Doctoral Dissertation

Search for lepton flavor violating muon decay mediated by a new light particle in the MEG experiment

By

MITSUTAKA NAKAO

Department of Physics
Graduate School of Science
THE UNIVERSITY OF TOKYO

A Dissertation Submitted for the Degree of DOCTOR OF PHILOSOPHY.

JANUARY 2020

ABSTRACT

Recently, in elementary particle physics research, exploring new physics beyond the standard model has been actively conducted. However, there is no clear evidence of such a new physics to date but of some anomalies. To tackle this situation, we have searched for lepton flavor violating muon decay mediated by a new light particle. The charged lepton flavor violation is one of the powerful tools to search for new physics beyond the standard model. On the other hand, light new physics has attracted a great deal of attention. In the analysis performed in this thesis, we combined these directions and have searched for the $\mu^+ \to e^+X, X \to \gamma\gamma$ decay using the full datasets (2009–2013) of the MEG experiment.

The MEG experiment was designed to search for lepton flavor violating muon decay, $\mu^+ \to e^+\gamma$, not for $\mu^+ \to e^+X, X \to \gamma\gamma$. However, we made full use of the resources developed for the $\mu^+ \to e^+\gamma$ search. We have newly developed reconstruction methods and dedicated corrections for the $\mu^+ \to e^+X, X \to \gamma\gamma$ search. The search analysis has been completely renewed; we combined blind, cut-counting, and maximum likelihood analysis. The full datasets of the MEG experiment, which corresponds to $7.5 \times 10^{14}\,\mu^+$s decay on the target, were analyzed. No significant excess was found in the mass region of 20–45 MeV, lifetime below 40 ps. Thus, we set the most stringent branching ratio upper limits in the mass region of 20–40 MeV. Especially, upper limits are pushed down to the level of $\mathcal{O}(10^{-11})$ for 20–30 MeV.

It is at most 60 times stringent result than the bound converted from the previous experiment, the Crystal Box experiment. Together with the previous analysis using the first-two-year (2009 and 2010) physics data of the MEG experiment, this is the first direct search of the $\mu^+ \to e^+X, X \to \gamma\gamma$ decay in the world.

1

INTRODUCTION

Physics is the natural science that tries to explain all the phenomena as simple as possible from the smallest number of laws. Elementary particle physics is one of the physics fields that studies the fundamental matter, elementary particles, and their mechanics in space and time. Huge numbers of theoretical and experimental studies have revealed a secret of the elementary particle physics and have established a milestone, the standard model.

The standard model of particle physics explains matter and forces well. In 2012, the last particle, the Higgs boson, which is predicted from the standard model, has been discovered in the ATLAS and CMS experiments [1, 2]. In the next year, the Nobel prize was awarded jointly to F. Englert and P. W. Higgs [3] for the theoretical discovery of the Brout-Englert-Higgs mechanism. With this discovery, the standard model has been "completed". However, there are unsolved mysteries in nature; we still have a long way to go towards the final goal of particle physics, *the theory of everything*.

The first extension of the standard model is the neutrino mass. Neutrinos are included in the standard model, but their masses are originally assumed to be 0. In 1998, neutrino oscillations have been discovered, which clearly shows a non-zero value of the neutrino mass. The nobel prize was awarded jointly to T. Kajita and A. B. McDonald [4] for this discovery. Recent topics in elementary particle physics research have been exploring new physics beyond the standard model. The neutrino mass was one of these topics.

However, there is no clear evidence of new physics to date but for some anomalies. In this thesis, we try to tackle this situation by combining two different directions: charged lepton flavor violation and light new physics.
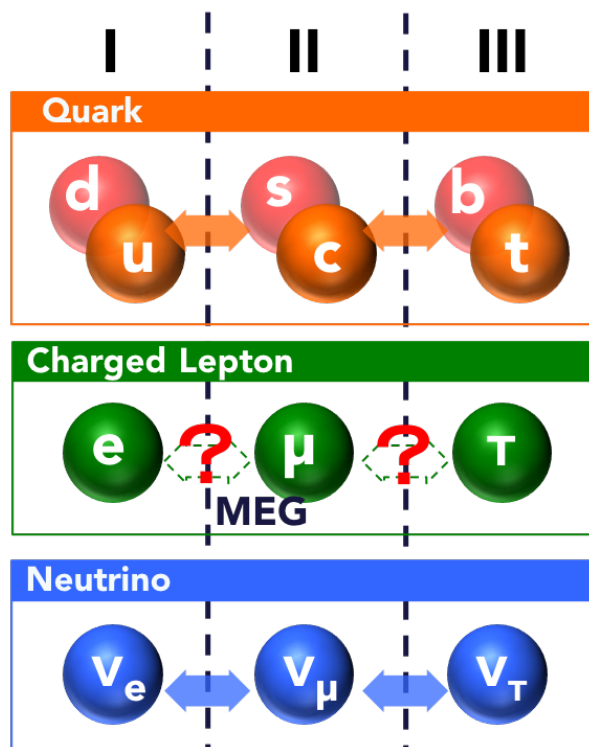
Figure 1.1: Fermions in the standard model. Quarks and leptons (charged leptons and neutrinos) have three generations (I, II, and III).

1.1 Charged lepton flavor violation

Fermions in the standard model compose matters in nature and they are summarized in Figure 1.1. These particles can be divided into two sectors: quarks and leptons. Each sector has six species, called flavor. Leptons are further divided into charged leptons and neutral leptons (neutrinos). There are three generations and the higher generations are just copies of the lowest one. The only difference is their masses.

One of the striking features of these particles is an inter-generational mixing, called flavor violation. In the quark and neutrino sectors, this phenomenon has already been discovered. However, in the charged lepton sector, it has not yet been discovered. The observation of neutrino oscillations suggests that the lepton flavor is not conserved in nature. Therefore, it is natural that the charged lepton flavor is violated even though the charged lepton flavor violation (CLFV) is forbidden in the standard model. It can occur at an experimentally observable rate in the framework of new physics beyond the standard model such as the grand unified theories. This is the reason why there are a lot of efforts to find a hint of CLFV in the world [5].

The MEG experiment at Paul Scherrer Institut (PSI) in Switzerland searched for one of the charged lepton flavor violating processes, $\mu^+ \rightarrow e^+ \gamma$ decay, with the highest sensitivity in the world. No evidence of the decay was found, leading to a new upper limit on the branching ratio of

4.2×10^{-13} (90%C.L.)[6]. The collaboration is preparing the upgraded experiment, the MEG II experiment, to achieve a sensitivity below 6×10^{-14} after three-year-data taking[7]. Other CLFV experiments such as Mu3e[8], Mu2e[9], and COMET[10] are planned in the 2020s.

These searches, however, assume new particles with a CLFV coupling whose mass is much heavier than that of the related particles (MeV range). No hints for new physics in the CLFV searches may suggest new physics exist in the lighter scale, that is, MeV scale.

1.2 Light new physics

There are two directions in particle physics experiments: intensity-frontier experiments and energy-frontier experiments. The MEG experiment is one of the intensity-frontier experiments. The Large Hadron Collider (LHC), one of the energy-frontier experiments, has also searched for new physics, but no hint has been observed. Therefore, the importance of searches for new physics in a completely different approach is increasing. Recently, unconventional searches such as displaced vertex searches and long-lived particle searches have been proposed[11–13] both in the ATLAS[14] and CMS[15] experiments in LHC. In addition, there are several proposals for new projects to search for these unconventional signals in LHC: MATHUSLA[16], MilliQan[17–19], Codex-b[20], SHiP[21–23], and FASER[24–27]. Some of these searches assume new particles exist in the lighter (MeV) range.

There are several implications of light (MeV range) new physics both experimentally and theoretically. In 2016, a Hungarian group reported a possibility of an existence of a new particle X17 with a mass of 17 MeV[28]. Furthermore, new evidence supporting the existence was reported in 2019[29]. Possible explanations of the observation are summarized in Ref. [29]. A Russian group reported a possibility of a new light scalar particle with a mass of 38 MeV in photon pairs spectra[30, 31] in 2015. The KOTO experiment reported an excess of events in the signal region of $K_L \to \pi^0 \nu \bar{\nu}$ decay[32]. Possible explanations of the excess including new light scalars are discussed in Ref. [33, 34]. Theoretically, there are several ideas to explain anomalies with the light particles: light scalars that carry lepton number and solve the $(g-2)_e$ anomaly are proposed[35]. Axion-like particles with lepton-flavor violating couplings can offer new ways to explain the anomalies related to the magnetic moments of muon and electron[36]. No observation of $\mu^+ \to e^+ \gamma$ does **not** directly indicate that $\mu^+ \to e^+ X, X \to \gamma\gamma$ does not exist, neither.

A possible search for such a new light particle in the MEG experiment is $\mu^+ \to e^+ X, X \to \gamma\gamma$ (MEx2G) decay process, where a (pseudo) scalar X is generated via a lepton flavor violating coupling and the on-shell X decays into standard model particles, two γs. In this thesis, we search for this decay process using the full dataset of the MEG experiment. In the search, we assume X is long-lived and the width of m_X is narrow. Axion-like particles[37–40], majoron[41, 42], familon[43–46], flavon[47, 48], flaxion[49], hierarchion[50] and strongly interacting massive particles[51, 52] can be possible candidates of X.

2

$\mu^+ \to \mathrm{e}^+\mathrm{X}, \mathrm{X} \to \gamma\gamma$ **DECAY**

I n this chapter, the information needed for the MEx2G search is summarized. First, we review muon and its decay modes which are needed for the analysis. Then we discuss possible parameter space for the search after summarizing constraints from the other experiments. Finally, kinematics and background are discussed, which is needed for the MEx2G event reconstruction and its search analysis.

2.1 Muon in the standard model

2.1.1 Basic proprieties

Table 2.1: Basic properties of muon[53].

Paramter	Value
Mass	$105.6583745 \pm 0.0000024$ MeV
Life	$(2.1969811 \pm 0.0000022) \times 10^{-6}$ s
Magnetic moment($\frac{g_\mu - 2}{2}$)	$(11659208.9 \pm 5.4 \pm 3.3) \times 10^{-10}$ s
Electric dipole moment	$(-0.1 \pm 0.9) \times 10^{-19}\, e$ cm

Muon was found in cosmic rays in 1936[54]. It has 200 times larger mass than electron, but the other properties are the same as electron. The discovery of muon was the first clue of generations of elementary particles. Figure 2.1 summarizes the properties of muons.

The standard model interaction of muon is given by

$$
\begin{aligned}
\mathscr{L} = \ & e\bar{\mu}\gamma^\mu \mu A_\mu \\
& - \frac{g}{\sqrt{2}}(\bar{\nu}_{\mu_L}\gamma^\mu \mu_L W_\mu^+ + \bar{\mu}_L \gamma^\mu \nu_{\mu_L} W_\mu^-) \\
& - \sqrt{g^2 + g'^2}\{\bar{\mu}_L \gamma^\mu (-\frac{1}{2} + \sin^2\theta_W)\mu_L + \bar{\mu}_R \gamma^\mu \sin^2\theta_W \mu_R\} Z_\mu^0 \\
& - \frac{m_\mu}{v}\bar{\mu}\mu H.
\end{aligned}
$$
(2.1)

Muon interacts through the electromagnetic (first line in Equation (2.1)) and weak interactions (second and third lines in Equation (2.1)). It also couples to the Higgs boson (fourth line in Equation (2.1)).

2.1.2 Decay modes

The decay modes of muon are summarized in Table 2.2.

Table 2.2: Decay modes of muon

Mode	Fraction	References
$\mu^{(-)} \to e^{(-)}\nu_{(\mu)}\bar{\nu}_{(e)}$	~100 %	[53]
$\mu \to e\nu\bar{\nu}\gamma$	(1.4 ± 0.4) %	[53]
$\mu \to e\nu\bar{\nu}_e\bar{e}e$	$(3.4 \pm 0.4) \times 10^{-5}$	[53]
$\mu^{(-)} \to e^{(-)}\nu_{(e)}\bar{\nu}_{(\mu)}$	< 1.2 %(90% C.L.)	[55]
$\mu \to e\gamma$	$< 4.2 \times 10^{-13}$ (90% C.L.)	[6]
$\mu \to e\bar{e}e$	$< 1.0 \times 10^{-12}$ (90% C.L.)	[56]
$\mu \to e\gamma\gamma$	$< 7.2 \times 10^{-11}$ (90% C.L.)	[57]
$\mu N \to eN$	$< \mathcal{O}(10^{-12})$	[58]
$\mu \to e\gamma X, X \to \text{invisible}$	$< \mathcal{O}(10^{-9})$	[57, 59]
$\mu \to eX, X \to \text{invisible}$	$< \mathcal{O}(10^{-5})$	[60]
$\mu \to eX, X \to ee$	$< \mathcal{O}(10^{-11})$	[61]
$\mu \to eX, X \to \gamma\gamma$	$< \mathcal{O}(10^{-10})$	[57, 62]

Michel decay   The dominant mode is $\mu \to e\nu\bar{\nu}$, which is called the Michel decay. A muon decays into an $e^{-(+)}$ and two neutrinos without γ emission. It is used as the normalization mode in the MEx2G analysis. e^+ coming from the decay is one of the sources of accidental backgrounds.

The differential decay rate in the standard model is given by ([63])

$$
\frac{d^2\Gamma\left(\mu^\pm \to e^\pm \nu\bar{\nu}\right)}{dx\,d\cos\theta_e} = \frac{m_\mu^5 G_F^2}{192\pi^3}x^2[(3-2x)\pm P_\mu \cos\theta_e(2x-1)],
$$
(2.2)

where $x = \frac{2m_\mu}{m_\mu^2 + m_e^2}\cdot E_e$; m_μ and m_e are mass of muon and $e^{-(+)}$, respectively; P_μ is polarization of muon and θ_e is angle between muon polarization and $e^{-(+)}$ momentum direction. Figure 2.1 shows
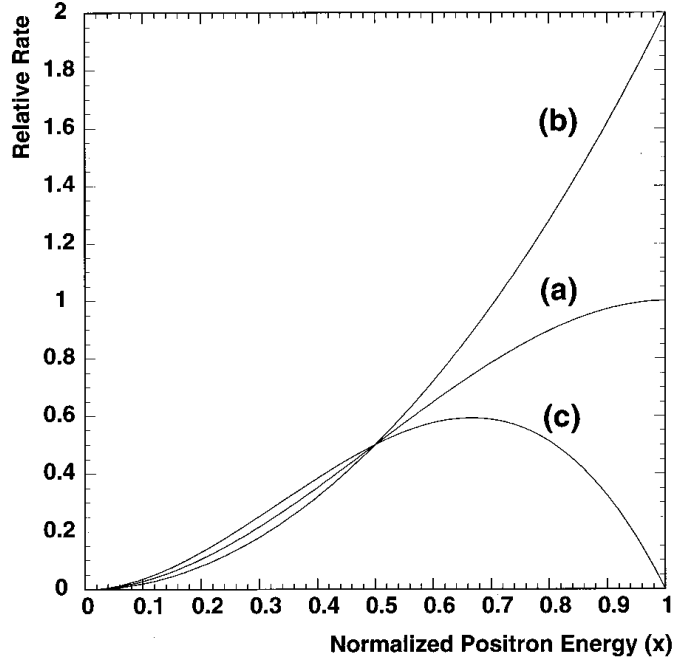
Figure 2.1: Michel spectrum of fully polarized $\mu^+ \to e^+ \nu_e \bar{\nu}_\mu$. (a) $\cos\theta_e = 0$, (b) $\cos\theta_e = +1$, (c) $\cos\theta_e = -1$ (from [63])

the e^+ energy spectrum for $\mu^+ \to e^+ \nu_e \bar{\nu}_\mu$ in the standard model, for the cases of $\cos\theta_e = 0, +1, -1$. $P_\mu = 1$ is assumed in all cases. A radiative correction to Equation (2.2) is given in Ref. [64], which is used in the analysis.

Radiative muon decay   The next to the dominant mode is $\mu \to e\nu\bar{\nu}\gamma$, which is called the radiative muon decay (RMD). RMD emits a γ ray in the final state. Thus this mode is also used in the analysis for calibration and performance evaluation purposes. A γ ray coming from RMD is one of the sources of accidental backgrounds. The differential branching ratio is given by ([63])

$$\frac{dB\left(\mu^\pm \to e^\pm \nu\bar{\nu}\gamma\right)}{dy\,d\cos\theta_\gamma} = \frac{1}{y}\left[J_+(y)\left(1 \pm P_\mu \cos\theta_\gamma\right) + J_-(y)\left(1 \mp P_\mu \cos\theta_\gamma\right)\right],$$

where θ_γ is the angle between muon polarization and γ momentum. y is the normalized γ energy given by $y = 2E_\gamma/m_\mu$. Terms suppressed by m_e/m_μ are neglected to get Equation (2.3). $J_+(y)$ and

7
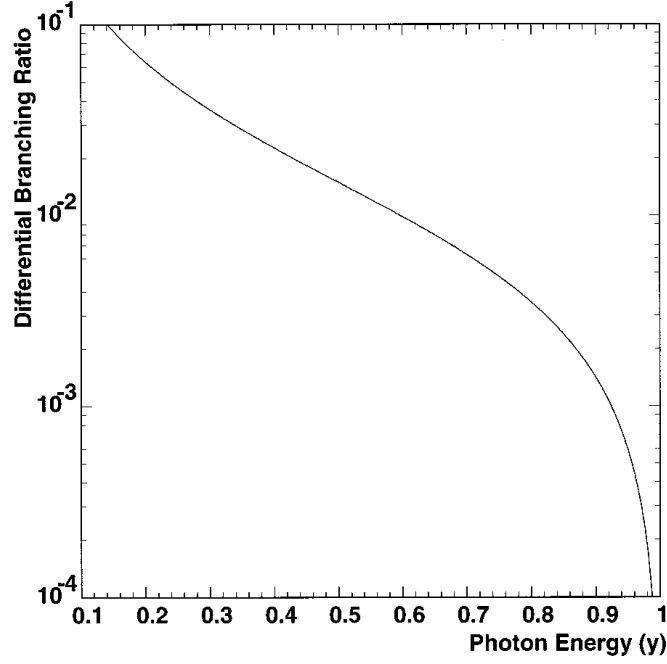
Figure 2.2: Differential branching ratio of RMD (from [63])

$J_-(y)$ are defined by

$$
\begin{aligned}
J_+(y) &= \frac{\alpha}{6\pi}(1-y)\left[\left(3\ln\frac{1-y}{r} - \frac{17}{2}\right)\right.\\
&\quad \left. +\left(-3\ln\frac{1-y}{r} + 7\right)(1-y) + \left(2\ln\frac{1-y}{r} - \frac{13}{3}\right)(1-y)^2\right],\\
J_-(y) &= \frac{\alpha}{6\pi}(1-y)^2\left[\left(3\ln\frac{1-y}{r} - \frac{93}{12}\right)\right.\\
&\quad \left. +\left(-4\ln\frac{1-y}{r} + \frac{29}{3}\right)(1-y) + \left(2\ln\frac{1-y}{r} - \frac{55}{12}\right)(1-y)^2\right],
\end{aligned}
$$

where $r = (m_e/m_\mu)^2$. Figure 2.2 shows the γ spectrum for unpolarized muons.

Non-standard decay modes   Decay modes except for the first three modes listed in Table 2.2 are not predicted from the standard model. Current upper limits are also listed. Any observation of these modes indicates new physics. In particular, the MEG collaboration gives the stringent limits on $\mu \to e\gamma$ and $\mu \to eX, X \to \gamma\gamma$. The latter one, the MEx2G decay, is the main topic of this thesis.

2.2 Decay rate

A full effective field theory description of $\mu^+ \to e^+ X, X \to \gamma\gamma$ assuming X to be axion-like particles can be found in Refs. [40, 65]. In the on-shell regime, the branching ratio of $\mu^+ \to e^+ X, X \to \gamma\gamma$ can be obtained using the narrow-width approximation:

$$\mathcal{B}(\mu^+ \to e^+ X, X \to \gamma\gamma) \approx \mathcal{B}(\mu^+ \to e^+ X) \times \mathcal{B}(X \to \gamma\gamma).$$

The effective Lagrangian of $X \to \gamma\gamma$ is given by:

$$\mathcal{L} \supset -\frac{g_{\gamma\gamma}}{4} X F_{\mu\nu} \tilde{F}^{\mu\nu},$$

where $g_{\gamma\gamma}$ is the coupling constant of $X \to \gamma\gamma$ vertex. It contains a loop contribution and has the inverse of the cut-off scale of the effective Lagrangian. Then the decay width Γ is given by

$$\Gamma(X \to \gamma\gamma) = \frac{g_{\gamma\gamma}^2 m_X^3}{64\pi}.$$

The boosted decay length of X in the lab frame is given by

$$l = \frac{c|P_X|}{m_X \Gamma}. \tag{2.3}$$

Assuming the dominant decay mode of X to be $X \to \gamma\gamma$ like electrophobic models proposed in Refs. [66, 67], the lifetime of X is given by $\tau = 1/\Gamma$.

2.3 Experimental searches

Current constraints of the MEx2G decay are classified into three types: direct search, constraints from related modes, and constraints from parameter space of X.

2.3.1 MEG2012

The MEx2G direct search requires experimental features in common with the $\mu^+ \to e^+\gamma$ search, and hence, the $\mu^+ \to e^+\gamma$ search experiments, like MEG, are suitable for this search. As will be reviewed in Chapter 3, the MEG experiment utilizes the world's highest muon beam and high-performance gamma and positron detectors. Therefore, it is a unique experiment enabling the best search of this decay to date. The MEx2G direct search was actually performed, for the first time, by the MEG collaboration using the datasets taken in 2009 and 2010. No significant excess was found and upper limits on the branching ratio for the mass range of 10–45 MeV and lifetime equal or less than 10 ns were set. The upper limits for a mass range below ~30 MeV have been significantly updated. The results are summarized in a Ph.D. thesis published in 2012 [62].
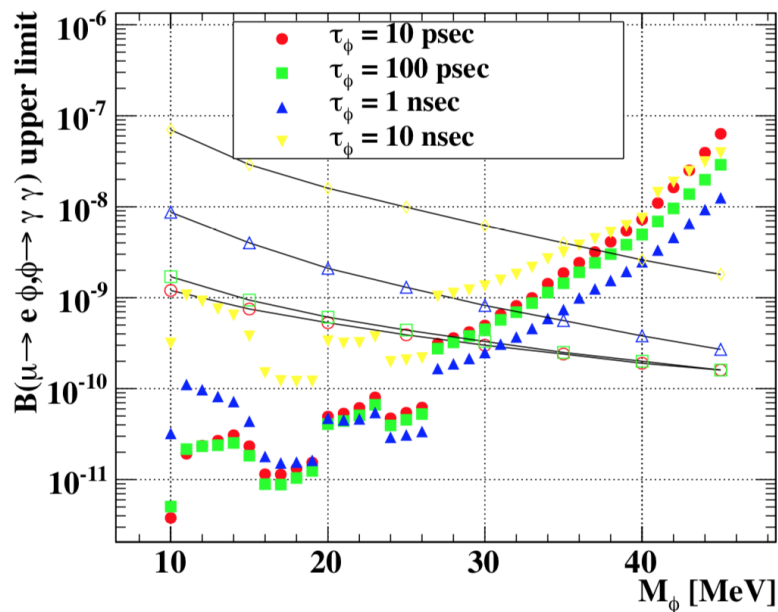
Figure 2.3: Upper limits estimated from the Crystal Box experiment (blank markers with solid lines) and the previous MEx2G search. Marker colors correspond to lifetime of the mediated particle (from [62]).

2.3.2 The Crystal Box experiment

The Crystal Box experiment [57] searched for the lepton-flavor-violating muon decays: $\mu \to e\gamma$, $\mu \to e\gamma\gamma$, $\mu \to eee$. Among them, $\mu \to e\gamma\gamma$, whose most stringent upper limit is given by this experiment, is relevant to the constraint on the MEx2G decay. This upper limit can be converted into the MEx2G equivalent upper limit taking the difference of the detector acceptance into account. The detailed calculation is found in Ref. [62]. In this calculation, relative differences of geometrical acceptance for both decays ($\mu \to e\gamma\gamma$ and $\mu \to eX, X \to \gamma\gamma$) are multiplied by the obtained upper limit for $\mu \to e\gamma\gamma$ (7.2×10^{-11}). The converted limits are shown in Figure 2.3 as the solid lines.

2.3.3 Constraints from other modes

As listed in Table 2.2, upper limit of $\mu \to eX, X \to$ invisible mode is $\mathcal{O}(10^{-5})$ in mass range of 13–80 MeV [60]. This level gives no constraints on the MEx2G decay considering the current upper limit of the decay. In the future, the Mu3e experiment has a possibility to push the upper limit down to 5×10^{-9} [68], which is comparable to the current upper limit of the MEx2G decay in the higher mass region.

$\mu \to eX, X \to ee$ can give a more stringent constraint on the MEx2G decay than its direct search if we assume X is more likely to decay into e^+e^- pairs. However, there is a possibility for X to be electrophobic as suggested in Refs. [66, 67] and searches for both decay modes give a hint to
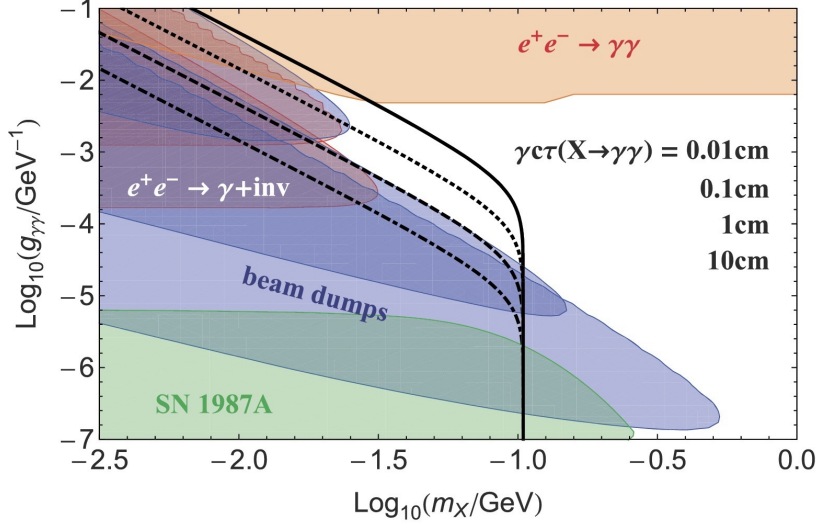
Figure 2.4: Excluded parameter regions for a scalar X with mass and coupling to γs from collider, beam dumps, and supernova (from [69]). Only $g_{e\mu}$ and $g_{\gamma\gamma}$ are assumed to be non-zero. The black lines corresponds to the boosted decay length of X; the solid one is 0.01 cm; the dotted one is 0.1 cm; the dashed one is 1 cm; the dot-dashed one is 10 cm. The region with decay length < 1 cm and $m_X > 20$ MeV have room for the MEx2G search[69].

determine the model behind these decay modes.

2.3.4 Constraints from other experiments

In axion-like particles' search, collider, beam dumps, and supernova observation give constraints on $X \to \gamma\gamma$ if the axion-like particles are generated from coupling with γ. Figure 2.4 summarizes the excluded parameter region by these experiments. A part of parameter space for the MEx2G decay is excluded by lepton collider experiments and beam dump experiments. The region with decay length < 1 cm and $m_X > 20$ MeV have room for the MEx2G search.

2.3.5 Target parameter space in this analysis

We define the target parameter space in τ–m_X surface as shown in Figure 2.5. τ is the lifetime of the mediated particle X. m_X is the mass of X. The decay length is converted into the lifetime of X(τ) by using Equation (2.3) and the subsequent assumption, which gives:

$$(2.4) \qquad l(= c\beta\gamma\tau) < 1\,(\text{cm}) \quad \Leftrightarrow \quad \tau < \frac{m_X}{cP_X} \cdot 1\,(\text{cm}),$$

where P_X is the momentum of X. The blue region has already been excluded and the red region ($> m_\mu/2$) cannot be searched from the two body decay of muon. Therefore we focus on the following parameter space:

$$m_X = (20, 25, 30, 35, 40, 45)\,\text{MeV}$$
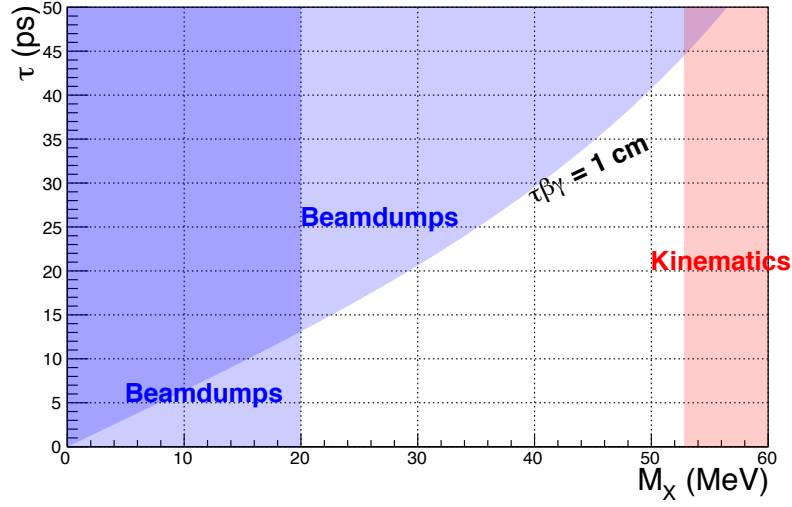
Figure 2.5: Allowed parameter space. We focus on the white region in this analysis.

$$\tau = (5, 20, 40)\,\text{ps}.$$

2.4 Kinematics

In this section, we calculate the kinematical dependence of output particles, particularly on 2γs.

In the muon's rest frame, four momentum of $\mu \to eX$ is written as

$$(2.5) \qquad \begin{pmatrix} E_\mu \\ 0 \end{pmatrix} = \begin{pmatrix} E_e \\ \vec{P}_e \end{pmatrix} + \begin{pmatrix} E_X \\ \vec{P}_X \end{pmatrix},$$

where E_μ, E_e, E_X are energy of μ, e, X, respectively. $\vec{P}_e$ is momentum of e and $\vec{P}_X$ is momentum of X. $\vec{P}_e + \vec{P}_X = 0$ folds by momentum conservation in the μ's rest frame. Thus we define $P_e = |\vec{P}_e| = |\vec{P}_X|$. Equation (2.5) can be rewritten as

$$\begin{pmatrix} m_\mu \\ 0 \end{pmatrix} = \begin{pmatrix} \sqrt{m_e^2 + P_e^2} \\ \vec{P}_e \end{pmatrix} + \begin{pmatrix} \sqrt{m_X^2 + P_e^2} \\ -\vec{P}_e \end{pmatrix},$$

where $E_\mu = m_\mu$ is used. By energy conservation, one can get

$$m_\mu = \sqrt{m_e^2 + P_e^2} + \sqrt{m_X^2 + P_e^2}.$$

Therefore, P_e can be written as a function of masses:

$$\begin{aligned} P = P_e = P_X &= |\vec{P}_e| = |\vec{P}_\times| \\ &= \frac{\sqrt{\left(m_\mu^2 + m_e^2 - m_X^2\right)^2 - 4m_\mu^2 m_e^2}}{2m_\mu}. \end{aligned}$$
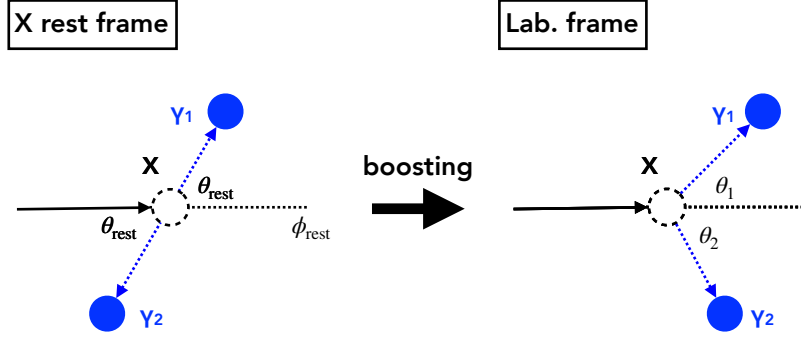
Figure 2.6: The $\mu^+ \rightarrow e^+ X, X \rightarrow \gamma\gamma$ decay topology.

Hereafter, we calculate $X \rightarrow \gamma\gamma$ kinematics. As shown in Figure 2.6, 2γs are emitted in the opposite direction each other in the X rest frame. The X-boosted frame, i.e. the lab. frame is shown on the right-hand side of the figure. We define the boosted direction as the x-axis. The y-axis is defined so that 2γs are decays into an x-y-surface. An emitted angle of γ from the x axis and a rotation angle around x axis are defined as θ_{res} and ϕ_{res}, respectively. θ_{res} is restricted from 0 to $\pi/2$ by defining the forward γ as γ_1 shown in Figure 2.6. ϕ_{res} can take the full range (0 to 2π).

The boosting parameters, β, γ are

$$(2.6) \qquad \beta = \frac{P_X}{\sqrt{m_X^2 + P_X^2}},$$

$$(2.7) \qquad \gamma = \frac{1}{\sqrt{1-\beta^2}} = \frac{\sqrt{m_X^2 + P_X^2}}{m_X}.$$

m_X dependence of P_X, β, γ, sum of 2γs' energy ($E_{\gamma 1} + E_{\gamma 2}$) are shown in Figure 2.7.

The relation of four momentum in the X rest frame is

$$\begin{pmatrix} E_X \\ 0 \end{pmatrix} = \begin{pmatrix} E_1 \\ \vec{P}_1 \end{pmatrix} + \begin{pmatrix} E_2 \\ \vec{P}_2 \end{pmatrix}.$$

By momentum conservation and $m_\gamma = 0$,

$$\begin{pmatrix} \sqrt{m_X^2 + 0^2} \\ 0 \end{pmatrix} = \begin{pmatrix} \sqrt{m_\gamma^2 + p^2} \\ \vec{p} \end{pmatrix} + \begin{pmatrix} \sqrt{m_\gamma^2 + p^2} \\ -\vec{p} \end{pmatrix}.$$

Thus,

$$\begin{pmatrix} m_X \\ 0 \end{pmatrix} = \begin{pmatrix} p \\ \vec{p} \end{pmatrix} + \begin{pmatrix} p \\ -\vec{p} \end{pmatrix}.$$

By energy conservation,

$$p = \frac{m_X}{2}.$$
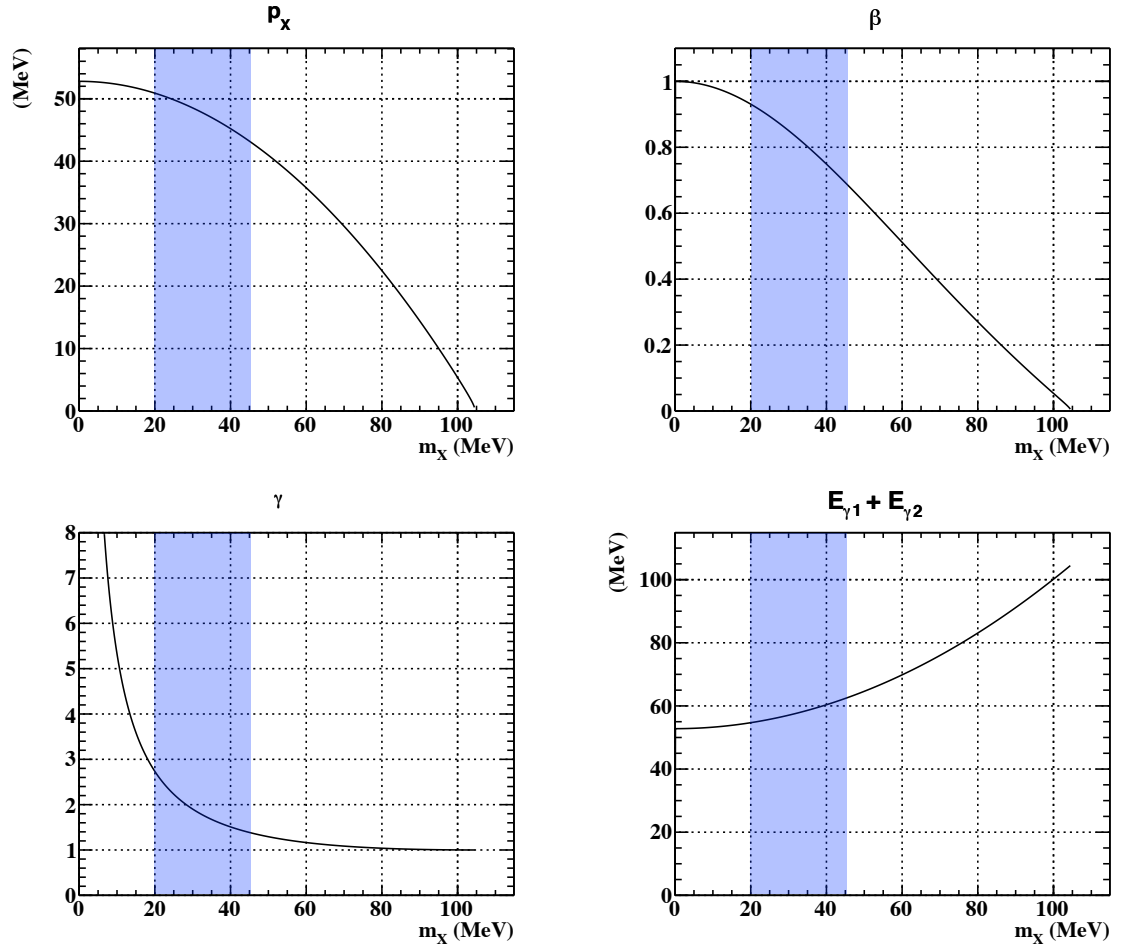
Figure 2.7: m_X dependence of $P_X(= P_e), \beta, \gamma, E_{\gamma 1} + E_{\gamma 2}$. The target parameter space is shown in shadow.

