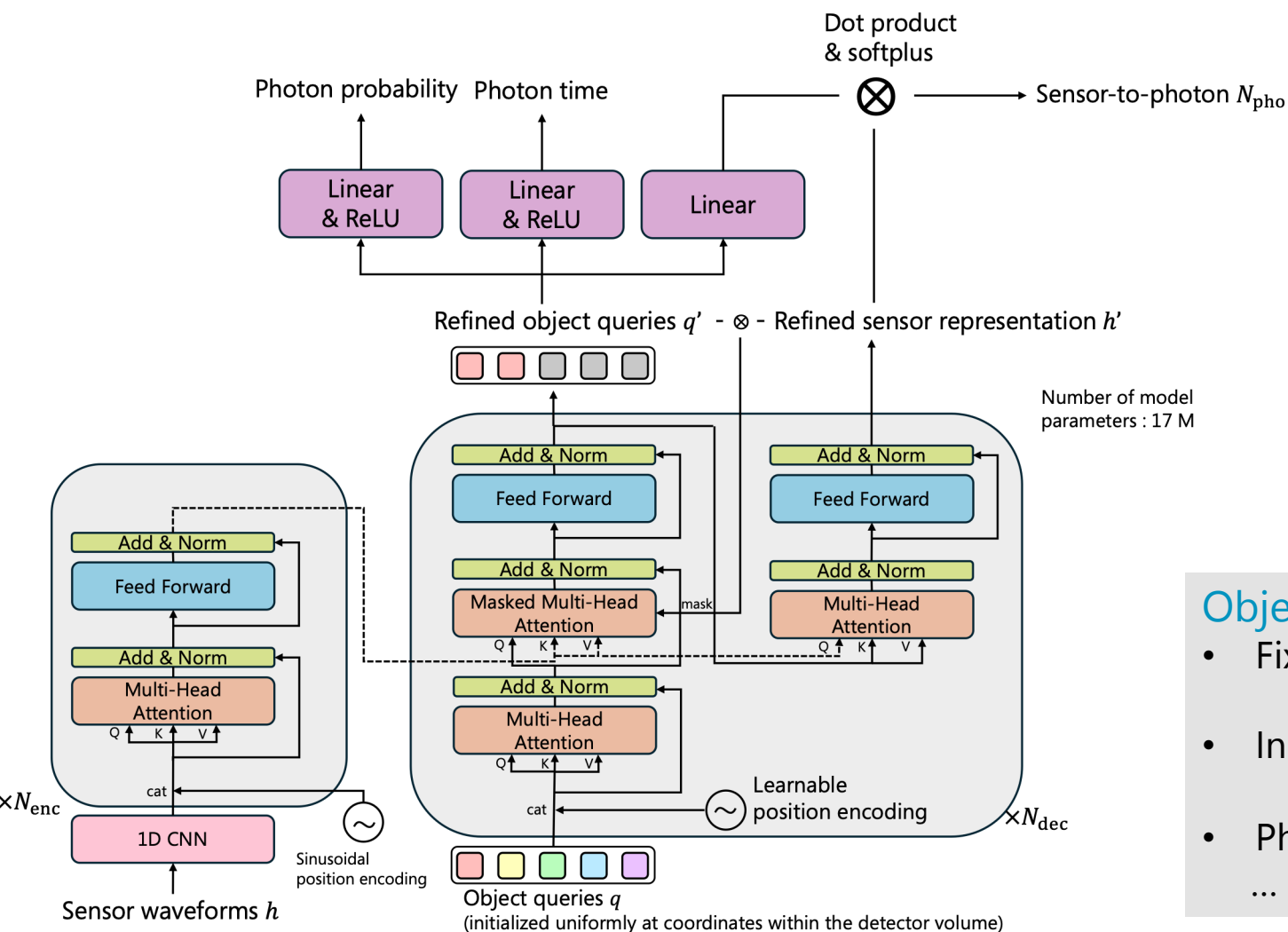
- C-shaped homogeneous calorimeter
- Scintillation light detected by
- 4092 SiPM (Inner face) + 668 PMTs (other faces)

Typical event display



Model architecture

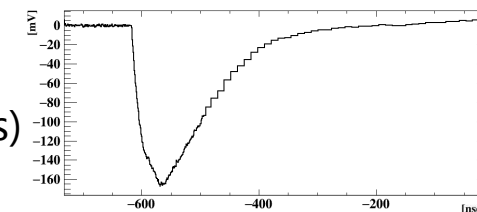
Model developed based on **Object Detection Transformer**



