

Pileup Analysis for the Liquid Xenon Detector of the MEG II experiment

16/03/2022, 16pA573-2

Rina Onda on behalf of MEG II Collaboration

The University of Tokyo

MEG II Experiment

- The MEG II experiment searches for a charged lepton flavor violating (cLFV) decay of a muon, $\mu \rightarrow e\gamma$.
- In the SM, the branching ratio of the decay is too small to be detected: $Br(\mu \rightarrow e_{\gamma}) \sim 10^{-54}$.
- Many BSM models predict experimentally detectable branching ratios for the decay: $Br(\mu \rightarrow e_{\gamma}) \sim 10^{-14} 10^{-12}$.
 - SUSY-GUT
 - SUSY-seesaw
- The MEG experiment gives the most stringent upper limit of 4.2×10^{-13} .
- The MEG II experiment plans to search for the decay with the higher sensitivity by one order of magnitude than MEG.





 * Ratio of Br is predicted to be

$$\begin{split} & \frac{\mathcal{B}(\mu \to \text{eee})}{\mathcal{B}(\mu^+ \to \text{e}^+ \gamma)} \approx 6 \times 10^{-3}, \\ & \frac{\mathcal{B}(\mu^- N \to \text{e}^- N)}{\mathcal{B}(\mu^+ \to \text{e}^+ \gamma)} \approx 2.6 \times 10^{-3} \text{ (for N = Al). } 1 \end{split}$$

Y-ray Detector of MEG II Experiment



Inside LXe



- Liquid xenon photon detector (LXe) detects energy, position and timing of a γ-ray.
- Scintillation lights from liquid xenon are detected with PMTs and MPPCs.
- In this talk, the pileup analysis for the LXe detector will be reported.

Signal & BG in MEG II

- $\mu \rightarrow$ ey signal event can be characterized by
 - $E_e = E_{\gamma} = 52.8 \text{ MeV}$
 - back to back
 - coincident in time
- The dominant background derives from the accidental coincidence of e^+ and γ -ray from different μ decays.
- The number of the accidental background is proportional to the square of the beam rate R_{μ} : $N_{bg}~\propto~{R_{\mu}}^2$





Source of y-ray Pileups

• The pileup γ -rays can greatly affect the energy reconstruction since it uses information of all channels.

 \leftrightarrow The effects on the position and the timing are limited since they are reconstructed using local information.

- The existence of the pileups increases the number of background events in the signal region.
 - 1. γ -ray from the same μ decay \rightarrow On-timing pileup
 - AIF 2 γ : 34% ($\sigma_{E\gamma}$ = 1.7%)
 - 2. γ -ray from different μ decays \rightarrow Off-timing pileup
 - RMD γ + accidental pileup γ
 - AIF 1γ + accidental pileup γ
- Therefore, the pileup elimination is crucial for the better sensitivity.

Fraction of Background γ-rays



Strategy for Pileup Elimination

The pileup γ -rays are spatially or temporally detectable.

- On-timing pileup: spatial search and event rejection
- Off-timing pileup: temporal search and unfolding of waveform



Pileup Elimination Algorithm

- A series of algorithms was developed for the pileup elimination.
- It consists of three steps:
 - 1. Pileup event identification with DL-based algorithm
 - 2. Peak search and clustering of channels in light and timing distributions
 - 3. Unfolding of sum waveform
- Detailed algorithm was reported in the Autumn meeting of 2021 (15pT3-7).



Topics of This Talk

- The performance of the algorithm was evaluated with MC.

 The assumption was too ideal.
- This talk focuses on the performance evaluation in more realistic situation.
 - Noise
 - Dead channel
 - PDE decrease of the MPPCs
 - Calibration precision

Noise/Dead Channel

For the realistic simulation, the effects of noise and dead channels were introduced.

- Noise: mixing noise data taken in 2021 to simulated waveforms. Coherent noise
- Dead channels: masking information of dead channels observed in 2021 (30 MPPCs + 28 PMTs).