Hereafter we calculate m_X dependence of each γ. Four momenta of 2γs are rewritten as

$$\begin{pmatrix} p \\ p_x \\ p_y \\ p_z \end{pmatrix}, \begin{pmatrix} p \\ -p_x \\ -p_y \\ -p_z \end{pmatrix}.$$

By definition, $p_z = 0$ and p_x, p_y are function of θ_{rest}.

$$\begin{pmatrix} p \\ p \cos\theta_{\text{rest}} \\ p \sin\theta_{\text{rest}} \\ 0 \end{pmatrix}, \begin{pmatrix} p \\ -p \cos\theta_{\text{rest}} \\ -p \sin\theta_{\text{rest}} \\ 0 \end{pmatrix}.$$

14

The Lorentz transformation matrix is

$$
\begin{pmatrix}
\gamma & \beta\gamma & 0 & 0 \\
\beta\gamma & \gamma & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix},
$$

where the lab frame is moving to the $-$X direction with a speed of β. Thus the four momenta of γs are transformed to

$$
\begin{pmatrix}
p \\
p\cos\theta_{\mathrm{rest}} \\
p\sin\theta_{\mathrm{rest}} \\
0
\end{pmatrix}
\mapsto
\begin{pmatrix}
\gamma p + \beta\gamma p\cos\theta_{\mathrm{rest}} \\
\beta\gamma p + \gamma p\cos\theta_{\mathrm{rest}} \\
p\sin\theta_{\mathrm{rest}} \\
0
\end{pmatrix}
$$

$$
\begin{pmatrix}
p \\
-p\cos\theta_{\mathrm{rest}} \\
-p\sin\theta_{\mathrm{rest}} \\
0
\end{pmatrix}
\mapsto
\begin{pmatrix}
\gamma p - \beta\gamma p\cos\theta_{\mathrm{rest}} \\
\beta\gamma p - \gamma p\cos\theta_{\mathrm{rest}} \\
-p\sin\theta_{\mathrm{rest}} \\
0
\end{pmatrix}.
$$

Therefore, energy of each γ is

$$
\begin{aligned}
E_1 &= \gamma p + \beta\gamma p\cos\theta_{\mathrm{rest}} \\
&= \gamma\frac{m_{\mathrm{X}}}{2} + \beta\gamma\frac{m_{\mathrm{X}}}{2}\cos\theta_{\mathrm{rest}},
\end{aligned}
\tag{2.8}
$$

$$
\begin{aligned}
E_2 &= \gamma p - \beta\gamma p\cos\theta_{\mathrm{rest}} \\
&= \gamma\frac{m_{\mathrm{X}}}{2} - \beta\gamma\frac{m_{\mathrm{X}}}{2}\cos\theta_{\mathrm{rest}}.
\end{aligned}
\tag{2.9}
$$

The opening angles $\theta_{1,2}$ in the lab. frame are calculated as follows:

$$
\begin{pmatrix}
E_1 \\
E_1\cos\theta_1 \\
E_1\sin\theta_1 \\
0
\end{pmatrix}
=
\begin{pmatrix}
\gamma p + \beta\gamma p\cos\theta_{\mathrm{rest}} \\
\beta\gamma p + \gamma p\cos\theta_{\mathrm{rest}} \\
p\sin\theta_{\mathrm{rest}} \\
0
\end{pmatrix},
$$

$$
\begin{pmatrix}
E_2 \\
E_2\cos\theta_2 \\
E_2\sin\theta_2 \\
0
\end{pmatrix}
=
\begin{pmatrix}
\gamma p - \beta\gamma p\cos\theta_{\mathrm{rest}} \\
\beta\gamma p - \gamma p\cos\theta_{\mathrm{rest}} \\
-p\sin\theta_{\mathrm{rest}} \\
0
\end{pmatrix}.
$$

Therefore,

$$
\cos\theta_1 = \frac{P_{\mathrm{X}}^2 + \left(E_1^2 - E_2^2\right)}{2E_1 P_{\mathrm{X}}},
\tag{2.10}
$$

$$
\cos\theta_2 = \frac{P_{\mathrm{X}}^2 - \left(E_1^2 - E_2^2\right)}{2E_2 P_{\mathrm{X}}},
\tag{2.11}
$$

where the following relations are used:

$$\beta\gamma = \frac{P_X}{m_X},$$
$$E_1 + E_2 = 2\gamma p = \sqrt{m_X^2 + P_X^2},$$
$$E_1 - E_2 = 2\beta\gamma p \cos\theta_{\text{rest}} = P_X \cos\theta_{\text{rest}}.$$

m_X and angle dependence are shown in Figure 2.8. When θ_{rest} is close to 90°, 2γs are orthogonal to the boosted direction and the energy difference becomes small. The difference between θ_1 and θ_2, emission angles in the lab frame, also becomes small. When θ_{rest} is close to 0°, on the other hand, one γ is the same direction of the boosted direction and the other is the opposite, resulting in the energy difference becomes large. For larger m_X, more γs are emitted in the opposite direction of the boosted direction in the larger θ_{rest} and the opening angle becomes larger, resulting in the less acceptance and less trigger efficiency (Section 3.5.2) for the MEx2G signal.
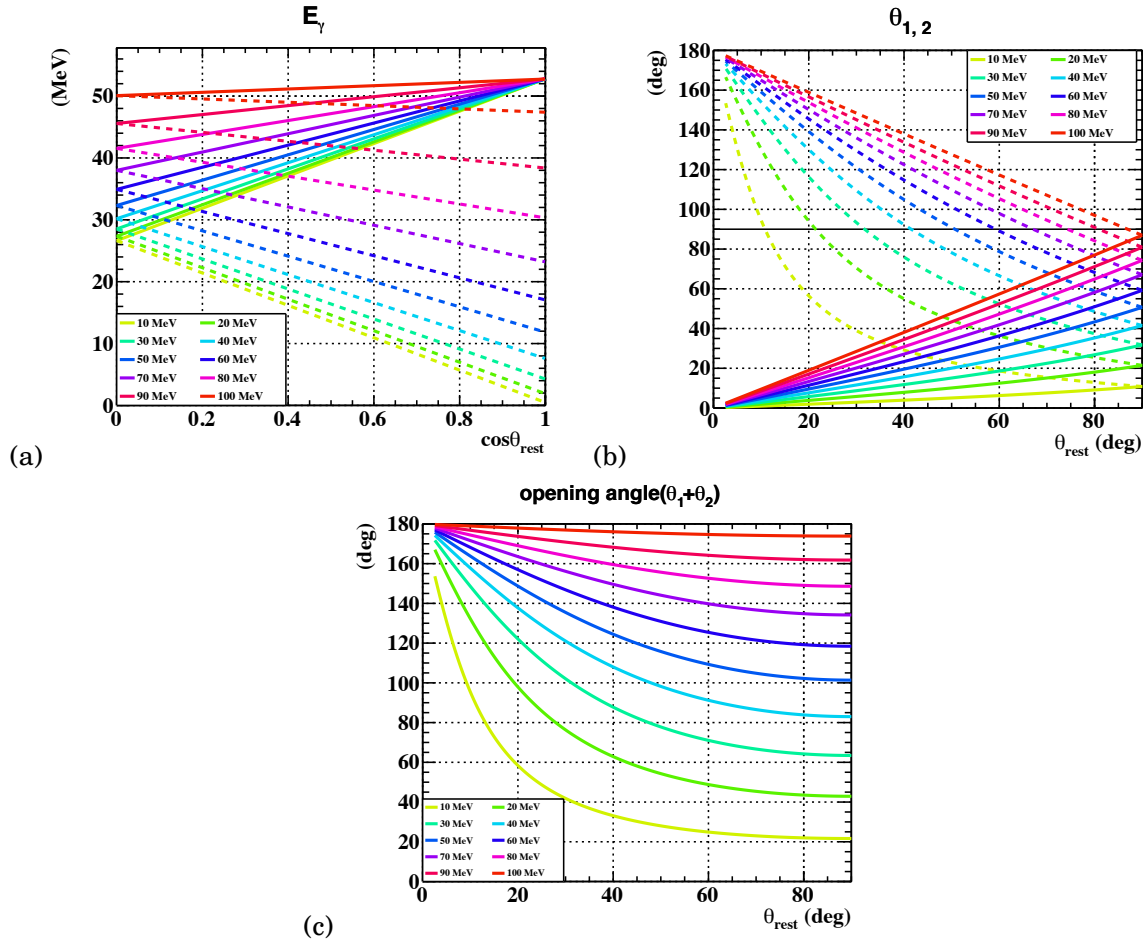
Figure 2.8: (a) E_γ as a function of $\cos\theta_{\text{rest}}$. Solid line: γ_1, dashed line: γ_2 (b) θ_1(solid line, γ_1), θ_2(dashed line, γ_2) as a function of θ_{rest}. (c) Opening angle as a function of θ_{rest}.

2.5 Backgrounds

Possible backgrounds are classified into four types as listed in Table 2.3. The symbols ($\bigcirc, \triangle, \diamond$) indicate origins and the same one has the same physics origin. The first three types are accidental backgrounds and the last one is a physics background. There are two candidates of the physics backgrounds: $\mu \to e\gamma\gamma\nu\nu$, which we call doubly radiative muon decay (DRMD) and a RMD ($\mu^+ \to e^+\gamma\nu\bar{\nu}$) associated background.

DRMD has one additional γ emission to RMD. DRMD was not considered seriously in the previous studies [62]. This mode has not yet been measured but it exists in the standard model framework. The partial branching ratio at tree level is calculated to be $\sim \mathcal{O}(10^{-15})$ by A. Signer and Y. Ulrich[1], which is negligible in our analysis. The other one is an RMD associated background. In an RMD event with e^+ energetic enough to trigger and a γ of large energy, the γ can convert into an e^+e^- pair before entering the γ calorimeter and both e^+ and e^- can reach the calorimeter mimicking two γs. Its probability is estimated to be $\sim \mathcal{O}(10^{-15})$ taking the RMD branching ratio and a probability for γ to convert into an e^+e^- pair before entering the calorimeter into account. This is also negligible.

Other accidental backgrounds (type1, 2, and 3) need to be considered in the analysis. In type1, e^+ and one of γ come from the same RMD event. The other γ comes from another RMD event or one of annihilation-in-flight (AIF, $e^+(+e^-) \to \gamma\gamma$) γ or bremsstrahlung of accidental e^+. In type2, e^+ comes from the Michel decay, and 2γ come from the same physics processes: the first process is AIF 2γ. In the second process, one γ comes from a RMD event and the other γ comes from AIF or bremsstrahlung from the RMD e^+. In type3, e^+ comes from the Michel decay and 2γ come from different physics processes. As we discuss in Section 7.2, the numbers of these background events are estimated using predefined off-timing sideband regions.

Table 2.3: Background types in the MEx2G analysis. See texts for the detailed description of each background source.

type	e^+	γ_1	γ_2	possible sources
1	$\bigcirc$	$\bigcirc$	$\triangle$	$\bigcirc$: RMD $\triangle$: RMD, AIF, brems.
2	$\bigcirc$	$\triangle$	$\triangle$	$\bigcirc$: Michel $\triangle$: AIF 2γ/ RMD, AIF or brems. from accidental e^+
3	$\bigcirc$	$\triangle$	$\diamond$	$\bigcirc$: Michel $\triangle, \diamond$: RMD, AIF, or brems.
4	$\bigcirc$	$\bigcirc$	$\bigcirc$	Doubly radiative muon decay, RMD accociated BG

---

[1]A related talk can be found in Ref. [70]. They calculated NLO of RMD [71]. In this calculation, the tree level of DRMD is included. This NLO contribution is negative and the branching ratio becomes a further smaller value.

I n this chapter, the experimental apparatus for the MEx2G search is introduced. We use data taken with the MEG detector. Thus we first overview the MEG experiment. Then the experimental apparatus of the MEG experiment related to the MEx2G search is presented. Finally, the analysis strategy for the MEx2G search with the MEG datasets is summarized.

3.1 Overview of the MEG experiment

The MEG experiment searched for the lepton flavor violating decay of muon, $\mu^+ \to e^+\gamma$, with an unprecedented sensitivity ever. In the standard model, this decay is not allowed to occur and the branching ratio calculated from the standard model extended with neutrino oscillation is very tiny ($\mathcal{O}(10^{-55})$ [72]). However, physics models beyond the standard model such as supersymmetric models and extra-dimensional models predict larger branching ratios ($\mathcal{O}(10^{-12}) - \mathcal{O}(10^{-14})$), which are experimentally accessible. To suppress background events and detect signal events, the following three points are required:

- Beam: high intensity and continuous muon beam.

- Gamma-ray detector: good time, energy, and position resolutions.

- Positron detector: operational under a high rate environment.

To get the required beam, the 2.2-mA proton cyclotron at Paul Scherrer Insititut (PSI) in Switzerland was used. We have developed a dedicated detector to meet the requirements above, the MEG detector [73]. Figure 3.1 shows a schematic view of the MEG detector. A muon beam comes from the left-hand side and stops at the stopping target placed at the center of the detector. An e^+ and a γ from a $\mu^+ \to e^+\gamma$ signal are emitted back-to-back from the target. The e^+ shown in the
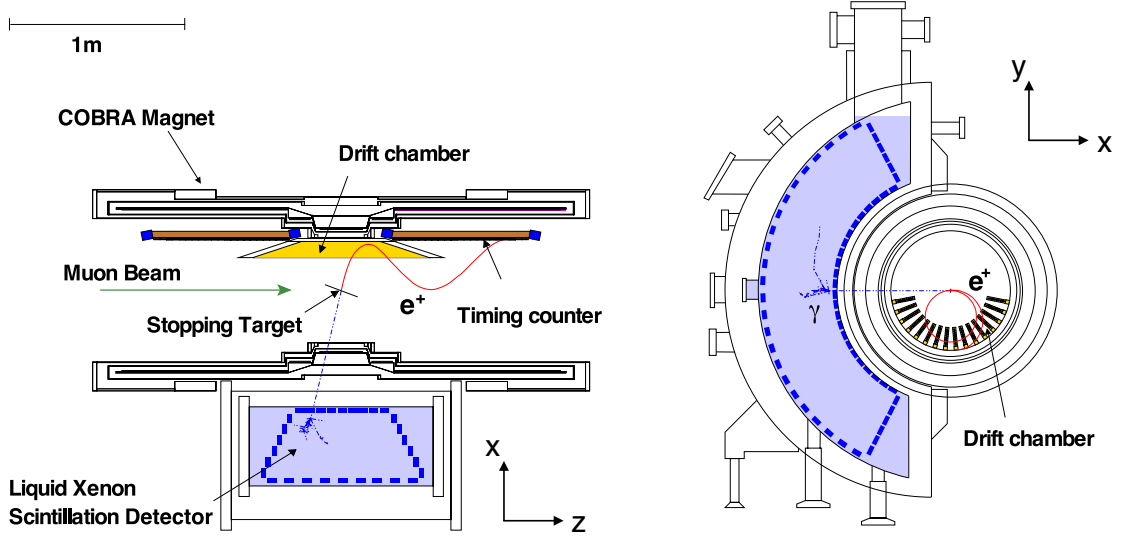
FIGURE 3.1. The MEG detector.

red curve in Figure 3.1 follows a helical trajectory inside the detector under a magnetic field produced by the COBRA magnet. Drift chamber (DCH) and timing counter (TC) detect the e$^+$. The γ shown in blue broken line goes into Liquid Xenon Scintillation Detector (LXe).

Figure 3.2 shows the $\mu^+ \to e^+ X, X \to \gamma\gamma$ signal (MC) with the MEG detector. The e$^+$ shown in the blue curve in Figure 3.2 follows a helical trajectory like the $\mu^+ \to e^+\gamma$ signal. The mediated particle, X is not shown in the figure, but it decays into 2γs shown in yellow line.

We define a cylindrical coordinate (r, ϕ, θ). The origin is set to the center position of the COBRA magnet and the z-axis is defined along the beam direction. The direction of $\phi = 0$ is defined as the opposite of the center of LXe. In the corresponding Cartesian coordinate, the x-axis points to the opposite to the center of LXe while the y-axis points to the upwards of the detectors.

The angular acceptance is defined by LXe acceptance, which corresponds to ϕ between $\frac{2}{3}\pi$ and $\frac{4}{3}\pi$ and $|\cos\theta| < 0.35$. Total acceptance is ~11%. The e$^+$ detector is designed to accept e$^+$ (52.8 MeV) from $\mu^+ \to e^+\gamma$ when γ (52.8 MeV) from $\mu^+ \to e^+\gamma$ is in the LXe acceptance. An e$^+$ from the MEx2G decay, however, has lower energy than 52.8 MeV depending on the mass of X, resulting in less acceptance because the TC position is optimized to detect energetic e$^+$(See Section 7.6.4).

The MEG data-taking started in 2008 and finished in 2013. No evidence of the decay was found, leading to a new upper limit on the branching ratio of 4.2×10^{-13}(90%C.L.) in 2016 [6]. From the next section, each apparatus used in the MEG experiment is reviewed.
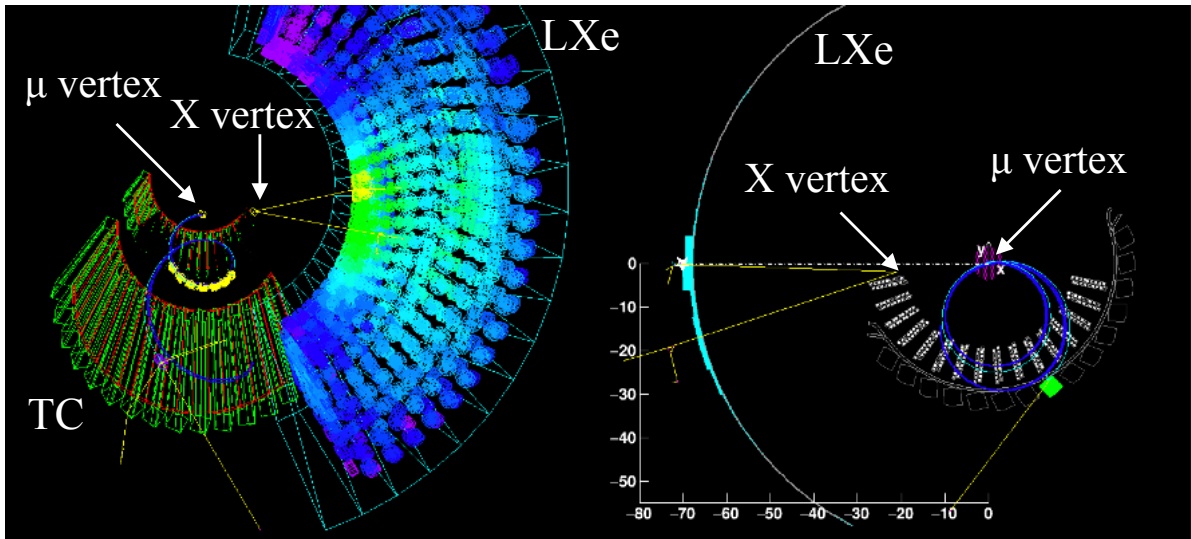
FIGURE 3.2. An example of the MEx2G signal event (MC).
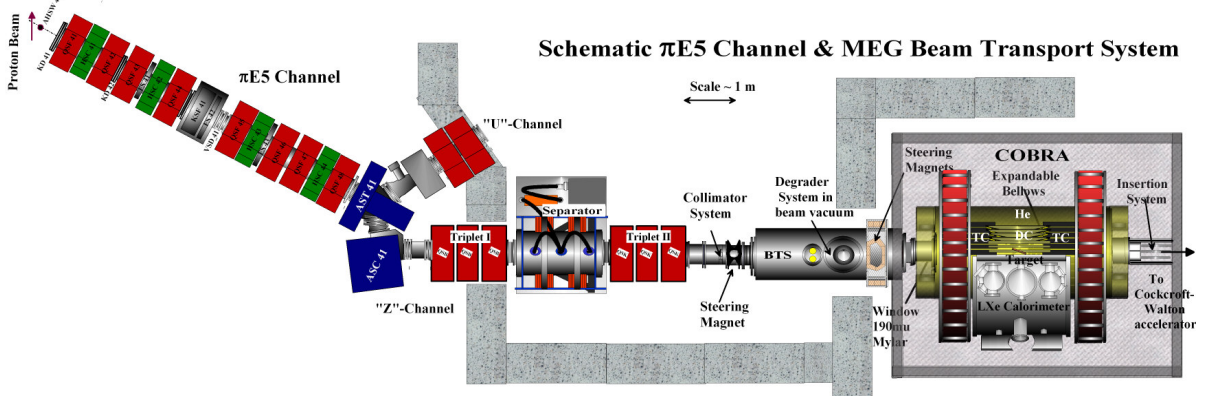
FIGURE 3.3. πE5 beamline.

3.2 The muon beam

PSI provides the highest intensity continuous muon beam in the world. A continuous beam is required to suppress the number of accidental background events, which is proportional to a square of the beam rate. The proton cyclotron of PSI provides a 590 MeV proton beam with a current of 2.2–2.4 mA, 1.4 MW. The beam structure is originally pulsed (RF: 50.7 MHz). In the MEG dedicated beamline, μ^+ beam with beam rate of $3 \times 10^8 \mu^+/s$ at 28 MeV/c in a momentum bite of 5–7% is available, but we reduced the beam rate down to $3 \times 10^7 \mu^+/s$ for stable detector operation. We will make the most of its capability in the MEG II experiment [7].

3.2.1 πE5 Channel

The proton beam passes through one of the pion production targets, Target E, where the beam energy is reduced to 570 MeV. The target is made of 4 cm long graphite. 70% of the beam is recaptured. μ^+s are produced at this target and guided to the MEG experimental area called the πE5 channel. Figure 3.3 shows a schematic view of πE5 channel. We use μ^+ not μ^- to avoid forming muonic atoms at the stopping target. As shown in the leftmost part of Figure 3.3, the muon beam is extracted with a slant angle of 165° from the original proton beam direction. This, so-called a surface muon beam [74], is produced through $\pi^+ \to \mu^+ \nu_\mu$ from pions stopped at the surface of Target E. The extracted muon beam is fully polarized ($P_{\mu^+} = -1$) and it is depolarized down to $P_{\mu^+} = -0.86 \pm 0.02(\text{stat})^{+0.05}_{-0.06}(\text{sys})$ [75] before it reaches the stopping target. The beam can be assumed to be continuous owing to the long lifetime of muon ($\sim 2\mu$s).

Quadrupole and sextupole magnets are used to guide the beam to the MEG experimental area (the right half of Figure 3.3 after a shielding wall made of concrete). There exist 10 times larger e^+s than μ^+s just after the shielding wall. We remove these e^+s using a Wien-filter ("Separator"),
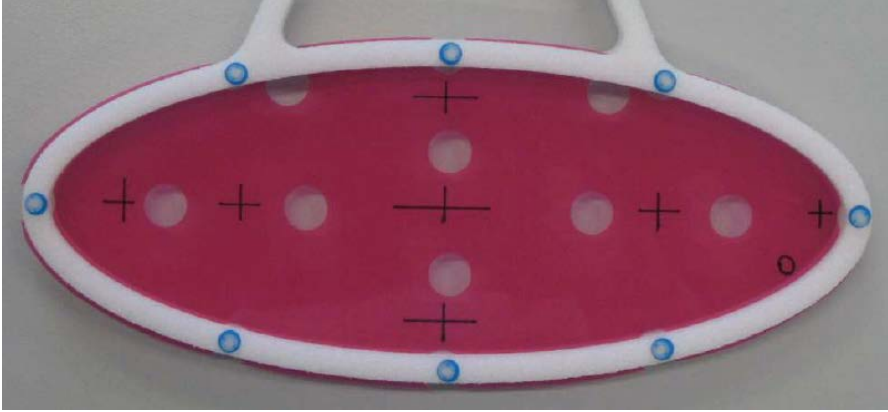
FIGURE 3.4. The stopping target. Markers and holes on the target are used to measure the shape and the position of the target.

quadrupole magnets (Triplet II), and a collimator made of lead. A separation power of 8.1 σ_μ from the beam originated e$^+$s is achieved after these systems. Finally, the superconducting Beam Transport Solenoid (BTS) is used to couple the beamline to the MEG detector. To directly couple to the beamline, the inside of the BTS is evacuated. A 300 μm thick Mylar® degrader is placed at the center of BTS to reduce the muon momentum to the low level at which the muon beam stops at the target. A 0.36 T magnetic field is applied for focusing. Gaussian beam-spot profile at the target has $\sigma_{x,y} \sim 10$ mm.

π^- beam is also available in this beamline and it is used for calibration as explained in Section 3.6. The π^- beam is tuned at 70.5 MeV/c to produce high energy γs through the charge exchange reaction.

3.2.2 Stopping target

The stopping target shown in Figure 3.4 (before installation) and in Figure 3.5 (after installation). It is made of a 205 μm thick polyethylene and polyester sheet (density: 0.895 g/cm^3). It has an elliptical shape with semi-major and semi-minor axes of 10 cm and 4 cm, respectively. It is placed on the beam axis at a slanted angle of 20° with respect to the beam direction. This geometry enables us to have a long length for incoming muons (to stop more than 80 % of them) and a short length for outgoing particles (to reduce multiple scattering and sources of background) at the same time.
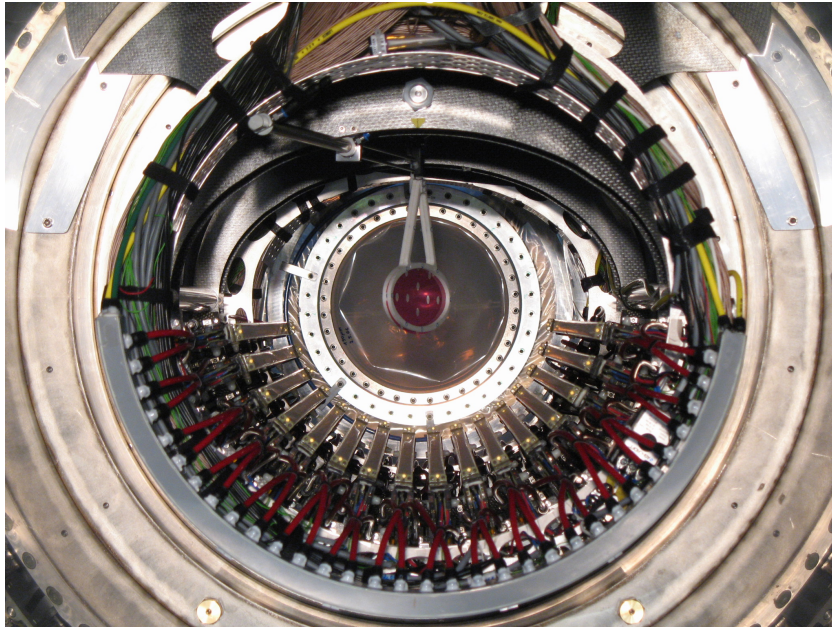
FIGURE 3.5. Stopping target and Drift chamber after installation.

(a)

(b)
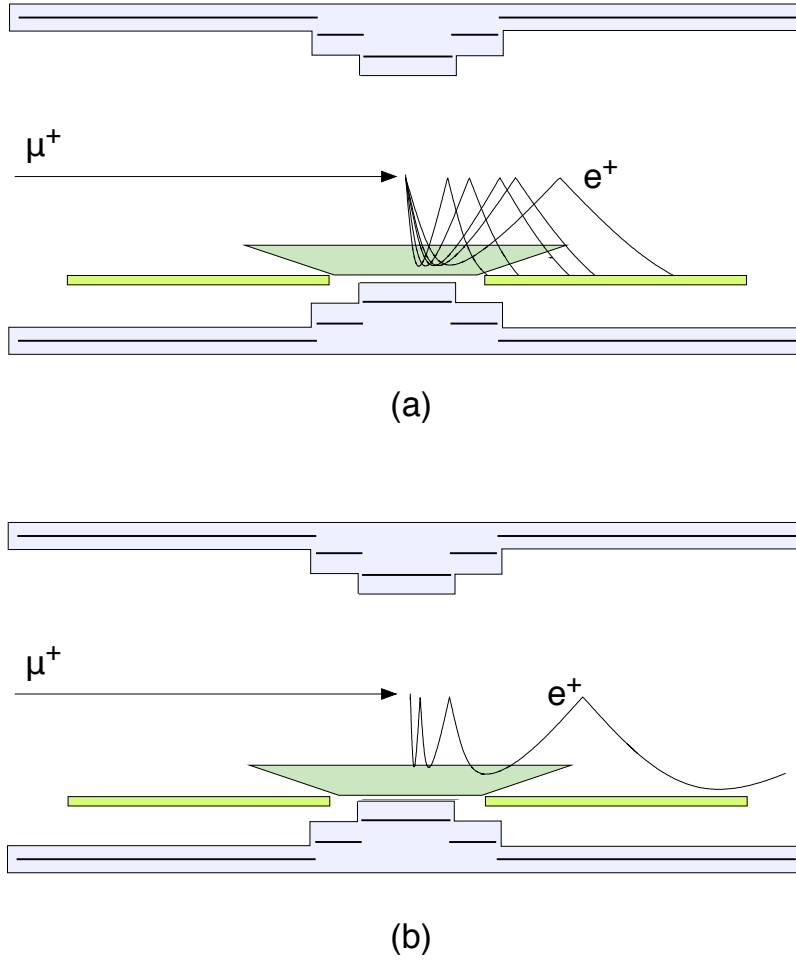
FIGURE 3.6. Cross sectional view of the COBRA magnet. See texts in detail.

3.3 Positron spectrometer

The MEG positron spectrometer consists of three components: COBRA magnet, Drift Chamber, and Timing Counter.

3.3.1 COBRA magnet

COBRA [76] stands for COnstant Bending RAdius. It is made of a superconducting magnet with 3 different radii. It generates a gradient magnetic field of 1.27 T at the center and 0.49 T at both edges. The characteristics are schematically shown in Figure 3.6. e⁺s emitted from the target follow helical trajectories under the magnetic field. The diameters of the trajectories depend on e^+ momenta independent of their emission angles thanks to the gradient field (Figure 3.6(a)). This means that we can select e⁺s with a specific momentum range by placing detectors in a

specific radius range; e$^+$s whose momenta are more than ~45 MeV come into the acceptance region of timing counter.

Furthermore, the gradient field prevents e$^+$s emitted to directions orthogonal to the beam direction from staying inside the spectrometer. This results in the suppression of hit rates at the detector.

The thickness of the center part of COBRA is 0.2 radiation length to minimize the effect on γ. 85% of the signal γs penetrate COBRA and reach the γ detector. Since the performance of photomultiplier tubes (PMT) deteriorates under a magnetic field, a compensation coil for COBRA is placed outside the detector, resulting in a reduction of the leak magnetic field around the LXe detector down to 50 Gauss.

3.3.2 Drift Chamber

Drift CHamber (DCH) [77] is designed to track e$^+$s. DCH is composed of 16 independent modules as shown in Figure 3.5. Each module has a trapezoid shape with base lengths of 40 cm (larger radius) and 104 cm (smaller radius, close to the stopping target). The directions of base lengths are parallel with the beam direction. These modules are installed along the bottom half of the circle at 10.5° interval. DCH covers the azimuthal region between 191.25° and 348.75° and the radial region between 19.3 cm and 27.9 cm. It has a two-layered structure and wires in each layer are stretched in the axial direction (beam direction). As shown in Figure 3.7 (a), the distance between adjacent cells is 9 mm and these cells are staggered among layers. Thanks to these moduled structures and the gradient magnetic field, DCH is operational under a high rate environment. The e$^+$ hit rate at the most inner part is suppressed down to 10 kHz while the original muon beam rate is 30 MHz. DCH is composed of low mass materials and helium-based chamber gas (He: C_2H_2 = 1: 1) is used to suppress multiple scattering. 2.0×10^{-3} radiation length is achieved for the $\mu^+ \rightarrow e^+\gamma$ signal e$^+$ (52.8 MeV).

As mentioned above, wires are stretched along the beam direction and position resolution along the beam axis is not good. To get a better position resolution along the beam axis, vernier pads are placed in each module. Figure 3.7 (b) shows how the pad works. The vernier pad has periodic structures and the charge ratio between cathode readouts depends on hit position inside the single periodic structure. Anode readout determines in which period the hit exists and vernier readouts determine in which position inside the period the hit exists.

3.3.3 Timing Counter

Timing Counter (TC) [78, 79] is designed to precisely measure the e$^+$ hit time. Figure 3.8 shows a schematic view of TC. It consists of two parts: scintillation bars and scintillation fibers. 15 scintillation bars are placed outside the scintillation fibers at each side of the stopping target. The scintillation bars are made of $4 \times 4 \times 80$ cm^3 plastic scintillator (Bicron BC404) and PMTs attached to both edges of the bars. Since the detector is placed in the high magnetic field, fine-mesh
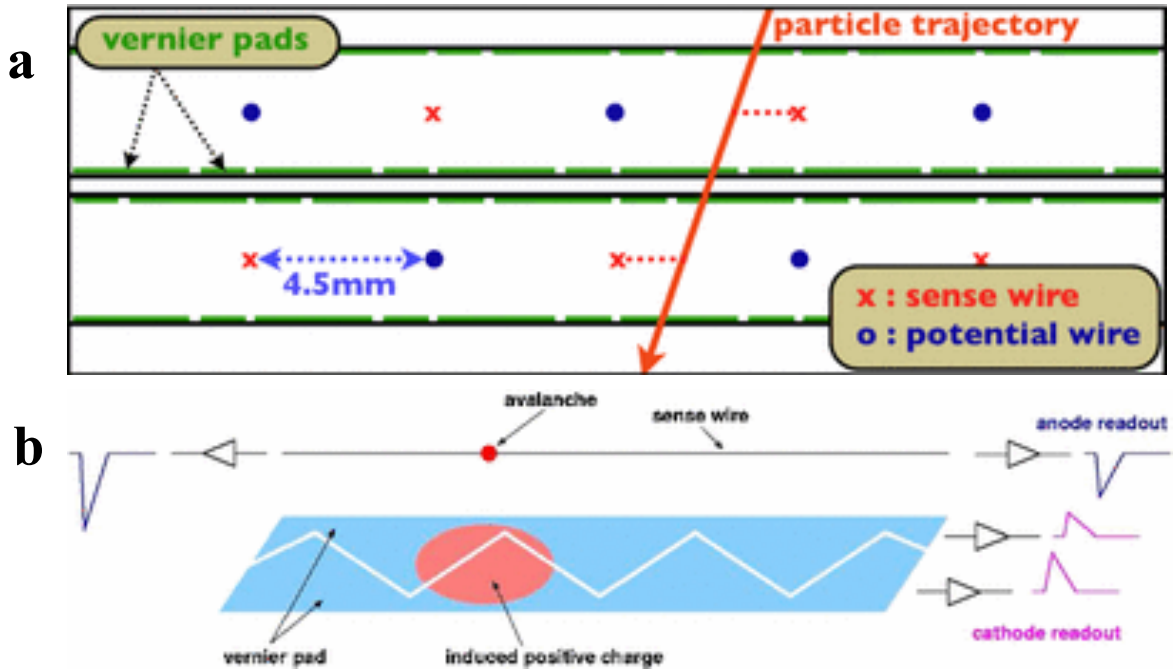
FIGURE 3.7. (a) Cross sectional view of the DCH module. (b) Schematic view of the varnier pad method. See text in detail.
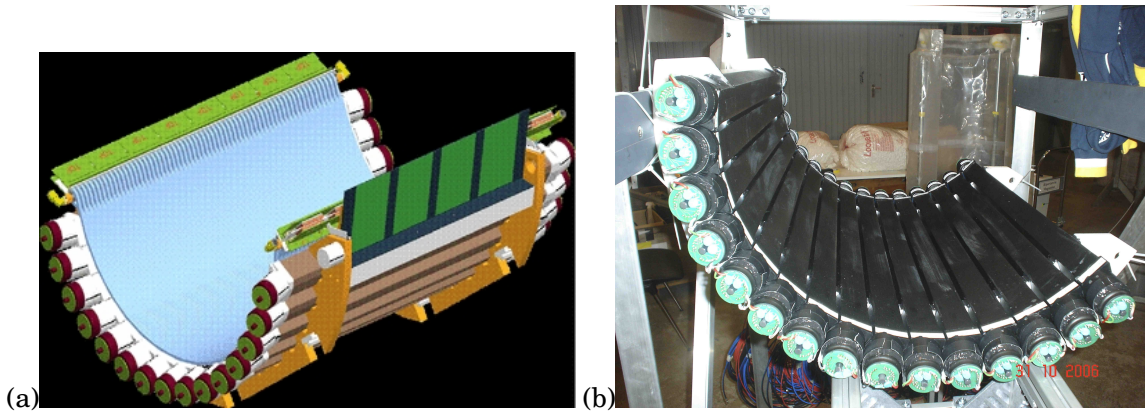
Figure 3.8: (a) Schematic view and (b) picture of TC. Scintillation fibers are not installed in the picture.

2" PMTs are used. The scintillation fibers readout by avalanche photodiodes were originally developed to get independent position information and improve time resolution, but they did not work in the experiment. TC is covered with a N_2 bag to prevent chamber gas (He) from entering TC PMTs.
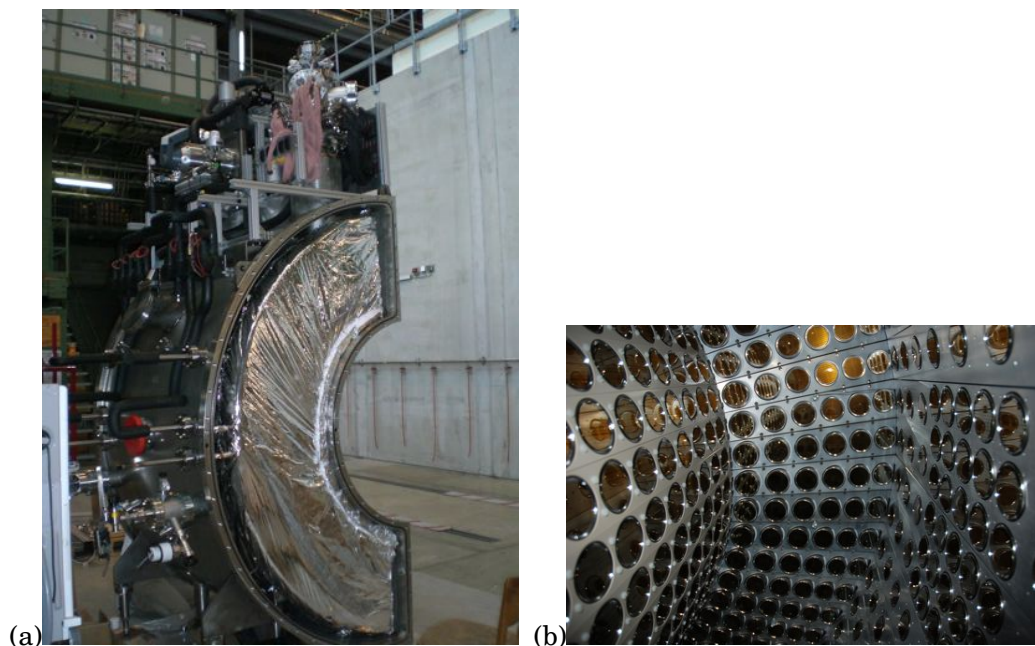
Figure 3.9: (a) Outer and (b) inner view of LXe

3.4 Liquid Xenon gamma calorimeter

Liquid Xenon gamma calorimeter (LXe)[80, 81] is designed to detect γ using liquid xenon (active volume 800 L) and the scintillation light are readout by 846 PMTs. Outer and inner views of LXe are shown in Figure 3.9. The interaction position and time inside LXe are reconstructed from the light distribution detected with the PMTs. The energy is reconstructed by summing up all the scintillation light detected with the PMTs. Liquid xenon is selected as a scintillation medium from the following reasons: LXe is operational under a high rate environment thanks to its fast time response (decay time: 45 ns). This also contributes to a good time resolution. The radiation length of liquid xenon is short (2.8 cm) enough to stop γ inside the active volume. Light yield is also large (75% of NaI). Furthermore, liquid scintillator has advantages over a solid one in terms of uniformity and purity. Liquid scintillators can be purified if needed. These characteristics contribute to a good energy resolution.

Problems to use liquid xenon are the wavelength of the scintillation and low-temperature operation. The wavelength of scintillation light from liquid xenon ($\sim$178 nm) is in the region of Vacuum Ultra-Violet (VUV), which is a shorter wavelength than that can be detected with usual PMTs. To keep xenon in the liquid phase, the detector should be operated under $\sim$165 K. The VUV-sensitive and low-temperature-operational PMT (R9869) was newly developed in cooperation with Hamamatsu Photonics K.K.

γ interacts with xenon atoms through three different processes: scattering, photoelectric absorption, and pair production. The pair production is the dominant process for γ around

MeV scale in LXe [82] and it makes electromagnetic shower in LXe. During the process, excited atoms (Xe^*) or ions (Xe^+) are produced. Then, the scintillation mechanism has two different processes [83, 84]: The first process is called a self-trapping process and it is given by

$$Xe^* + Xe + Xe \rightarrow Xe_2^* + Xe,$$
$$Xe_2^* \rightarrow 2Xe + h\nu, \tag{3.1}$$

where Xe_2^* is called excimer, excited state of molecule, and $h\nu$ is a VUV photon. These photons are emitted from the excimer. The excimer has two molecular states, singlet and triplet, which have shorter (4.2 ns [85]) and longer (22 ns [85]) decay times, respectively.

The second process is called a recombination process and it is given by

$$Xe^+ + Xe \rightarrow Xe_2^+,$$
$$Xe_2^+ + e^- \rightarrow Xe^{**} + Xe,$$
$$Xe^{**} \rightarrow Xe^* + heat,$$
$$Xe^* + Xe + Xe \rightarrow Xe_2^* + Xe,$$
$$Xe_2^* \rightarrow 2Xe + h\nu. \tag{3.2}$$
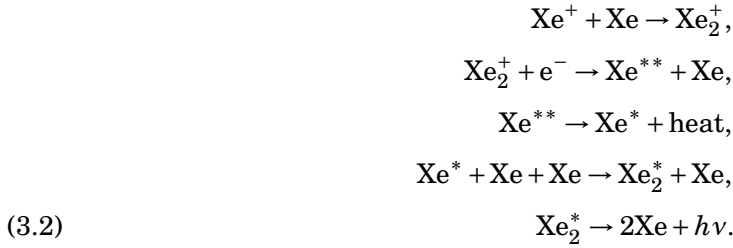
This process is slower (45 ns [85]) than the first one. In both cases, scintillation photons are emitted from the excited dimers at the same process (Equation (3.1) and Equation (3.2)), not from the excited atom itself. Owing to this characteristic, the scintillation photons are not absorbed by Xe itself.