$\times 4760$

Input

Sensor waveform
(4760 sensors $\times$ 512 bins)



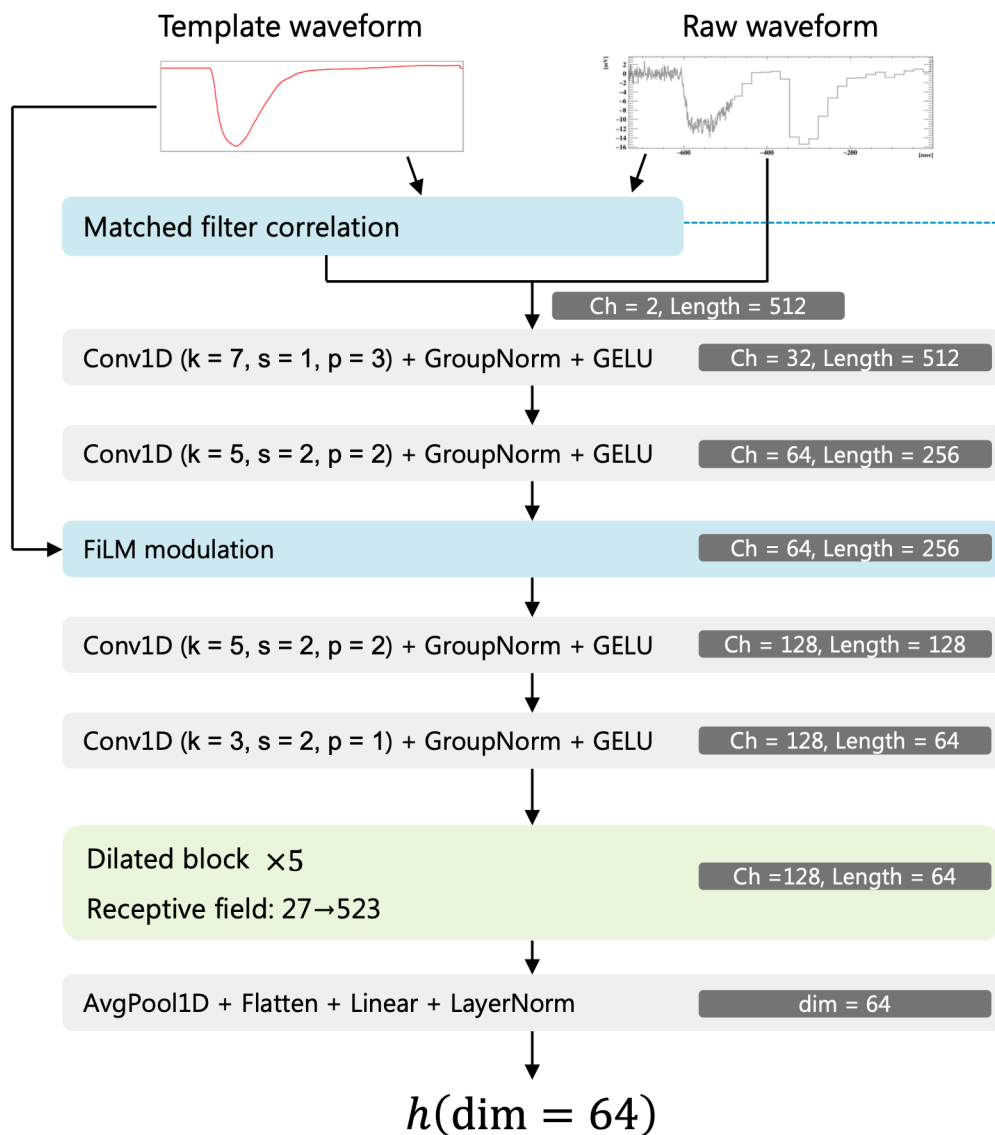
Output

- For each object queries,
- Photon probability
 - Sensor-to-photon # of photoelectrons
 - Photon time