Simulated Noise (Gaussian White Noise)



Real Noise

165.96

72.44

31.62

12.02

5.25

Performance - DL-based Pileup Identification -

- Less background at the same signal efficiency was achieved.
 - Higher detection efficiency especially at deep region
 - Tolerance to the fake peak
- The DL model is superior in terms of **optimized filtering and threshold**.

 \leftrightarrow The re-binning and low-pass filter are applied for the peak search to avoid picking up fake peaks, and threshold is set by hand.

• The threshold was optimized to maximize the signal-tobackground ratio.



Performance - Peak Search and Unfolding of Sum Waveform -

- The peak search finds more pileup events.
 - The background events are reduced by 59%.
 - The signal efficiency is also decreased by 27%.
- The unfolding recovers the signal efficiency by 19%.
 ↔ The increase of the backgrounds is only 8%.
- In total, the backgrounds are reduced by 51% keeping the signal efficiency of 93%.

Signal Efficiency vs. Normalized Nbg

Effect on Sensitivity

The DL-based event rejection improves the sensitivity by 18%.

 \leftarrow Reduction of on-timing pileups.

- The peak search and the unfolding improves it by 4%.
 ← Less backgrounds and recovery of the signal efficiency.
- In total, 22% improvement is achieved at $7 \times 10^7 \ \mu^+$ stops/s and 26% at $3.5 \times 10^7 \ \mu^+$ stops/s.

Relative Branching Ratio Sensitivity

Effect of PDE Degradation

- The PDEs of the MPPCs were found to be degraded probably due to the radiation damage.
- The sensitivity was evaluated assuming different PDEs.
- No tendency due to the PDE degradation.
- Fluctuation from -4% to +7%. $\overleftarrow{}$ Given by the balance between $N_{\rm bg}$ and signal efficiency
- The performance does not depend on the PDE at training of DL.

Table 5.4 Number of backgrounds and signal efficiency for different MPPC PDEs ($E_{\gamma} = 51.5-54 \text{ MeV}$) when the deep learning model is trained at a fixed PDE of 13% or each PDE.

PDE	Trained at PDE 13%		Trained at each PDE		
	$N_{ m bg}$	signal efficiency	$N_{ m bg}$	signal efficiency	
13%	0.491	0.930	0.491	0.930	
8%	0.468	0.950	0.486	0.956	
4%	0.489	0.944	0.485	0.942	
2%	0.584	0.962	0.576	jpsSpring2022 (16p/ 0.962	4573-2

Relative Branching Ratio Sensitivity

Effect of Calibration Precision

- The number of photons can fluctuate and be biased due to the statistical and systematic uncertainties of the calibration.
- The performance was evaluated including the statistical uncertainty up to 10% and the systematic uncertainty.
 - \leftarrow The statistical uncertainty was estimated to be 4% for the calibration.
- No degradation was observed.

Systematic Uncertainty of Calibration

Systematic uncertainty estimated by two different calibration methods

MEG II Projected Sensitivity

- The branching ratio sensitivity was calculated assuming two scenarios for PDE degradation:
 - Optimistic: $7 \times 10^7 \ \mu^+$ stops/s, PDE saturation at 2%
 - Pessimistic: $3.5 \times 10^7 \ \mu^+$ stops/s, No PDE saturation
- The sensitivity was estimated to be
 - 5.6 \times $10^{\text{-14}}$ at $7{\times}10^7~\mu^{\text{+}}\text{stops/s}$
 - 5.8 \times $10^{\text{--}14}$ at 3.5 $\times 10^7$ $\mu^{\text{+}}\text{stops/s}$

Equivalent even with halved statistics mainly due to better position efficiency (65% \rightarrow 74%) and less backgrounds.