Figure 3.10 shows a schematic view of the LXe detector. It has a C-shaped structure and distance between the center of stopping target and the inner surface of LXe is 67.85 cm. The depth of LXe is 38.5 cm, which corresponds to ∼14 radiation length. Green markers in Figure 3.10 show PMTs. These PMTs are directly put in the liquid xenon. They are placed in six faces as shown in the development view of the detector in Figure 3.11.
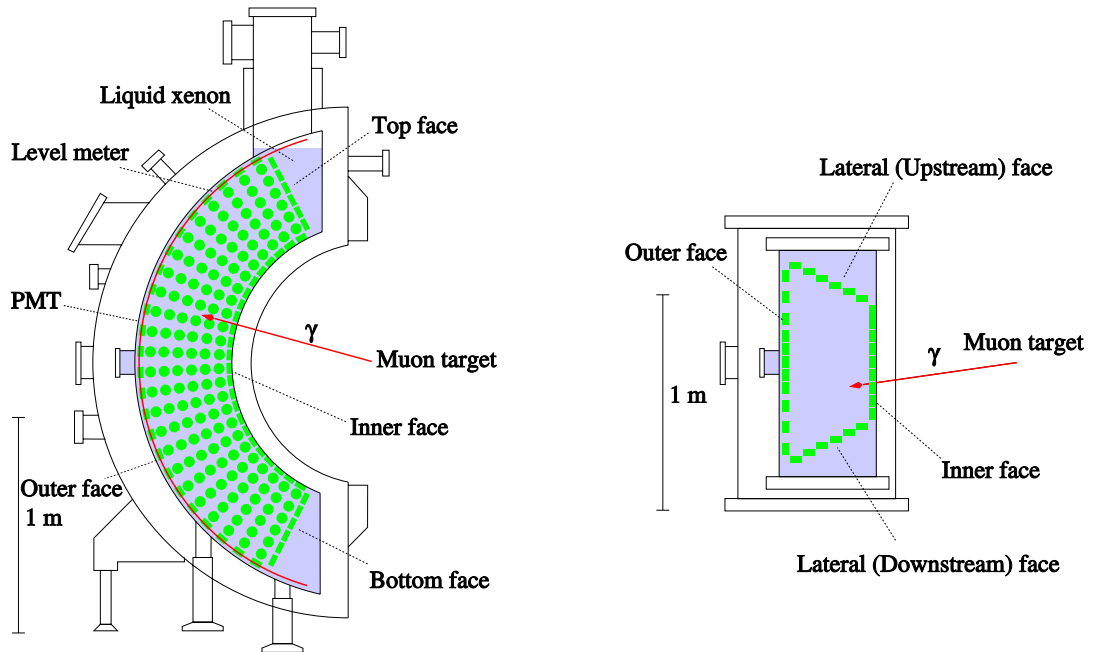
FIGURE 3.10. Schematic view of the LXe detector.
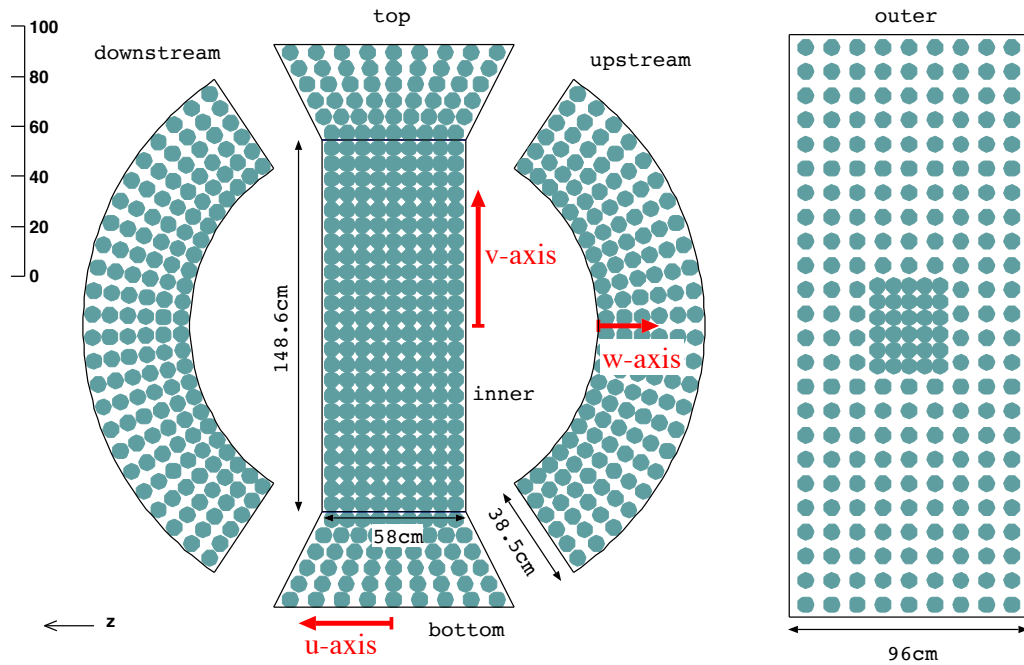
FIGURE 3.11. Development view of the LXe detector (from [86]).

3.5 DAQ and Trigger

3.5.1 Front-end electronics

One of the distinctive features of the MEG experiment is to record all the waveforms from the detectors. We used Domino Ring Sampler v4 (DRS4) chip [87] for digitizing and recording of the waveforms. It enables us to apply complicated algorithms to the acquired data and it makes it easy to reanalyze the data when we change algorithms or parameters. The sampling frequencies are set to 1.6 GHz for TC and LXe and 0.8 GHz for DCH. This lower frequency value for DCH is determined to match the drift velocity.

3.5.2 Trigger

The trigger rate should be kept below 10 Hz in order not to lose events. Table 3.1 shows a list of the most important triggers in the MEG experiment. Note that the dedicated trigger for the $\mu^+ \to e^+ X, X \to \gamma\gamma$ events was not prepared. Thus we use $\mu^+ \to e^+ \gamma$ triggered data in the $\mu^+ \to e^+ X, X \to \gamma\gamma$ search analysis (Section 3.9). We developed 2 different types of triggers. One is used in physics data taking and the other is used in calibration data taking. In the physics data taking, several triggers are mixed with their own prescaling factors. The main trigger is the MEG trigger, which is shown in the first line of Table 3.1 with a prescaling of 1. The trigger names starting with MEG uses the following observables:

- γ energy

- Time difference between e^+ and γ

- Relative direction of e^+ and γ

A total charge of PMTs in the LXe detector is used to select high energy γ events. An e^+ time is calculated from TC. The time difference is used to select coincident events. The relative direction is used to select back-to-back events, which is called direction matching (DM) trigger. To calculate the relative direction, the PMT that has the largest amount of scintillation photons is used at the γ side assuming the emission point locates at the center of the target. Since DCH information is too slow to be used in the trigger, an e^+ hit position at TC is used making use of the fact that TC bar IDs and z position correlate with the e^+ emission angle. As we discuss in Section 7.6.6, this DM trigger loses the MEx2G signal. Unlike the $\mu^+ \to e^+ \gamma$ analysis, the MEx2G signal has 2γs with a finite opening angle and this makes the signal unlikely to be back-to-back events. This $\mu^+ \to e^+ \gamma$ dedicated trigger loses 10%–50% of the MEx2G signal events depending on the mass of the mediated particle.

Table 3.1: The most important triggers in the MEG experiment

name	Prescaling factor	Selection criteria
MEG	1	$E_\gamma > 45\,\text{MeV} \wedge |\Delta t_{e\gamma}| < 10\,\text{ns} \wedge$ narrow DM
MEG Low E_γ	50	$E_\gamma > 40\,\text{MeV} \wedge |\Delta t_{e\gamma}| < 10\,\text{ns} \wedge$ narrow DM
MEG Wide Wide DM	500	$E_\gamma > 45\,\text{MeV} \wedge |\Delta t_{e\gamma}| < 10\,\text{ns} \wedge$ wide DM
MEG Wide $|\Delta t_{e\gamma}|$	200	$E_\gamma > 45\,\text{MeV} \wedge |\Delta t_{e\gamma}| < 20\,\text{ns} \wedge$ narrow DM
Radiative decay	1000	$E_\gamma > 45\,\text{MeV} \wedge |\Delta t_{e\gamma}| < 10\,\text{ns}$
LXe alone	-	Threshold on E_γ
DCH alone	10^7	DCH hit multiplicity
TC alone	10^7	TC hit multiplicity
π^0 decay	-	Coincidence between LXe and NAI/BGO
Pedestal	20000	Clock for pedestal
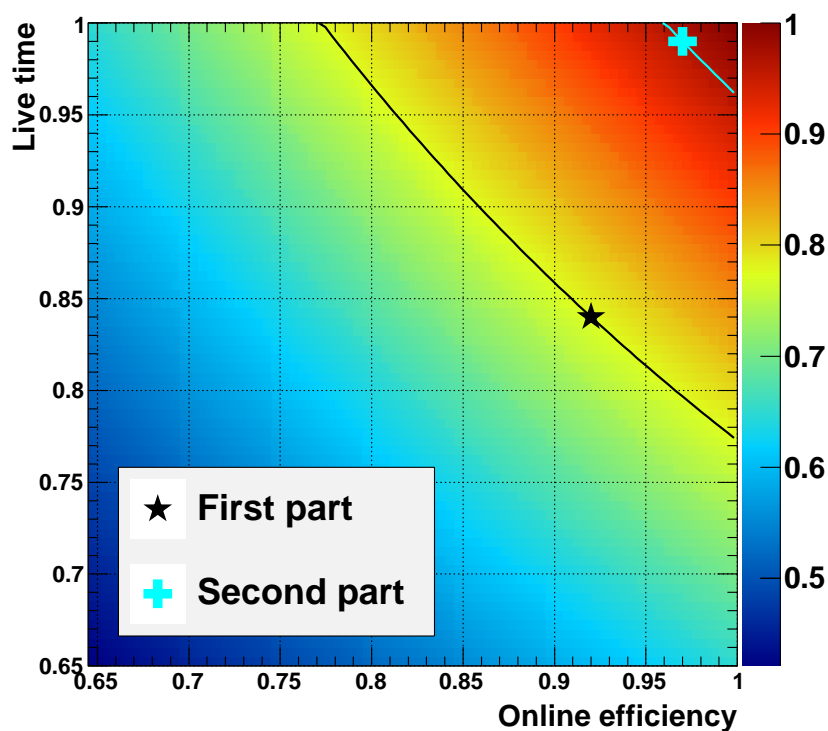
FIGURE 3.12. DAQ efficiency during different run periods. First part corresponds to runs in 2009, 2010 and second part corresponds to runs in 2011–2013 (from [6]).

3.5.3 DAQ

DAQ systems should have long live time and high online efficiency as much as possible. DAQ efficiency is defined as the product of these two items. Figure 3.12 shows a contour plot of the DAQ efficiency. A multiple-buffer scheme to efficiently retain waveforms was implemented in 2011 and the DAQ efficiency was improved from 75% to 97%.

3.6 Calibration tools

Calibration tools in the MEG experiment are summarized in Table 3.2. Processes in bold font are explicitly used in this analyais. In this section, we focus on the Cockcroft-Walton accelerator and charge exchange reaction. Other calibration tools are mentioned in this thesis as necessary. Detailed descriptions of all the calibration tools can be found in Ref. [73].

Table 3.2: Calibration tools in the MEG experiment

Name	Process	Energy	Main Purpose	Frequency
Cosmic rays	$\mu^{\pm}$ from atmospheric showers	Wide spectrum $\mathcal{O}$(GeV)	LXe-DCH relative position	annually
			DCH alignment	
			TC energy and time offset calibration	on demand
Charge exchange (CEX)	$\pi^- p \rightarrow \pi^0 n$ $\pi^0 \rightarrow \gamma\gamma$	55, 83, 129 MeV photons	LXe purity	annually
			LXe energy scale/resolution	
			2γ calibration	
Radiative μ−decay (RMD)	$\mu^+ \rightarrow e^+\gamma\nu\bar{\nu}$	photons >40 MeV, positrons >45 MeV	LXe-TC relative timing	continuously
Normal μ−decay (Michel)	$\mu^+ \rightarrow e^+\nu\bar{\nu}$	52.83 MeV end-point positrons	DCH energy scale/resolution	continuously
			DCH and target alignment	
			Normalisation	
Mott positrons	e^+ target $\rightarrow e^+$ target	≈ 50 MeV positrons	DCH energy scale/resolution	annually
			DCH alignment	
Cockcroft-Walton (CW) accelerator	^{7}Li$(p,\gamma)^8$Be	14.8, 17.6 MeV photons	LXe uniformity/purity/2γ calibration	weekly
	^{11}B$(p,\gamma)^{12}$C	4.4, 11.6, 16.1 MeV photons	TC interbar/ LXe–TC timing/2γ calibration	weekly
Neutron generator	^{58}Ni$(n,\gamma)^{59}$Ni	9 MeV photons	LXe energy scale	weekly
Radioactive source	^{241}Am$(\alpha,\gamma)^{237}$Np	5.5 MeV α's, 56 keV photons	LXe PMT calibration/purity	weekly
Radioactive source	^{9}Be$(\alpha_{^{241}\text{Am}},n)^{12}$C* ^{12}C$^\star(\gamma)^{12}$C	4.4 MeV photons	LXe energy scale	on demand
LED			LXe PMT calibration	continuously
Laser tracker			detector alignment	on demand
3D scan			target alignment	on demand

34
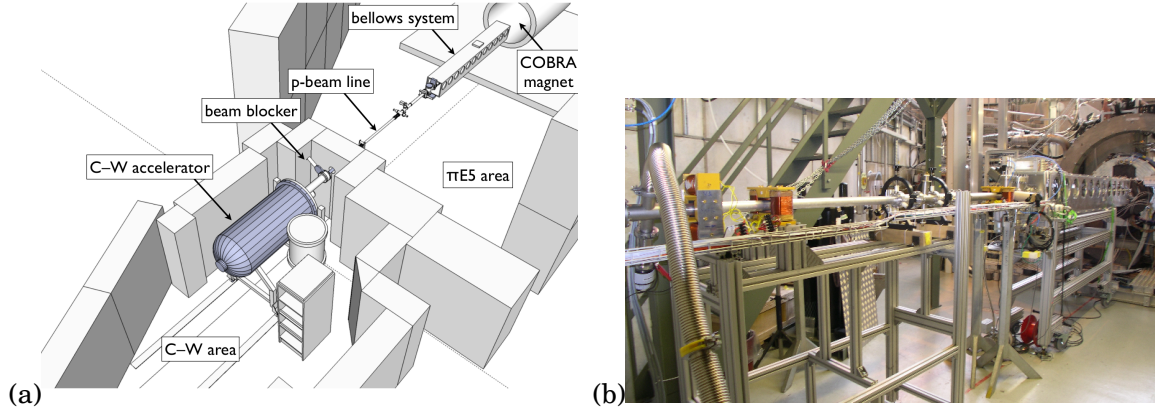
Figure 3.13: (a) Schematic layout of CW area and πE5 area, (b) CW beamline inside πE5 area.
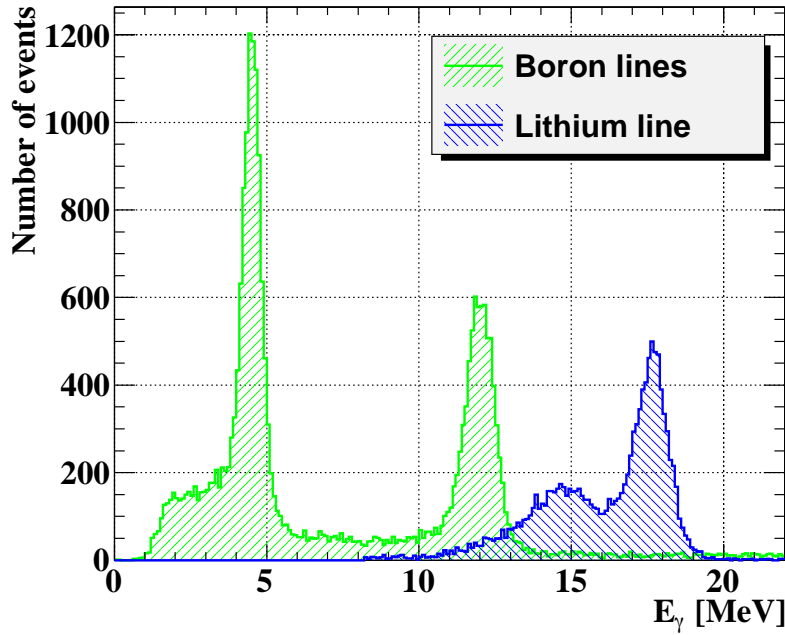
Figure 3.14: CW spectrum [73]

3.6.1 The Cockcroft-Walton accelerator

A dedicated Cockcroft–Walton (CW) accelerator [88] is installed next to the πE5 area as shown in Figure 3.13 (a). Nuclear reactions excited by protons from CW are used to calibrate and monitor LXe energy scale and relative timing between LXe and TC.

Two nuclear reactions are used. $^{7}_{3}\mathrm{Li}(\mathrm{p},\gamma)^{8}_{4}\mathrm{Be}$ produces 17.6 MeV and 14.6 MeV γ. The higher one is monochromatic and the lower one has broader resonance. $^{5}_{11}\mathrm{B}(\mathrm{p},\gamma)^{12}_{6}\mathrm{C}$ produces 11.7 MeV and 4.4 MeV γ. These two γs are emitted simultaneously. Thus they are used not only for energy

calibration but also for time calibration between different detectors: LXe and TC.

A lithium tetraborate ($Li_2B_4O_7$) target is installed to excite both reactions. A proton current is tuned at the threshold where each reaction occurs. An E_γ low threshold trigger is used for $^7_3Li(p,\gamma)^8_4Be$ events and a time coincidence (LXe and TC) trigger is used for $^{11}_5B(p,\gamma)^{12}_6C$ events. Figure 3.14 shows measured γ energy spectra by the CW calibration.

3.6.2 The π^-p charge exchange reaction

We use the Charge EXchange (CEX) reaction $\pi^- + p \rightarrow \pi^0 + n$ to calibrate the LXe detector. The following performances were evaluated with the CEX calibration:

- Determination of the energy scale

- γ energy resolution

- γ time resolution

- γ position resolution

- γ efficiency

A π^- beam instead of μ^+ beam is injected into a dedicated target for the CEX calibration, a liquid-hydrogen (LH_2) target instead of the μ^+ stopping target. The LH_2 is used because of its higher proton density[1]. The π^- captured by protons inside the LH_2 target reacts in two different ways: one is the charge exchange reaction (CEX), $\pi^- + p \rightarrow \pi^0 + n$ followed by $\pi^0 \rightarrow \gamma + \gamma$. The other is the radiative capture reaction (RC), $\pi^- + p \rightarrow n + \gamma$. The relative probability (Panofsky ratio) between these reactions was measured to be

$$(3.3) \qquad P = \frac{\Gamma\left(\pi^- + p \rightarrow \pi^0 + n\right)}{\Gamma\left(\pi^- + p \rightarrow n + \gamma\right)} = 1.546 \pm 0.009$$

in Ref. [90]. The energy of γ in the RC reaction is ~ 129 MeV, which is higher than our use and the γs in CEX reaction are mainly used for the LXe calibration. 2γs from the π^0 are emitted back-to-back and have the energy of ~ 67.5 MeV at the rest frame:

$$(3.4) \qquad E_\gamma^{\text{rest}} = \frac{m_{\pi^0}}{2} \simeq 67.5 \text{MeV}.$$

In the lab frame, the γs have

$$(3.5) \qquad E_{\gamma_{1,2}} = \gamma \frac{m_{\pi^0}}{2}\left(1 \pm \beta \cos\theta^{\text{rest}}\right),$$

where β is the π^0 velocity ($\beta \simeq 0.2$). The energy depends on its emission angle θ^{rest} in the rest frame. Therefore, 2γs from the CEX reaction have energy from E_{min} to E_{max}:

$$(3.6) \qquad E_{\text{min}} = \gamma \frac{m_{\pi^0}}{2}(1 - \beta) \simeq 54.9 \text{MeV},$$

$$(3.7) \qquad E_{\text{max}} = \gamma \frac{m_{\pi^0}}{2}(1 + \beta) \simeq 82.9 \text{MeV}.$$

---

[1]Detailed considerations on the CEX target can be found in Ref. [89]
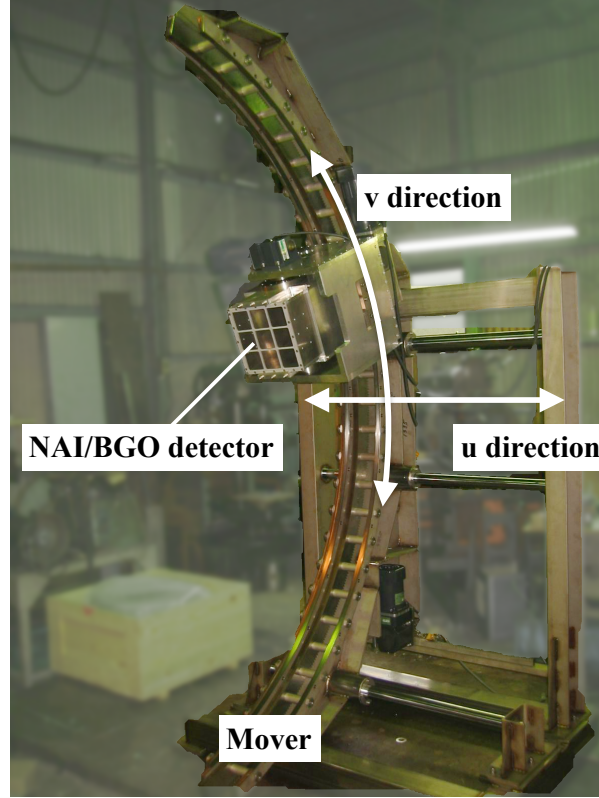
Figure 3.15: NAI/BGO mover placed at the opposite side of the LXe detector.

The nearly monochromatic γ energies of 54.9 MeV and 82.9 MeV are available by selecting back-to-back events. For this purpose, a NAI (in 2009 and 2010) or BGO (in 2011 – 2013) calorimeter was installed at the opposite side of the LXe detector when the CEX calibration was performed. The calorimeter is mounted on a moving stage shown in Figure 3.15. The detector can move in the u and v directions defined in Figure 3.15 to take back-to-back events. Figure 3.16 shows correlation of reconstructed energy between the LXe and BGO detector. One can get almost monochromatic energy of 54.9 MeV (82.9 MeV) in LXe by selecting 82.9 MeV (54.9 MeV) at the opposite side of the detector, BGO.

A timing detector (plastic scintillator plates) together with a lead converter is placed in front of the NAI/BGO detector for the evaluation of time resolution. A 1.8 cm thick lead collimator is placed right in front of the LXe detector to evaluate the position resolution.

The CEX calibration was performed only once per year because a frequent change of the target and beam setting is difficult. A dedicated trigger is fired by a coincidence between LXe and the NAI/BGO detector. It is used for energy calibration. Coincidence trigger between LXe and NAI/BGO detector is used for CEX data taking.
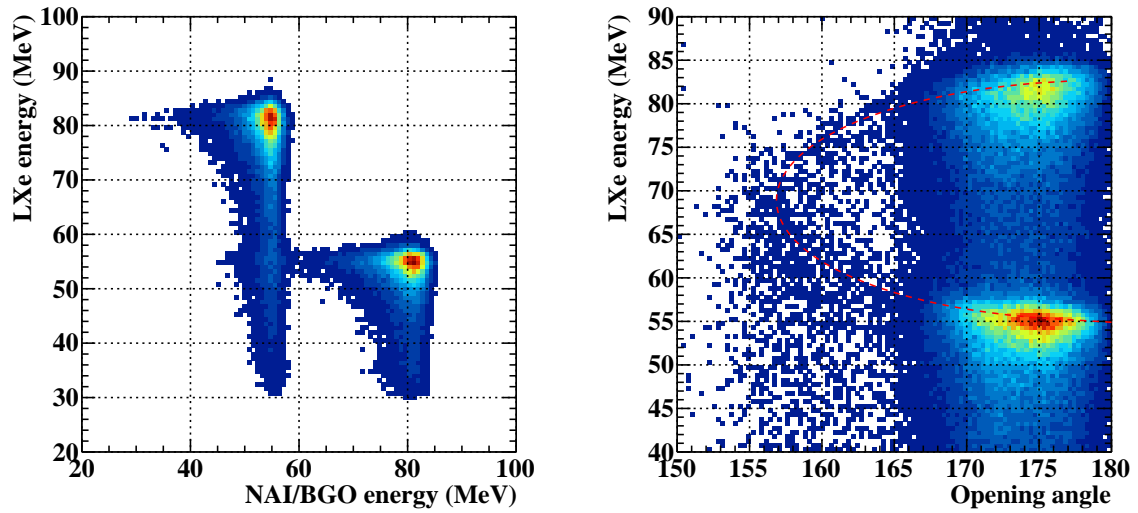
Figure 3.16: Left: LXe energy vs. BGO energy. Right: LXe energy vs. Opening angle. Back-to-back events are actively taken by moving the NAI/BGO detector.
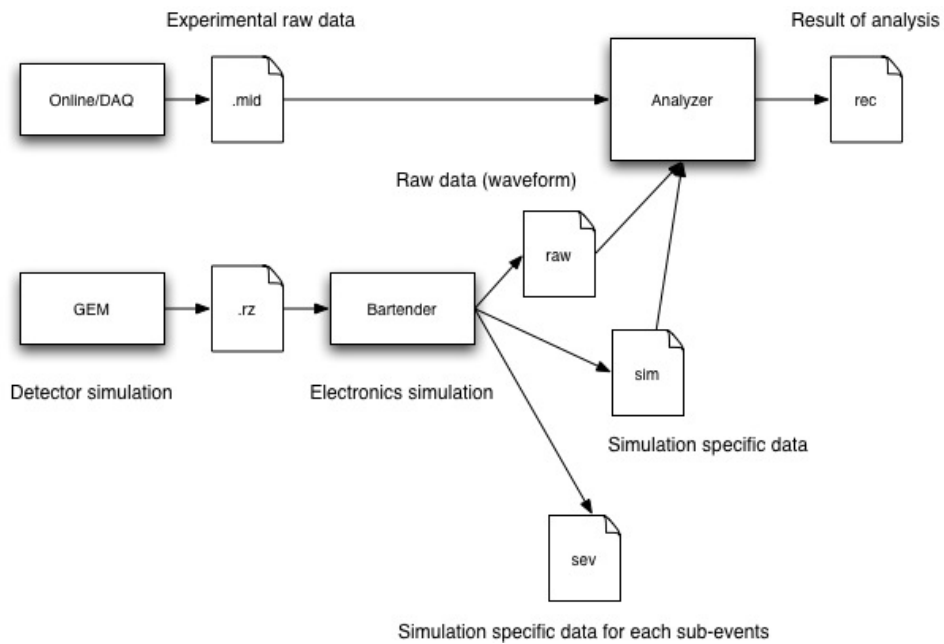
Figure 3.17: Analysis framework in MEG. See text in detail.

3.7 Simulation and analysis tools

Figure 3.17 shows the analysis framework in the MEG experiment and the bottom half corresponds to the Monte Carlo (MC) simulation framework.

MC starts from the **GEM** part in Figure 3.17. In this part, detector responses up to readouts

Table 3.3: The most important MC generations in the MEG MC framework. Particles written in brackets are not simulated in the generation.

MEG event	$\mu^+ \to e^+\gamma$
MEx2G event	$\mu^+ \to e^+X, X \to \gamma\gamma$
Michel e^+	$\mu^+ \to e^+(\nu\bar{\nu})$ [63]
RMD $e^+\gamma$	$\mu^+ \to e^+\gamma(\nu\bar{\nu})$ [63, 91]
CEX 2γ	$\pi^-p \to \pi^0(n), \pi^0 \to \gamma\gamma$
Boron	$^{11}\mathrm{B}(p,2\gamma)^{12}\mathrm{C}$
Lithium	$^7\mathrm{Li}(p,\gamma)^8\mathrm{Be}$

are simulated based on GEANT3 [92], which is an MC simulation tool written in FORTRAN77. The LXe detector simulation includes the propagation of scintillation photons with reflection, refraction, and polarization effects. GARFIELD [93] is used to simulate drift lines inside the DCH cell. Table 3.3 shows a part of primary physics events available in the MEG MC framework. Generated events are written in ZEBRA format developed at CERN (.*rz* files shown in Figure 3.17). Then, in the Bartender part, conversion of ZEBRA files to ROOT files, electronics simulation, and event mixing are performed. From this part, programs are written in C++. **Bartender** simulates readout electronics and makes waveforms (*raw* files). This raw data is encoded in the same way as the experimental raw data (.*mid* files). Both raw data are processed in the **Analyzer** and event reconstructions are performed. The results of the analysis are written in *rec* files. Parameters such as the geometry of systems and calibration factors depend on runs. We stored these parameters in the MySQL based database.

Further detailed descriptions of the framework are available from [73, 94, 95].

3.7.1 Pseudo 2γ event

Pseudo 2γ data To evaluate performance on the reconstruction, we need 2γ data. However, there is no 2γ data which can be used for the calibration or performance evaluation. Therefore we made pseudo 2γ data using the existing calibration data such as CW-B, CW-Li, and CEX. We select two events from these calibration data and they are overlaid to make pseudo 2γ events. Technically, the number of photoelectrons and photons are summed up.

Figure 3.18 shows an example of the formed pseudo 2γ events. The left two events show the original events coming from calibration data. The source of original events is one of the followings:

- 54.9 MeV, 82.9 MeV from CEX

- 17.6 MeV, 14.6 MeV from $^7_3\mathrm{Li}(p,\gamma)^8_4\mathrm{Be}$ in CW

- 11.7 MeV from $^{11}_5\mathrm{B}(p,\gamma)^{12}_6\mathrm{C}$ in CW

The right figure in Figure 3.18 is the overlaid event.
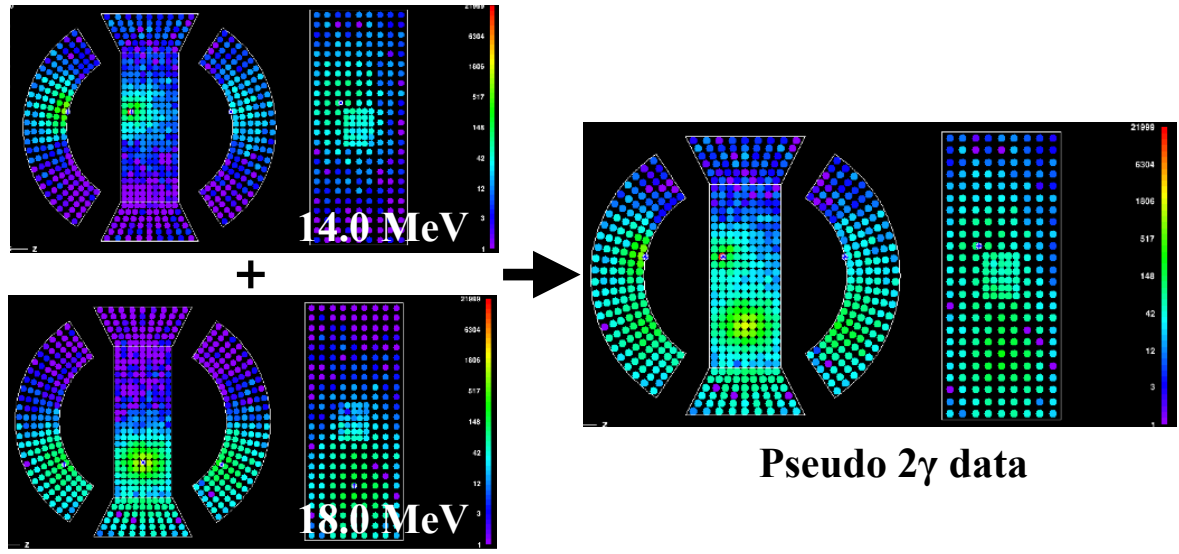
Figure 3.18: Pseudo 2γ data. The left 2 γs are combined to make the right event.
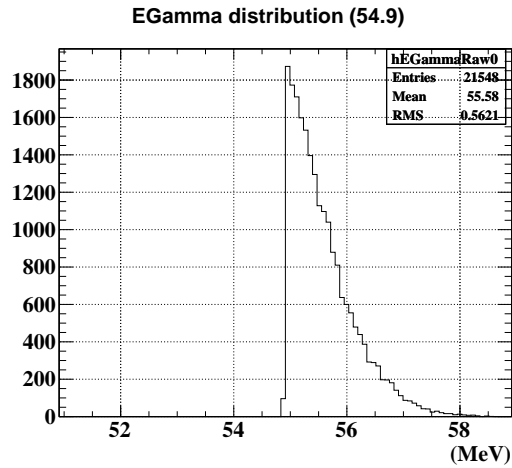
Figure 3.19: Initial energy distribution for CEX 54.9 MeV events. MC events used to make pseudo 2γ events are generated based on this distribution.

Pseudo 2γ MC   To compare data and MC, pseudo 2γ MC is generated with the same procedure as the pseudo 2γ data. The 11.7 MeV peak from CW-Boron and 17.6 MeV peak from CW-Li are generated as a monochromatic energy source. The 54.9 MeV peak from CEX is generated based on the finite energy distribution shown in Figure 3.19. The lower edge is 54.9 MeV. The higher tail is due to the finite angle resolution.

Table 3.4: Summary of detector resolutions in the $\mu^+ \to e^+\gamma$ search (from [96])

		year				
		2009	2010	2011	2012	2013
γ	$E_\gamma(\text{w} > 2\text{cm})(\%)$	1.8	1.9	1.6	1.6	1.6
	$E_\gamma(\text{w} < 2\text{cm})(\%)$	2.4	2.5	2.3	2.3	2.3
	t_γ (ps)	96	67	67	64	66
	u,v (cm)	5	5	5	5	5
	w (cm)	6	6	6	6	6
e^+	E_{e^+} (MeV)	0.31	0.31	0.30	0.34	0.33
	θ (mrad)	9.2	10.3	10.4	10.9	11.2
	ϕ (mrad)	8.5	9.5	9.4	10.1	10.4
	y (mm)	1.2	1.2	1.3	1.3	1.4
	z (mm)	2.3	2.9	3.0	3.2	3.3
combined	$t_{e^+\gamma}$ (ps)	143	126	117	119	111
	$\theta_{e^+\gamma}$ (mrad)	14.5	14.4	14.5	14.8	14.9
	$\phi_{e^+\gamma}$ (mrad)	9.5	9.8	9.5	9.5	9.9

3.7.2 MC with pedestal data

To reproduce the background γ events like the 2γ in MC, data taken with the pedestal trigger (Table 3.1) is used. We can randomly record data with the pedestal trigger. These data are overlaid with MC events on the photoelectrons-basis.

3.8 Detector performance in the $\mu^+ \to e^+\gamma$ search

Before we discuss event reconstruction performance for the $\mu^+ \to e^+X, X \to \gamma\gamma$ search, we summarize the performance for the $\mu^+ \to e^+\gamma$ search in Table 3.4. Detail methods to estimate these resolutions are summarized in Refs. [6, 73, 96].

3.9 Overview of the MEx2G search analysis

3.9.1 Analysis strategy

The MEx2G search analysis is a combination of blind, cut-counting, and maximum likelihood analysis. The signal region is defined beforehand and blinded to reduce the experimenter's bias. After applying optimized cut conditions, we count survived events. The number of survived events, normalization, and its uncertainty are simultaneously fitted to find the number of the signal using maximum likelihood fitting. The Feldman–Cousins method is used to get the confidence interval.

We make full use of resources in the $\mu^+ \to e^+\gamma$ search analysis in terms of the dataset and event reconstruction point of view as we describe in the following subsections.

3.9.2 Dataset

All the detectors of the MEG experiment were ready in 2007 and we performed engineering run. After fixing several problems found in the engineering run, physics data taking had started in 2008. From 2009 to 2013, data were accumulated as shown in Figure 3.20 except for the annual beam shutdown period. The data in 2008 was not used for the final analysis of the MEG analysis because of a discharge problem of the drift chamber, which resulted in a bad e^+ performance. It is not used for the MEx2G analysis, either. In total, $7.5 \times 10^{14}\,\mu^+$s were stopped on the target. The MEx2G analysis based on the first $1.8 \times 10^{14}\,\mu^+$s (2009 and 2010) were presented in [62]. In this analysis, we use the full dataset accumulated in five years.

In the MEG data taking, a dedicated trigger for $\mu^+ \to e^+X, X \to \gamma\gamma$ events was not prepared (Section 3.5.2). We use $\mu^+ \to e^+\gamma$ triggered events in the MEx2G search analysis. One of the $\mu^+ \to e^+\gamma$ trigger conditions requires back-to-back $e^+\gamma$ events, but it has more chance to lose the $\mu^+ \to e^+X, X \to \gamma\gamma$ signal for the larger mass of X. This inefficiency needs to be taken into account.

Then, events were preselected at the first stage of the $\mu^+ \to e^+\gamma$ decay analysis, requiring at least one positron track and the time difference between LXe and TC to be $-6.9 < t_{\text{LXe–TC}} < 4.4$ns. The dataset was reduced down to ~16% at this stage. The event reconstruction for the $\mu^+ \to e^+\gamma$ decay analysis was performed on the preselected dataset. Some reconstructions are common between the $\mu^+ \to e^+\gamma$ and the $\mu^+ \to e^+X, X \to \gamma\gamma$ decay analysis. We applied further event selection on the $\mu^+ \to e^+\gamma$ reconstructed dataset before the $\mu^+ \to e^+X, X \to \gamma\gamma$ dedicated reconstruction. The following selections were applied: existence of multiple γs, at least one high quality e^+ track [2] , $|m_\mu - E_{\text{Total}}| < m_\mu \times 20\%$, and $E_\gamma > 40\,\text{MeV}$ where E_{Total} is energy sum of e^+ and γ. We apply event reconstruction for the $\mu^+ \to e^+X, X \to \gamma\gamma$ decay search on these selected datasets.

---

[2]Selection conditions for the high quality track is defined in Section 4.1.6
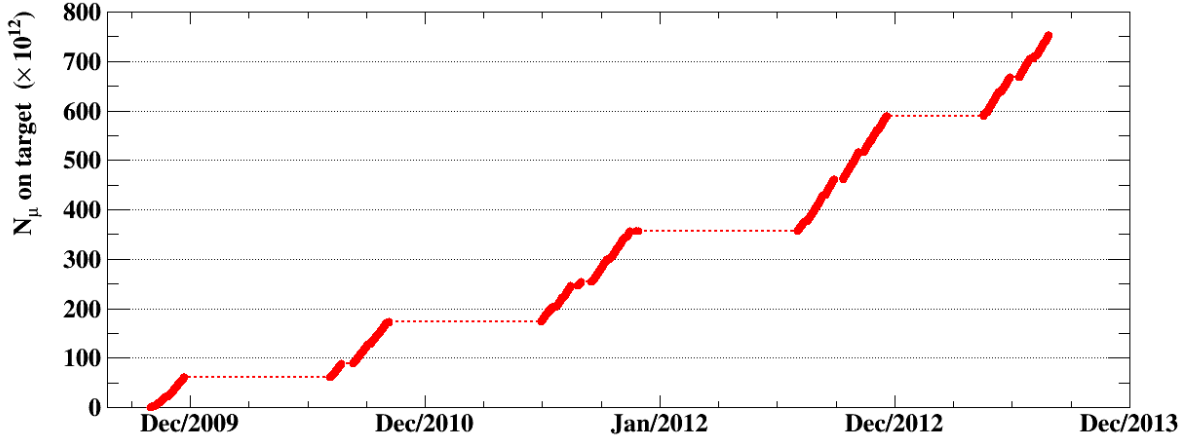
FIGURE 3.20. Number of stopped muons on target [6].

3.9.3 Event reconstruction

Event reconstruction methods for the $\mu^+ \to e^+\gamma$ decay analysis are summarized in [6, 73]. All the reconstructions related to the $\mu^+ \to e^+X, X \to \gamma\gamma$ decay analysis are summarized in Figure 3.21 and described in the subsequent chapters.

In Chapter 4, Chapter 5, Chapter 6, we discuss reconstruction methods and their performances. e^+ reconstruction is the same as the $\mu^+ \to e^+\gamma$ decay analysis (Chapter 4). 2γs' reconstruction has been newly developed for the $\mu^+ \to e^+X, X \to \gamma\gamma$ decay analysis (Chapter 5). After reconstruction of e^+ and 2γs, reconstructed variables are combined to reconstruct the X decay vertex (Chapter 6).

We also discuss the difference between MC and data in these chapters. In the MEx2G analysis, we use MC to estimate the signal efficiency because events with a similar topology of signal do not exist in data and it is difficult to extract information only from the data. Estimated differences are used as smearing parameters to reproduce data distributions in MC.

43
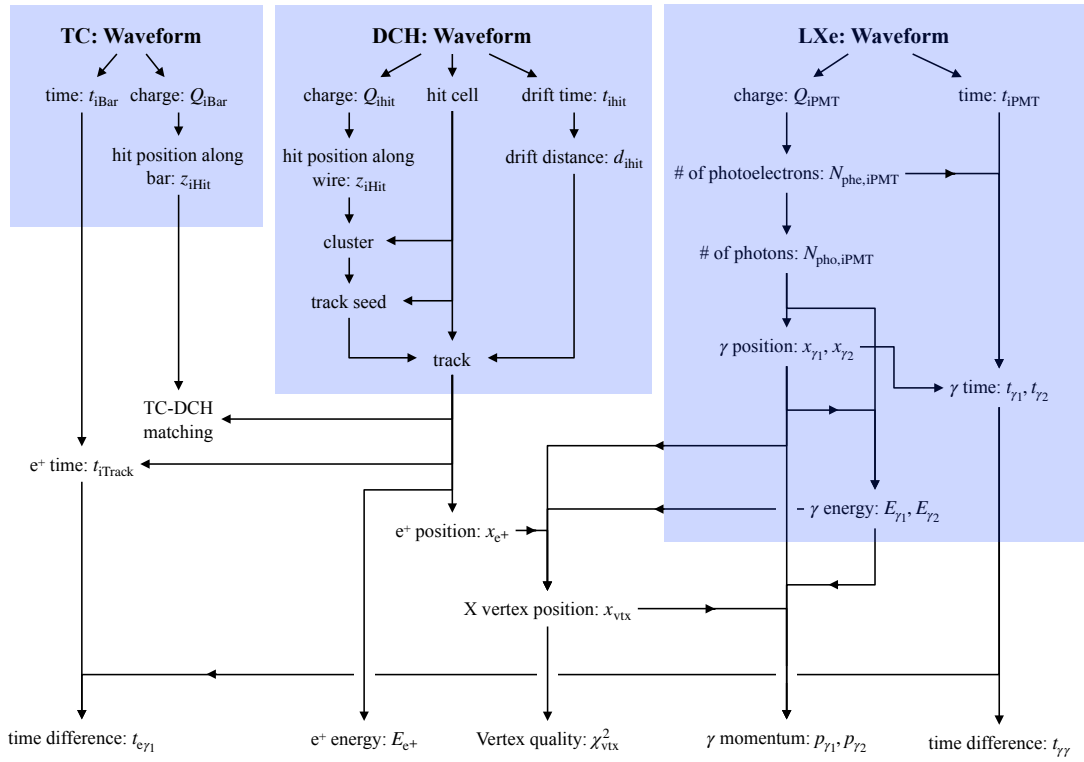
Figure 3.21: Overview of the event reconstruction. All the reconstructions start from waveform analysis. The blue regions indicate detector-wise event reconstruction while the other regions indicate inter-detector reconstruction. Detailed reconstruction methods will be described in the corresponding chapters (Chapter 4, Chapter 5, and Chapter 6).