Object query

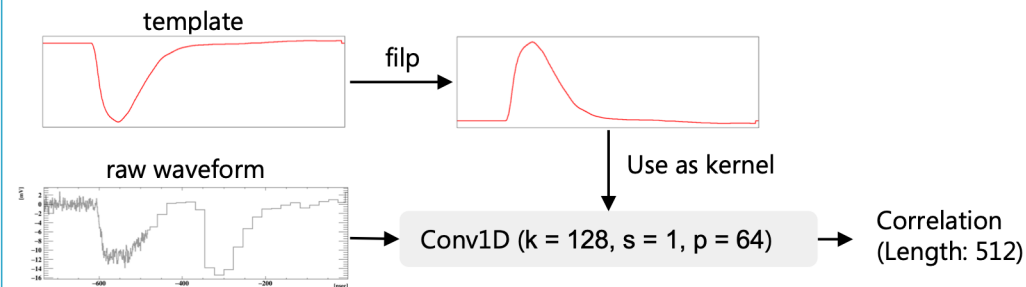
- Fixed number of vectors, each representing photon slot
- Initialized randomly and input to decoder
- Photon probability $\sim 1 \rightarrow$ Represents photon ... Enable prediction of variable # of objects (photons)

Feature extraction by 1D-CNN



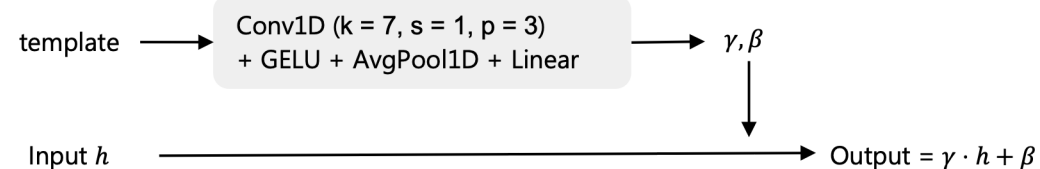
Matched filter correlation

Standard convs only combine channels linearly and cannot perform element-wise multiplication $\text{raw}(t) \times \text{template}(t)$. This product is thus precomputed and passed as an extra input channel.



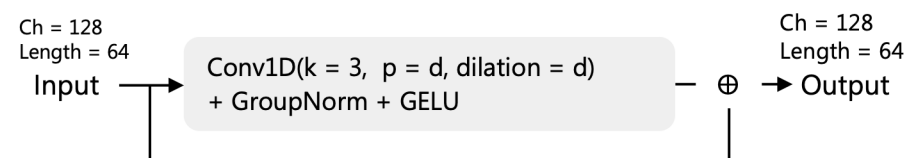
FiLM modulation

Modulates a feature map with a scale (γ) and shift (β) generated from template waveform, applying an affine transform per channel