 ${\sim}10$ times higher sensitivity than that of MEG

Summary

- The MEG II experiment searches $\mu \rightarrow e\gamma$ decay.
- The pileup analysis for the LXe detector is important to reduce the γ -ray background events in the signal region.
- The new algorithm for the pileup elimination was developed.
- It consists of three steps:
 - 1. Pileup event identification with DL-based algorithm
 - 2. Peak search and clustering of channels in charge and timing distributions
 - 3. Unfolding of sum waveform

Summary

- The performance of the new algorithm was evaluated in more realistic situation.
 - Sensitivity improvement by 22-26% compared to the previous algorithm.
 - No effect from PDE decrease and calibration uncertainty
- The branching ratio sensitivity of the MEG II experiment was evaluated with the updates.
 - 5.6 \times 10⁻¹⁴ at 7×10⁷ $\mu^{+}stops/s$ for three years
 - 5.8 \times 10⁻¹⁴ at 3.5×10⁷ μ^{+} stops/s for three years
 - ${\sim}10$ times higher sensitivity than that of MEG
- The performance will be evaluated using data in 2021.

Backup Slides

Pre-processing is applied for the inputs.

• Dead channel recovery

Values of dead channels are estimated by the mean of surroundings.

 \leftarrow Tolerance to the effect of dead channels

• Normalization

Normalized by the maximum value, i.e. all input values are no more than 1.

 \leftarrow Suppress the energy dependence

• Cut off

Negative charges are set to 0, i.e. all input values are no less than 0.

 \leftarrow Due to a failure of the baseline calculation

EfficientNet

- A type of CNN
- <u>https://arxiv.org/pdf/1905.11946.pdf</u>
- The performance of DL models can be improved by scaling up the original model.
- The optimal scaling method was investigated, and they introduced the efficiently scaled models. \rightarrow A better performance with less parameters was achieved compared to other models.

jpsSpring2022 (16pA573-2)

Model Architecture

Layer (type:depth-idx)	Input Shape	Output Shape	Kernel Shape
Model -EfficientNet: 1-1 -Sequential: 2-1 -ConvBN: 3-1 -Swish: 3-2 -BMConvBlock: 3-3 -BMConvBlock: 3-4 -BMConvBlock: 3-5 -BMConvBlock: 3-6 -BMConvBlock: 3-7 -BMConvBlock: 3-8 -BMConvBlock: 3-9 -BMConvBlock: 3-10 -BMConvBlock: 3-11 -BMConvBlock: 3-12 -BMConvBlock: 3-12 -BMConvBlock: 3-13 -BMConvBlock: 3-14 -BMConvBlock: 3-15 -BMConvBlock: 3-16 -BMConvBlock: 3-16 -BMConvBlock: 3-17 -BMConvBlock: 3-18 -ConvBN: 3-19 -Swish: 3-20 -Sequential: 2-2 -AdaptiveAvgPool2d: 3-21 -Flatten: 3-22 -Sequential: 1-2 -Dropout: 2-3 -Linear: 2-4 -ReLU: 2-5 -Linear: 2-6 -Sigmoid: 2-7	 [1, 1, 93, 44] [1, 1, 93, 44] [1, 1, 93, 44] [1, 32, 46, 21] [1, 32, 46, 21] [1, 24, 23, 11] [1, 24, 23, 11] [1, 40, 12, 6] [1, 40, 12, 6] [1, 80, 6, 3] [1, 80, 6, 3] [1, 112, 6, 3] [1, 112, 6, 3] [1, 112, 6, 3] [1, 192, 3, 2] [1, 192, 3, 2] [1, 192, 3, 2] [1, 1280, 3, 2] [1, 1280, 3, 2] [1, 1280, 3, 2] [1, 1280, 1, 1] [1, 1280] [1, 1] [1, 1]	<pre>[1, 1280] [1, 1280, 3, 2] [1, 32, 46, 21] [1, 32, 46, 21] [1, 32, 46, 21] [1, 16, 46, 21] [1, 24, 23, 11] [1, 24, 23, 11] [1, 40, 12, 6] [1, 40, 12, 6] [1, 80, 6, 3] [1, 80, 6, 3] [1, 80, 6, 3] [1, 112, 6, 3] [1, 129, 3, 2] [1, 192, 3, 2] [1, 192, 3, 2] [1, 1280, 3, 2] [1, 1280, 3, 2] [1, 1280] [1, 1280] [1, 1280] [1, 1280] [1, 256] [1, 256] [1, 1] [1, 1]</pre>	
Total params: 4,335,165 Trainable params: 4,335,165 Non-trainable params: 0 Total mult-adds (M): 38.45			
Input size (MB): 0.02 Forward/backward pass size (MB): 9.42 Params size (MB): 17.34 Estimated Total Size (MB): 26.78	insSpring2022 (16pA	573-2)	