E$^+$ RECONSTRUCTION AND PERFORMANCE

In this chapter, the e$^+$ reconstruction and its performance are discussed. As mentioned in Section 3.9.3, estimation of MC smearing parameters is an important task. e$^+$ efficiency relevant to the MEx2G decay search is discussed in Section 7.6.4.

4.1 e$^+$ reconstruction

The flow of the e$^+$ reconstruction is summarized in top left half of Figure 3.21. e$^+$ hit time measured with TC and its trajectories under the magnetic field is measured with DCH. Then, both information is combined to get e$^+$ time, position, and momentum.

4.1.1 Waveform analysis

The event reconstruction starts from waveforms. We calculate the hit time and charge from the waveform as follows: the raw waveforms are filtered using FFT to reduce known noise contributions. We obtain six waveforms per single drift cell as shown in Figure 4.1: two are from both ends of anode wires. The others are from adjacent cathode pads. Baselines are calculated from the waveforms themselves in the region before the signal. The hit time is taken from an anode waveform, which crosses a threshold determined by baselines of each waveform. Charge integration range around the peak is optimized to minimize electrical noise contributions.
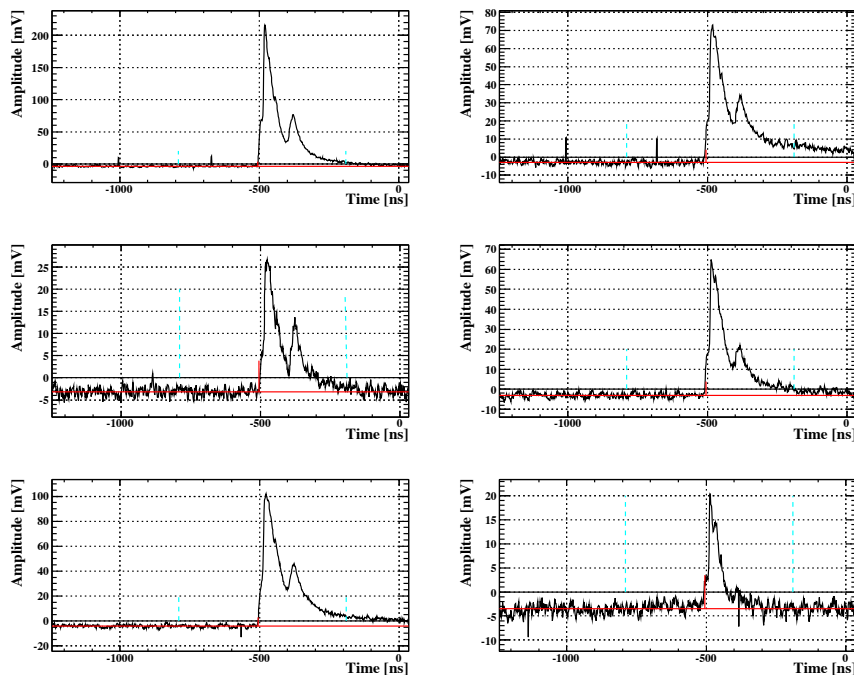
Figure 4.1: Waveforms of DCH. Top two waveforms are from both ends of anode wires and the others are from cathode pads [97]. Red horizontal lines are calculated baselines and red vertical lines are calculated hit times.

4.1.2 Hit position

z position Hit position along beam axis (z direction) reconstruction has two steps. First, z position is reconstructed only using the anode information:

$$z_0 = \left(\frac{L}{2} + \frac{Z}{\rho} \right) \cdot \epsilon_0,$$

where L is the length of the anode wire and Z is the input impedance and ρ is the resistivity of the anode wire. ϵ_0 is the asymmetry of charges given by

(4.1)
$$\epsilon_0 \equiv \frac{Q_\mathrm{U} - Q_\mathrm{D}}{Q_\mathrm{U} + Q_\mathrm{D}},$$

where Q_U and Q_D are calculated charges from the upstream (U) and downstream (D) ends of the anode wire, respectively. The resolution of the charge ratio method is ~1 cm.

The z position is then refined down to below 1 mm using the vernier pads as follows. The refined z position is calculated by

$$z = l_\mathrm{pad} \cdot \left(\frac{\alpha}{2\pi} + i - \frac{n}{2} \right),$$

where $l_\mathrm{pad} = 5$ cm is the length of one period of the vernier pattern; n is the total number of the periods in a cathode pad while i indicates which period the hit position is located in (i-th period
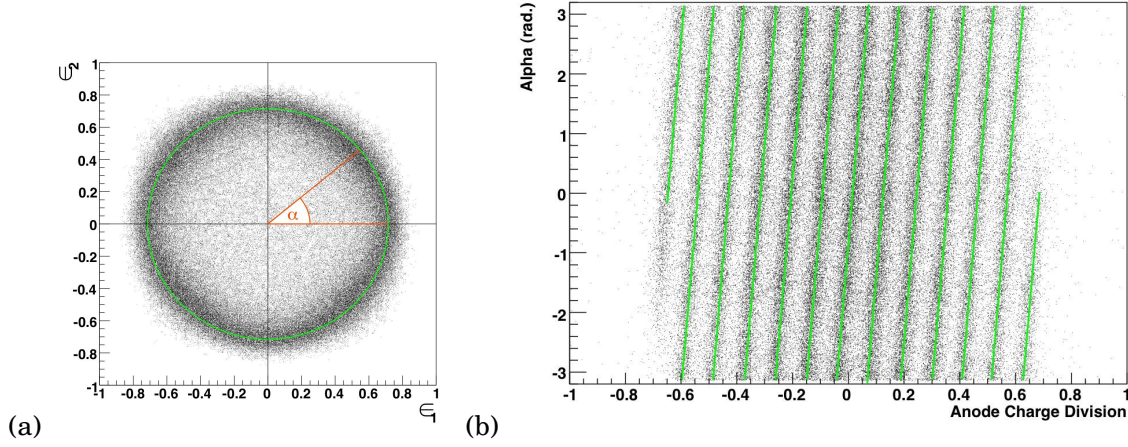
Figure 4.2: Vernier pad calculations. (a) Charge ratio of cathode pads (ϵ_1, ϵ_2) and α (b) α and anode charge division (ϵ_0).

from downstream side); α is the phase of the vernier circle calculated from charge ratios of two cathodes:

$$\alpha = \tan^{-1}\left(\frac{\epsilon_2}{\epsilon_1}\right),$$

where ϵ_1 and ϵ_2 are the charge asymmetries like Equation (4.1) calculated with inner and outer cathode pads, respectively. Figure 4.2 (a) shows relation between these variables (the vernier circle). Figure 4.2 (b) shows α as a function of anode charge division (ϵ_0), which is used to determine i (in which number of period the hit exists).

Drift distance Drift distance inside the drift cell is calculated from the anode hit time. TC hit time is set to time zero. The time difference (drift time) between hit time and the time zero is converted to the drift distance using a pre-calculated (based on the GARFIELD software [93]) and calibrated functions.

4.1.3 Clustering and track finding

Reconstructed hits in nearby cells are clustered to remove accidental hits. The clusters are combined to make seeds for track fitting. The seeding starts from the outer cells because track occupancy is less in the outer region while high momentum e$^+$s are expected to hit outer cells. The wire coordinates are used as the first estimation of track curvature and momentum. The seeds are extended to find other clusters by using an invariant variable p_T^2/B_z (p_T: transverse momentum). Most of the left/right ambiguity inside the drift cell is also solved at this step. Then a circle fit in x-y plane gives the first estimate of track time and the improvement of solving the left/right ambiguity.

47

4.1.4 Track fitting

The Kalman fitter technique is used to reconstruct tracks [98, 99] based on the GEANE software [100]. The effect of materials can be taken into account by using the technique. After this first track fitting in DCH, the track is propagated to the TC region to test matching with the TC hits. The matched TC hits are connected to the track and then the track time is refined using the TC hit time. Finally, the fitted track is propagated back to the stopping target and the point of intersection of the target is defined as the decay vertex position and angle.

4.1.5 TC reconstruction

The signals from TC PMTs are processed using Double Thresholds Discriminators (DTD) to minimize the time-walk effect. The DTD outputs a NIM pulse at the timing when the input signal crosses the lower threshold if the signal is higher than the higher threshold. A TC hit is reconstructed if both PMTs in a bar have signals higher than the higher threshold of the DTD. The output NIM pulses are fitted with a template waveform to extract the timing information (t_{IN} and t_{OUT}), where IN (OUT) corresponds to the PMT close to (far from) the target. Then TC hit time is given by

$$t_{\text{TC}} = \frac{t_{\text{IN}} + t_{\text{OUT}}}{2} - \frac{b_{\text{IN}} + b_{\text{OUT}}}{2} - \frac{w_{\text{IN}} + w_{\text{OUT}}}{2} - \frac{L}{2v_{\text{eff}}},$$

where b is the time offset; w is a correction value of the time-walk effect; v_{eff} is the effective velocity of the scintillation light inside the scintillator bar; L is the length of the bar. The hit position along the bar is given by

$$z_{\text{TC}} = \frac{v_{\text{eff}}}{2} \left\{ (t_{\text{IN}} - t_{\text{OUT}}) - (b_{\text{IN}} - b_{\text{OUT}}) - (w_{\text{IN}} - w_{\text{OUT}}) \right\}.$$

4.1.6 e$^+$ selection

We defined selection criteria to select a single track per event from pileup events and duplicated events from a e$^+$. Duplicated events can occur because all the possible track combinations remain in the reconstruction step. The criteria use tracking information[1]:

- The number of turn (less than 2).

- Sufficient matching quality with TC: difference of reconstructed z, r, t between TC and tracking is small.

- Good fitting quality: the number of hits in the DCH is large enough, small χ^2 of track fitting, small uncertainty of reconstructed energy and angles.

- Target constraint: backward track propagation to the target is within its fiducial volume.

---

[1]See [6, 97] for the detailed conditions.

- Best ranked track: select one track from several reconstructed tracks that has the best rank defined by a principal component analysis. The rank is a linear combination of the involved observables.

4.2 Performance

4.2.1 Energy, angle, and position

Smearing parameters   To estimate MC smearing parameters for e$^+$s, a double turn method was used. There is no direct way to estimate angular and position resolutions for e$^+$s using data. To overcome this difficulty, the double turn method was developed [73]. In this method, we select specific events that have two turns inside the drift chamber (while typical events have ~1.5 turns). Two turns are divided into two single turns and each turn is reconstructed individually. The difference between these turns is compared at an imaginary plane placed after the first turn at the beamline.

The resolutions estimated from the double turn method can differ from the true resolution. Thus this difference between the signal resolution and the double-turn resolution is estimated from MC and corrected. Uncertainties related to this correction are assigned to systematic uncertainties of the estimation of smearing parameters. The quadratic difference of resolutions between MC and data is defined as the smearing parameter. This comparison is performed using 2010 MC/data and summarized in Table 4.1.

Table 4.1: Smearing parameters

Item	Smearing	Systematics
E (MeV)	0.145	± 0.008
ϕ (mrad)	5.34	$^{+2.45}_{-4.20}$
θ (mrad)	4.55	$^{+1.44}_{-1.99}$
z (cm)	0.0474	$^{+0.0942}_{-0.0474}$
y (cm)	0.0772	$^{+0.0078}_{-0.0083}$

Resolutions   e$^+$ resolutions for the $\mu^+ \to e^+\gamma$ signal (52.8 MeV) are summarized in Table 3.4. Resolutions for lower energy e$^+$, which are expected for the MEx2G signals, are worse than these values. In particular, energy and angle (ϕ and θ) resolutions are relevant to the later analysis. Thus, we focus on the expected resolutions of these variables.

Expected energy resolution for the MEx2G signal e$^+$s is shown in Figure 4.3. The energy dependence is estimated using MC with smearing and the curve is scaled to fit its resolution at 52.8 MeV. The energy resolution is below 0.5 MeV in our momentum range (Figure 2.7). We set e$^+$ energy window to be 1 MeV in the event selection for each m_X in Chapter 7.

Energy dependence of ϕ and θ are estimated based on MC with smearing and scaled to match resolutions at 52.8 MeV. The estimated dependencies are shown in Figure 4.4. These dependencies are partially used in the vertex reconstruction as discussed in Chapter 6.
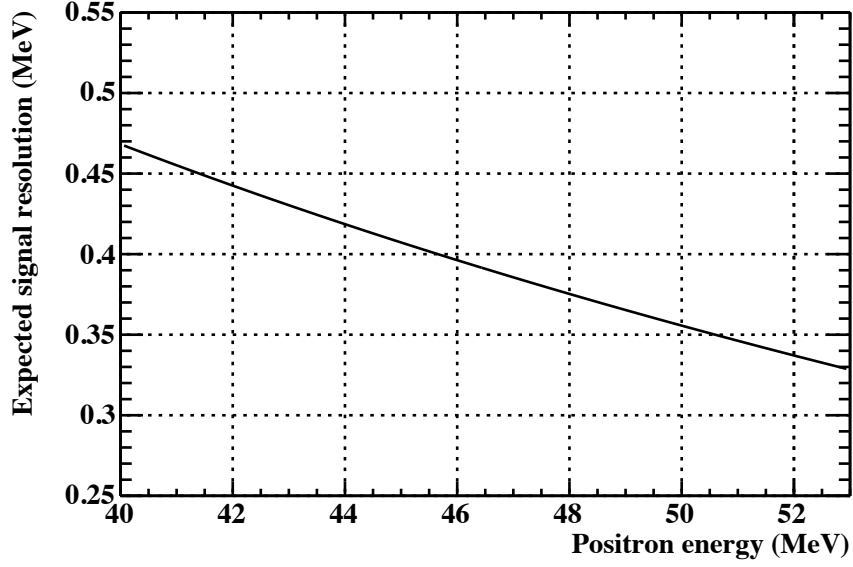
Figure 4.3: e$^+$ energy resolution
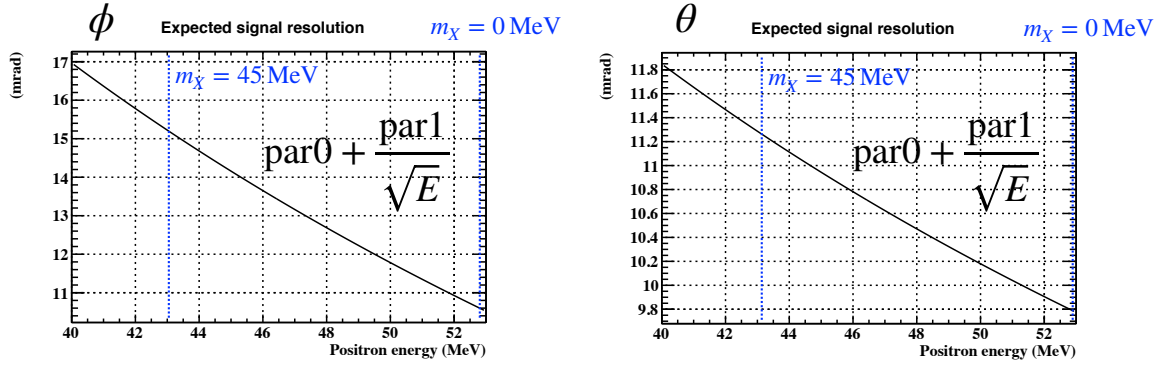
Figure 4.4: Expected e$^+$ θ (left) and ϕ (right) resolution

4.2.2 Timing

The time resolution of $t_{e\gamma}$ is estimated from the RMD peak. Then that of t_e is estimated by subtracting $t_\gamma{}^2$ contribution from $\sigma(t_{e\gamma})$. The e$^+$ time resolution is summarized in Table 4.2.

To estimate smearing parameter of t_e, first we evaluate the time resolution using the signal MC as shown in Figure 4.5. The energy dependent curve shown in red in the figure is fitted to this points. 52.8 MeV-equivalent time resolution is estimated from the fitted curve and is summarized in the second row of Table 4.2 (**resolution (MC)**). Then, smearing parameter is estimated from

---

[2]Section 5.3.3

Table 4.2: e$^+$ timing summary. "+" in the last row indicates the smearing parameter only have the positive side of its systematics because the parameter is estimated to be 0.

(ps)	2009	2010	2011	2012	2013
resolution(Data)	106	107	96	100	89
resolution(MC)	127	93	93	92	91
smearing	0	64	28	49	0
systematics	75(+)	30	25	28	29(+)

the quadratic difference between data and MC at 52.8 MeV:

$$\sigma_{\text{diff.}} \equiv \sqrt{|\sigma_{\text{data}}^2 - \sigma_{\text{MC}}^2|}$$

Since the MEx2G signals have different e$^+$ energy depending on the mass of X, we also consider energy dependence (lower than 52.8 MeV). The e$^+$ time resolution can have a larger value at $m_X = 20$ MeV by a factor of $1.47 (\equiv f_E)$, estimated from the fitted curve in Figure 4.5. We assume the smearing parameter evolves up to f_E.

If $\sigma_{\text{data}} - \sigma_{\text{MC}} > 0$ (like in 2010, 2011, and 2012), the mean of the smearing parameter is defined as

(4.2)
$$\frac{f_E + 1}{2} \times \sigma_{\text{diff.}},$$

by taking average between possible maximum ($f_E \times \sigma_{\text{diff.}}$) and minimum ($1 \times \sigma_{\text{diff.}}$) values. Systematics of the mean is calculated by taking a quadratic sum of the following two contributions:

- Measurement error of the resolution: $\sqrt{(\sigma_{\text{data}} + \delta_{\text{data}})^2 - \sigma_{\text{data}}^2}$ where δ is measurement error of σ_{data}[3].

- Contribution from possible energy dependence from the mean value (Equation (4.2)) ~ a half of the difference between maximum and minimum values: $\frac{f_E - 1}{2} \times \sigma_{\text{diff.}}$

In other cases (2009 and 2013), the mean of the smearing parameter is defined as 0. The systematics of the mean is calculated by taking a quadratic sum of the same two contributions above. The only difference is the sign of $\sigma_{\text{data}} - \sigma_{\text{MC}}$. The second contribution reflects our incomplete understanding of the data and MC difference.

---

[3]3.3% is used for δ. This value is originally estimated for $t_{e\gamma}$ in the MEG final result [6]. However we use this value both for e$^+$ and gamma as a conservative estimation.
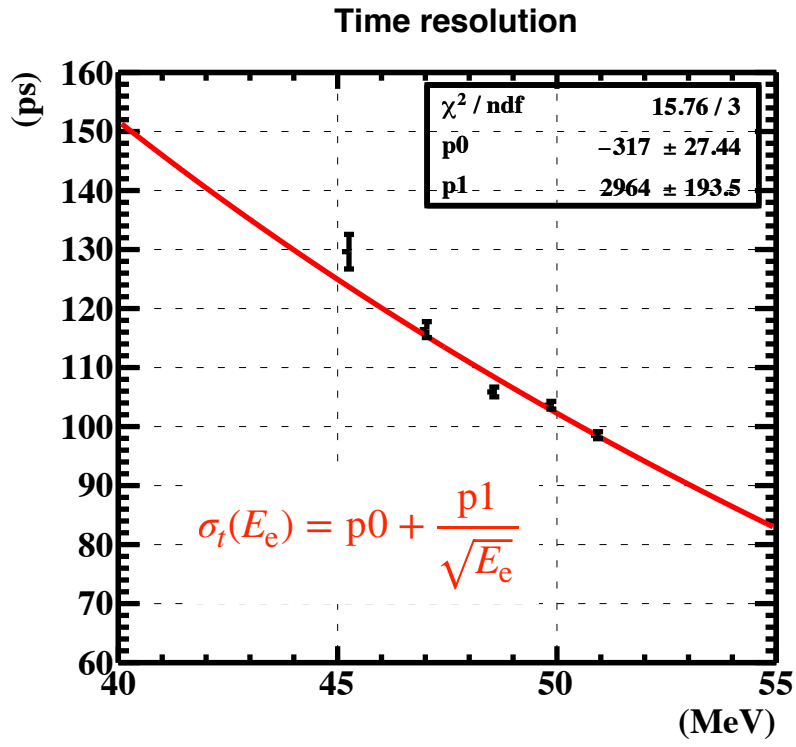
Figure 4.5: e$^+$ time resolution for the MEx2G signal (MC) in black points as function of e$^+$ energy. The energy dependent curve in red is fitted to this points.

γ **RECONSTRUCTION AND PERFORMANCE**

I n this chapter, 2γ reconstruction and its performance are discussed. There are two reconstruction methods: one is developed for the $\mu^+ \rightarrow e^+\gamma$ analysis and the other is for the MEx2G analysis. First, we discuss basics ideas of γ reconstruction common to the $\mu^+ \rightarrow e^+\gamma$ [6, 73] and MEx2G reconstruction. Then, we focus on the MEx2G reconstruction dedicated to 2γ analysis. In Section 5.3, we estimate MC smearing parameters. Finally, we discuss γ efficiencies. These estimation is used in the MEx2G decay search analysis in Chapter 7.

5.1 Generalities

The LXe coordinates are defined in Figure 5.1. The coordinates (u,v,w) are used as LXe local coordinates instead of MEG global coordinates (x,y,z).

Waveform analysis   The analysis starts from waveform analysis. The charge and time of each PMT are obtained from the waveform of each PMT. A digital constant fraction method is used to determine the time of the signal. The crossing time is calculated by interpolating adjacent two points.

The charge is calculated by integrating the waveform. A high-pass filter (moving average[1]) is applied to the original waveform to suppress noise contribution. The integration range is determined to be 48 ns and the same range is used for all the PMTs. 15% of events interact with liquid xenon where is close (below 1 cm) to the inner surface of the detector. PMTs close to the interaction points have larger pulse height and the waveforms are saturated. For these waveforms, the Time Over Threshold (TOT) is used: Figure 5.2 illustrates how to calculate the charge for

---

[1] 89 points ~ cutoff frequency: 11 MHz.
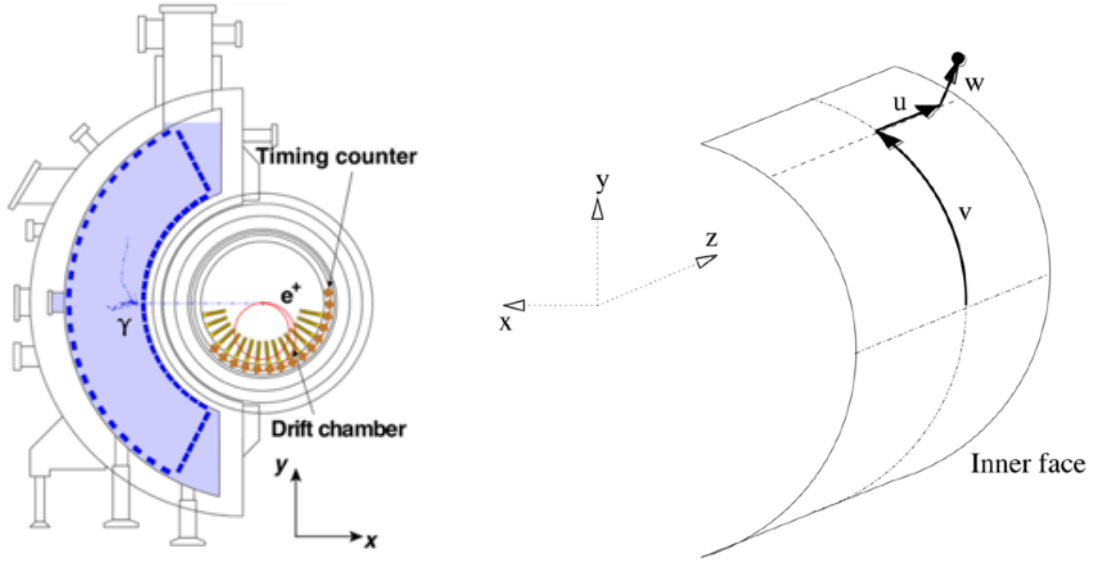
Figure 5.1: LXe coordinates

saturated waveforms by TOT. First, we calculate the duration time when the waveform is below a threshold (150 mV). This TOT correlates with charge. The correlation is calculated using a template waveform beforehand. Finally, we convert TOT to charge for saturated waveforms.

The charge is converted into the number of photoelectrons by using a measured PMT gain. The number of photoelectrons of i-th PMT ($N_{\mathrm{pe},i}$) is given by

$$N_{\mathrm{pe},i} = \frac{Q_i}{e \cdot G_i},$$

where Q_i is the measured charge and G_i is the gain of the PMT. The gain is determined by LED data listed in Table 3.2. Then, the number of photoelectrons is converted into the number of photons ($N_{\mathrm{pho},i}$) by using a measured PMT quantum efficiency (QE):

$$N_{\mathrm{pho},i} = \frac{N_{\mathrm{pe},i}}{QE_i}.$$

The QE is determined by a radioactive source (α source) listed in Table 3.2.
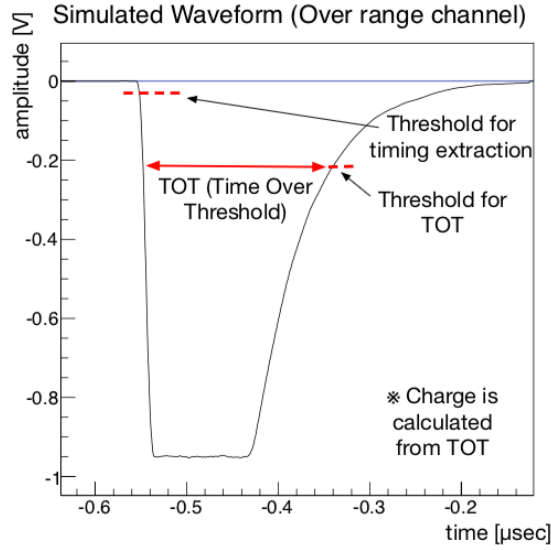
Figure 5.2: Time over threshold [101]

5.2 2γ reconstruction

The reconstruction flow is summarized in the right half of Figure 3.21.

5.2.1 Pileup

The LXe detector is designed to detect γ from the $\mu^+ \to e^+\gamma$ decay. More than 1γ can exist in a single event. There are two cases in the origin of these γs: different μs and the same μ. The MEx2G signal γs should be (at least) the latter case. In any case, we call these events pileup. In the $\mu^+ \to e^+\gamma$ analysis, these events are identified for the later analysis. We actively use these events in the MEx2G analysis.

To identify these pileup events, a peak search is performed based on the inner and outer light distribution by using *TSpectrum2* [102]. Peaks in the two-dimensional histogram of the number of photons are searched for. The threshold of the peak search is set to 200 photons. Events that have more than 1 peak are identified as pileup events. When the number of pileup events is more than 2, the following pre-fittings are iterated and the largest two peaks are selected.

5.2.2 Position and energy

Hereafter, only pileup events are analyzed. Figure 5.3 shows a typical event display of a 2γ event. Each PMT has a contribution from each γ. The point in the 2γ reconstruction is how to divide the number of photons observed at a selected PMT into the contribution from each γ.

Calculation of initial values   First, results from the peak search are used for the initial value of u and v while w is set to 1.5 cm. Once the interaction point of each γ is determined, the contribution from each γ is calculated for each PMT. Assuming

$$E_{\gamma_1} : E_{\gamma_2} = r : 1 - r \, (0 < r < 1),$$

where the initial value of r is set to 0.5. The ratio of the number of photons from γ source 1 is

$$(5.1) \qquad\qquad r_1' = \frac{r\Omega_1}{r\Omega_1 + (1-r)\Omega_2}$$

or

$$(5.2) \qquad\qquad r_1' = \frac{rd_1}{rd_1 + (1-r)d_2}$$

where Ω is the solid angle between each γ and the selected PMT, which is calculated numerically, and d is distance between each γ and the selected PMT. Equation (5.1) is used for shallower (close to the inner surface of the LXe detector) events and Equation (5.2) is used for deeper events because the solid angle is less sensitive for deeper events. By using this ratio and the number of photons at each PMT, the initial value of the number of photons ($M_{\text{pho},1(2)}$) emitted from the interaction point of $\gamma_{1(2)}$ is calculated.

Position pre-fitting   Inner PMTs with reasonable light yield is selected (nUsedPMT) and a position pre-fitting is performed to reproduce light distributions of the LXe detector. This fitting is performed 1γ by 1γ and iterated; first, $\vec{x}_1, M_{\text{pho},1}$ is fitted while the other parameters are fixed. Then $\vec{x}_2, M_{\text{pho},2}$ is fitted while the other parameters are fixed.

In the fitting, the following χ^2 is minimized:

$$(5.3) \qquad \chi^2_{\text{pre}} = \frac{1}{\text{nUsedPMT}} \sum_i^{\text{nUsedPMT}} \frac{(N_{\text{pho},i} - M_{\text{pho},1} \times \Omega_i(\vec{x}_1) - M_{\text{pho},2} \times \Omega_i(\vec{x}_2))^2}{\sigma_{\text{pho},i}(N_{\text{pho},i})^2},$$

where $N_{\text{pho},i}$ is the number of photons measured at i-th PMT; $\sigma_{\text{pho},i}(N_{\text{pho},i})$ is statistical uncertainty of the number of photons ($\propto \sqrt{N_{\text{pho}}}$). The minimization is performed by a grid-search in $\vec{x}_{1(2)} = (u, v, w)_{1(2)}$ space while the other parameters are fixed to the initial values.

Next, to find the better parameter sets for each γ, the following χ^2 is minimized 1γ by 1γ:

$$(5.4) \qquad \chi^2 = \sum_i^{\text{nPMT}} \frac{(N_{\text{pho},i} - M_{\text{pho},1} \times \Omega_i(\vec{x}_1) - M_{\text{pho},2} \times \Omega_i(\vec{x}_2))^2}{\sigma_{\text{pho},i}(N_{\text{pho},i})^2}$$

If there are 3 or more peaks, the two largest peaks are selected and the largest one is defined as γ_1 and the other is defined as γ_2 in the later analysis.

Energy pre-fitting   The same χ^2 (Equation (5.4)) is used, but in this time $M_{\text{pho},1(2)}$ is fitted while other parameters including positions are fixed. All the PMTs except for those with lower light yields[2] are used.

---

[2]The number of photoelectrons is less than 200.
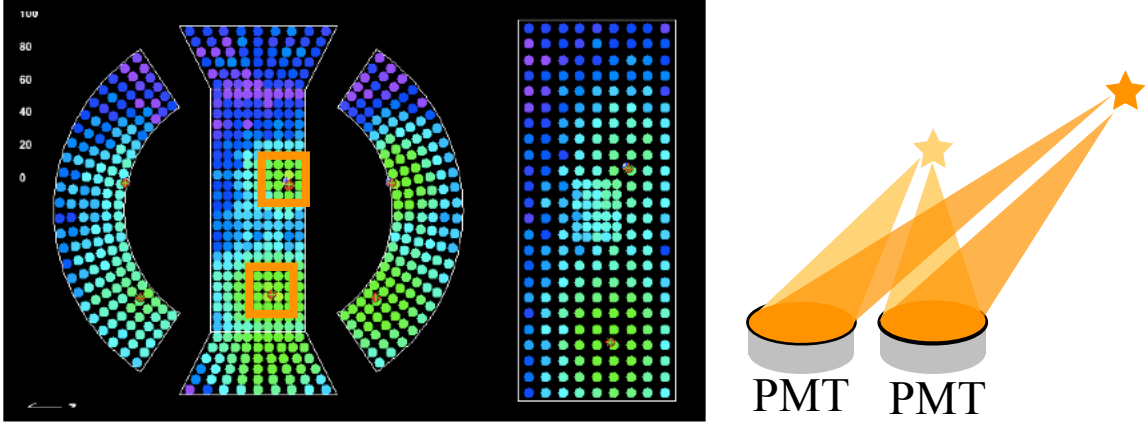
Figure 5.3: Position and energy reconstruction for 2γ

Position and energy fitting At the final step, 2γs are fitted simultaneously using the same χ^2(Equation (5.4)). All the parameters are fitted to find the final values. The position is fitted again at this step because position pre-fitting depends on the initial value of $M_{\text{pho},1(2)}$.

Finally, $E_{\gamma1(2)}$ is calculated from the the best-fitted value of $M_{\text{pho},1(2)}$:

$$(5.5) \qquad E_{\gamma1(2)} \quad = \quad F(u,v,w) \times T(t) \times C \times M_{\text{pho},1(2)}$$

where $F(u,v,w)$ is a uniformity correction factor, $T(t)$ is a time variation correction factor, and C is a conversion factor from the number of photon to energy. These factors are calibrated using various calibration methods listed in Table 3.2. The detailed description of these factors are summarized in Refs. [6, 103].

In addition, we newly implemented the following three corrections.

5.2.3 Solid angle correction

We observed a bias on the distribution of the difference between the reconstructed energy and the MC truth vs. the reconstructed position (u) shown in Figure 5.4. A zigzag structure is observed for shallower events. The pitch of the structure corresponds to the interval of the PMTs (design value: 62 mm). For shallower events, the light yield is largely different depending on the γ position. More scintillation light is collected when γ interacts directly above a PMT than the γ interacting between PMTs. Thus the collected light is position sensitive for shallow events and this causes the structure shown in Figure 5.4.

As shown in Figure 5.5, there is a correlation between the normalized number of photons and a solid angle of γ interaction point covered by a PMT for shallow events. Especially, Ω is defined as its maximum value. This correlation is fitted with the following function to find parameters to
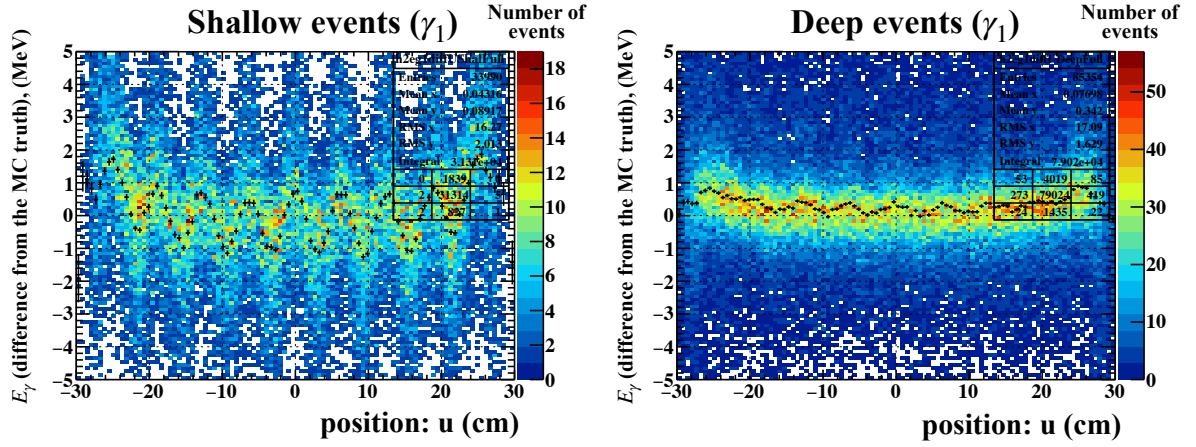
Figure 5.4: Solid angle bias (MC). Left: shallow events (w < 3 cm), right: deep events. The MEx2G signal MC of m_X =(20, 25) MeV, τ =(5, 10) ps are used for this plots.
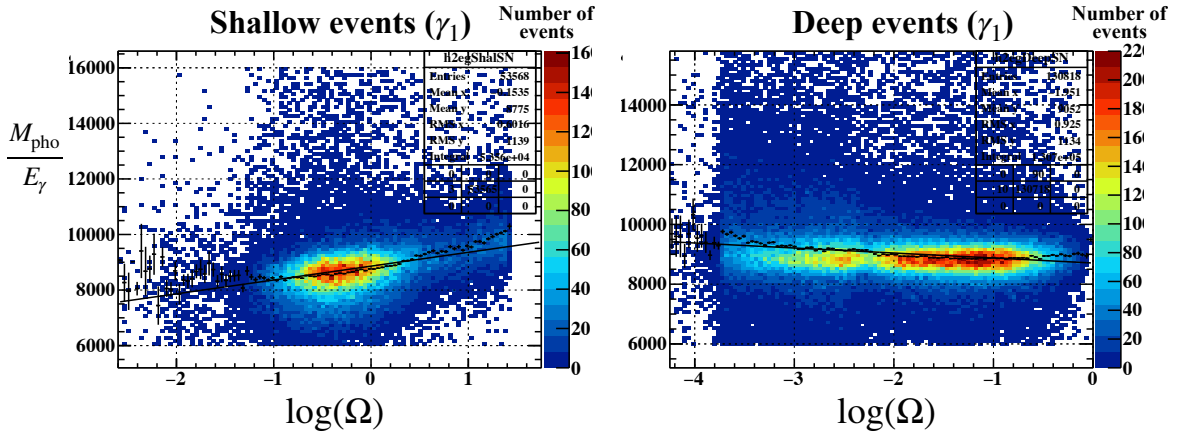
Figure 5.5: Correlation between the normalized number of photons and solid angle (MC). Left: shallow events (w < 3 cm), right: deep events.

be used for the correction:

$$\frac{M_{\text{pho}}}{E_\gamma} = f(E_\gamma) \times \log(\Omega) + a$$

where M_{pho} is the number of total photons and it is normalized using E_γ because the MEx2G signal is not mono-energetic. $f(E_\gamma)$ and a are fitted parameters. The coefficient of the solid angle ($f(E_\gamma)$) depends on the energy because the solid angle bias should have energy dependence due to the size of shower development.

By using the fitted parameters, the number of photons is corrected by applying the following factor to M_{pho} in the last step of the energy reconstruction:

$$M_{\text{pho}} = M_{\text{pho}} \times \frac{a}{f(E_\gamma) \times \log(\Omega) + a}$$
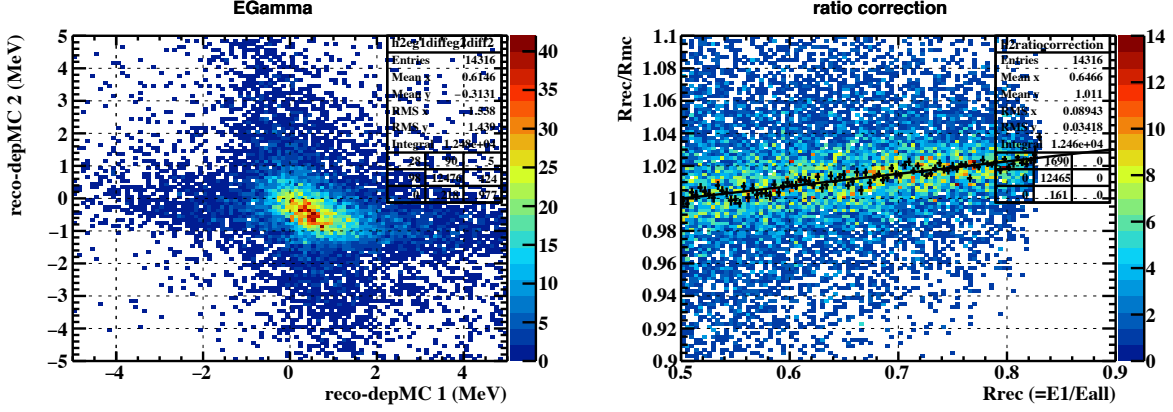
Figure 5.6: Ratio bias (MC). See text in detail.

This correction reduces solid angle dependence of $\frac{M_{\mathrm{pho}}}{E_\gamma}$ and the scale is set to $\frac{M_{\mathrm{pho}}}{E_\gamma}(\log(\Omega))|_{\log(\Omega)=0}$ (the edge of shallow and deep events). The pseudo 2γ data (Section 3.7.1) is used to find these parameters for data.

5.2.4 Ratio correction

The left plot of Figure 5.6 shows the correlation of reconstructed energy − MC truth among 2γs. E_{γ_1} (higher γ) tends to be reconstructed to be larger than the truth. At the same time, E_{γ_2} tends to be reconstructed to be lower than the truth. The sum of energies is not biased. These observations imply that the energy ratio R defined below has energy dependence.

$$R = \frac{E_{\gamma 1}}{E_{\gamma 1} + E_{\gamma 2}}$$

The ratio is calculated using both truth (R_{MC}) and reconstructed variables (R_{rec}) and its correlation is shown in the right plot of Figure 5.6. The ratio of R_{MC} and R_{rec} should be 1 regardless of E_γ. However, the plot shows clear dependence on $R_{\mathrm{rec}} \propto E_\gamma$. The idea of the ratio correction is to correct $R_{\mathrm{rec}}/R_{\mathrm{MC}}$ to be 1.

The pseudo 2γ data (Section 3.7.1) is used to find correction coefficients. The following five datasets with different ratios are used to obtain a wider ratio range:

- 18 MeV + Boron(11.7 MeV): R ~0.61

- 18 MeV + CW(11.7, 14.6, 17.6 MeV): R ~0.61, 0.55, 0.5

- 18 MeV + CEX(54.9 MeV, 82.9 MeV): R ~0.75, 0.82

In the data, R_{MC} is not available. Thus we use reconstructed results from original γ events before making pseudo 2γ data ($R_{\mathrm{original\,rec.}}$) instead of R_{MC}. Before making pseudo 2γ data, original γ

events are reconstructed. This result can be used as reference because this ratio bias can occur only in the 2γ reconstruction (we do not need to divide energy into 2 γ sources). $R_{\text{rec}}/R_{\text{original rec}}$ vs. R_{rec} is fitted with a linear function and R_{rec} is corrected accordingly to make $R_{\text{rec}}/R_{\text{original rec}}$ independent of R_{rec}.