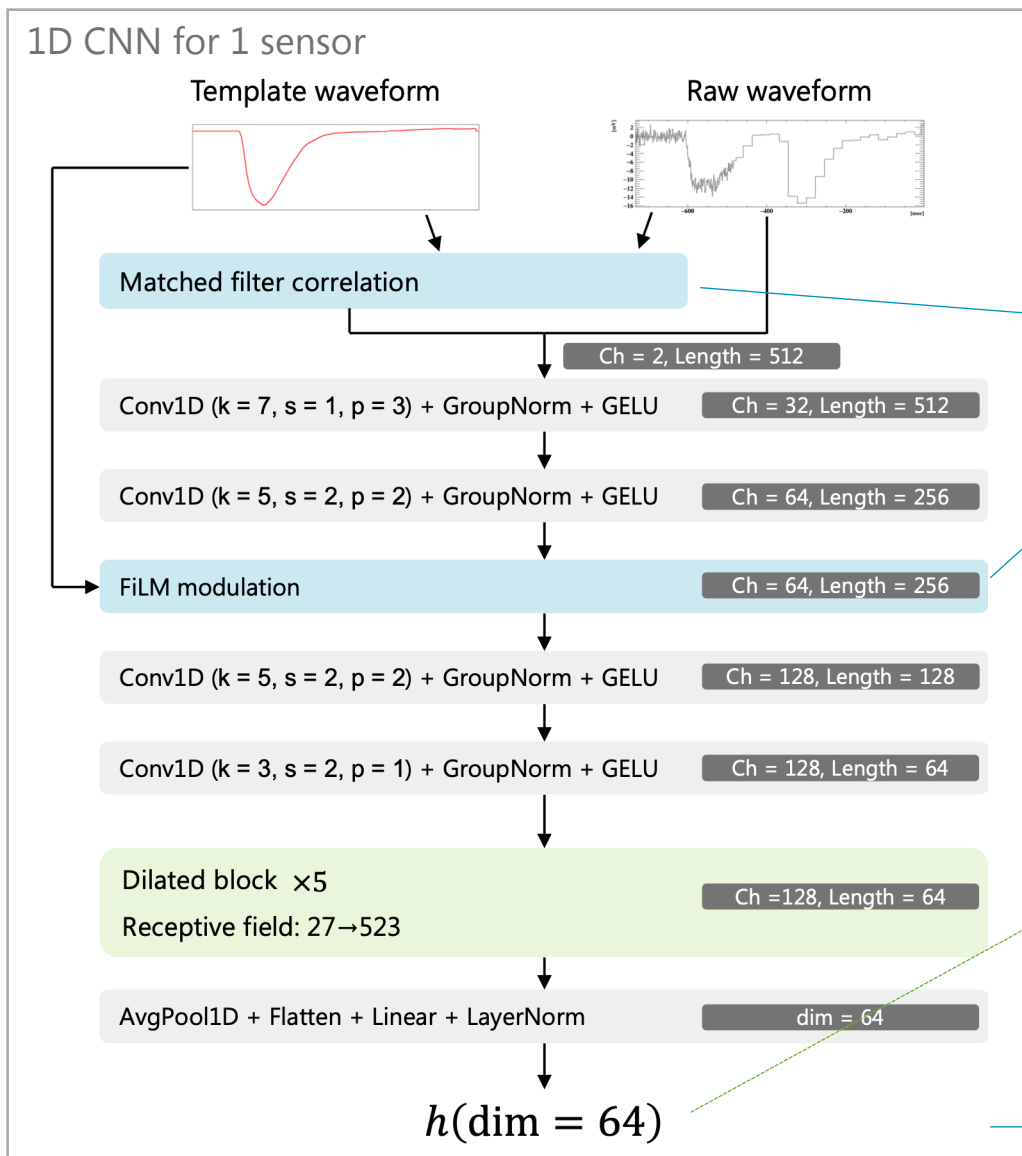


Dilated block

Expand receptive field without changing shape



Feature extraction by 1D-CNN

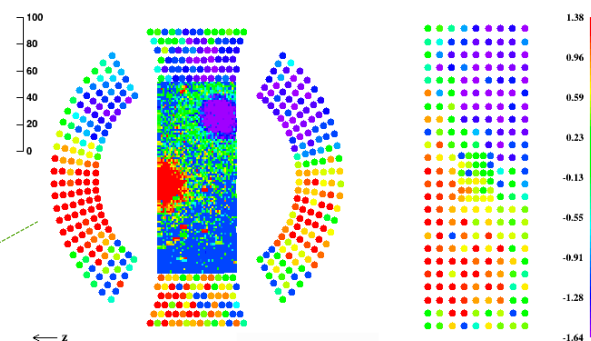


Template waveform input

Waveform shapes against single gamma vary sensor by sensor
 → Give typical template waveform per sensor

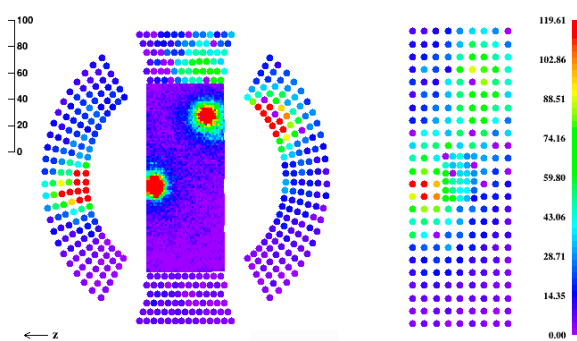
- Take correlation between template and raw waveform
- Intermediate modulation by template information