Pileup γ-rays

• A small energy and a shallow conversion point are dominant.

Pileup γ-rays

The pileup analysis can find pileups whose energies are more than 0.2 MeV. The event rate of γ -ray hits for $E_{\gamma} > 0.2$ MeV is 0.7 MHz.

 $(lul < 25 cm) \land (lvl < 71 cm) \land (0 < w < 38.5 cm)$ jpsSpring2022 (16pA573-2)

22

Signal Efficiency

• The number of backgrounds for the new unfolding method is equivalent to the DL-based rejection at the same signal efficiency.

Previous Algorithm for Pileup Elimination

- There were algorithms already implemented.
- It consists of two steps:
 - 1. Unfolding with sum waveform fitting
 - Take sums of MPPC and PMT channels
 - Fit a template waveform
 - The waveforms are unfolded.
 - Sensitive to off-timing pileups

2. Rejection with peak search in charge distribution

- Search peaks whose charges are larger than a threshold on inner face.
- The events with pileups are rejected.
- Sensitive to on-timing pileups
- They are processed independently.

Example of Charge Distribution

- The deep learning-based pileup identification method was implemented.
- The DL model judge whether the event likely has pileup γ-rays.
- Model architecture
 - Based on Convolutional Neural Network
 - Input: light distribution on inner face (93 \times 44 pixels)
 - Output: Probability to include pileup γ-rays

Dataset

- Generated with MC
 - Main γ (uniform energy in 20-100 MeV, 1.6×10^5 events) \leftarrow Suppress the energy dependence
 - Pileup γ (resampled from the original pileup $\gamma,~1.2\times10^5$ events)
- Two types of data are prepared by mixing them.
 - Only main γ : labeled as "0"
 - Main and pileup γ : labeled as "1"

jpsSpring2022 (16pA573-2)

- The DL model was trained to predict
 - Single γ -ray event: 0
 - Multiple γ-ray event: 1
 ← The peak around 0 is due to low energy pileup γ-rays which are too difficult to be identified.
- The threshold to decide whether an event has pileup γ-rays or not was set to maximize the signal-to-background ratio.

- Two peak search and clustering methods are implemented.
 - Search in light distribution
 - \rightarrow search for on-timing pileups
 - Search in timing distribution
 - \rightarrow search for off-timing pileups

- Pileups are searched for in the light distribution on the inner and outer faces.
- A peak search is performed and channels whose light yields are larger than a threshold are clustered.
- Sensitive to the on-timing pileups.

Light Distribution

- Pileups are searched for in the timing χ^2 distribution on all the faces.
- The χ^2 of i-th channel is defined as

$$\chi_i^2 = \frac{\left(t_{\gamma} - t_i\right)^2}{\sigma_i^2},$$

where t_{γ} is the reconstructed γ timing, and t_i and σ_i are the timing and its uncertainty of the channel.

- Channels whose χ^2 are larger than a threshold are clustered.
- Sensitive to the off-timing pileups.

Timing χ^2

Step3: Unfolding of Sum Waveform

Step3: Unfolding of Sum Waveform

- Template waveforms are fit to the total sum waveforms using the extracted timings and energies as initial values.
- The energy of the main γ -ray is reconstructed from the fitting result.
- Finally, an event status is assigned for each event depending on the unfolding results.