5.2.5 Position correction

Figure 5.7 shows difference between reconstructed (u, v) position and MC truth as a function of the reconstructed position. A global bias is observed for u while it is not for v. The bias comes from an oblique incidence of γ with respect to the surface of PMTs and its shower development in the LXe detector. When u is large, a γ is more likely to enter into the detector with an oblique angle. In this case, the shower of such a γ develops to a larger area in the projected (u, v) surface and this makes the bias. As for v, PMTs are installed on the cylindrical surface and γ enters vertically without causing any bias.

This bias can have energy dependence because its origin relates to shower development. In fact, Figure 5.7 shows that there is smaller bias for γ_2 (smaller energy). Thus these distributions are fitted with an energy dependent function like the solid angle correction (Section 5.2.3).

5.2.6 Time

To calculate a hit time of each γ, we use PMTs whose light yields from one γ are 5 times higher than those from the other γ and remove PMTs whose light yields are less than 100 photons or which give large χ^2 value in the following fitting process. $\gamma_{1,2}$ hit times ($t_{\gamma 1(2)}$) are reconstructed by minimizing the following χ^2:

$$(5.6) \qquad \chi^2_{\text{time}}(t_{\gamma 1(2)}) = \sum_i^{\text{nPMT}} \frac{(t_{\text{hit},i,\gamma 1(2)} - t_{\gamma 1(2)})^2}{\sigma_{t,i}(N_{\text{pe},\gamma 1(2)})^2},$$

where $t_{\text{hit},i,\gamma 1(2)}$ is a hit time of each γ at i-th PMT and calculated from the waveform time ($t_{\text{PMT},i}$):

$$t_{\text{hit},i,\gamma 1(2)} = t_{\text{PMT},i} - t_{\text{delay},i,\gamma 1(2)} - t_{\text{offset},i},$$

where $t_{\text{delay},i,\gamma 1(2)}$ includes contributions from a propagation time between the interaction point and PMT and a time-walk effect; $t_{\text{offset},i}$ is a time offset inside the readout electronics. $\sigma_{t,i}(N_{\text{pe},\gamma 1(2)})$ is time resolution of each PMT as a function of the number of photoelectrons.
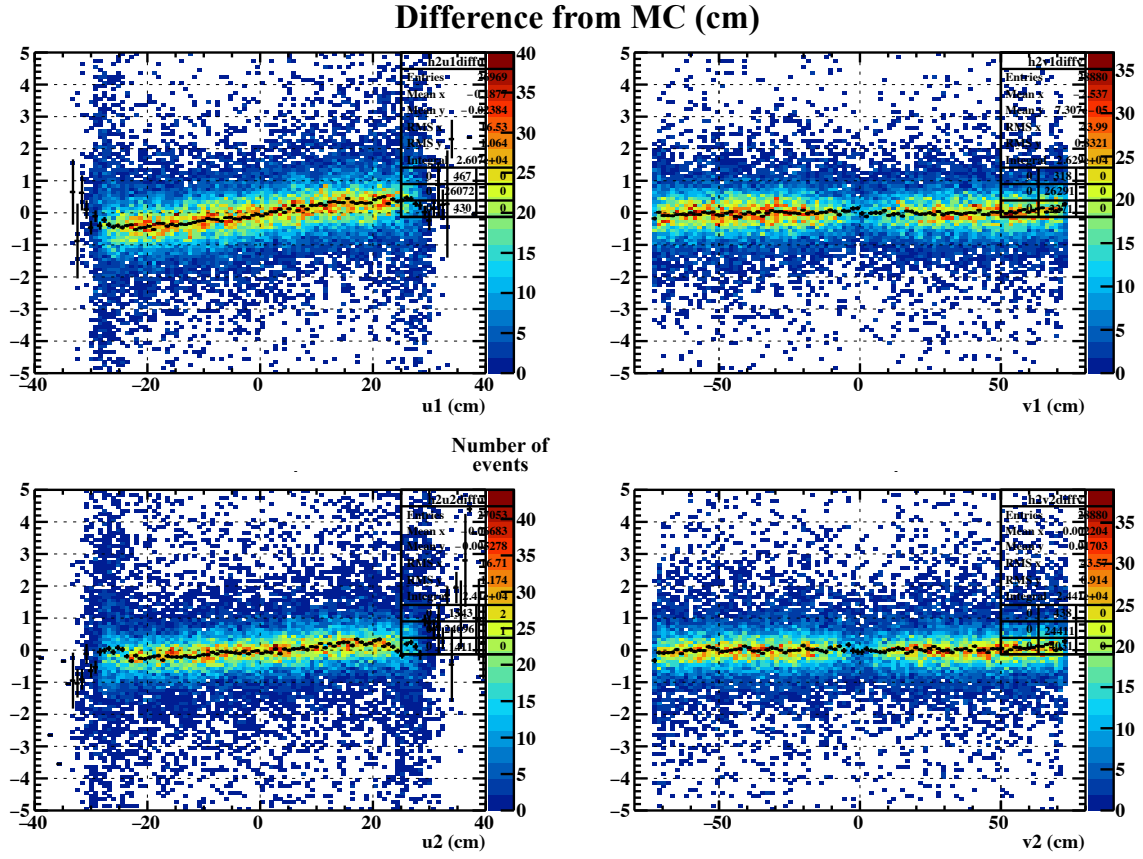
Figure 5.7: Reconstructed u, v dependence (MC). Top: γ_1, bottom: γ_2, left: u, right: v. The y-axis is difference between reconstructed position and MC truth and the x-axis is reconstructed position.

5.3 Performance

5.3.1 Position

MC smearing To estimate MC smearing parameter on γ position, pseudo 2γ data and MC (Section 3.7.1) are used. However, there is no position reference in the pseudo events. Thus we use results from 1γ reconstruction, which is originally developed for the $\mu^+ \to e^+\gamma$ analysis. Its position resolution is confirmed by using a lead collimator, as the position reference in a calibration run. Differences between 2γ reconstruction (developed for the MEx2G analysis) and 1γ reconstruction have similar distribution among pseudo data and MC. In addition, these differences are not so large compared with the absolute resolutions. Therefore, the data/MC difference in 1γ reconstruction can also be used for that in 2γ reconstruction. Differences of $\{2\gamma$ results $- 1\gamma$ results$\}$ between data and MC are estimated and treated as a systematic uncertainty of the data/MC difference.

In 1γ reconstruction, the data/MC difference is estimated by using a lead collimator in CEX run. The collimator (the right one was used) is shown in Figure 5.8. The collimator was placed just in front of the LXe detector. The number of γ events is reduced at positions on which the γ rays impinge across the bulk of the collimator, and hence, the event distribution shows the shadow of the collimator. The reconstructed distribution is shown in the left plot of Figure 5.9. Three peaks around the center region correspond to three slits in the collimator. These peaks are fitted with triple Gaussians and two error functions (for the outer region) to extract its width. The same analysis was performed on MC and the result is shown in the right plot of Figure 5.9. The difference of position resolution between data and MC is estimated to be 1.8 mm. This value is used for the MC smearing. A possible reason for the difference is incomplete PMT calibration. Its systematic uncertainty is estimated and summarized in Table 5.1 by comparing pseudo data and MC. These values are averaged over different energy range (Boron and CEX data). Larger systematics in w comes from a peak shift. Note that this uncertainty is expected to have little effect on the signal efficiency. It is because the smeared position is used as an initial value in maximum likelihood fitting in the vertex reconstruction and does not directly affect its performance.

Figure 5.8: The lead collimator used in 2008. The thickness is 1.8 cm, width of the slit is 1 cm, and a 2-inch PMT is superimposed as a reference.

Table 5.1: Smearing parameters

Item	Smearing	Systematics
u (cm)	0.18	±0.10
v (cm)	0.18	±0.06
w (cm)	0.18	±0.13
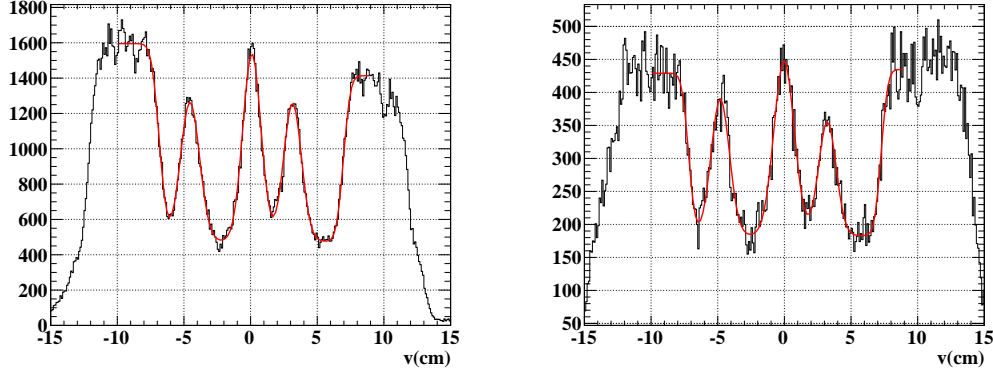
Figure 5.9: Event distribution with the lead collimator along v for data (left) and MC (right)[103].

Table 5.2: γ position resolution

Item	Core Resolution	Tail Resolution
u1 (cm)	0.54±0.01	2.32±0.11
v1 (cm)	0.49±0.01	1.58±0.07
w1 (cm)	0.66±0.01	4.45±0.17
u2 (cm)	0.53±0.01	3.47±0.08
v2 (cm)	0.44±0.02	1.84±0.10
w2 (cm)	0.63±0.01	4.57±0.11

Performance   Position resolution for the MEx2G signal is estimated using MC for parameter sets of (20, 25, 30, 35, 40, 45) MeV × 20 ps. Then its resolution is smeared with 1.8 mm to take the data/MC difference into account. The smeared distributions are fitted with double Gaussian. Figure 5.10 shows typical position resolution. Core and tail resolution do not depend on m_X, τ. Its standard deviation is estimated to be 1–3% and it is treated as an uncertainty when we quote single resolution value. Its average is quoted as position resolution for the MEx2G signal. These results are summarized in Table 5.2. Core to tail ratios are ~0.8.

5.3.2   Energy

MC smearing   Next, we estimate MC smearing parameter on γ energy. For this purpose, pseudo 2γ data and pseudo 2γ MC of Boron(11.7 MeV) and CEX (54.9 MeV) are used because there are energy references. Thus we can directly compare data and MC.

The response function of the LXe detector for monochromatic γs has an asymmetric shape;
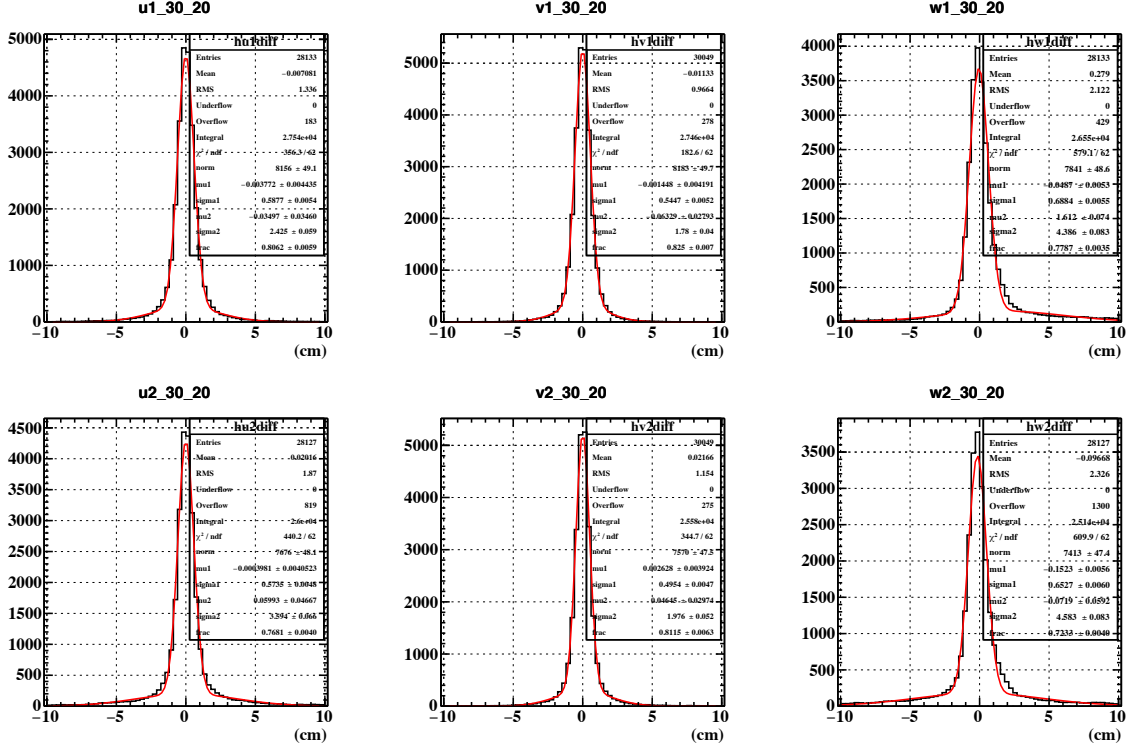
Figure 5.10: Differences between reconstructed position and MC truth. (30MeV, 20 ps) is selected and plotted.

low energy tail. There are two reasons for the tail: first, γs interact materials before the LXe detector and lose their energy. Second, a part of the γ shower escapes from the fiducial volume. To include this lower tail, the reconstructed spectrum is fitted using an exponential Gaussian function. The exponential Gaussian is described as:

$$(5.7) \qquad F(x) = \begin{cases} A \exp\left(-\dfrac{(x-x_0)^2}{2\sigma_{E_\gamma}^2}\right) & (x > x_0 + \tau) \\[2ex] A \exp\left(\dfrac{\tau}{\sigma_{E_\gamma}^2}(\tau/2 - (x - x_0))\right) & (x \le x_0 + \tau) \end{cases}$$

where A is a scale parameter, x_0 is a peak position, σ_{E_γ} energy resolution, and τ is a transition parameter. The exponential Gaussian has an exponential component when x is lower than $x_0 + \tau$ and Gaussian component when x is higher than $x_0 + \tau$. The peak position and the transition position is not necessarily the same and the difference is included in τ. In the fitting, we use a double exponential Gaussian to fully include tail resolution.

Reconstructed Boron spectra are shown in Figure 5.11. Color solid lines are fitted curves. The left plot is from pseudo-MC and the right plot is from pseudo data. Fitting is performed based on the double exponential Gaussian. We added a background spectrum in the fitting function as exponential for pseudo data whereas we simply use double exponential Gaussian for pseudo-MC.
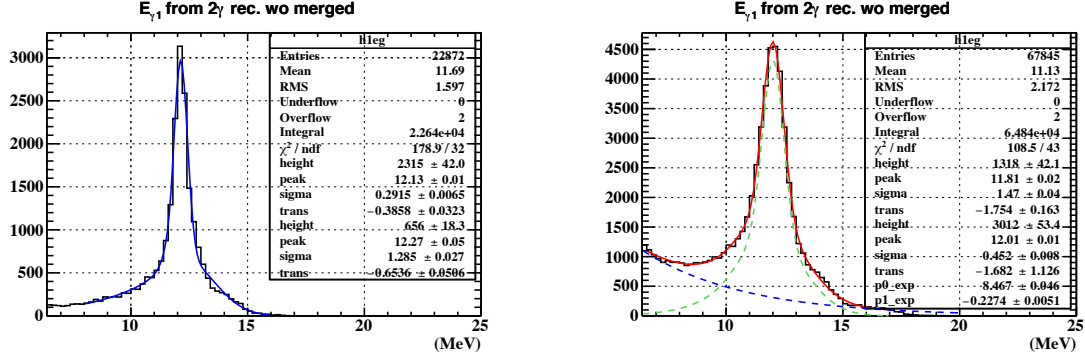
Figure 5.11: Reconstructed Boron (11.7 MeV) spectrum. Left: MC, right: data.
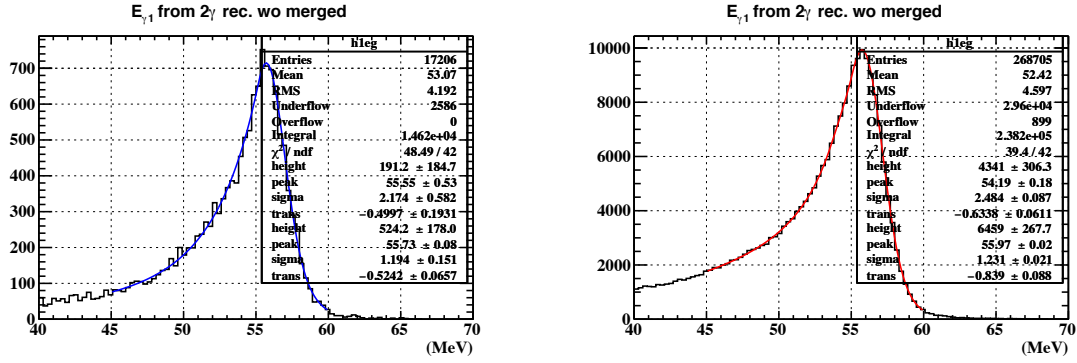
Figure 5.12: Reconstructed CEX (54.9 MeV) spectrum. Left: MC, right: data.

It is because we only simulate the monochromatic Boron peak; we do not simulate a background γ from the lower events whose energy is less than 10 MeV. For the fair comparison, in the fitting, we added the background function only in data. To extract the smearing parameter from these two plots, we first compare the resolution of the core and tail, respectively. Then, these two quadratic differences are weighted with the core fraction. The smearing parameter for Boron events is estimated to be 0.46 MeV.

To evaluate the energy resolution using CEX data, first, we apply an angle correction. There is a correlation between opening angle and reconstructed energy in LXe detector as shown in the right plot of Figure 3.16. Thus we correct the energy based on the reconstructed opening angle: the energy of ~55 MeV γ becomes lower by this angle correction. The smearing parameter for CEX events is estimated to be 0.66 MeV in the same way as the Boron case.

Combining two results described above, the energy smearing parameter is defined to be $0.56^{+0.10}_{-0.21}$ MeV by averaging over these two estimations. The energy smearing does not affect so much on the signal efficiency, so that we define the single value for simplicity. Its systematic uncertainty is assigned to this value taking into account the core resolution of Boron and the

Expected signal energy resolution
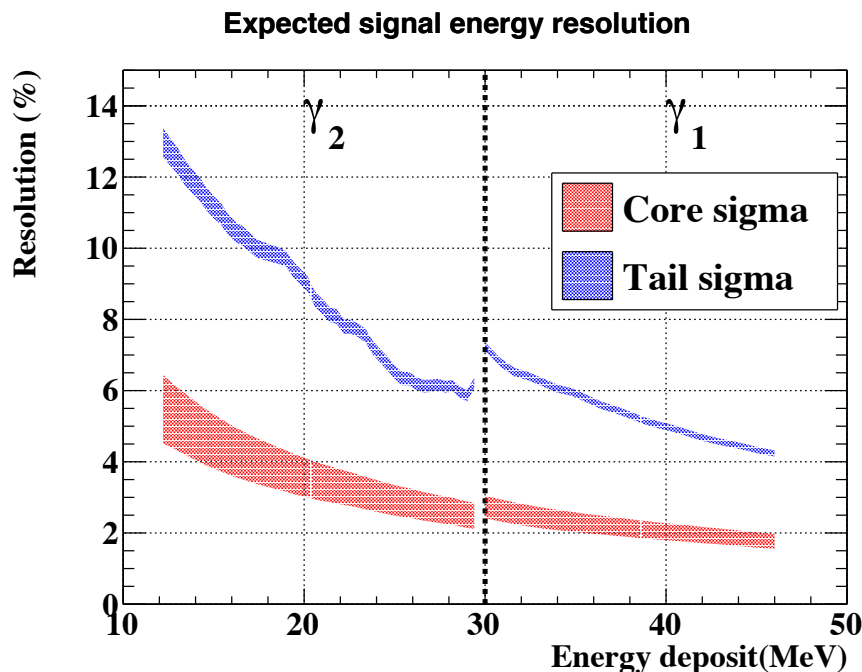
Figure 5.13: Expected signal gamma energy resolution. Band comes from systematics of the smearing parameter. Statistical error of the mean resolution estimation is small compared with the systematics.

average resolution of CEX. [3]

Performance   Expected signal energy resolution can be estimated using signal MC smeared with the parameter obtained above. The results are shown in Figure 5.13. The left half is for the lower γ and the right half is for the higher one. Core resolution and tail resolution are plotted individually. To sum up, expected signal energy core (tail) resolution is 2–3% (4–7%) for γ_1 and 2–6% (6–13%) for γ_2. These values should be compared with the CEX resolution 2% (5%) for the higher γ and the Boron resolution 4% (13%) for the lower γ. The expected energy resolution from MC agrees with the pseudo data.

Finally, the linearity of the energy scale is confirmed using pseudo 2γ data as shown in Figure 5.14. Uncertainty of the energy scale is estimated to be 1% from residuals of a linear fitting.

---

[3]Comments on this individual smearing: this smearing is applied on γ_1 and γ_2, individually. What we want to smear here is the difference between data and MC and this is possibly due to an incomplete calibration of photosensors and incomplete implementation of the light distribution inside the detector. The first effect is treated individually, and the second effect can have a correlation between two γs. However, the possible correlation is already taken into account in the ratio correction (Section 5.2.4) in the 2γ reconstruction framework.
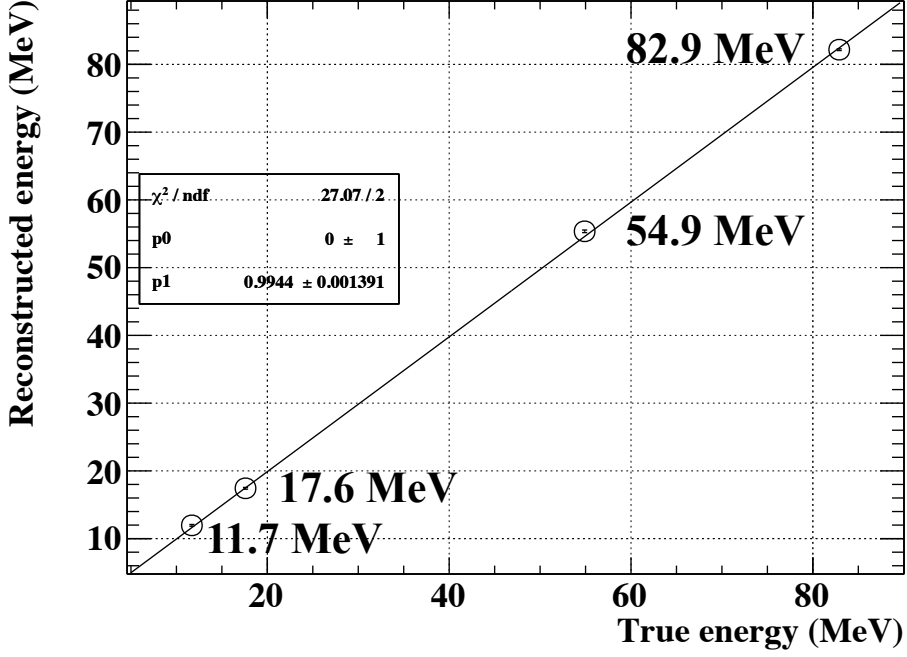
Figure 5.14: Linearity of reconstructed energy. pseudo 2γ data is used.

5.3.3 Time

The time resolution of t_γ is estimated using the CEX data and summarized in Table 5.3. In the CEX run, we took nearly back-to-back 2γ events. These 2γs are emitted at the same time and one enters into the LXe detector, and the other enters into the NAI/BGO detector. By subtracting the time resolution of the NAI/BGO detector, we can get the time resolution of the LXe detector. Figure 5.15 shows energy dependence of the time resolution. Black points show measured time resolution including NAI/BGO contribution in the CEX calibration. Red points show LXe intrinsic resolution; time resolution a preshower counter placed in front of NAI/BGO for timing measurement is subtracted. In the CEX calibration, we obtain the time resolution at 55 MeV and 83 MeV. By fitting them with an energy-dependent function shown in Figure 5.15, its energy dependence and the time resolution at lower energy can be calculated.

A smearing parameter and its systematics are estimated in the same way as positron (Section 4.2) and summarized in Table 5.3.

Table 5.3: Gamma time summary. "+" in the last row indicates the smearing parameter only have the positive side of its systematics because the parameter is estimated to be 0. In 2010, the readout board was modified, resulting in the improvements of timing accuracy from the previous year.

	2009	2010	2011	2012	2013
resolution (ps, Data)	96	67	67	64	66
resolution (ps, MC)	69	69	69	69	69
smearing (ps)	94	0	0	0	0
systematics (ps)	36	34(+)	34(+)	49(+)	40(+)
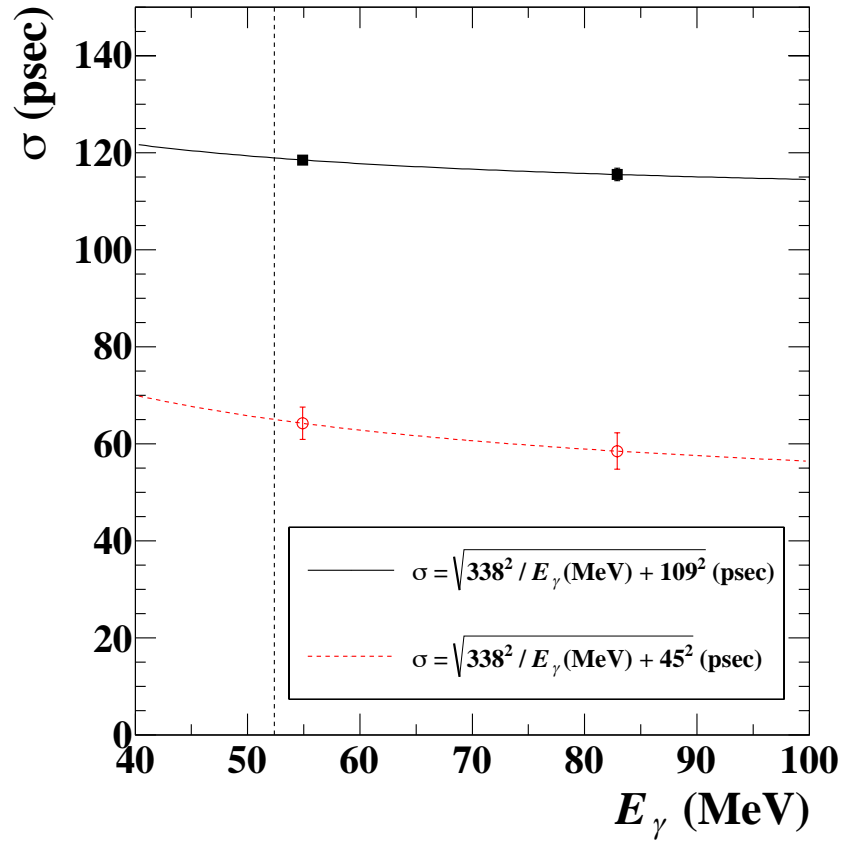
Figure 5.15: Energy dependence of time resolution. Black points show measured timing resolution including NAI/BGO contribution in the CEX calibration. Red points show LXe intrinsic resolution [73].

5.3.4 Efficiency

The 2γ efficiency is used in the estimation of normalization in the MEx2G decay search analysis (Section 7.6). It is defined as

$$(5.8) \qquad \epsilon_\gamma \;\equiv\; P(2\gamma \in \text{acceptance} \,\&\, \gamma\,\text{triggered} \,\&\, 2\gamma\,\text{reconstructed}\,|\,e^+ \in \text{acceptance}).$$

There are three conditions in γ efficiency. It is estimated using the signal MC. First, 2γs should be in the detector acceptance, and then they have suitable energy to fire the trigger. Finally, they should be successfully reconstructed. The definition of each step is summarized as follows:

5.3.4.1 Acceptance

For the MC, 2 γs are supposed to enter the LXe acceptance when they satisfy the following conditions[4]:

- Each energy deposit is more than 5 MeV.

- Total energy deposit is more than 35 MeV.

- γ direction is in an extended detector volume. ($|u| < 28.1\,\text{cm} \wedge |v| < 74.1\,\text{cm}$)

- Position of $X \to \gamma\gamma$ decay vertex is before the LXe (the decay length is less than 67.85 cm).

Figure 5.16 shows gamma acceptance for 20 ps as a function of m_X. Larger mass makes the opening angle between 2γs larger and it makes it difficult for both γs being in the detector acceptance.

5.3.4.2 E_γ trigger efficiency

E_γ trigger requires the energy of γ to be larger than $\sim 40\,\text{MeV}$ (Q_γ). In practice, the threshold was set against the sum of PMT charges (Qsum) depending on the detector status. When triggering with the MEG trigger, we also require a cosmic veto threshold. For the $\mu^+ \to e^+\gamma$ event (52.8 MeV), the threshold ($\sim 65\,\text{MeV}$, Q_veto) is large enough and the veto does not affect the $\mu^+ \to e^+\gamma$ event. However, it does affect the $\mu^+ \to e^+ X, X \to \gamma\gamma$ event. Therefore we need to estimate E_γ trigger efficiency taking both of Q_γ and Q_veto into account.

Table 5.4 shows the history of E_γ trigger threshold (Q_γ and Q_veto). The run periods are determined by the cosmic veto threshold (Q_veto) and beam conditions. To make trigger efficiency curve from the data, we use CEX calibration data of each year. There are high energy ($\sim 60\,\text{MeV}$) γs in the CEX data. Figure 5.17 illustrates how to estimate gamma trigger efficiency. The offline energy spectrum without any cuts is shown in the blue histogram of Figure 5.17. The Qsum value used in the online trigger logic is also recorded in the data and the spectrum is shown in the right

---

[4]At this step, looser conditions are applied to take detector resolutions into account.

Gamma acceptance (20 ps)
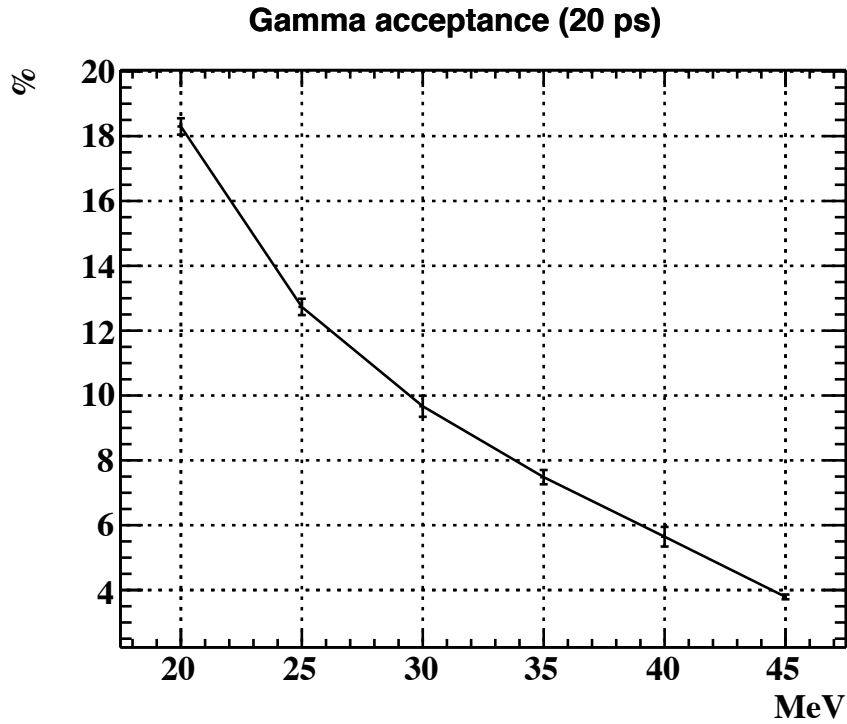
Figure 5.16: Gamma acceptance (20 ps). The error bar includes systematic uncertainty of the beam position.

histogram of Figure 5.17. By applying the Q_γ and Q_{veto} cuts for the $\mu^+ \to e^+ \gamma$ trigger to the CEX data, we can simulate the trigger effects. The red histogram in Figure 5.17 shows the spectrum after the cuts. Finally, the γ trigger efficiency curve is obtained by taking the ratio between these energy spectra. The left bottom figure in Figure 5.17 shows the curve with a run period of 2009a. The drop in the lower region corresponds to Q_γ trigger, and that in the higher region corresponds to Q_{veto} trigger.

Two different functions depending on the run periods are fitted to the obtained efficiency curves. Two sided error function is used for 2009a, 2009b, 2010a, 2010b, and 2011a (left figure in Figure 5.18). Error function with a line (*pol1*) is used for 2011b, 2012, and 2013 (right figure in Figure 5.18). The efficiency curves for all run periods are shown in Figure 5.19.

Table 5.4: Definition of run periods.

| | | | thresholds | |
run period	start run#	end run#	Q_γ(a. u.)	Q_{veto}(a. u.)
2009a	51823	61178	10340	15000
2009b	61218	64487	10100	15000
2010a	73691	74410	11900	16800
2010b	74518	92348	11900	20000
2011a	100035	100070	10100	17000
	100100	123630	10000	17000
2011b	123745	141661	10000	25000
2012	144893	145567	10000	26000
	145617	148750	10700	26000
	148752	152745	10400	26000
	152746	197087	10700	26000
2013	199928	233334	10250	26000
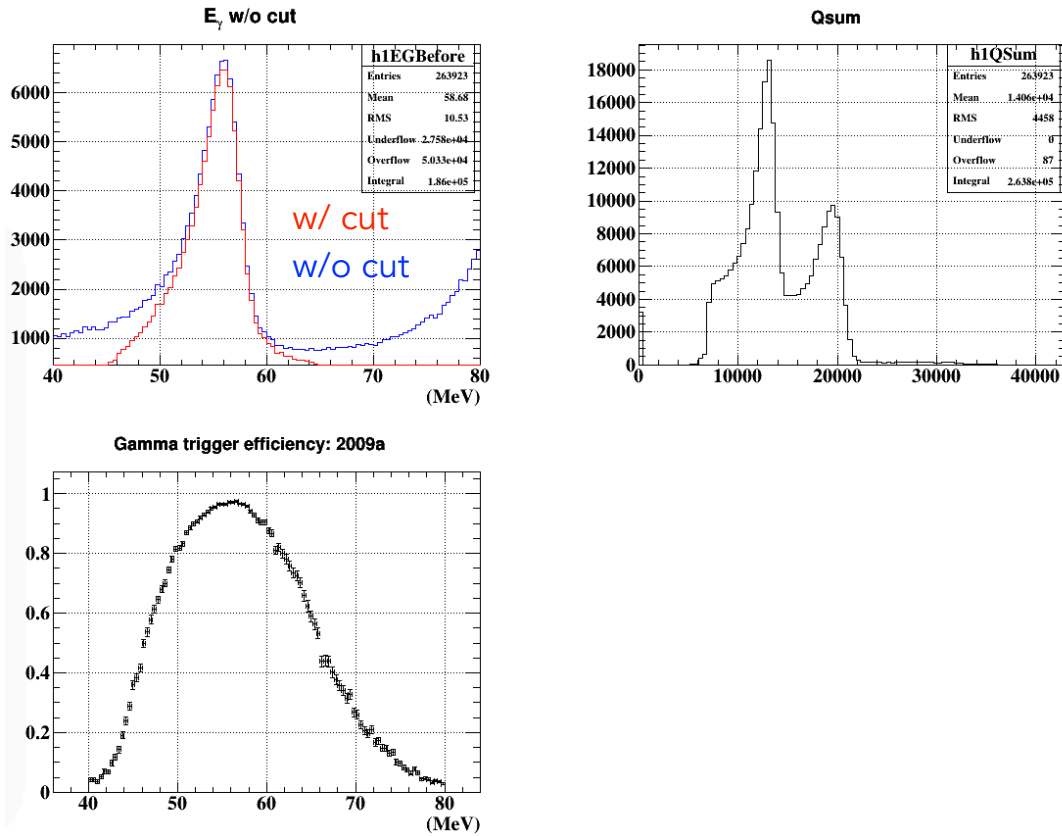
Figure 5.17: An example to estimate gamma trigger efficiency. Top left figure is E_γ spectrum; top right figure is corresponding charge spectrum. 2 peaks come from 55 MeV and 83 MeV in the CEX data; bottom figure is estimated trigger efficiency curve. See text in detail.
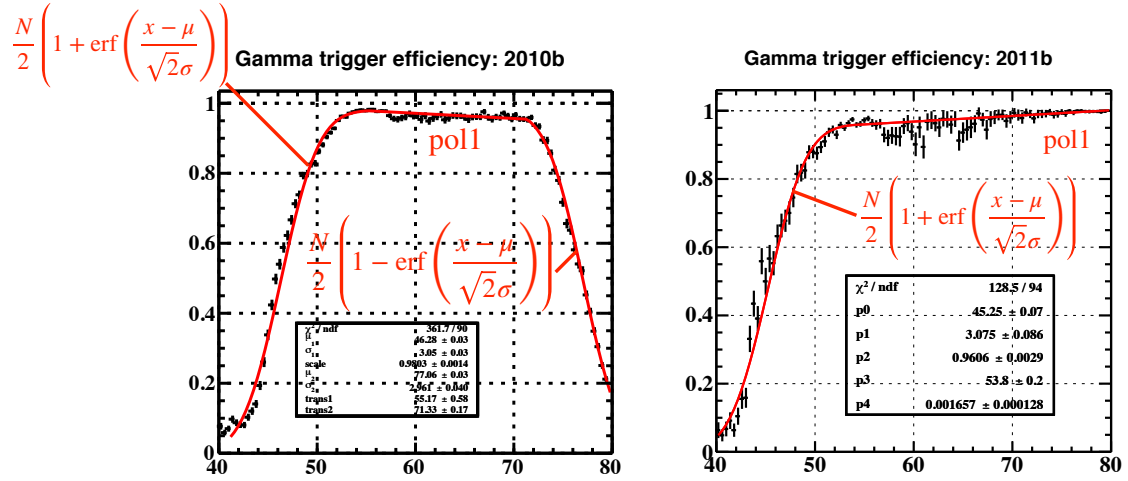
Figure 5.18: Gamma trigger efficiency curve fitting.
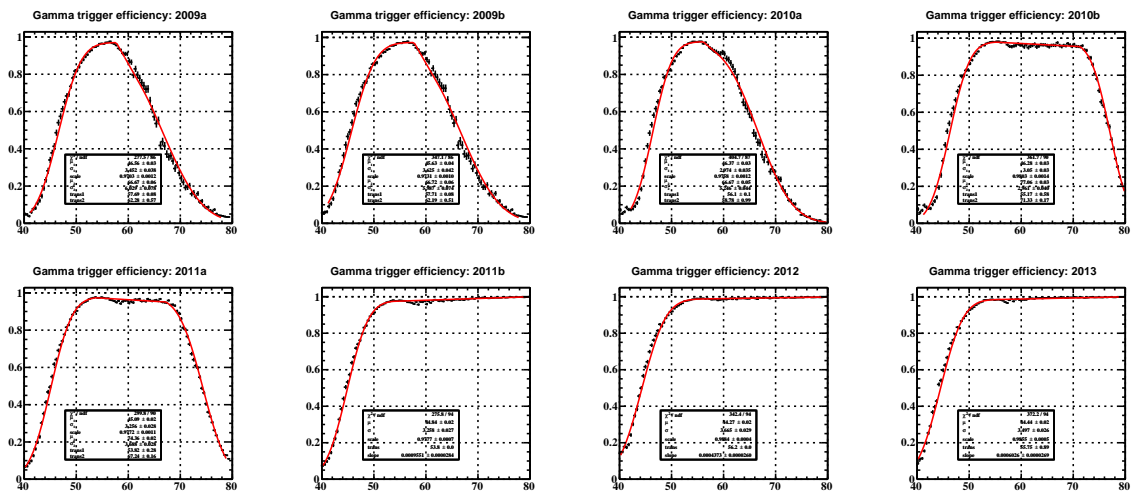
Figure 5.19: Gamma trigger efficiency curve fitting (all run periods)

Since the threshold for the trigger depends on the energy scale, its shift and uncertainty need to be taken into account

Gain drop effect During the MEG data taking, we observed that the PMT gains gradually decreased (at most $\sim 0.2\%$/day) under the μ beam. To keep the effective threshold stable, we updated the online trigger threshold once in 10 days. However, this means that the effective threshold can differ at most 2% before and after changing the online threshold. Therefore we assigned $\pm 1\%$ systematic error on the energy scale to take this effect into account.

Effective threshold difference between the CEX and $\mu^+ \to e^+\gamma$ trigger The gain decrease rate under π beam (CEX trigger) was $\sim$3 times higher than that under the μ beam ($\mu^+ \to e^+\gamma$ trigger). The efficiency curves are estimated using data taken under the π beam in one day without updating the threshold values. Therefore we need to take these differences into account as a correction of the energy scale shift and estimated to be $-0.35\% \pm 0.25\%$.

5.3.4.3   Reconstruction efficiency

2γs are successfully reconstructed if the following conditions for the reconstructed variables are true.

- The 2γ energy and position fitting are converged.

- Each energy deposit is more than 10 MeV.

- Total energy deposit is more than 40 MeV.

- γ directions are in the detector fiducital volume($|u| < 25\,\text{cm} \wedge |v| < 71\,\text{cm}$)

- Distance between 2γ positions in the inner surface is more than $20\,\text{cm}$.

5.3.4.4   Pileup inefficiency

In the evaluation of γ efficiency based on MC, we use the signal MC. However, other γs may come into the detector and it may be reconstructed as one of the signals in the experimental environment. We estimate this effect by using pedestal MC (Section 3.7.2). The pedestal trigger events were taken using the pedestal trigger in the MEG run. The number of photoelectrons and photons are superimposed onto the signal MC to reproduce the background γ environment. We estimate the pileup inefficiency to be at most $\sim 5\%$ and this is incorporated into the γ efficiency.