Example of one element of feature vector h



$\times 64$ elements

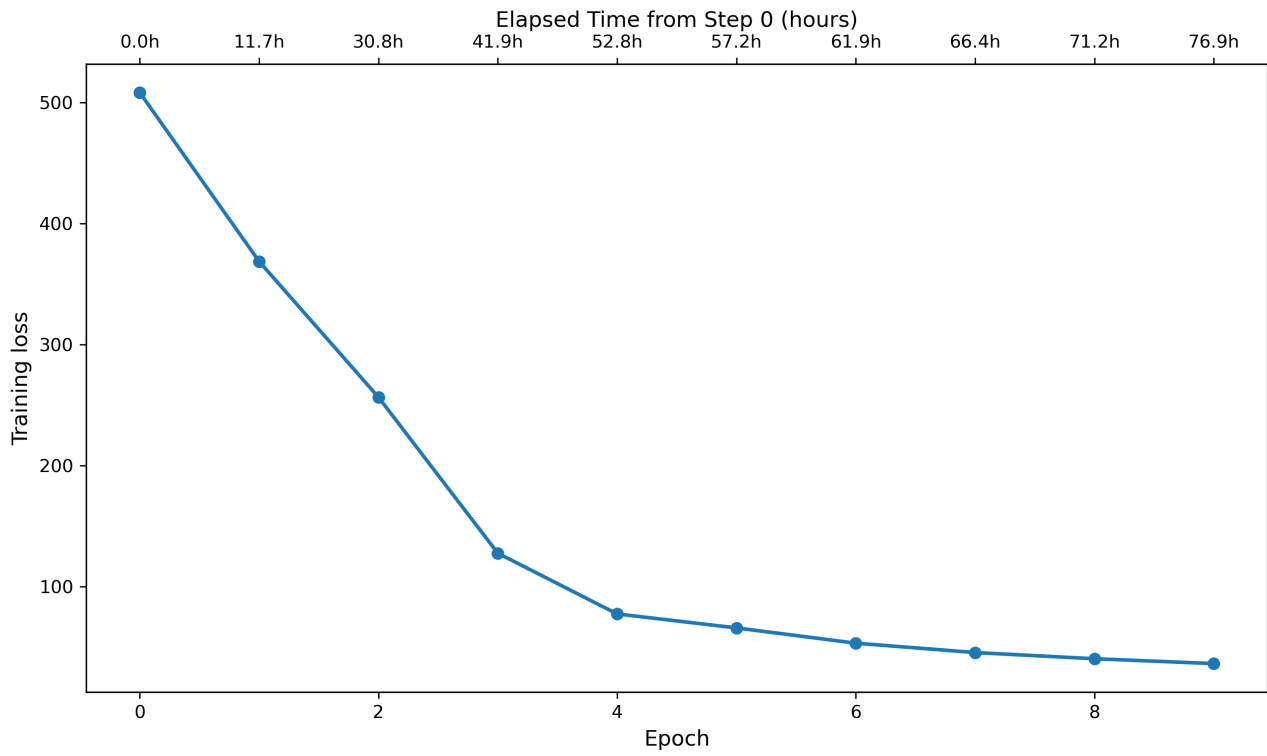
(Original light distribution)



$\times 4760$ sensors

To Transformer

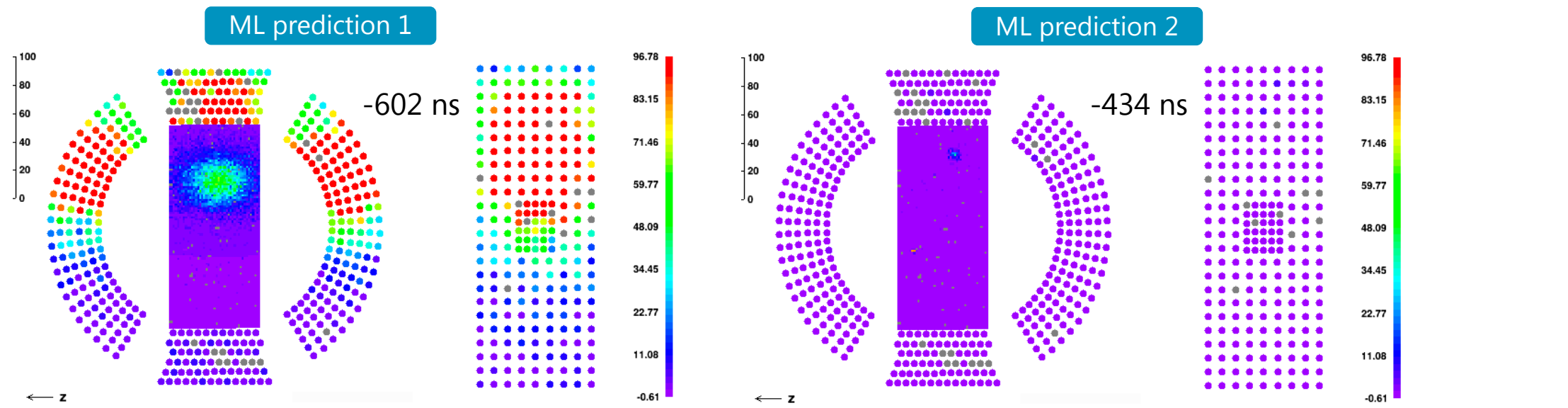
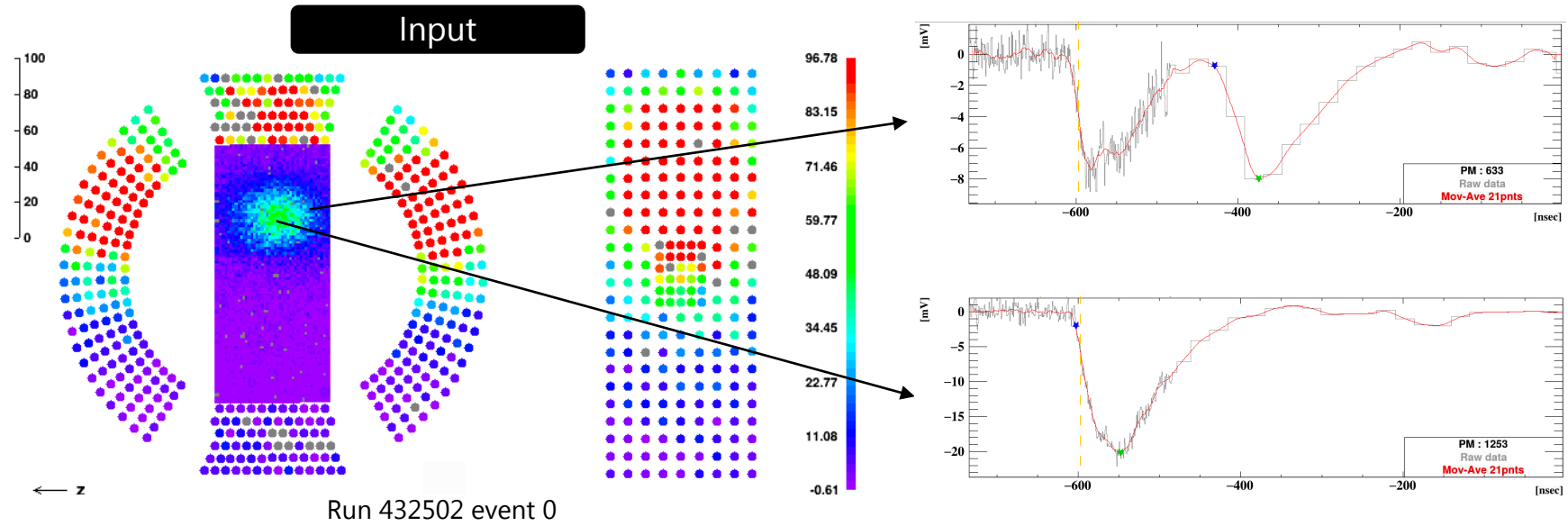
Training