33

Training

- Optimizer: SGD, Ir=0.01
- Loss: Binary cross entropy
- Scheduler: CosineAnnealing(max_T=500)
- Batch size: 200
- The number of epochs: 500

Threshold Scan

- The probability given by the model is translated to a binary flag by setting a certain threshold.
- The threshold was set to maximize the signal-tobackground ratio of the relative signal likelihood as

$$R_{\text{sig}} = \log_{10} \left(\frac{S(x_i)}{f_R R(x_i) + f_A A(x_i)} \right)$$

• The threshold was set to be 0.40.

Example of Event

- Single γ -ray event
- Typical light distribution
- Pileup probability: 0.01

Example of Event

- Single γ -ray event
- Spread due to escaped γ -rays
- Pileup probability: 0.31

Example of Events

- Single γ-ray event
- Fake peaks from escaped γ-rays
- Pileup probability: 0.99

Example of Event

- Two γ-ray event
- Pileup γ-ray in deep region
 - depth: 16 cm
 - energy: 4.51 MeV
- Pileup probability: 0.65

Example of Event

- Two γ-ray event
- Pileup γ-ray in deep region
 - depth: 22 cm
 - energy: 1.80 MeV
- Pileup probability: 0.02

• The peak search in "inverted" light distribution is also implemented.

 \leftarrow Light yields are calculated to be negative if there is a pileup in the baseline region.

• The peak search is performed at the same way after multiplying -1.

Step3: Unfolding of Sum Waveform

Assignment of Event Status

- Finally, an event status is assigned for each event depending on the unfolding results.
 - NoPileup: Only main γ -ray is found by all the algorithms.
 - Unfolded: Pileup γ-rays are found, and they are successfully unfolded.
 - Coincidence: On-timing pileup γ -rays are found, which cannot be unfolded.
 - **DLRejected**: DL model identifies pileups, but no pileup γ -ray is found with the others.
 - NotConverged: Fitting fails to converge.
- Events with Coincidence, DLRejected or NotConverged are rejected.

Performance - DL-based Pileup Identification -

• The performance of the DL algorithm was compared to that of peak search in the light distribution on the inner face.

 \leftarrow The same inputs are used.

• The DL model outperforms the peak search in the deep region.

 \leftarrow A peak structure is not required by utilizing the global distribution.

iraction

0.9

0.7

0.6

0.5

0.4

Detection Efficiency of Pileup γ-rays

+

-

- The noise situation greatly affects the prediction.
- The model trained with simulated noise misidentifies the single γ-ray events as pileup events.

Prediction for Single γ-ray Event

- One possible reason is the coherent noise.

 Not included in simulated noise.
- Noise coherently arises in the same WDB.

Channel Map

• The cluster-like structures in the light distribution derive from the coherent noise. \rightarrow Can be erased by shuffling the channel assignment.

Real Noise

Real Noise (Shuffled)

The shuffling of the channel assignment slightly decreases the entries in the misidentification peak.
 → Partly contributes, but not explained completely.

Prediction for Single γ-ray Event

• The training with the appropriate noise can vanish the difference of the prediction result.

Prediction for Single γ-ray Event

Dead Channel

Dead channels were masked based on the observation in 2021:

- 30 channels for the MPPCs
- 28 channels for the PMTs

Effect of PDE Degradation

- The PDE decrease can deteriorate the pileup elimination performance.
 Low signal-to-noise ratio
- The performance was evaluated assuming lower PDEs.
- At PDE 2%, N_{bg} increases due to the lower detection efficiency. \leftrightarrow Signal efficiency also gets higher.

Effect of PDE Degradation

- The PDE decrease can deteriorate the pileup elimination performance.
 ← Low signal-to-noise ratio
- The performance was evaluated assuming lower PDEs.
- N_{bg} is slightly less when the model is trained at PDE 2%.
 ↔ Signal efficiency is similar.