Gamma efficiency (20 ps)
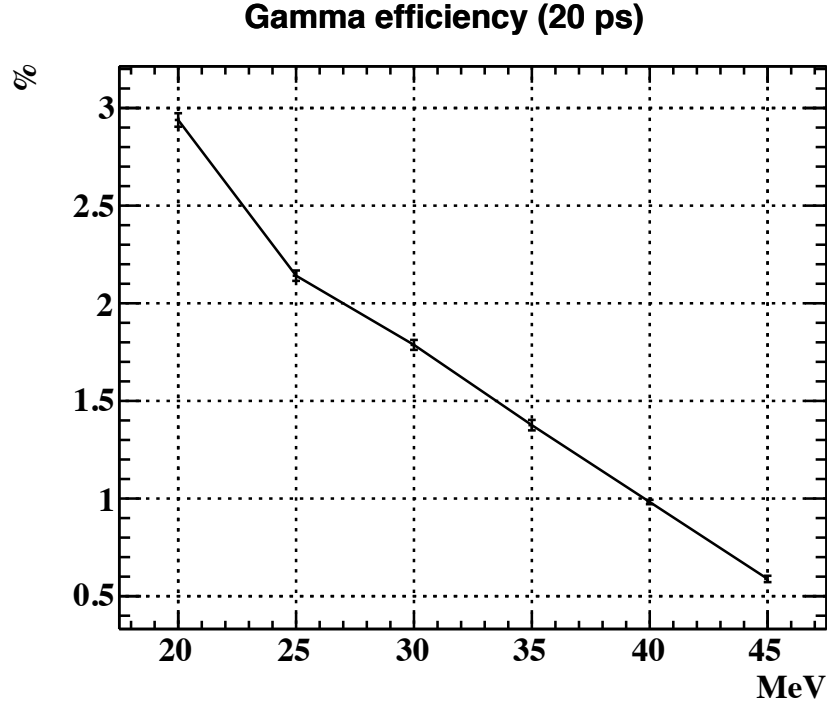
Figure 5.20: Two gamma efficiency (20 ps). The error bar only includes statistical uncertainty. Other systematic uncertainties are included in the estimation of normalization factor (Section 7.6.9).

5.3.4.5 Detection efficiency

γ detection efficiency in the $\mu^+ \rightarrow e^+\gamma$ analysis is estimated to be $62.5 \pm 2.3\%$ (relative uncertainty: $r_{1\gamma} = 3.7\%$) [6]. This relative uncertainty estimated for the $\mu^+ \rightarrow e^+\gamma$ event (1γ) is used to estimate the relative uncertainty for the MEx2G event (2γ). The 2γ efficiency can be changed from $(1-r_{1\gamma})^2$ to $(1+r_{1\gamma})^2$ and the difference from 1 is assumed to be its relative uncertainty. 7.4% is assigned to the systematic uncertainty of 2γ efficiency.

5.3.4.6 Summary

After applying all the conditions above, 2γ efficiencies are estimated and shown in Figure 5.20. These values are weighted averaged efficiencies using normalization of each year. No significant difference among lifetimes is observed and the results of 5 ps and 40 ps are summarized in Appendix B. m_X dependence comes from gamma acceptance; efficiencies except for the gamma acceptance is $\sim 16\%$ for all m_X.

COMBINED RECONSTRUCTION AND PERFORMANCE

I n this chapter, a combined reconstruction, especially, $X \to \gamma\gamma$ decay vertex reconstruction is discussed. A maximum likelihood fitting is used for the vertex estimation.

6.1 Combined reconstruction

The main goal for the combined analysis is to reconstruct the decay vertex position of X ($X \to \gamma\gamma$ vertex). Hereafter we assume m_X in the reconstruction. A different mass assumption gives a different reconstructed result. m_X is assumed to be 20–45 MeV at 1 MeV intervals, which is small enough not to lose the signal in between[1].

6.1.1 X decay vertex
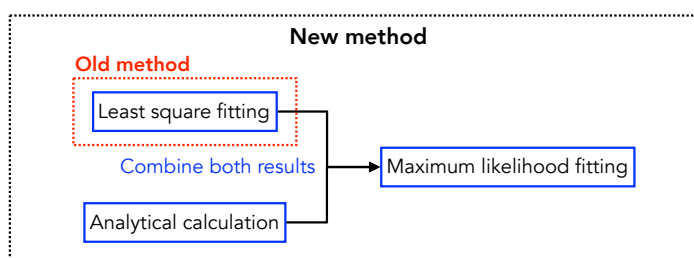
Figure 6.1: Overview of the vertex reconstruction.

Figure 6.1 shows the overview of the vertex reconstruction. There are two parts: the calculation of initial values and maximum likelihood fitting. There are two ways to calculate initial

---

[1]This interval is revisited in Section 7.1.
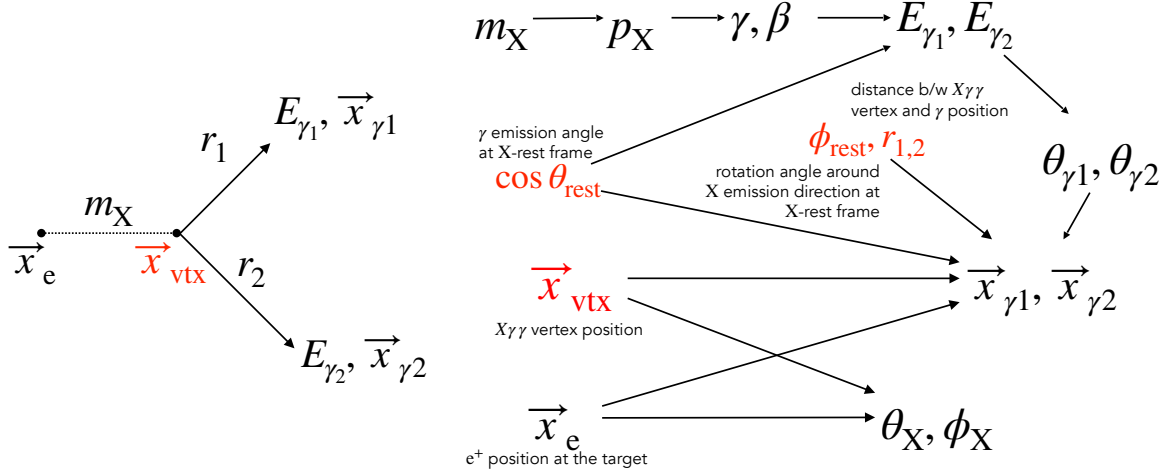
Figure 6.2: Left: definition of variables. Right: relations between variables in the vertex reconstruction. Fitting parameters are shown in red.

values: least-square fitting and analytical calculation. These two methods have a different bias (tail distribution) on the vertex reconstruction. Thus it is better to combine both results as an initial value. In the previous study in 2010, the least square fitting was used[62]. Analytical calculation is newly implemented in this study. The maximum likelihood fitting enables us to use the more realistic probability density function (PDF) in the reconstruction than the least square fitting. This is also new.

6.1.1.1 Calculation of initial values

Least square fitting   The variables below are calculated kinematically when $m_X, \theta_{rest}, \phi_{rest}, r_{1,2}, \vec{x}_{vtx}, \vec{x}_e$ are given:

$$E_{\gamma 1,2}(m_X, \theta_{rest})$$
$$\vec{x}_{\gamma 1,2}(m_X, \theta_{rest}, \phi_{rest}, r_{1,2}, \vec{x}_{vtx}, \vec{x}_e)$$
$$\theta_X(\vec{x}_{vtx}, \vec{x}_e)$$
$$\phi_X(\vec{x}_{vtx}, \vec{x}_e),$$

where θ_X, ϕ_X are X direction. The relation between variables are shown in Figure 6.2. $\theta_{rest}, \phi_{rest}$ are defined in Figure 2.6; θ_{rest} is γ emission angle at X-rest frame (in this frame, 2γs are emitted in the opposite direction); ϕ_{rest} is rotation angle around X-emitted direction at X-rest frame; $r_{1(2)}$ is distance between X decay vertex ($X \to \gamma\gamma$) and γ interaction position in the LXe detector (the left figure of Figure 6.2); $\vec{x}_{vtx}, \vec{x}_e$ are vertex position and e^+ position at the target, respectively.

Our task is to find $\vec{x}_{vtx}$. m_X is fixed and measured value is used for $\vec{x}_e$. $\theta_{rest}, \phi_{rest}, r_{1,2}$ are
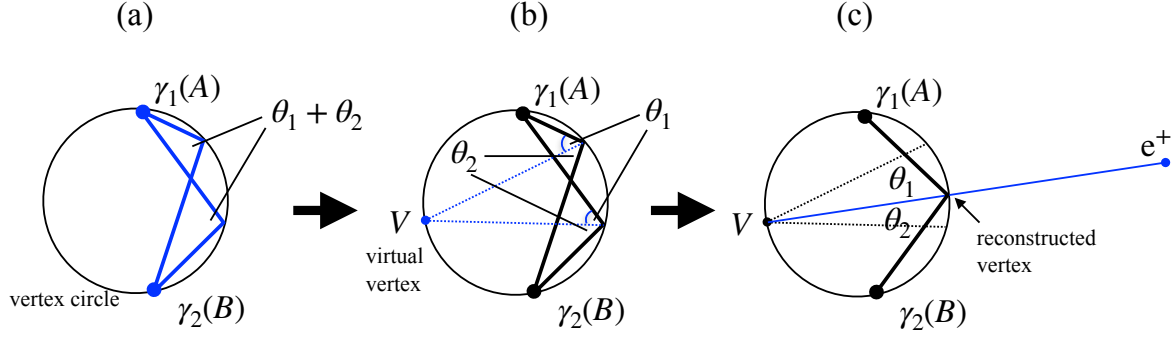
Figure 6.3: Analytical calculation of the X decay vertex. See text in detail.

fitted but not to be used in the later analysis. χ^2 to be minimized is given by

(6.1) χ^2_{LS}

$$= \chi^2_{\mathrm{LS}}(E_{\gamma 1,2}, \vec{x}_{\gamma 1,2}, \theta_{\mathrm{X}}, \phi_{\mathrm{X}}) = \chi^2_{\mathrm{LS}}(\theta_{\mathrm{rest}}, \phi_{\mathrm{rest}}, r_{1,2}, \vec{x}_{\mathrm{vtx}})$$

$$= \sum_{\gamma=\gamma 1,\gamma 2} \left(\frac{E_\gamma^{\mathrm{measured}} - E_\gamma}{\sigma_{E_\gamma}} \right)^2 + \sum_{\gamma=\gamma 1,\gamma 2} \left(\frac{\vec{x}_\gamma^{\mathrm{measured}} - \vec{x}_\gamma}{\sigma_{\vec{x}}} \right)^2 + \left(\frac{\theta_{\mathrm{X}}^{\mathrm{measured}} - \theta_{\mathrm{X}}}{\sigma_\theta} \right)^2 + \left(\frac{\phi_{\mathrm{X}}^{\mathrm{measured}} - \phi_{\mathrm{X}}}{\sigma_\phi} \right)^2,$$

where variables with $^{\mathrm{measured}}$ in Equation (6.1) indicates are measured variables from e+ and γ reconstruction. $\theta_{\mathrm{X}}^{\mathrm{measured}}$ and $\phi_{\mathrm{X}}^{\mathrm{measured}}$ are defined as the opposite direction of e^+. Resolutions of each variable is assigned to its uncertainty (σ) in the χ^2.

Analytical calculation   Another method to find the vertex position is an analytical calculation. We have enough information to analytically calculate the vertex position. The calculation can be summarized as follows:

- Given: mass of X (m_X), energy of 2γ ($E_{\gamma 1,2}$), position of 2γ ($(u,v,w)_{1,2}$), position of e^+ at the target ($(x,y,z)_{e^+}$)

- Find: position of the X$\rightarrow \gamma\gamma$ vertex ($(x,y,z)_X$)

Figure 6.3 shows how to calculate the vertex. The calculation below works on the e^+-γ_1-γ_2-surface.

1. Calculate θ_1 and θ_2 from $E_{\gamma_1}, E_{\gamma_2}$,and m_X (Equation (2.11) and Equation (2.11)).

2. Calculate vertex circle with the same inscribed angle (Figure 6.3 (a)).

3. Calculate virtual vertex (V, in blue) which divides circumference AB with $AV : VB = \theta_1 : \theta_2$ (Figure 6.3 (b))

4. Make line (in blue) between X and e^+ at the target to find the reconstructed vertex (Figure 6.3 (c)).
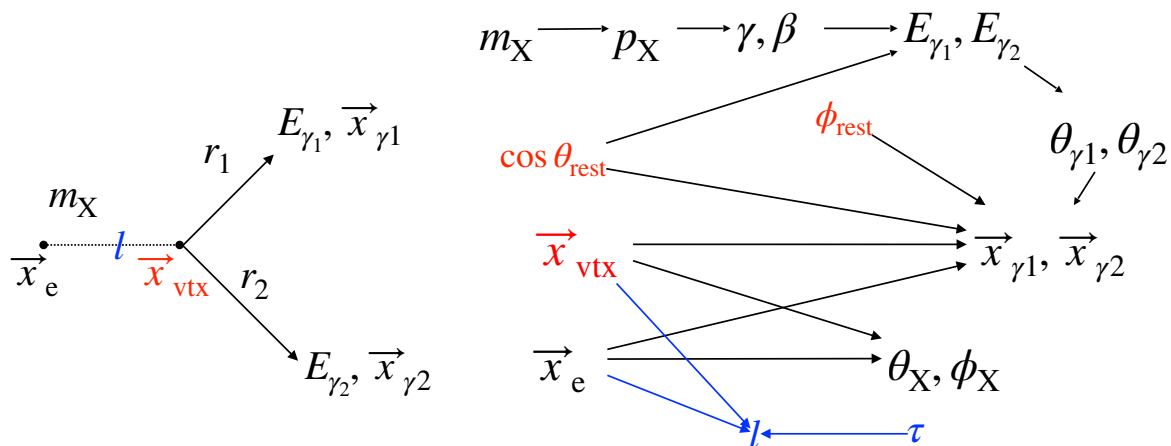
Figure 6.4: Left: definition of variables. Right: relations between variables in the vertex reconstruction. Fitting parameters are shown in red. Differences from Figure 6.2 are shown in blue.

6.1.1.2 Maximum likelihood fitting

The least-square fitting and the analytical calculation can roughly find the vertex position but the uncertainty of the reconstructed variables cannot be incorporated correctly. Thus both results are used as the initial value of the maximum likelihood fitting. The maximum likelihood fitting is performed 50 times for each initial value by changing its initial value randomly at each time to avoid a local minimum of the minimization. The parameter set which provides the maximum likelihood is used as the final result. Residuals are also calculated to use it for vertex quality cut. Relations between parameters are illustrated in Figure 6.4. The differences from the least square fitting are:

- Implement energy dependent and asymmetric PDF for $E_{\gamma1}$ and $E_{\gamma2}$.

- Implement double Gaussian position PDF for (u, v, w).

- Implement decay length (l) information into the fitting.

Observables are the following:

$$X = (E_{\gamma1}, E_{\gamma2}, \vec{x_1}, \vec{x_2}, \vec{x_e})$$

Fitting parameters are the following:

$$\Theta = (\cos\theta_{\text{rest}}, \phi_{\text{rest}}, \vec{x}_{\text{vtx}})$$
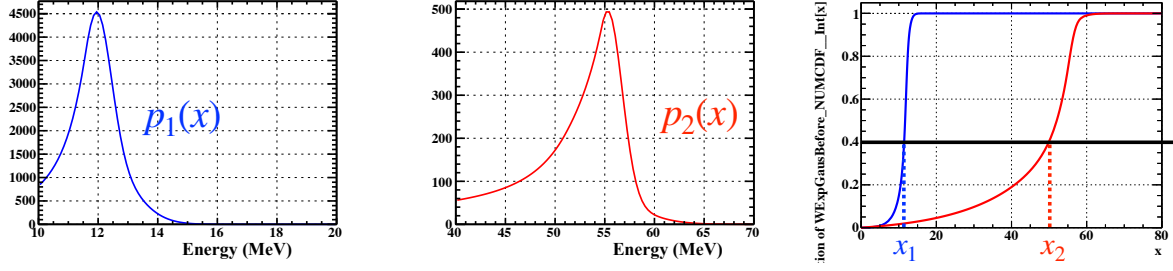
Figure 6.5: PDF morphing technique. Left: pseudo Boron 2γ data. Center: pseudo CEX 2γ data. Right: cumulative distribution of the other two distributions.

The likelihood function $L(\Theta)$ is defined as below:

$$
\begin{aligned}
L(\Theta) \;=\;\; & P(E_{\gamma 1}|\cos\theta_{\rm rest}, m_X) \\
\times\;\; & P(E_{\gamma 2}|\cos\theta_{\rm rest}, m_X) \\
\times\;\; & P(\vec{x_1}|\cos\theta_{\rm rest}, \phi_{\rm rest}, \vec{x}_{\rm vtx}, \vec{x}_e, m_X) \\
\times\;\; & P(\vec{x_2}|\cos\theta_{\rm rest}, \phi_{\rm rest}, \vec{x}_{\rm vtx}, \vec{x}_e, m_X) \\
\times\;\; & P(\theta_e|\vec{x}_{\rm vtx}, \vec{x}_e) \\
\times\;\; & P(\phi_e|\vec{x}_{\rm vtx}, \vec{x}_e) \\
\times\;\; & P(l|\vec{x}_{\rm vtx}, \vec{x}_e, \tau)
\end{aligned}
$$

$E_{\gamma 1}$ and $E_{\gamma 2}$ are assumed to be independent thanks to the ratio correction (Section 5.2.4) in the 2γ reconstruction. $P(\vec{x}_e|\vec{x}_e{}^{\rm true})$ term is neglected by using measured value $\vec{x}_e$ to reduce fitting parameters. PDFs used for the maximum likelihood fitting are assumed to be single Gaussian functions only for positron angles. Descriptions of other PDFs are the following.

Gamma energy PDF   The E_γ PDF has the following information:

- Energy-dependent PDF

- Lower tail

To make full use of this information, a PDF morphing technique is used to make the E_γ PDF. Figure 6.5 and Figure 6.6 illustrate how the PDF morphing [104] works . The first two figures in Figure 6.5 are the original PDFs. Both are fitted spectrum of E_γ distribution of different energies using pseudo data. The task here is to interpolate and find PDFs between these two PDFs.

First, we normalize the original distribution (left and center distributions in Figure 6.5) and make the cumulative distribution form these distributions as shown in the right figure in Figure 6.5. Then we find x_1, x_2 such that

$$
\int_0^{x_1} p_1(x)dx = \int_0^{x_2} p_2(x)dx = y,
$$
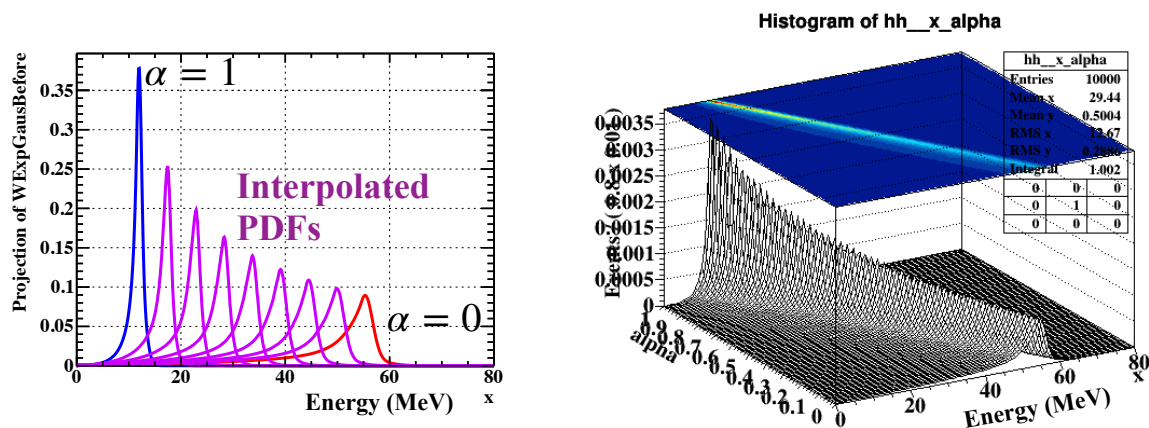
Figure 6.6: PDF morphing.

where y ranges from 0 to 1. Cumulative distribution function of the interpolated PDF $p(x)$ is defined as

$$(6.2) \qquad \int_0^{\alpha x_1 + (1-\alpha)x_2} p(x)dx = y,$$

where α is a interpolation parameter ranging from 0 to 1. $\alpha = 0$ and $\alpha = 1$ correspond to the original distributions. Equation (6.2) gives the form of interpolated PDF:

$$p_{\text{interpolate}}(\alpha x_1 + (1-\alpha)x_2) = \frac{p_1(x_1)p_2(x_2)}{\alpha p_2(x_2) + (1-\alpha)p_1(x_1)}.$$

Figure 6.6 shows the resulting PDFs. By definition, α is proportional to the energy. We can get any PDF with selected energy between the original PDFs by selecting α.

Interpolated PDFs are not exactly the same as the true PDF. It is because this interpolation procedure is linear while the width of the true PDF is not linear to its energy. However, it is good enough for our use on the fact that the original purpose for the maximum likelihood fitting is already achieved.

Gamma position PDF The difference between reconstructed position and true position has tail components and it is fitted with a double Gaussian function. There are two ways to implement gamma position PDFs. The first one is using the product of three independent double Gaussian PDFs assuming no correlation. The second one is a 3D histogram to take all the correlations into account. There is no significant difference in reconstructed distributions between these two ways. For the simplicity, double Gaussian PDF is used as gamma position PDFs.

Decay length PDF The likelihood function up to here has small information on the direction along X. The decay length l of X is also implemented into the likelihood function:

$$(6.3) \qquad p(l, \text{m}_{\text{X}}, \tau) = \gamma \beta c \tau \cdot e^{-\frac{l}{\gamma \beta c \tau}}.$$

This PDF is defined and normalized in $l \geq 0$. We fix $\tau = 20\,\text{ps}$ because no significant change in vertex performance by assuming different τ for the assumed τ range in this analysis.

6.1.2 Goodness of the vertex fitting, momentum, and time difference

The following items are calculated when the X decay vertex position is determined.

6.1.2.1 Goodness of the vertex fitting

We define goodness of the vertex fitting as the following χ^2. This value is to be used in a event selection of the MEx2G signal (Section 7.4).

$$\chi^2 = \chi^2_{\text{LS}} + \chi^2_l, \tag{6.4}$$

where the first term comes from Equation (6.1) and the second term comes from the decay length and defined as follows:

$$\chi^2_l = \left(\frac{l_{\text{best}}}{\gamma\beta c\tau}\right)^2. \tag{6.5}$$

Since we use maximum likelihood fitting, the χ^2 is not minimized during the fitting. Instead, we calculate the χ^2 using maximum likelihood fitted parameters by assigning a single Gaussian resolution to σ of each variable.

6.1.2.2 Momentum

After the vertex reconstruction, we can calculate the momentum of each γ from their positions and energies. Finally, the sum of these momenta together with the e^+ momentum is calculated and this is used for the signal selection by requiring the momentum conservation.

$$\vec{P}_{\text{sum}} \equiv \vec{P}_{e^+} + \vec{P}_{\gamma_1} + \vec{P}_{\gamma_2} \tag{6.6}$$

6.1.2.3 Time difference

As we discuss in Chapter 7, blinding and background estimation are performed on the $t_{\gamma\gamma}$–$t_{\gamma_1 e}$ surface, where $t_{\gamma\gamma}$ is time difference between 2γs and $t_{\gamma_1 e}$ is time difference between γ_1 and positron. These time differences are calculated from the reconstructed time: $t_{\gamma_1}, t_{\gamma_2}, t_e$. t_{γ_1} and t_{γ_2} are reconstructed time inside the LXe detector and t_e is reconstructed time at the target.

Figure 6.7 illustrates schematics of the MEx2G decay. $r_{1(2)}$ is distance between $\gamma_{1(2)}$ and the vertex position of $X \to \gamma\gamma$. l is distance between the vertex position of $X \to \gamma\gamma$ and that of $\mu^+ \to e^+ X$. $t_{\gamma\gamma}$ should be 0 at the vertex position of $X \to \gamma\gamma$. Thus $t_{\gamma\gamma}$ is calculated as

$$t_{\gamma\gamma} = (t_{\gamma_1} - \frac{r_1}{c}) - (t_{\gamma_2} - \frac{r_2}{c}). \tag{6.7}$$
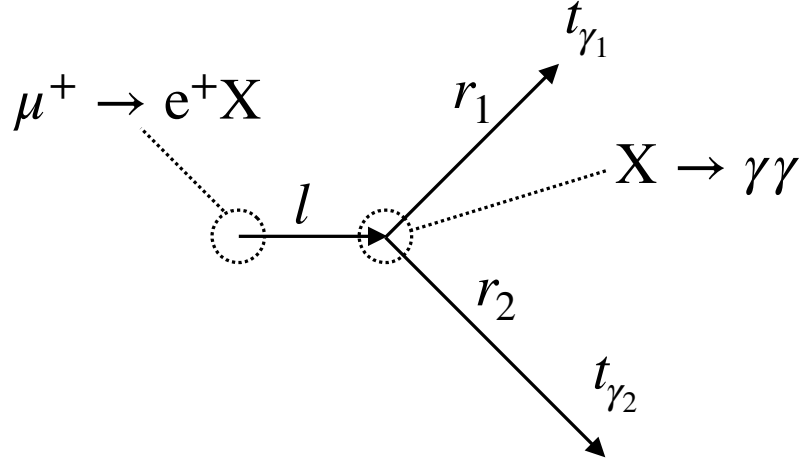
Figure 6.7: Calculation of time differences.

$t_{\gamma_1 e}$ should be 0 at the vertex position of $\mu^+ \rightarrow e^+ X$. Thus $t_{\gamma_1 e}$ is calculated as

$$(6.8) \qquad t_{\gamma_1 e} = (t_{\gamma_1} - \frac{r_1}{c} - \frac{l}{\beta c}) - t_e.$$

6.2 Performance

6.2.1 X decay vertex

Vertex resolution is estimated using MC. These results are used in the vertex quality cut. All the inputs in the vertex reconstruction are smeared to reproduce data and thus the vertex resolutions evaluated in the MC should reproduce expected performance in data. Therefore, the following performances are not directly used in the MEx2G decay search (Chapter 7).

Vertex position   The vertex resolution is evaluated by the difference between the reconstructed position and the MC truth. Figure 6.8 shows these distributions in (x, y, z). The distributions in red are the final results and those in blue are the results with the old reconstruction method in 2010 without using the maximum likelihood. All the distributions become sharper than before; RMS of the distributions become smaller by 35–45%. In addition, tail events are reduced and this is expected to increase the signal efficiency.

Vertex quality   An important variable to reduce the background events while keeping the signal is one related to the vertex quality. The vertex quality is defined as a χ^2 (Equation (6.4)) of all the related variables after the maximum likelihood fitting in the vertex reconstruction. The χ^2 distribution as a function of assumed mass is shown in Figure 6.9. To show its dependence
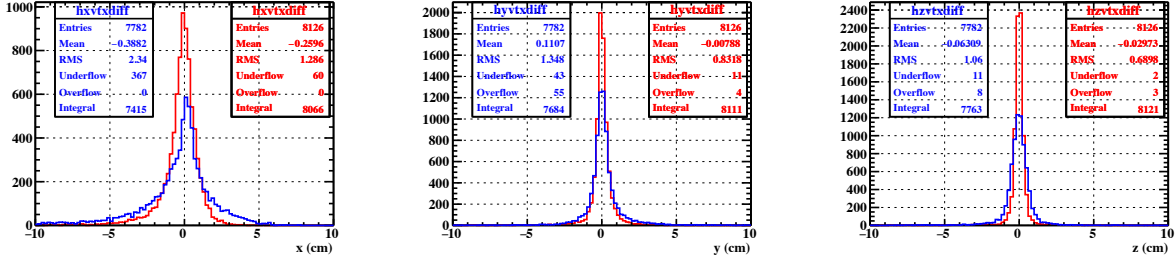
Figure 6.8: An example of vertex position resolution. (m_X, τ)=(30 MeV, 20 ps). The results from the old reconstruction are shown in blue and those from the new one are shown in red.
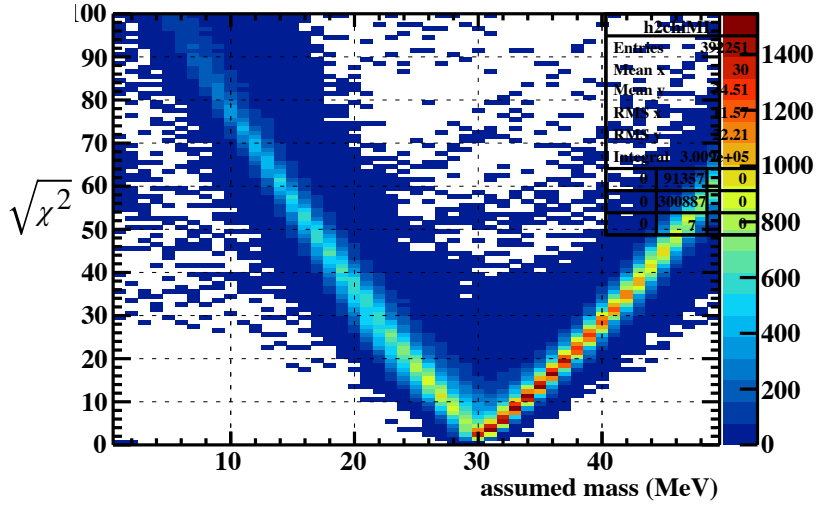
Figure 6.9: $\sqrt{\chi^2}$ distribution. $\sqrt{\chi^2}$ is plotted to clearly see the mass dependence. If the assumed mass and true mass differ, $\sqrt{\chi^2}$ becomes larger. (m_X, τ)=(30 MeV, 20 ps)

clearly, we take a square root of χ^2. The MC truth is (m_X, τ)=(30 MeV, 20 ps) in this case. If the assumed mass is the same as the truth, the χ^2 takes its minimum value. If the assumed mass becomes far from the truth, the χ^2 gets larger and larger. We discuss signal and background separation on the χ^2 in Section 7.4.

6.2.2 Time difference

Both $t_{\gamma\gamma}$ and $t_{\gamma_1 e}$ are explained in Section 6.1.2.3. $t_{\gamma_1 e}$ includes contribution from different detectors and thus time difference between detectors needs to be calibrated. After the detector based calibrations, t_e is adjusted (t_{shift}) by using the position of the RMD peak. Figure 6.10 shows a distribution of $t_{e\gamma}$ for the $\mu^+ \to e^+ \gamma$ trigger after the adjusting. Therefore, $t_{e\gamma}$ is given by

$$t_{\gamma_1 e} = (t_{\gamma_1} - \frac{r_1}{c} - \frac{l}{\beta c}) - (t_e + t_{shift})$$

85
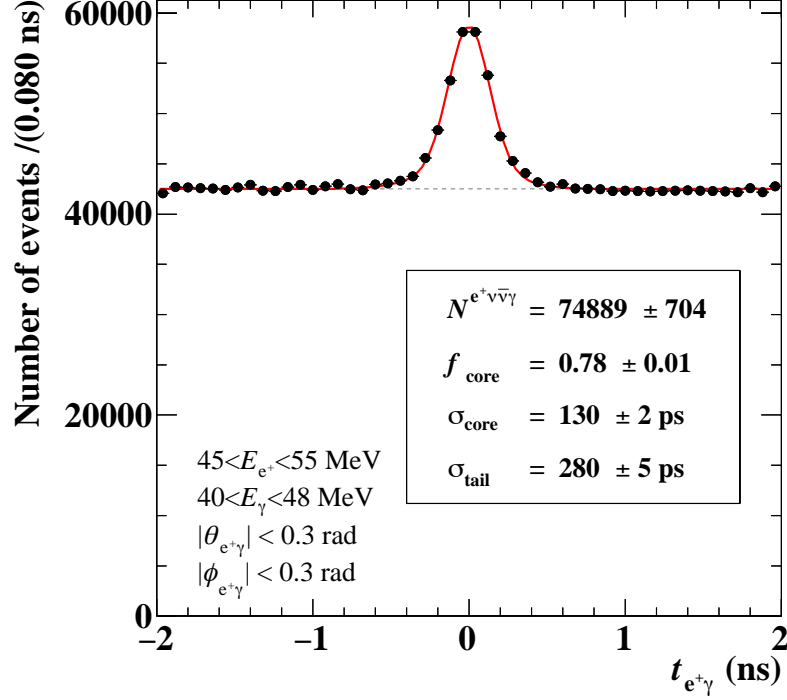
Figure 6.10: RMD peak [6]. All the data from 2009 to 2013 are accumulated and the peak position is already adjusted to 0.

$t_{\gamma_1 e}$ needs to be further shifted to take a time reconstruction bias into account. MC studies show that there is a bias between $\mu^+ \to e^+\gamma$ time reconstruction and MEx2G time reconstruction as shown in Figure 6.11. The bias depends on the energy of γ and the difference of energy-dependent time correction can cause bias. However, this correction is not applied on the MEx2G time reconstruction because the correction factor is calculated for $\mu^+ \to e^+\gamma$ energy range and not suitable for the MEx2G case. $t_{\gamma_1 e}$ is shifted by the mean value of this distribution (the final term is added):

$$t_{\gamma_1 e} = (t_{\gamma_1} - \frac{r_1}{c} - \frac{l}{\beta c}) - (t_e + t_{\text{shift}}) + \Delta(t_{1\gamma} - t_{\gamma 1}),$$

where $t_{1\gamma}$ is reconstructed time from the MEG (1γ) reconstruction and $t_{\gamma 1}$ is reconstructed time from the MEx2G reconstruction; $\Delta(t_{1\gamma} - t_{\gamma 1})$ is calculated using time sideband data (defined in Section 7) and summarized in Table 6.1.

To confirm the blind box defined in Chapter 7) is large enough, we roughly estimate the time resolution of $t_{\gamma_1 e}$ and $t_{\gamma\gamma}$. In MC, difference between reconstructed time and MC truth is fitted with Gaussian and its sigma is quated as time resolution. Then, data/MC difference is also taken into account. $t_{\gamma_1 e}$ resolution is estimated to be ~140 ps. We also observe a long tail in the negative region. This suggests that the blinding window should be large enough to cover most of the tail.

T1g-T2g (gamma1)

ht1g2g	
Entries	30011
Mean	−4.27e−10
RMS	5.7e−10
Underflow	4967
Overflow	1910
Integral	2.31e+04
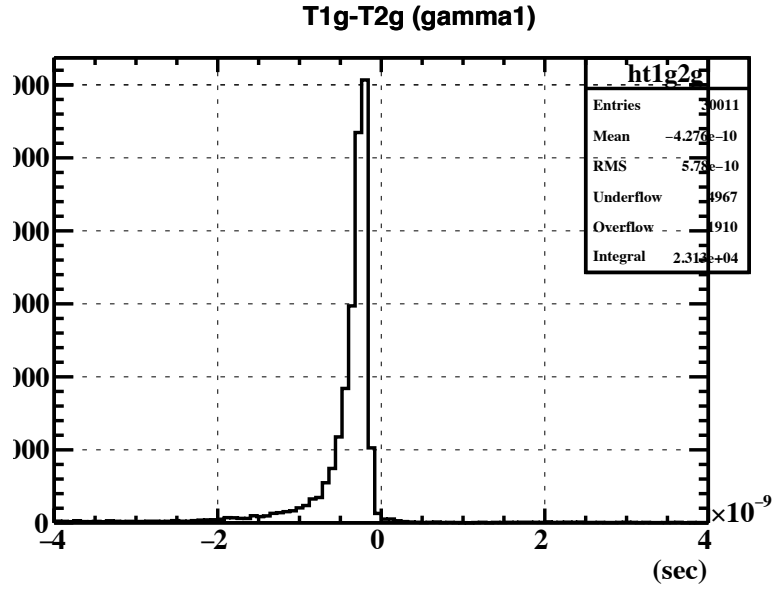
Figure 6.11: Time difference between MEG and MEx2G time reconstruction (MC: 30 MeV, 10 ps).

Table 6.1: Time shift calculated from the time sideband

year	mean (ns)
2009	−0.243
2010	−0.231
2011	−0.209
2012	−0.216
2013	−0.261

We set the window size as 1 ns, which is > 5 times larger value than this resolution. $t_{\gamma\gamma}$ resolution is estimated to be ∼160 ps. We also set the window size as 1 ns.

87

$\mu^+ \rightarrow \text{e}^+\text{X}, \text{X} \rightarrow \gamma\gamma$ **DECAY SEARCH ANALYSIS**

7.1 Overview

As we discussed in Chapter 6, we assume m_X in the X decay vertex reconstruction. Thus the $\mu^+ \rightarrow \text{e}^+\text{X}, \text{X} \rightarrow \gamma\gamma$ decay search analysis is also performed assuming m_X. A mass range of 20–45 MeV (Section 2.3.5) is searched at an interval of 1 MeV. This interval is set small enough not to miss signals between the interval.[1] Therefore, adjacent mass bins are not statistically independent; a single event can be observed in several bins. We also assume the lifetime of X to be 5, 20, and 40 ps, but the lifetime difference only affects the signal efficiency.

A blind analysis is used to reduce the experimenter's bias. It means that we determine conditions of event selection and estimate the number of background events without looking at a signal region. Figure 7.1 schematically shows definitions of the blind region, the signal region and time sideband regions. As we discussed in Section 2.5, we only need to consider accidental backgrounds. Thus, the signal region and the blind region are defined on a time surface, $t_{\text{e}\gamma_1}$–$t_{\gamma\gamma}$ surface[2]. The signal region is defined at the center region of the $t_{\text{e}\gamma}$–$t_{\gamma\gamma}$ surface as shown in red shadow in Figure 7.1 Its box size is optimized in Section 7.4. The blind region is predefined to be $|t_{\text{e}\gamma_1}| < 1\,\text{ns} \wedge |t_{\gamma\gamma}| < 1\,\text{ns}$, which is large enough to blind the signal. It is shown in blue shadow in Figure 7.1. We also define time sideband regions A, B, and C and these regions are used as control samples to estimate backgrounds and check the analysis method. See Section 7.2 and Section 7.4 for the other definitions in Figure 7.1.

We adopt a cut-counting analysis for the signal search; we apply several cuts to reduce background events as much as possible while keeping the number of signal events. Before applying these signal selections, all the other selections including trigger effects have been already

---

[1] Effective mass resolution in the vertex reconstruction is ~ 2.5 MeV.

[2] See Section 6.1.2.3 for the definitions. Hereafter we use $t_{\text{e}\gamma}$ instead of $t_{\text{e}\gamma_1}$
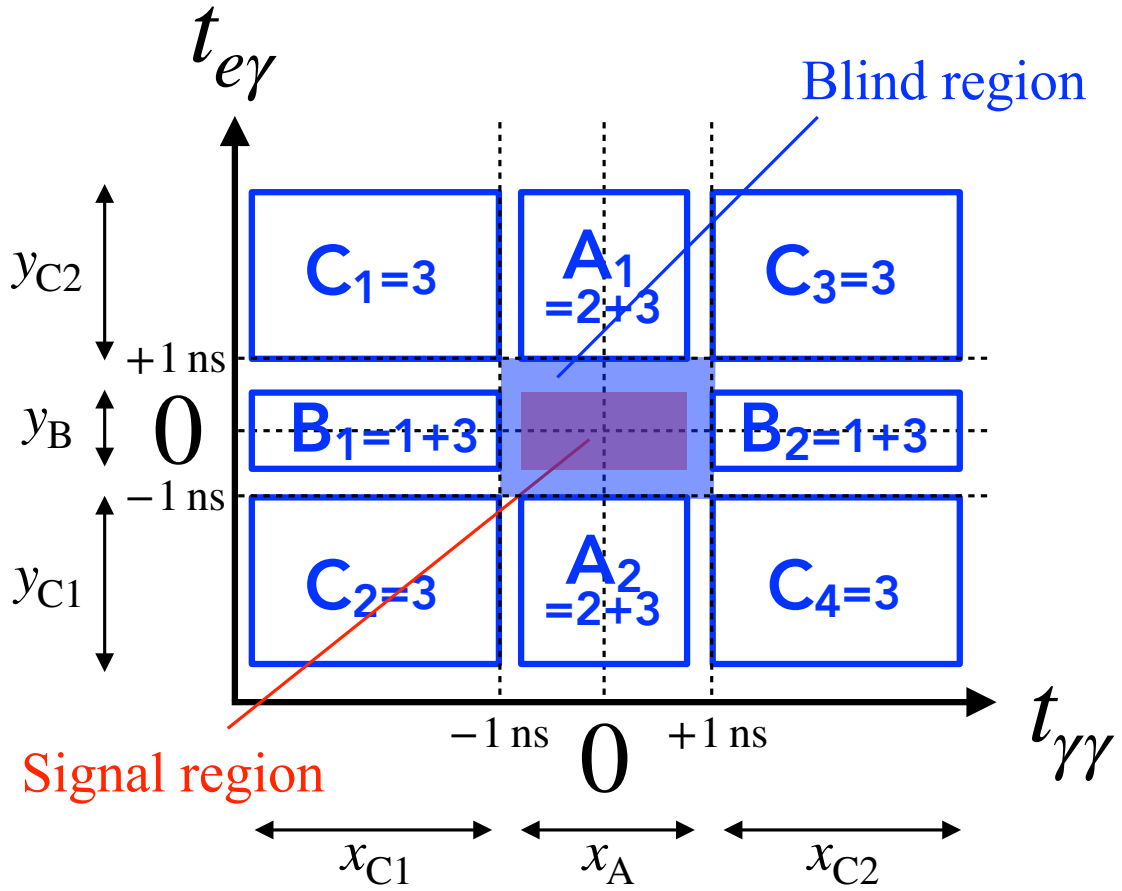
Figure 7.1: Definitions of blind region, signal region, and time sideband regions.

applied. The number of signals and backgrounds in the signal region are **simultaneously estimated by a maximum likelihood fit** using events observed in the sidebands and the signal region (Section 7.2, Section 7.3). The confidence interval of the number of signal events and the significance of the signal, if excess is found, are calculated in a frequentist approach.