No Pileup		Off-time pileup			On-time pileup			Total
No mix	mix at rate	2 gamma mix	2 gamma mix (Difficult)	3 gamma mix	2 gamma mix	2 gamma mix (Difficult)	3 gamma mix	
0.090 M	0.060 M	0.039 M	0.240 M	0.044 M	0.087 M	0.104	0.022	0.4 M

Performance on real pileup

"no pileup"
in conventional



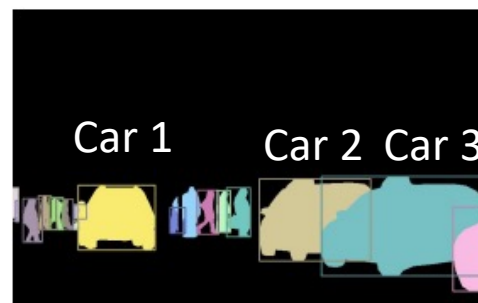
Other backups

Machine Learning Approach

Machine learning can be used to comprehensively handle the space & time correlations of sensors

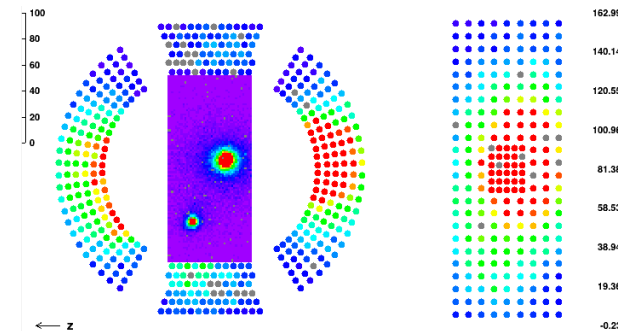
- Pileup detection is similar to object detection in computer vision
 - Treat photosensor as pixel, photon as object

Object detection



https://openaccess.thecvf.com/content_CVPR_2019/papers/Kirillov_Panoptic_Segmentation_CVPR_2019_paper.pdf