Effect of Real Noise

The real noise was found to increase the number of backgrounds in the high energy region.

• Previous algorithm : +11% (48-58 MeV)

+28% (51.5-54 MeV)

• New algorithm : + 2% (48-58 MeV)

+ 5% (51.5-54 MeV)

10 10^{-1} Previous (simulated noise) Previous (real noise) 10^{-2} New (simulated noise) New (real noise 10^{-3} 50 55 60 45 (MeV)

BG E_v spectrum

Nominal Branching Ratio before Update

- The previous paper reported the sensitivity of 6×10^{-14} for three years of data-taking.
- The nominal setting of this estimation includes
 - Positron updates: $k = 1.03 \times 10^{14} \rightarrow k = 9.38 \times 10^{13}$
 - Real noise: increase of the background events by 28% $(E_v = 51.5-54 \text{ MeV}).$
- The nominal value is calculated to be 7.8×10^{-14} .

Background E_γ Spectrum

Previous Algorithm for Pileup Elimination

- Used in "Shinji Ogawa. Liquid xenon detector with highly granular scintillation readout to search for µ+ → e+γ with sensitivity of 5 × 10–14 in the MEG experiment. PhD thesis, The University of Tokyo, 2020. URL: https://meg.web.psi.ch/docs/theses/ogawa_ph d. pdf."
- Pileup elimination with waveform
 - Only off-timing pileups can be detected.
 - Instability of fitting
 - Not tolerant to noise with hard codded parameters

Pileup elimination by waveform

2020/11/17

PHD DEFENSE SHINJI OGAWA

Discussion on Projected Sensitivity

- The sensitivity at the halved intensity was found to be equivalent to the other:
 - 5.6 \times 10⁻¹⁴ at 7 \times 10⁷ μ +stops/s
 - 5.8 \times 10^{-14} at 3.5 $\times 10^7~\mu^{+} stops/s$
 - \leftarrow Reduction of accidental backgrounds

Improvement of positron performance* at lower intensity

Variables	Values				
variables	$7 imes 10^7 \mu^+ { m stops/s}$	$3.5 imes 10^7 \mu^+ { m stops/s}$			
$\sigma_{p_{\mathrm{e}^+}}$	$100{ m keV}/c$	$90\mathrm{keV}/c$			
$\sigma_{ heta_{\mathrm{e}^+}}$	$6.7\mathrm{mrad}$	$6.2\mathrm{mrad}$			
$\sigma_{\phi_{\mathrm{e}^+}}$	$4.9\mathrm{mrad}$	$4.7\mathrm{mrad}$			
σ_{E_γ}	1.7%				
σ_{u_γ}	$2.5-8.1\mathrm{mm}~(\mathrm{for~different}~w_\gamma)$				
σ_{v_γ}	2.5 – $7.4\mathrm{mm}~(\mathrm{for~different}~w_\gamma)$				
σ_{w_γ}	$2.4{-}13.9\mathrm{mm}~(\mathrm{for~different}~w_\gamma)$				
$\sigma_{t_{\mathrm{e}^+\gamma}}$	$70\mathrm{ps}$	$68\mathrm{ps}$			
$\epsilon_{ m e^+}$	65%	74%			
ϵ_γ	69%				
k	9.38×10^{13} (for three years) 5.34×10^{13} (for three years)				

* Without the improvement of the positron performance, the sensitivity at the higher intensity is better thanks to the sufficient reduction power of the γ -ray pileup elimination.

Table 7.1 Parameters for sensitivity calculation as the nominal setting.

Discovery Sensitivity Sensitivity in Design Paper

- At design stage,
 - 90% C.L. Exclusion: 6×10⁻¹⁴ for three years
 ← same as the value of this thesis
 - 3σ Discovery: 1×10^{-13} for three years
 - 5 σ Discovery: 2×10⁻¹³ for three years