The branching ratio (BR) of the MEx2G decay is obtained as the product of the number of signals and a single event sensitivity (SES). SES is defined as a BR which gives 1 event observation and is inverse of the effective number of observed muons decaying at the target. This is explained in Section 7.6.

Finally, we open the blind region after all the preparations above are fixed and obtain the results (Section 7.8).

7.2 Background estimation

In this section, we discuss how we estimate the expected number of background events in the signal region from the number of events in the sidebands. There are 8 time sideband regions ($A_{1,2}$, $B_{1,2}$, $C_{1,2,3,4}$) as shown in Figure 7.1. Different alphabets indicate different off-timing sidebands. As we discussed in Section 2.5, there are three types of accidental background events: type 1, type 2, and type 3. The numbers written in right-hand side of the each box in Figure 7.1 show which type contributes to each sideband; $A_{1,2}$ has contributions from type 2 and type 3. $B_{1,2}$ has contributions from type 1 and type 3. $C_{1,2,3,4}$ has contribution from type 3.

The expected number of background events (N_{BG}) in the signal region is given by

$$N_{BG} = N_1 + N_2 + N_3,$$

where N_1, N_2, N_3 are the expected number of background events in the signal region from type 1, type 2, and type 3, respectively. The likelihood on N_{BG} is given by

$$\mathscr{L}(N_{BG}|N_A, N_B, N_C)$$
$$(7.1) \qquad = \quad \text{Poi}(N_A|N_A^{exp})\text{Poi}(N_B|N_B^{exp})\text{Poi}(N_C|N_C^{exp}),$$

where $\text{Poi}(n|\mu)$ is defined as

$$\text{Poi}(n|\mu) = \frac{\mu^n e^{-\mu}}{n!}$$

We define N_A, N_B, and N_C as the observed number of events in A, B, and C, respectively. For instance, N_A is sum of the number of events in A_1 and A_2. We also define the expected number of N_A, N_B, and N_C as N_A^{exp}, N_B^{exp}, and N_C^{exp}, respectively. N_A^{exp}, N_B^{exp}, and N_C^{exp} can be calculated from N_1, N_2, and N_3 assuming the background distribution is linear[3]:

$$(7.2) \qquad N_A^{exp} \quad = \quad N_2 \times \frac{y_{C1} + y_{C2}}{y_B} + N_3 \times \frac{y_{C1} + y_{C2}}{y_B}$$

$$(7.3) \qquad N_B^{exp} \quad = \quad N_1 \times \frac{x_{C1} + x_{C2}}{x_A} + N_3 \times \frac{x_{C1} + x_{C2}}{x_A} + N_2 \times f_{escape}$$

$$(7.4) \qquad N_C^{exp} \quad = \quad N_3 \times \frac{y_{C1} + y_{C2}}{y_B} \times \frac{x_{C1} + x_{C2}}{x_A} + N_2 \times f_{escape} \times \frac{y_{C1} + y_{C2}}{y_B}$$

$x_{A,C1,C2}$ and $y_{B,C1,C2}$ are the size of signal regions and sideband regions defined in Figure 7.1. f_{escape} is a fraction of events in $|t_{\gamma\gamma}| > 1\,\text{ns}$ escaping from $|t_{\gamma\gamma}| \leq 1\,\text{ns}$ whose origin is $t_{\gamma\gamma}$ peak. This factor is estimated to be 0.171. Its uncertainty is negligible. Figure 7.2 shows an example of the sideband distribution. The linearity of the sideband regions are discussed in Appendix A.

The best estimate of N_{BG} is obtained by maximizing the likelihood function defined in Equation (7.1). The values of N_{BG} in this search are discussed in Section 7.5 after defining the event selection in Section 7.4. Note that, as we discuss in Section 7.3, N_{BG} is not inferred separately from the signal but inferred together with the signal in the search analysis.

---

[3]We confirmed that a deviation from this assumption was negligible compared with statistical uncertainties of N_A, N_B, N_C. The detailed discussion is described in Appendix A.
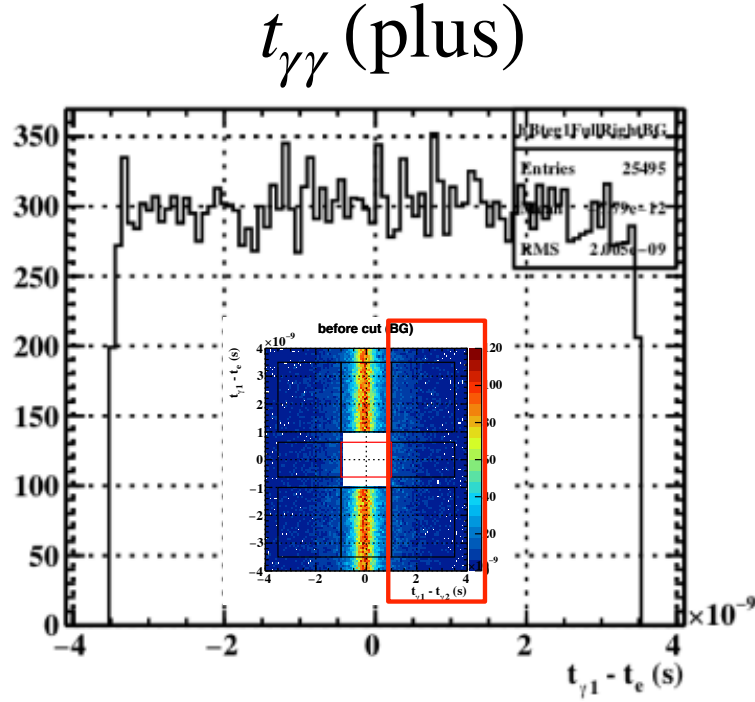
Figure 7.2: An example of the sideband distribution. The small figure with red box shows which part of sideband data is used to make the distribution. See Appendix A in detail.

7.3 Signal estimation with background extraction

We estimate the signal number (N_{MEx2G}) and it is converted to the branching ratio (limit) of the $\mu^+ \to e^+X, X \to \gamma\gamma$ decay ($\mathscr{B}_{\mathrm{MEx2G}}$) using a normalization factor (k)[4].

$$(7.5) \qquad\qquad \mathscr{B}_{\mathrm{MEx2G}} = \frac{1}{k} \times N_{\mathrm{MEx2G}}.$$

We use the following items for the estimation of the signal number:

- the number of events in time sidebands (Section 7.5)

- the normalization factor k and its uncertainty (Section 7.6.10)

- the number of events in the signal region (Section 7.8.1).

We perform a maximum likelihood fitting to estimate the number of signal. The likelihood function Equation (7.1) is extended to include the branching ratio (BR) as a parameter and the number of events in the signal region (N_{obs}) as an observable. In addition, to incorporate the

---

[4]We discuss k in Section 7.6

uncertainty of normalization into the BR, the estimated normalization (k_0) and the true value (k) are included into the likelihood function:

$$(7.6) \qquad \mathscr{L}(\text{BR}, N_{\text{BG}}, k | N_{\text{obs}}, N_{\text{A}}, N_{\text{B}}, N_{\text{C}}, k_0)$$

The expected number of events (N_{exp}) in the signal region has four different components: three different backgrounds (N_1, N_2, and N_3) and signal (N_{S}), with which BR $= N_{\text{S}}/k$:

$$(7.7) \qquad N_{\text{exp}} = N_1 + N_2 + N_3 + N_{\text{S}}.$$

Using N_1, N_2, N_3 and $N_{\text{A}}, N_{\text{B}}, N_{\text{C}}$ defined in Section 7.2, Equation (7.6) can be written as follows:

$$
\begin{aligned}
&\mathscr{L}(\text{BR}, N_1, N_2, N_3, k | N_{\text{obs}}, N_{\text{A}}, N_{\text{B}}, N_{\text{C}}, k_0) \\
= \; &\text{Poi}(N_{\text{obs}} | N_{\text{exp}}) \text{Poi}(N_{\text{A}} | N_{\text{A}}^{\text{exp}}) \text{Poi}(N_{\text{B}} | N_{\text{B}}^{\text{exp}}) \text{Poi}(N_{\text{C}} | N_{\text{C}}^{\text{exp}}) \text{Gaus}(k_0 | k) \\
(7.8) \quad = \; &\text{Poi}(N_{\text{obs}} | k\text{BR} + N_1 + N_2 + N_3) \text{Poi}(N_{\text{A}} | N_{\text{A}}^{\text{exp}}) \text{Poi}(N_{\text{B}} | N_{\text{B}}^{\text{exp}}) \text{Poi}(N_{\text{C}} | N_{\text{C}}^{\text{exp}}) \text{Gaus}(k_0 | k).
\end{aligned}
$$

If BR is fixed, the number of signal can be obtained from the estimation of the normalization factor. We define s to be the signal number assuming the normalization factor to be k_0 with satisfying BR$= s/k_0$ and $N_{\text{S}} = rs$ where $r = k/k_0$. Dividing by a factor k_0 (a constant value) does not change the form of the likelihood function and gives:

$$
\begin{aligned}
&\mathscr{L}(s, N_1, N_2, N_3, r | N_{\text{obs}}, N_{\text{A}}, N_{\text{B}}, N_{\text{C}}, k_0) \\
(7.9) \quad = \; &\text{Poi}(N_{\text{obs}} | rs + N_1 + N_2 + N_3) \text{Poi}(N_{\text{A}} | N_{\text{A}}^{\text{exp}}) \text{Poi}(N_{\text{B}} | N_{\text{B}}^{\text{exp}}) \text{Poi}(N_{\text{C}} | N_{\text{C}}^{\text{exp}}) \text{Gaus}(r_0 | r),
\end{aligned}
$$

where r_0 is defined by $r_0 = k_0/k_0 (= 1)$, respectively.

We assume a Gaussian distribution for the normalization PDF:

$$\text{Gaus}(r_0 | r) \sim \mathscr{N}(1 | r, \sigma_r^2),$$

where σ_r is the relative uncertainty of k (Table 7.11) and $\mathscr{N}(x | \mu, \sigma^2)$ is expressed as

$$\mathscr{N}(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Once we measure $N_{\text{obs}}, N_{\text{A}}, N_{\text{B}}, N_{\text{C}}, k_0$, the best estimated values of the parameter set s, N_1, N_2, N_3, r are obtained by maximizing the likelihood function. For the MEx2G search, only s is an interesting parameter, while the others are regarded as nuisance parameters. Once s is obtained, BR is calculated by BR $= s/k_0$.
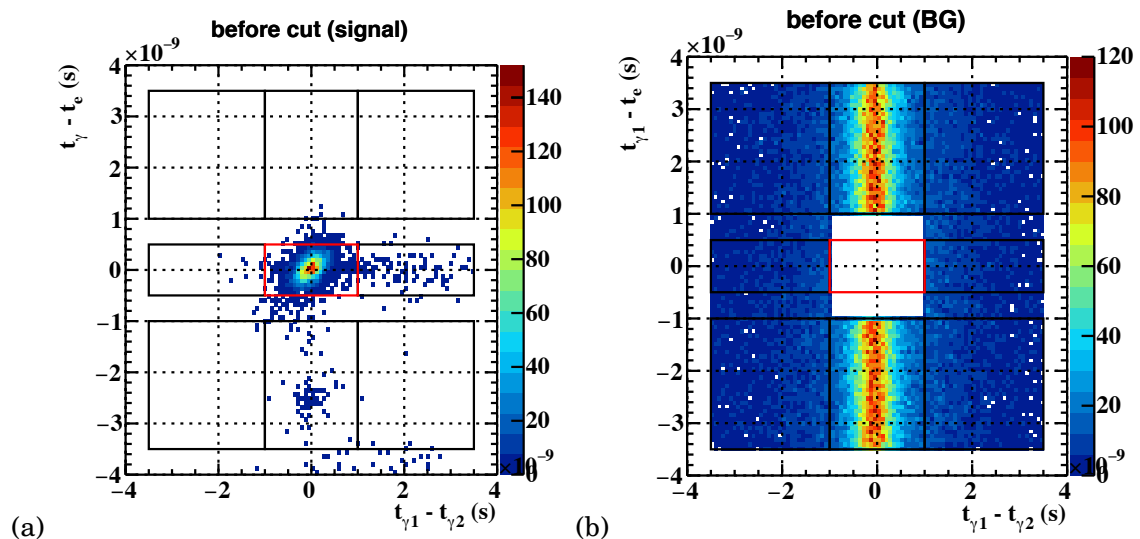
Figure 7.3: (a) Signal (MC) for (20 MeV, 20 ps) and (b) BG (sideband data) event distributions as a function of time differences.

7.4   Event selection

7.4.1   Cut optimization

The purpose of this optimization is to find the threshold, which maximizes experimental outcomes without assuming the signal number. For this purpose, Punzi's sensitivity [105] is used as a criterion of the cut optimization. It requires no assumption on signal counts whether 0 or not. The sensitivity $\sigma_{\mathrm{Punzi}}^{-1}$ is defined as

$$(7.10) \qquad \sigma_{\mathrm{Punzi}}^{-1} = \frac{\epsilon_{\mathrm{signal}}}{b^2 + 2a\sqrt{N_{\mathrm{BG}}} + b\sqrt{b^2 + 4a\sqrt{N_{\mathrm{BG}}} + 4N_{\mathrm{BG}}}},$$

where a is significance and b is power, which is set to the same value as confidence level (CL). $\epsilon_{\mathrm{signal}}$ is the signal efficiency, and N_{BG} is the number of the background events. a and b should be defined before the analysis, and we set $a = 3, b = 1.28$ (corresponds to 90%CL). The task is to find the optimal cuts to maximize $\sigma_{\mathrm{Punzi}}^{-1}$. In general, $\epsilon_{\mathrm{signal}}$ should be large and N_{BG} should be small to maximize $\sigma_{\mathrm{Punzi}}^{-1}$. The detail procedure is explained in Section 7.4.3.

Figure 7.3 shows time distribution of signal and background events. MC samples are used as the signal distribution, and time sideband events of data are used as the background distribution. All the other selections including trigger effects are already applied for these distributions.

7.4.2   Variables used for the event selection

At first, we set positron energy selection to be ± 1 MeV, which corresponds to $P_X = P_e$. The equivalent selection is applied when we count Michel events for normalization (Section 7.6.1). Then
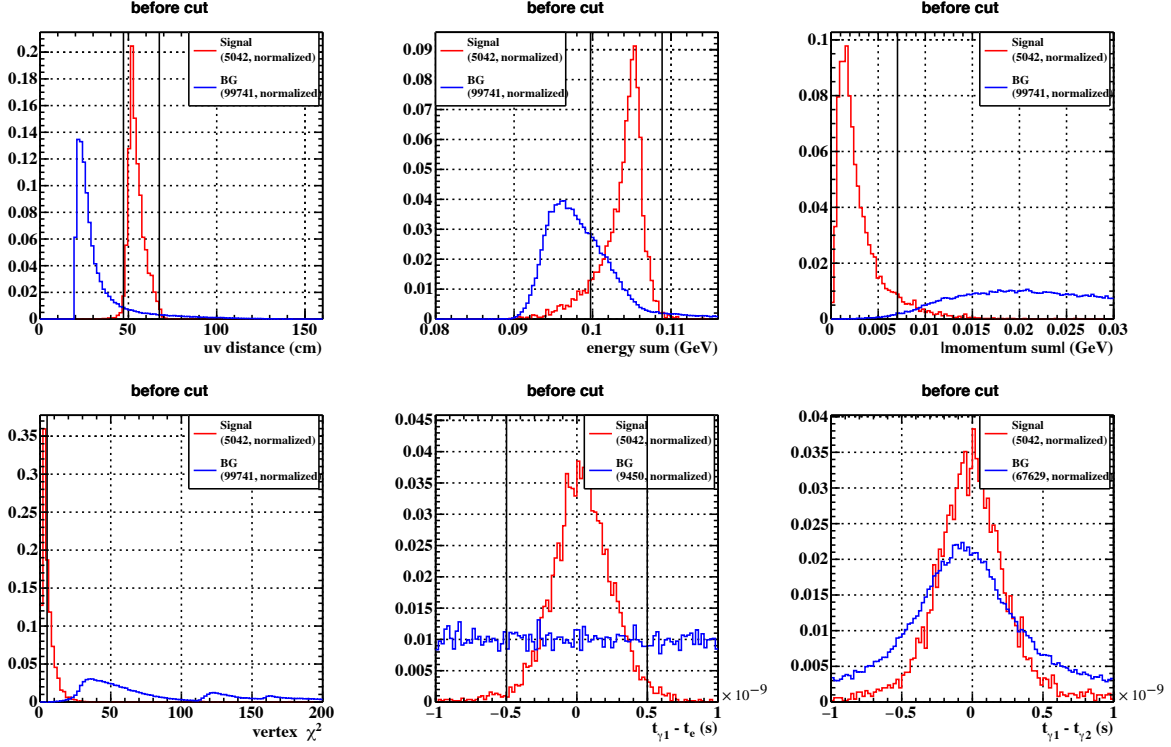
Figure 7.4: Distribution of variables used for signal selection for (20 MeV, 20 ps). Both signal (MC) in red and background (sidebands) in blue are shown in the figures. Each distribution is normalized. The black lines are optimized thresholds.

six variables are used for the signal selection and these variables are optimized. Distributions of these variables for both signal and background at two parameter sets ((20 MeV, 5 ps) and (45 MeV, 40 ps)) are shown in Figure 7.4 and Figure 7.5. The optimized thresholds are shown in the black lines.

uv distance uv distance is defined as the distance between reconstructed 2γs on the LXe inner face. It is given by

$$\sqrt{(u_{\gamma 1} - u_{\gamma 2})^2 + (v_{\gamma 1} - v_{\gamma 2})^2},$$

where $u_{\gamma 1,2}$ is reconstructed u position of each γ and $v_{\gamma 1,2}$ is reconstructed v position of each γ. Distributions of signal and background events are shown in the top left figure in Figure 7.4 and Figure 7.5. The uv distance becomes large for the higher mass because the opening angle between 2γs becomes large for the higher mass as shown in Figure 2.8. Cut-off at 20 cm in the background distribution comes from one of the 2γ reconstruction conditions (Section 5.3.4.3).

energy sum   Energy sum is defined as total energy of the three particles. It is given by

$$E_{\gamma 1} + E_{\gamma 2} + E_{e^+}.$$

95
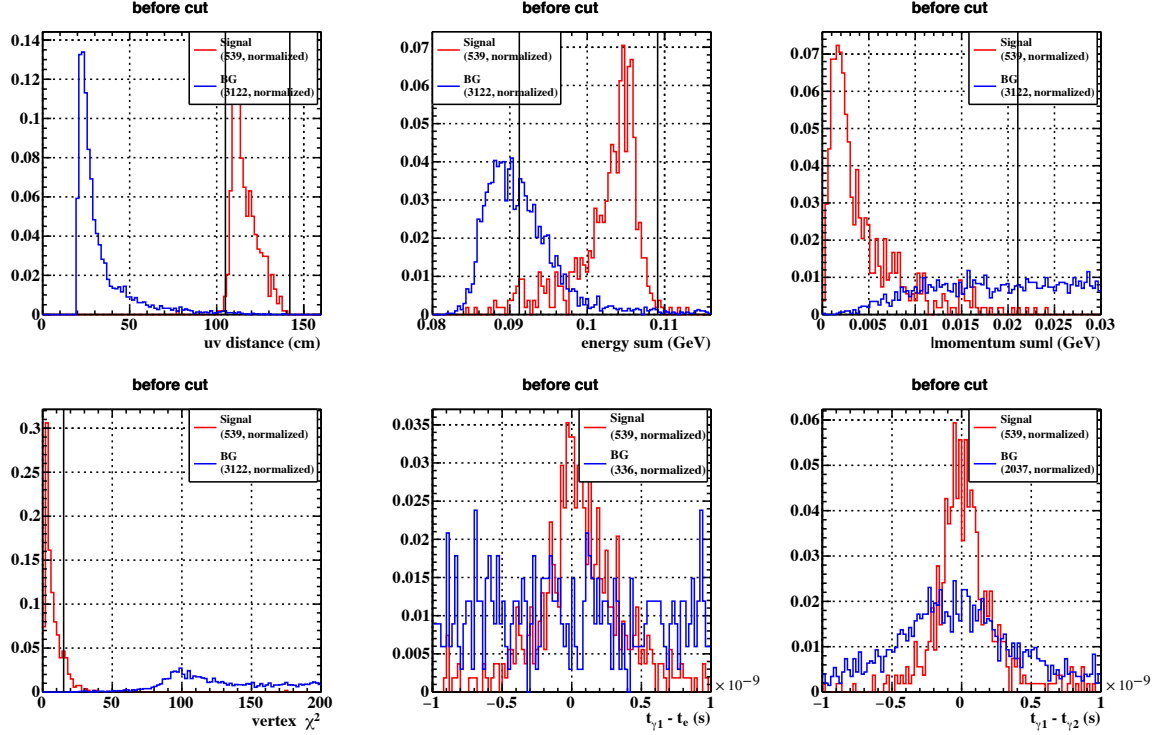
Figure 7.5: Distribution of variables used for signal selection for (45 MeV, 20 ps). Both signal (MC) in red and background (sidebands) in blue are shown in the figures. Each distribution is normalized. The black lines are optimized thresholds.

It should be equal to the mass of muon from the energy conservation. Distributions of signal and background events are shown in the top center figure in Figure 7.4 and Figure 7.5. The peak value of the signal distribution corresponds to the mass of muon.

momentum sum   Momentum sum is defined as sum of the momentum of three particles. It is given by

$$\vec{P}_{e^+} + \vec{P}_{\gamma_1} + \vec{P}_{\gamma_2}.$$

The absolute value should be 0 from the momentum conservation. Distributions of signal and background events are shown in the top right figure in Figure 7.4 and Figure 7.5.

vertex χ^2   Vertex χ^2 is calculated in Equation (6.4) after the vertex reconstruction. Distributions of signal and background events are shown in the bottom left figure in Figure 7.4 and Figure 7.5.

$t_{\gamma_1 e}$ $t_{\gamma_1 e}$ is time difference between γ_1 and e^+. It is defined in Equation (6.8). Distributions of signal and background events are shown in the bottom center figure in Figure 7.4 and Figure 7.5.

$t_{\gamma\gamma}$ $t_{\gamma\gamma}$ is time difference between 2γs. It is defined in Equation (6.7). Distributions of signal and background events are shown in the bottom right figure in Figure 7.4 and Figure 7.5.

Figure 7.3 is a two-dimensional plot of $t_{\gamma_1 e}$ and $t_{\gamma\gamma}$ shown in Figure 7.4.

7.4.3 Optimization procedure

We optimize thresholds for the variables in two steps. As we discuss below, this procedure enables us to determine selections while reducing a possible bias on the selection. Generally, there exists only 1 (or 0) background event around the optimal point (because we try to do so). Since we use the data itself for the optimization, the optimal cut may suffer from a bias coming from the low statistics.

We check the signal and background distributions of six variables as shown in Figure 7.4 and Figure 7.5. In order to reduce the bias, we determine thresholds independently at the first step to use larger statistics. However, it gives a suboptimal selection. Therefore we select one variable, which has the maximum discriminating power and determine its threshold after applying all the other selections at the second step. The vertex χ^2 gives the highest discriminating power measured with the ROC curve[5].

First step Thresholds of five variables except for vertex χ^2 were determined independently. Figure 7.6–Figure 7.10 show how to determine thresholds for each variable. The left plots show signal (in red) and background (in blue) distributions of each variable and black lines in these plots are determined thresholds. The right plots in Figure 7.6 and Figure 7.7 show Punzi's sensitivity as a function of the lower and upper thresholds. The right plots in Figure 7.8, Figure 7.9, and Figure 7.10 are the cumulative distributions of the corresponding left plots. Black curves in these plots show Punzi's sensitivity as a function of threshold. The number of background in the Punzi's calculation in Equation (7.10) is replaced with a background efficiency defined is by a ratio between before and after applying a selection under consideration, where number of background is normalized.

First, we set a threshold for $t_{\gamma\gamma}$ to be 1 ns (maximum value) because the sensitivity in Figure 7.10 (b) does not change so much compared with the other variables. Then, a threshold for uv distance is fixed where it gives the maximum Punzi's sensitivity. Concerning the other variables, we loosen the thresholds by 10%–30% from the point giving the maximum sensitivity in order to get enough events to estimate the number of background events.

---

[5]ROC stands for Receiver Operating Characteristic, which is calculated from true positive rate and false-positive rate with scanned thresholds.
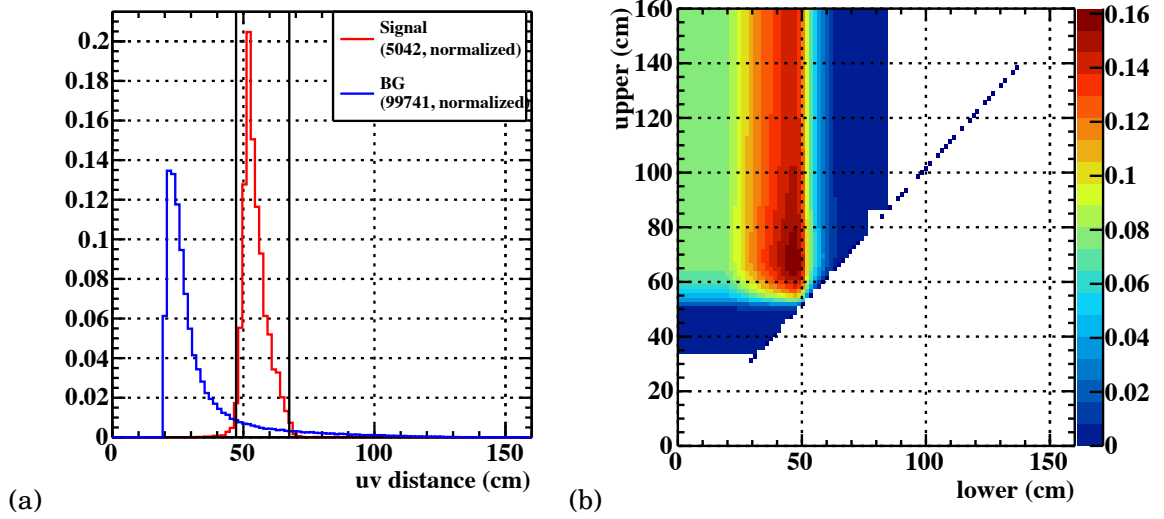
Figure 7.6: (a) signal and background distributions and (b) Punzi's sensitivity as a function of upper and lower threshold for uv distance. (20 MeV, 20 ps) case is shown as an example.
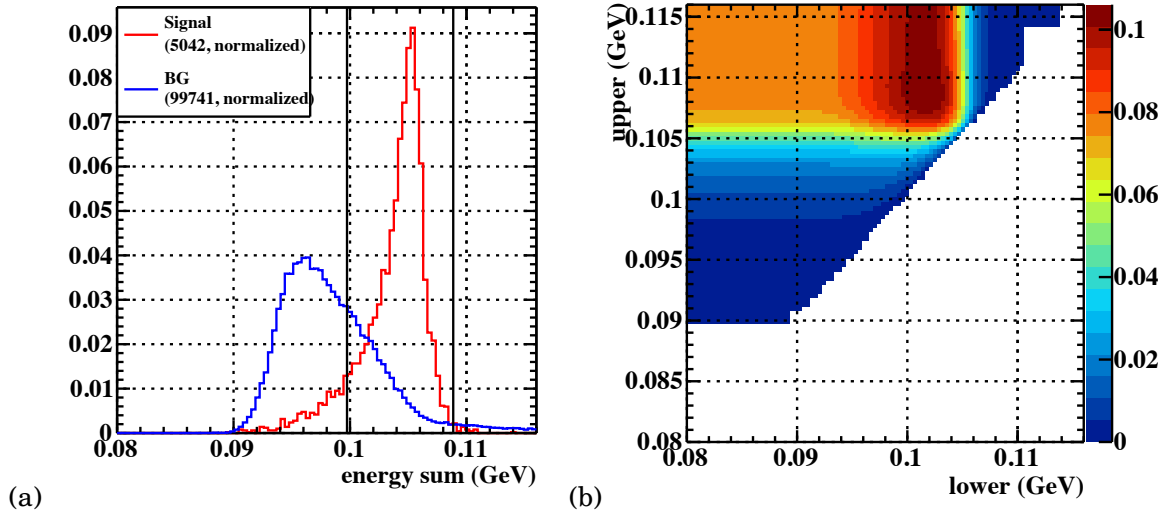
Figure 7.7: (a) signal and background distributions and (b) Punzi's sensitivity as a function of upper and lower threshold for energy sum. (20 MeV, 20 ps) case is shown as an example.
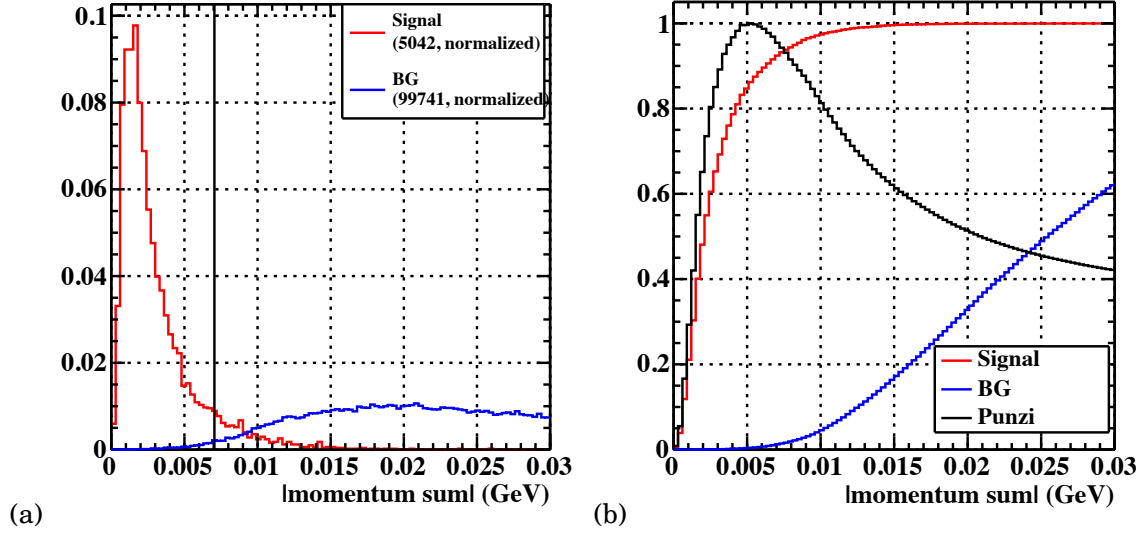
Figure 7.8: (a) signal and background distributions and (b) integrated spectrum of (a) and Punzi's sensitivity as a function of threshold for absolute value of momentum sum. (20 MeV, 20 ps) case is shown as an example.
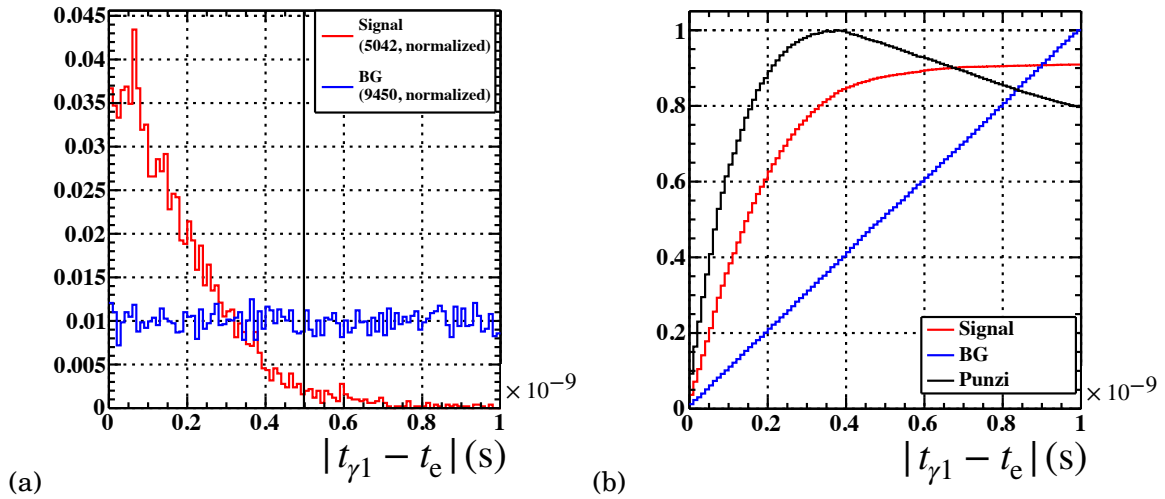
Figure 7.9: (a) signal and background distributions and (b) integrated spectrum of (a) and Punzi's sensitivity as a function of threshold for $t_{\gamma 1 e}$. (20 MeV, 20 ps) case is shown as an example.

99

(a)

(b)
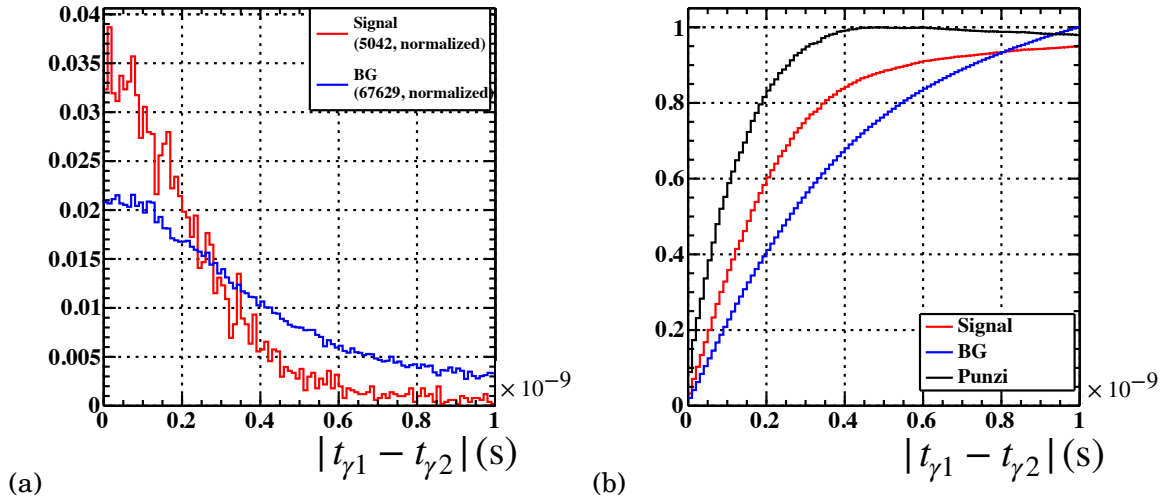
Figure 7.10: (a) signal and background distributions and (b) integrated spectrum of (a) and Punzi's sensitivity as a function of threshold for $t_{\gamma\gamma}$. (20 MeV, 20 ps) case is shown as an example.
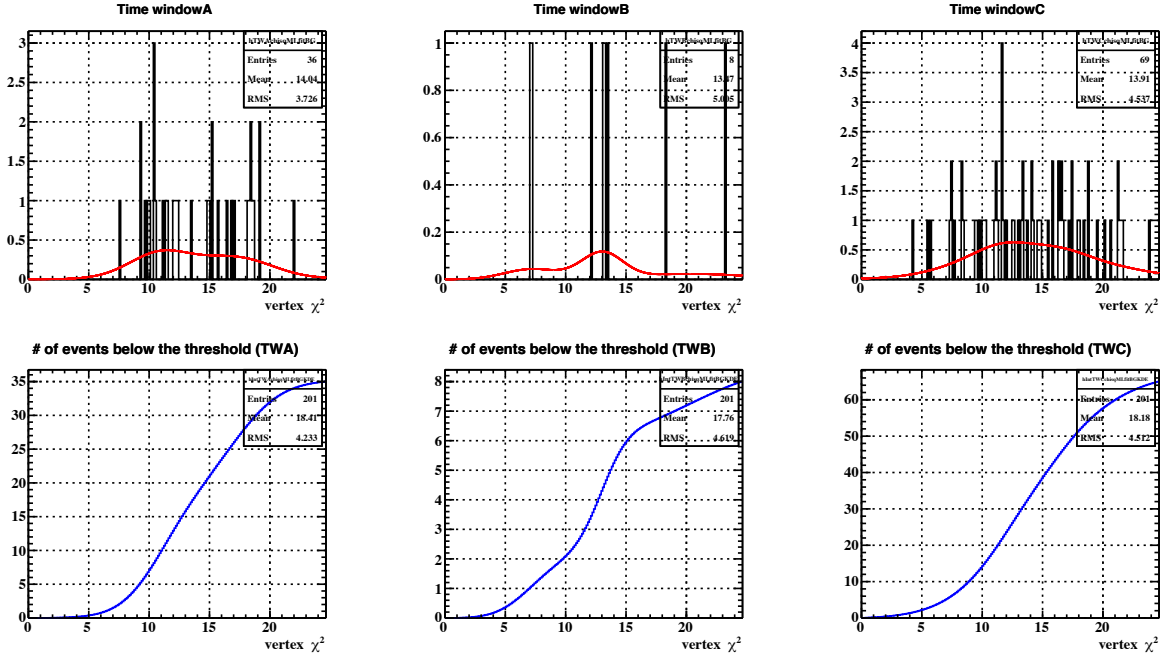
Figure 7.11: Top: event distributions after event selections except for vertex χ^2. Three figures correspond to time sidebands. Red curves are modeled curves using KDE. Bottom: cumulative curves of the red curves.

Second step After applying all the selections except for vertex χ^2, event distribution in the time sideband is modeled using kernel density estimation method[106, 107] to get its PDF. This modeling also contributes to reduce the bias due to low statistics. Figure 7.11 shows the KDE modeling for each time sideband. The red curves are modeled curves and the blue curves are cumulative ones. These are used to calculate the number of background events in the signal region as a function of vertex χ^2. Figure 7.12 (b) shows the expected number of background events in the signal region estimated from the time sidebands' modeling. Figure 7.12 (a) shows signal efficiency. Figure 7.12 (a) and (b) give Punzi's sensitivity as shown in Figure 7.12 (c). A threshold for vertex χ^2 is determined to be the point where it gives the maximum sensitivity in this plot.

These optimized thresholds are averaged over assumed lifetimes to reduce bias. Then, mass-dependent thresholds are determined by performing the optimization on each 5 MeV; for instance, thresholds for $m_X = 23, 24, 25, 26, 27$ MeV are the same ones which is optimized at $m_X = 25$ MeV.

As shown in Figure 7.13 (c), signal ratio compared with that without applying the signal selection is ~50% for 20 MeV. In Figure 7.13, six selection conditions are applied subsequently. On the other hand, the number of BG is reduced to (at least) a factor of $\sim \mathcal{O}(10^{-5})$ as shown in Figure 7.13 (b). Note that timing selections is not applied on Figure 7.13 (b) since no sideband events are survived by applying these selections.