Pileup detection

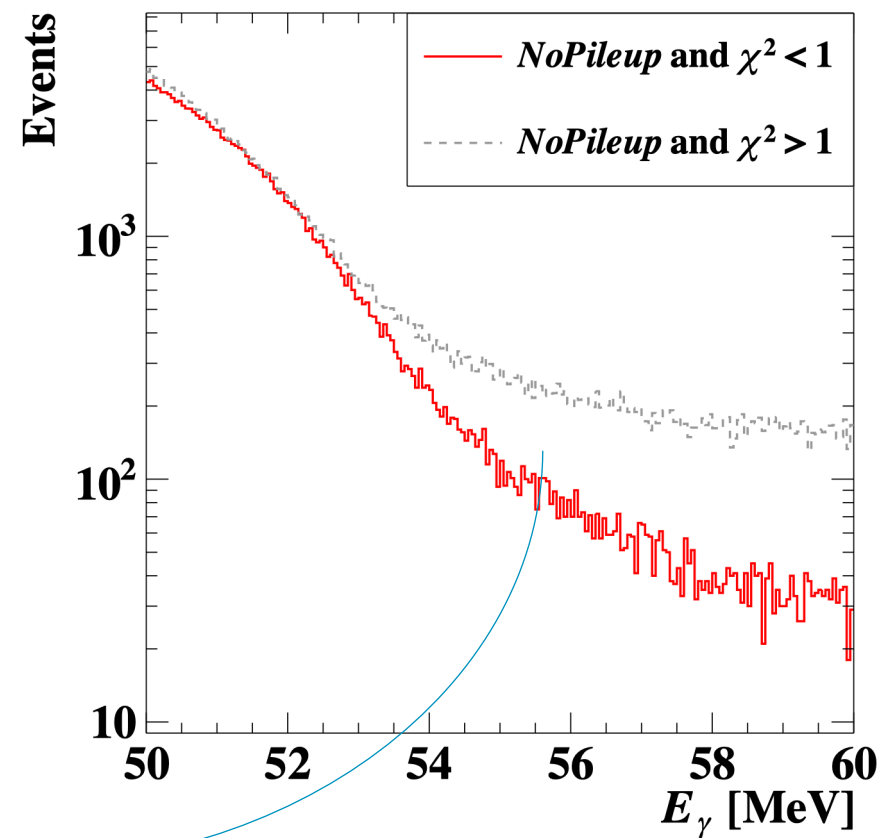
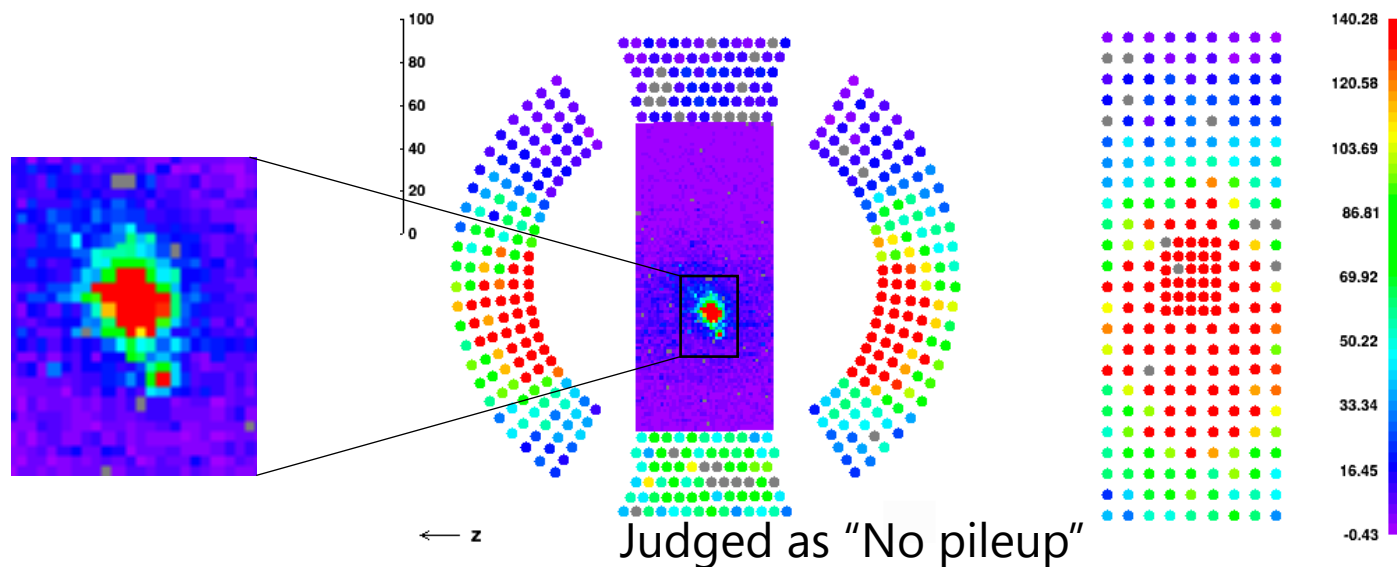


Pileup 1,
Pileup 2

- Usual CNN-based model is difficult to apply on LXe → Use Transformer
 - Transformer: utilize Attention mechanism (widely used in LLM, such as ChatGPT)
 - Sensors' 3D absolute coordinates can be directly given via positional encoding

Limitation of the current method

- Events with large χ^2 value in template fit are making high-energy tail
- Some of them seems to be pileups, even though they were judged as "No pileup"



Needs improvements

Challenge in ML development on LXe detector

MC vs data discrepancy

When trained with simulation data,

- High accuracy achieved on MC validation samples
- Performance degrades significantly on real experimental data

Estimated causes in MC

- Waveform variation: sensor-by-sensor differences are not modeled
- Light propagation: LXe optical processes may be imperfectly simulated

Possible solutions

Solution 1

Provide waveform features to model

Feed typical sensor-specific waveforms directly into input features

Solution 2

Train directly on real data samples

Construct training data from real experiments to eliminate simulation bias

Solution 3

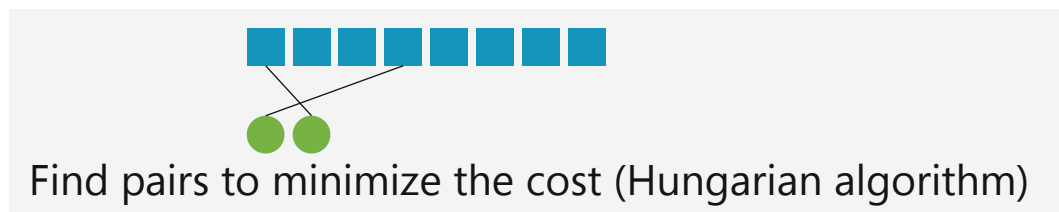
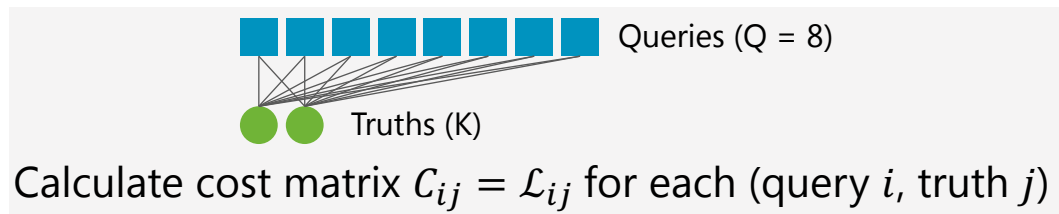
Update MC model

(Long-term) Improve simulation for LXe light propagation and sensor response

Training loss

$$\mathcal{L} = \alpha_{\text{cls}} \mathcal{L}_{\text{cls}} + \alpha_{\text{time}} \mathcal{L}_{\text{time}} + \alpha_{\text{nphe}} \mathcal{L}_{\text{nphe}} + \alpha_{\text{shape}} \mathcal{L}_{\text{shape}} + \alpha_{\text{pos}} \mathcal{L}_{\text{pos}} + \alpha_{\text{sum}} \mathcal{L}_{\text{sum}} + \alpha_{\text{empty}} \mathcal{L}_{\text{empty}}$$

Matching between truth and query



Loss calculation & back propagation

Class loss	$\mathcal{L}_{\text{cls}}$	Binary Cross Entropy
Time loss	$\mathcal{L}_{\text{time}}$	Huber loss
n_{phe} loss	$\mathcal{L}_{\text{nphe}}$	MSE of $\sqrt{n_{\text{phe}}}$ relative to truth, for matched queries
Shape loss	$\mathcal{L}_{\text{shape}}$	Dice loss of $\sqrt{n_{\text{phe}}}$
Position loss	$\mathcal{L}_{\text{pos}}$	Huber loss of n_{phe} -weighted centroid
n_{phe} sum loss	$\mathcal{L}_{\text{sum}}$	Huber loss of total n_{phe}
Empty loss	$\mathcal{L}_{\text{empty}}$	MSE of $\sqrt{n_{\text{phe}}}$ relative to 0, for unmatched queries

α_{cls}	α_{time}	α_{nphe}	α_{shape}	α_{pos}	α_{sum}	α_{empty}
300	10	200	100	2	0.1	35

Performance on real pileup

"no pileup"
in conventional