Signal efficiency

hOptIntchisqML.fitSig	
Entries	201
Mean	14.32
RMS	6.159

vertex χ^2

(a)

Number of BG (rated)

vertex χ vertex vertex χ

(b)

Punzi's sensitivity

hOptIntchisqML.fitPunzi2	
Entries	201
Mean	10.8
RMS	6.377

vertex χ^2

(c)
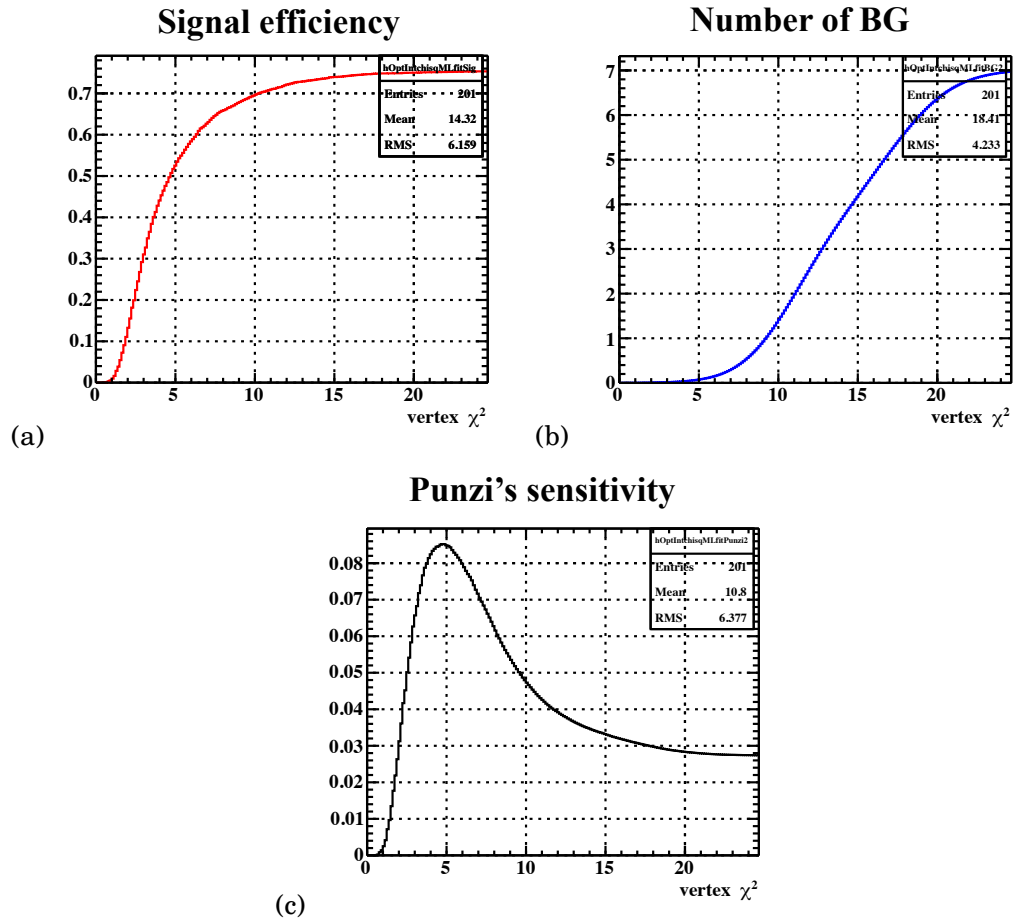
Figure 7.12: (a) signal efficiency, (b) the number of background events in the analysis region, and (c) Punzi as a function of a threshold for vertex χ^2. (20 MeV, 20 ps) case is shown as an example.
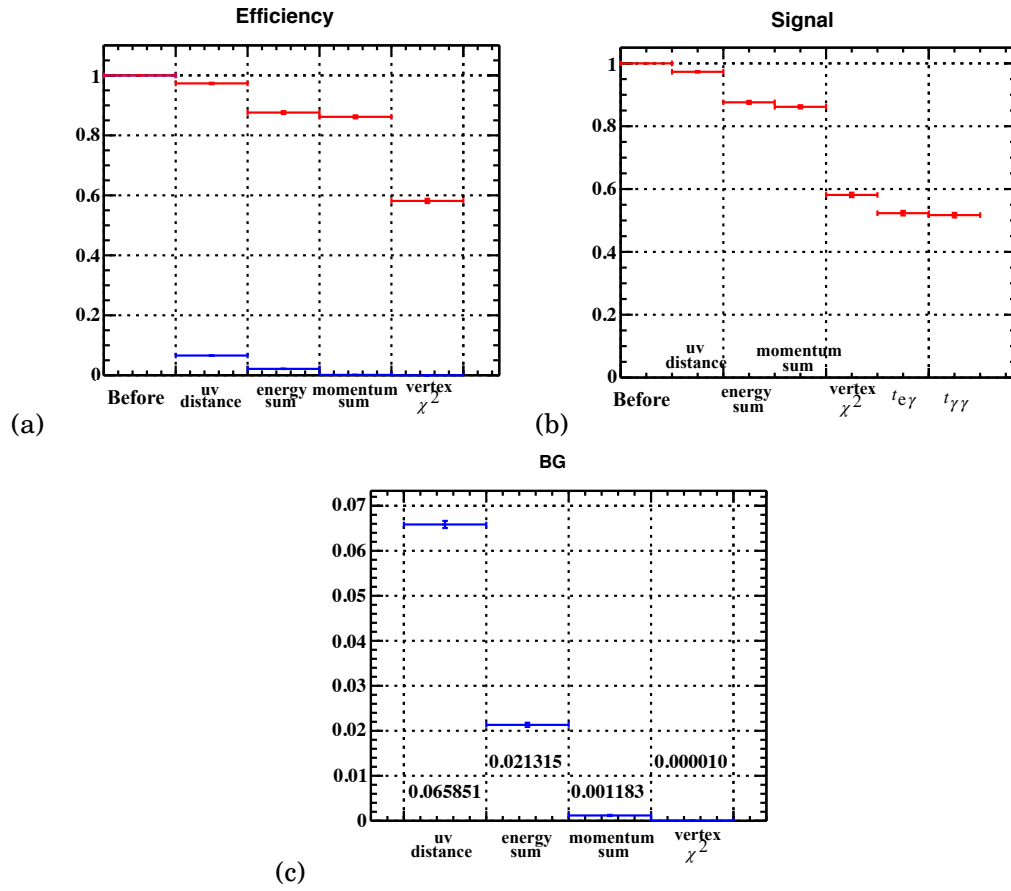
Figure 7.13: (a) Efficiency of signal (in red) and BG (in blue). (b) and (c) are zoomed-in plot of (a). The signal efficiency is estimated using the signal MC and the BG efficiency is estimated using time sideband events. (20 MeV, 20 ps) case is shown as an example. In this plot, events just after all the othere selections is set to 1 ("Before").

Table 7.1: Optimized thresholds

	uv distance (cm)		energy sum (MeV)		momentum sum (MeV)	vertex χ^2	$t_{\gamma 1e}$ (ns)	$t_{\gamma\gamma}$ (ns)
	low	up	low	up				
20 MeV (20 − 22 MeV)	47.2	67.5	99.7	109	7.05	4.93	0.498	1.00
25 MeV (23 − 27 MeV)	58.4	86.1	100	109	7.35	5.71	0.518	1.00
30 MeV (28 − 32 MeV)	71.2	104	99.7	109	7.05	7.79	0.498	1.00
35 MeV (33 − 37 MeV)	82.4	129	96.0	109	12.3	7.37	1.00	1.00
40 MeV (38 − 42 MeV)	94.1	145	94.9	109	14.6	7.73	1.00	1.00
45 MeV (43 − 47 MeV)	105	142	91.2	109	21.1	15.1	1.00	1.00

7.4.4 Summary of event selection

Table 7.1 summarizes thresholds of variables used for the signal selection. The same threshold sets are used for the same mass regardless of the lifetime.

Table 7.2: Definition of sideband and signal regions

mass (MeV)	x_{C1} (ns)	x_A (ns)	x_{C2} (ns)	y_{C1} (ns)	y_B (ns)	y_{C2} (ns)
20 MeV (20 − 22 MeV)	2.5	2.000	2.5	2.5	0.996	2.5
25 MeV (23 − 27 MeV)	2.5	2.000	2.5	2.5	1.036	2.5
30 MeV (28 − 32 MeV)	2.5	2.000	2.5	2.5	0.996	2.5
35 MeV (33 − 37 MeV)	2.5	2.000	2.5	2.5	2.000	2.5
40 MeV (38 − 42 MeV)	2.5	2.000	2.5	2.5	2.000	2.5
45 MeV (43 − 47 MeV)	2.5	2.000	2.5	2.5	2.000	2.5

7.5 Background expectation using only sidebands

We estimate the number of background events in the signal region based on the methods discussed in Section 7.2 with the event selections determined in Section 7.4. Definitions of sideband and signal regions are summarized in Table 7.2. The numbers of events in the sidebands are counted and the number of background events in the signal region is estimated and summarized in Table 7.3. In the last column of Table 7.3, we listed the best fitted values of the expected background in the signal region and their statistical uncertainties. These values depend on sideband events, that is, the box size of the sidebands. As discussed in Section 7.2 and Appendix A, systematic uncertainties related to the scale method (the use of Equation (7.2)–(7.4)) and the escape factor are negligible. Systematic uncertainty of the normalization factor is taken into account when estimating signal numbers.

Table 7.3: The number of events in sidebands and the expected number of BG in signal region

mass (MeV)	sideband								expected N_{BG} in signal region
	A_1	A_1	B_1	B_2	C_1	C_2	C_3	C_4	
20	0	0	0	0	0	0	1	0	$0.048^{+0.202}_{-0.046}$
21	0	0	0	0	1	1	0	1	$0.146^{+0.198}_{-0.084}$
22	1	0	0	0	2	2	0	1	$0.292^{+0.211}_{-0.140}$
23	2	1	0	0	1	1	0	1	$0.622^{+0.425}_{-0.330}$
24	1	1	0	0	0	1	0	0	$0.414^{+0.346}_{-0.260}$
25	1	1	0	0	0	2	0	1	$0.414^{+0.346}_{-0.261}$
26	0	0	0	0	0	1	1	1	$0.150^{+0.189}_{-0.091}$
27	0	0	0	0	1	0	0	0	$0.050^{+0.200}_{-0.049}$
28	0	0	0	0	0	1	0	0	$0.048^{+0.202}_{-0.046}$
29	0	0	0	0	0	1	0	0	$0.048^{+0.202}_{-0.046}$
30	0	0	0	0	0	0	0	0	$0.000^{+0.170}_{-0.000}$
31	0	0	0	0	0	1	0	0	$0.048^{+0.202}_{-0.046}$
32	0	0	0	0	0	0	0	0	$0.000^{+0.170}_{-0.000}$
33	0	0	0	0	0	0	0	0	$0.000^{+0.210}_{-0.000}$
34	0	0	0	0	0	0	0	0	$0.000^{+0.210}_{-0.000}$
35	0	0	0	0	0	0	0	0	$0.000^{+0.210}_{-0.000}$
36	0	0	0	0	0	0	0	0	$0.000^{+0.210}_{-0.000}$
37	1	0	0	0	0	0	0	0	$0.400^{+0.517}_{-0.301}$
38	0	0	0	0	1	0	0	1	$0.168^{+0.183}_{-0.105}$
39	0	0	0	0	0	0	0	1	$0.084^{+0.201}_{-0.084}$
40	0	0	0	0	0	0	0	0	$0.000^{+0.210}_{-0.000}$
41	0	0	0	0	0	0	0	0	$0.000^{+0.210}_{-0.000}$
42	0	0	0	0	0	0	0	0	$0.000^{+0.210}_{-0.000}$
43	0	0	0	0	0	1	0	0	$0.084^{+0.201}_{-0.084}$
44	0	0	0	0	0	0	0	0	$0.000^{+0.210}_{-0.000}$
45	0	0	0	0	0	0	0	0	$0.000^{+0.210}_{-0.000}$

7.6 Normalization and single event sensitivity

To calculate the branching ratio of the MEx2G mode, the number of muon decays effectively measured during the data taking is used. We use Michel e^+ events for this normalization. The number of Michel events are counted using a pre-scaled dedicated trigger during the physics data taking; signal events and Michel events were taken at the same time with different trigger scalings. The Michel normalization is beneficial for the following reasons. First, systematics coming from the beam are canceled. Instability of the beam is included in both of the Michel trigger events and the $\mu^+ \to e^+\gamma$ trigger. Moreover, we do not need to know μ^+ stopping rate nor live DAQ time. Second, most of the systematics coming from e^+ detection are also canceled. The absolute value of e^+ efficiency is not needed because e^+ is required in both events. The number of Michel events is given by

$$(7.11) \qquad N_{\mathrm{Michel}} \;=\; N_\mu \times \mathscr{B}_{\mathrm{Michel}} \times f_{\mathrm{Michel}} \times \frac{1}{P_{\mathrm{Michel}} \times P_{\mathrm{correction}}} \times \epsilon_{\mathrm{Michel}},$$

where

N_{Michel} the number of observed Michel events.

N_μ the number of stopped μ^+.

$\mathscr{B}_{\mathrm{Michel}}$ branching ratio of the Michel decay (~ 1).

f_{Michel} branching fraction of the selected energy region.

P_{Michel} prescaling factor of the Michel trigger ($= 10^7$).

$P_{\mathrm{correction}}$ correction factor of P_{Michel}. It depends on the beam intensity.

$\epsilon_{\mathrm{Michel}}$ the overall efficiency for Michel events. Its detail definition comes later.

The number of signal MEx2G events is given by

$$(7.12) \qquad N_{\mathrm{MEx2G}} \;=\; N_\mu \times \mathscr{B}_{\mathrm{MEx2G}} \times \frac{1}{P_{\mathrm{MEG}}} \times \epsilon_{\mathrm{MEx2G}},$$

where

N_{MEx2G} the number of observed signal MEx2G events

$\mathscr{B}_{\mathrm{MEx2G}}$ branching ratio of the MEx2G decay (unknown).

P_{MEG} prescaling factor of the $\mu^+ \to e^+\gamma$ trigger ($= 1$).

$\epsilon_{\mathrm{MEx2G}}$ the overall efficiency for MEx2G events. Its detailed definition comes later.

A normalisation factor k is defined as follows (the same equation as Equation (7.5)):

$$(7.13) \qquad \mathscr{B}_{\text{MEx2G}} = \frac{1}{k} \times N_{\text{MEx2G}}.$$

Thus, using Equation (7.11), (7.12), and (7.13), the normalisation factor k's measured value k_0 can be calculated as follows:

$$(7.14) \qquad k_0 \;=\; N_{\text{Michel}} \times \frac{1}{\mathscr{B}_{\text{Michel}}} \times \frac{1}{f_{\text{Michel}}} \times \frac{P_{\text{Michel}}}{P_{\text{MEG}}} \times P_{\text{correction}} \times \frac{\epsilon_{\text{MEx2G}}}{\epsilon_{\text{Michel}}}.$$

Detailed definition of the efficiencies   Efficiencies are written down using conditional probabilities. Detection of a particle means that the particle is in geometrical acceptance, triggered, and reconstructed. ϵ_{Michel} is given by

$$
\begin{aligned}
\epsilon_{\text{Michel}} \;&=\; P(e^+\,\text{detected}) \\
&=\; P(e^+ \in \text{acceptance}) \times P(e^+\,\text{triggered}, e^+\,\text{reconstructed}|e^+ \in \text{acceptance}).
\end{aligned}
$$

ϵ_{MEx2G} is given by

$$
\begin{aligned}
&\epsilon_{\text{MEx2G}} \\
=\;\;& P(e^+\,\text{detected}, 2\gamma\,\text{detected}, \text{selected}) \\
=\;\;& P(e^+ \in \text{acceptance}) \\
\times\;\;& P(e^+\,\text{triggered}, e^+\,\text{reconstructed}, 2\gamma \in \text{acceptance}, \gamma\,\text{triggered}, 2\gamma\,\text{reconstructed}, \\
& \text{DM} - \text{triggered}, \text{selected}|e^+ \in \text{acceptance}) \\
=\;\;& P(e^+ \in \text{acceptance}) \times P(e^+\,\text{triggered}, e^+\,\text{reconstructed}|e^+ \in \text{acceptance}) \\
\times\;\;& P(2\gamma \in \text{acceptance}, \gamma\,\text{triggered}, 2\gamma\,\text{reconstructed}|e^+ \in \text{acceptance}) \\
\times\;\;& P(\text{DM} - \text{triggered}, \text{selected}| \\
& e^+, 2\gamma \in \text{acceptance}, e^+\,\text{triggered}, e^+\,\text{reconstructed}, \gamma\,\text{triggered}, 2\gamma\,\text{reconstructed}) \\
=\;\;& P(e^+ \in \text{acceptance}) \times P(e^+\,\text{triggered}, e^+\,\text{reconstructed}|e^+ \in \text{acceptance}) \\
\times\;\;& P(2\gamma \in \text{acceptance}, \gamma\,\text{triggered}, 2\gamma\,\text{reconstructed}|e^+ \in \text{acceptance}) \\
\times\;\;& P(\text{DM} - \text{triggered}| \\
& e^+, 2\gamma \in \text{acceptance}, e^+\,\text{triggered}, e^+\,\text{reconstructed}, \gamma\,\text{triggered}, 2\gamma\,\text{reconstructed}) \\
\times\;\;& P(\text{selected}|e^+, 2\gamma \in \text{acceptance}, e^+\,\text{triggered}, \\
& e^+\,\text{reconstructed}, \gamma\,\text{triggered}, 2\gamma\,\text{reconstructed}, \text{DM} - \text{triggered}) \\
=\;\;& \epsilon_{e^+} \times \epsilon_\gamma \times \epsilon_{\text{DM}} \times \epsilon_{\text{selection}}
\end{aligned}
$$

where

$$\epsilon_{e^+} \equiv P(e^+ \in \text{acceptance}) \times P(e^+ \text{triggered}, e^+ \text{reconstructed} | e^+ \in \text{acceptance})$$

$$\epsilon_\gamma \equiv P(2\gamma \in \text{acceptance}, \gamma \text{triggered}, 2\gamma \text{reconstructed} | e^+ \in \text{acceptance})$$

$$\epsilon_{\text{DM}} \equiv P(\text{DM} - \text{triggered} | e^+, 2\gamma \in \text{acceptance},$$
$$e^+ \text{triggered}, e^+ \text{reconstructed}, \gamma \text{triggered}, 2\gamma \text{reconstructed})$$

$$\epsilon_{\text{selection}} \equiv P(\text{selected} | e^+, 2\gamma \in \text{acceptance}, e^+ \text{triggered},$$
$$e^+ \text{reconstructed}, \gamma \text{triggered}, 2\gamma \text{reconstructed}, \text{DM} - \text{triggered})$$

ϵ_{e^+} positron efficiency

ϵ_γ gamma efficiency

ϵ_{DM} direction match efficiency

$\epsilon_{\text{selection}}$ analysis selection efficiency

In summary, the normalization factor k can be written as follows from Equation (7.14):

$$
\begin{aligned}
(7.15) \qquad k_0 =\ & N_{\text{Michel}} \\
& \times \frac{1}{\mathscr{B}_{\text{Michel}}} \\
& \times \frac{1}{f_{\text{Michel}}} \\
& \times \frac{P_{\text{Michel}}}{P_{\text{MEG}}} \times P_{\text{correction}} \\
& \times \frac{\epsilon_{e^+}}{\epsilon_{\text{Michel}}} \\
& \times \epsilon_\gamma \\
& \times \epsilon_{\text{DM}} \\
& \times \epsilon_{\text{selection}}
\end{aligned}
$$

We will estimate each component from the next sections.

7.6.1 The number of observed Michel events

All the runs are divided into 8 periods. A different year correspond to different run periods. Within the same year, runs are divided in case the beam condition or trigger condition is different. 2009 is divided into 2 periods because of different (beam) degrader thickness. 2010 and 2011 are divided into 2 periods, respectively because of different γ veto trigger settings (see Section 5.3.4 for detail). The number of Michel events is counted period by period.

Figure 7.14 is e^+ time distribution of Michel triggered and high quality track events defined in Section 4.1.6. Triggered Michel events have a peak in the time distribution and its region
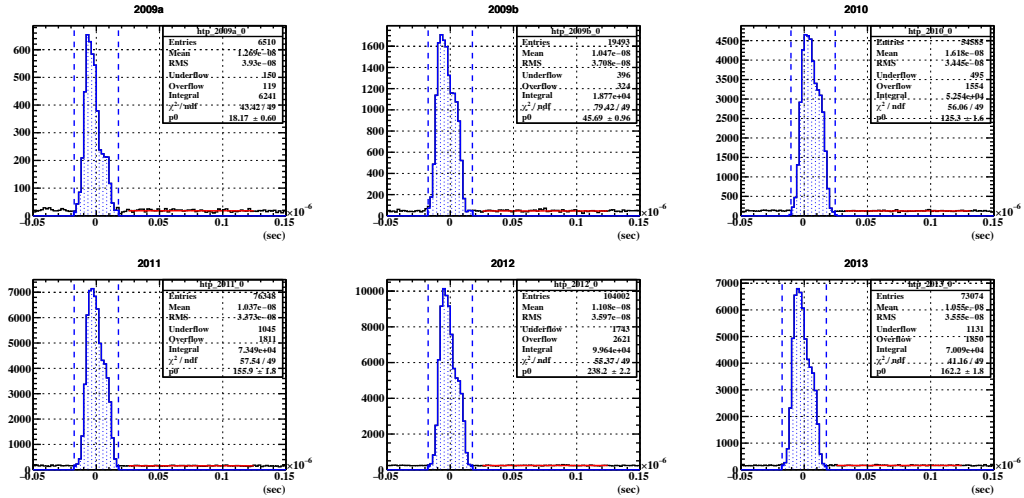
Figure 7.14: Timing distribution of Michel positrons for all years in the energy range of 50 MeV to 56 MeV.

Table 7.4: The number of observed Michel events. Statistical error is assgined to each value ($\sqrt{N}$ is assigned for N).

| m_X | **20** | **25** | **30** | **35** | **40** | **45** |
E_{e^+} (MeV)	**50–56**	**48.9–50.9**	**47.6–49.6**	**46.0–48.0**	**44.3–46.3**	**42.2–44.2**
2009a	11750.9	7258.8	5670.3	3040.3	1713.3	469.0
2009b	6740.7	4111.5	3285.0	1774.5	1001.9	256.9
2010a	1506.9	870.5	676.0	319.6	155.9	9.1
2010b	38396.7	21371.4	15294.7	7326.9	3517.7	776.9
2011a	32696.0	18193.2	13041.5	6102.9	2848.3	670.2
2011b	25151.7	13928.0	9887.1	4716.9	2279.5	505.9
2012	75812.3	43114.7	31800.9	15421.4	7488.4	1711.4
2013	53841.9	31666.7	23794.0	12276.4	6235.2	1394.0
Total	245897.1	140514.8	103449.5	50978.8	25240.1	5793.5

is defined by $-17.5\,\mathrm{ns} < t_e < 17.5\,\mathrm{ns}$[6] (blue broken lines). This region includes an accidental background. Thus the number of Michel events is estimated from time sideband (25 ns < t < 125 ns) and subtracted from the Michel region. The blue regions in Figure 7.14 are selected events. The red line is the estimated level of accidental background.

The results are summarized in Table 7.4. The energy region of 50–56 MeV is the same definition as the analysis of $\mu^+ \to e^+\gamma$ decay. We confirm the numbers of Michel events in this energy region are consistent with the results reported in Ref. [6].

---

[6]+ 7 ns shifted for 2010.

Theory ⊗ Response$_{2009}$
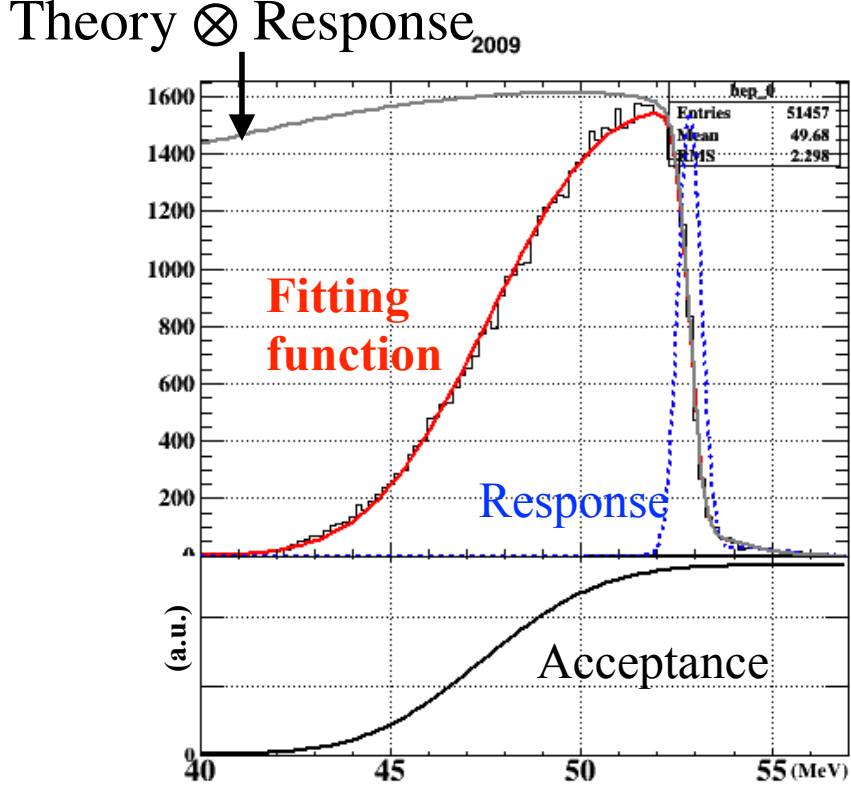
Figure 7.15: An example of the Michel spectrum (e$^+$) in 2009. A function in red is fitted to the histogram.

7.6.2 Fraction of Michel events

f_{Michel} is a fraction of Michel spectrum (e$^+$) within a certain energy range. A Michel spectrum used in this calculation is shown in Figure 7.15. This spectrum is made by applying the same selection as Section 7.6.1. The red solid line is a function fitted to the spectrum. A fitting function corresponds to a polarized Michel spectrum, which is modeled by a theoretical spectrum multiplied by an acceptance function and convoluted with a resolution function.:

$$(\mathrm{Theory} \times \mathrm{Acceptance}) \otimes \mathrm{Response}.$$

The theoretical spectrum comes from Ref. [64]. The acceptance function is defined as

$$\mathrm{Acceptance}\left(E_e^{\mathrm{true}}\right) = \frac{1 + \mathrm{erf}\left(\frac{E_e^{\mathrm{true}} - \mu_{\mathrm{acc}}}{\sqrt{2}\sigma_{\mathrm{acc}}}\right)}{2}.$$

We use double Gaussian as a response function. A positron energy range is selected depending on m_{X}. Thus a fraction of the selected energy range of grey curve (Theory ⊗ Response) is f_{Michel}. Our fitting range is extended to 40–56 MeV to include lower energy contribution. The results are

Table 7.5: Fractions of Michel events

m_X	20	25	30	35	40	45
E_{e^+} (MeV)	50–56	48.9–50.9	47.6–49.6	46.0–48.0	44.3–46.3	42.2–44.2
2009	10.4±0.01%	7.2± < 0.01%	7.2± < 0.01%	7.1± < 0.01%	7.0± < 0.01%	6.8± < 0.01%
2010	10.4±0.01%	7.2± < 0.01%	7.2± < 0.01%	7.1± < 0.01%	7.0± < 0.01%	6.8± < 0.01%
2011	10.3± < 0.01%	7.2± < 0.01%	7.2± < 0.01%	7.1± < 0.01%	7.0± < 0.01%	6.8± < 0.01%
2012	10.5± < 0.01%	7.2± < 0.01%	7.2± < 0.01%	7.1± < 0.01%	7.0± < 0.01%	6.8± < 0.01%
2013	10.4±0.01%	7.2± < 0.01%	7.2± < 0.01%	7.1± < 0.01%	7.0± < 0.01%	6.8± < 0.01%

Table 7.6: Prescale correction factor [96]

year	correction factor	relative uncertainty
2009	1.109	0.90%
2010	1.125	0.88%
2011	1.108	0.90%
2012	1.125	0.88%
2013	1.12	0.89%

summarized in Table 7.5. Uncertainties from the fitting, which is statistical uncertainties, are assigned as systematic uncertainties.

7.6.3 Prescale factor and its correction

$$\frac{P_{\text{Michel}}}{P_{\text{MEG}}} \times P_{\text{correction}}$$

$P_{\text{Michel}} = 10^7$ and $P_{\text{MEG}} = 1$ are the prescale factors of Michel trigger and MEG trigger, respectively. P_{Michel} needs a correction in order to account for pileup events [96]. This factor depends on the beam intensity and is estimated year by year summarized in Table 7.6.

7.6.4 Relative positron efficiency

Relative e^+ efficiency between the MEx2G e^+ and the Michel e^+ is given by

$$\frac{\epsilon_{e^+}}{\epsilon_{\text{Michel}}}$$

By taking a ratio between these two efficiencies, we do not need to estimate absolute values of e^+ efficiency. This is one of the advantages of the Michel normalization. However, a few corrections are needed depending on m_X. In the $\mu^+ \to e^+\gamma$ analysis, a similar correction was applied [96]. Thus we make use of it and the difference between the $\mu^+ \to e^+\gamma$ analysis and $\mu^+ \to e^+ X, X \to \gamma\gamma$ analysis is taken into account for an e^+ energy range of 50–56 MeV ($m_X = 20$ MeV).

Figure 7.16 illustrates how to estimate the relative e^+ efficiency. In the $\mu^+ \to e^+\gamma$ analysis (top plot in Figure 7.16), the signal e^+ energy is fixed at 52.8 MeV. There is a difference between ϵ_{MEG} and ϵ_{Michel} because of their acceptances due to different e^+ energy. Thus we need to correct

Table 7.7: Relative e$^+$ efficiencies

m_X(MeV)	**20**	**25–45**
2009	1.023±0.014	1
2010	0.999±0.013	1
2011	1.005±0.013	1
2012	1.021±0.013	1
2013	1.024±0.014	1

as follows:

$$\frac{\epsilon_{\text{MEG}}}{\epsilon_{\text{Michel}}} = \frac{A(52.8) \times (52.8 - 50)}{\int_{50}^{52.8} A(E)dE} \sim 1.15,$$

where $A(E)$ is acceptance curve as a function of e$^+$ energy. This acceptance curve is estimated from the fitting of the Michel spectrum (Section 7.6.2).

In the $\mu^+ \to e^+ X, X \to \gamma\gamma$ case, the correction can differ depending on the assumed m_X. There are two cases: when E_{e^+} is close to the Michel edge (middle plot in Figure 7.16), the similar correction like MEG [96] is needed. The only difference is signal energy. Therefore, the correction is modified to be

$$
\begin{aligned}
\frac{\epsilon_{e^+}}{\epsilon_{\text{Michel}}} &= \frac{A(E_{e^+}) \times (52.8 - 50)}{\int_{50}^{52.8} A(E)dE} \\
&= \frac{\epsilon_{\text{MEG}}}{\epsilon_{\text{Michel}}} \times \frac{A(E_{e^+})}{A(52.8)}.
\end{aligned}
$$

(7.16)
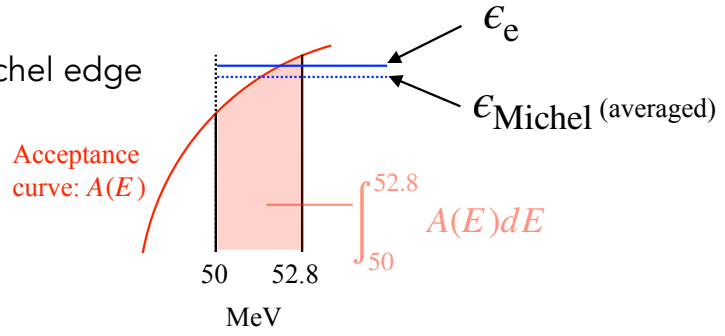
In the second case, when E_e is far from the Michel edge (bottom plot in Figure 7.16), no correction is needed, assuming acceptance curves in the selected region are the same. Thus, the correction is

$$\frac{\epsilon_{e^+}}{\epsilon_{\text{Michel}}} = 1$$

(7.17)

Table 7.7 summarizes e$^+$ relative efficiencies for each m_X. The uncertainty comes from acceptance curve fitting [96].
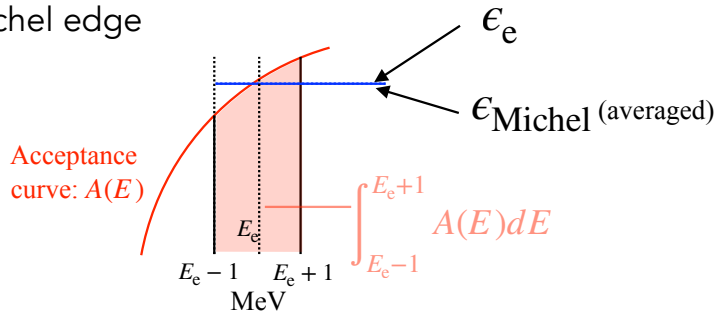
Figure 7.16: How to estimate relative e^+ efficiencies. An averaged line of ϵ_{Michel} corresponds to the integrated area (in red) divided by its energy range (not to accurate, just to illustrate the idea).

$$t_{\text{rec.}} - t_{\text{MC truth}}$$
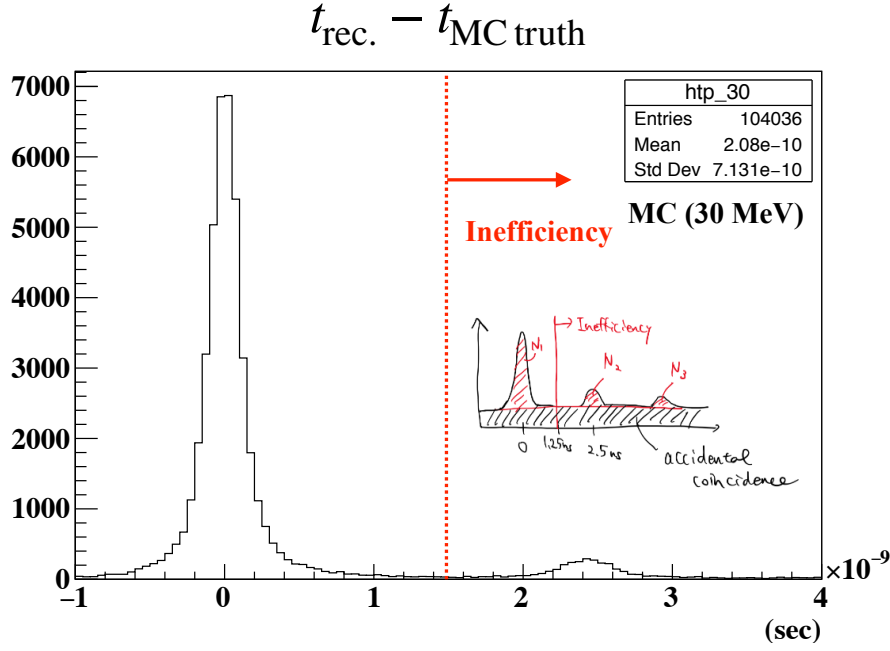
Figure 7.17: Difference between e^+ reconstructed time and truth time (MC).

Missing turn inefficiency   Another item we need to take into account for the e^+ efficiency is inefficiency coming from missing turns. Some e^+ tracks are reconstructed with one or more missing turns. In these cases, the track length is mistakenly estimated by $\sim 2.5\,\text{ns}$ (for one missing turn). Such e^+s are counted as Michel e^+s. However, they do not pass the signal selection because the analysis window is narrower than 2.5 ns (Section 7.6.7). Therefore, the number of Michel events should be corrected by a factor less than 1 according to the missing turn inefficiency. This inefficiency becomes larger for e^+ with smaller energy because of less tracking quality.

We estimate the inefficiency using the signal MC. Figure 7.17 shows the difference between e^+ reconstructed time and MC truth time. Missing turn events have peak shifted by $\sim 2.5\,\text{ns}$ from the main peak. The purpose here is to estimate a fraction of these missing turn events with respect to all events. First, the same e^+ selection is applied to the MC events as the Michel counting. Then the accidental coincident events are subtracted shown in the small figure in Figure 7.17. The final inefficiency is calculated by setting threshold as a half of 1 turn (1.25 ns).

$$(7.18) \qquad \text{Inefficiency} = 1 - \frac{N_1}{N_1 + N_2 + N_3 + \cdots}$$

MC events are generated with the several mass of X setups. Each setup has different e^+ energy and all the results are summarized in Figure 7.18. Larger error in the lower e^+ energy is due to limited statistics. By fitting an energy dependent function, $p_0 + \frac{p_1}{\sqrt{E}}$, one can get the missing turn inefficiency at selected e^+ energy shown in Table 7.8. The estimated inefficiency is reasonable, compared with previous results in the MEG analysis[97]. The relative positron efficiency is
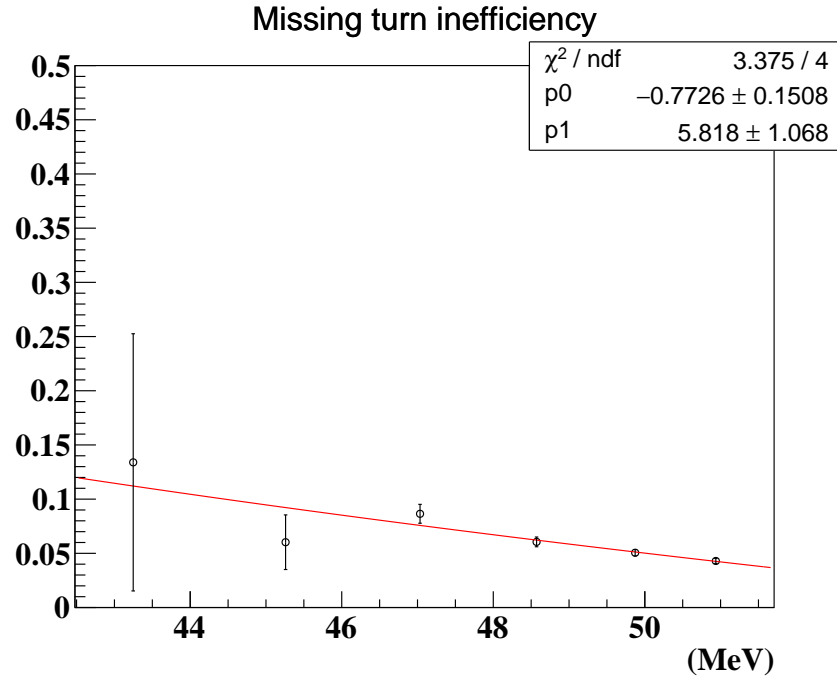
Figure 7.18: Missing turn inefficiency as a function of e^+ energy.

Table 7.8: Missing turn inefficiency. Uncertainties show the fitting errors.

m_X (MeV)	e^+ energy (MeV)	inefficiency (%)	uncertainty (%)
45	43.2477	11.21	1.16
40	45.2589	9.22	0.81
35	47.0334	7.57	0.51
30	48.5714	6.22	0.29
25	49.8728	5.12	0.17
20	50.9375	4.26	0.2

multiplied by the inefficiency.

7.6.5 γ efficiency

This is described in Section 5.3.4 and summarized in Figure 5.20.
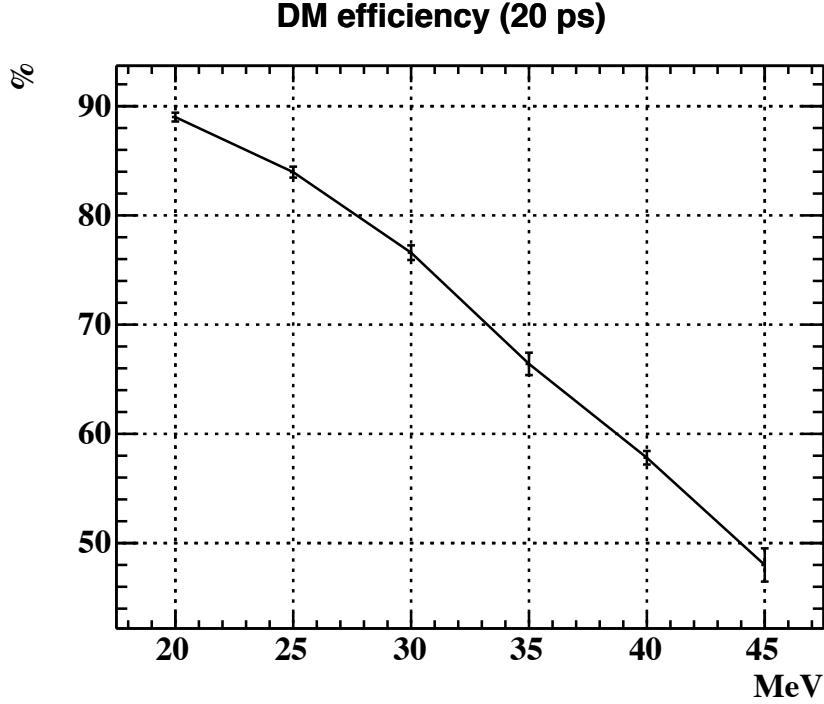
DM efficiency (20 ps)

Figure 7.19: DM efficiency (20 ps). The error bar only includes statistical uncertainty. Other systematic uncertainties are included in the estimation of the normalization factor (Section 7.6.9).

7.6.6 Direction match efficiency

$$\epsilon_{\mathrm{DM}} \quad \equiv \quad P(\mathrm{DM-triggered}|e^+, 2\gamma \in \mathrm{acceptance},$$

$$e^+ \,\mathrm{triggered}, e^+ \,\mathrm{reconstructed}, \gamma \,\mathrm{triggered}, 2\gamma \,\mathrm{reconstructed})$$

The direction match (DM) trigger is prepared to select back-to-back events of $\mu^+ \to e^+\gamma$ decay. For the back-to-back events like $\mu^+ \to e^+\gamma$, there are correlations between hit position of γ and e^+. Look-up tables in which PMT ID of the LXe detector and matched ID and z of the TC detector are listed are prepared. Only matched events are triggered to reduce accidental backgrounds. The trigger efficiency including the DM trigger for $\mu^+ \to e^+\gamma$ decay was, for example, 97% in 2013. However, γ and e^+ in $\mu^+ \to e^+X, X \to \gamma\gamma$ are not back-to-back. Thus the direction match trigger efficiency for $\mu^+ \to e^+X, X \to \gamma\gamma$ becomes worse.

We used two different look-up tables in the 5-year data taking. One is used in 2009 and 2010, and the other was used after 2010. In the later setup, a multi buffer read-out scheme is implemented [6, 73], which enables us to loose the direction match conditions. We estimated the DM efficiency using the signal MC and look-up tables. In MC, we check if PMT of LXe with the highest number of photoelectrons and TC hit position ID (ϕ, z) of generated events matches the direction match condition. Figure 7.19 shows the DM efficiency as a function of m_X for 20 ps.

117
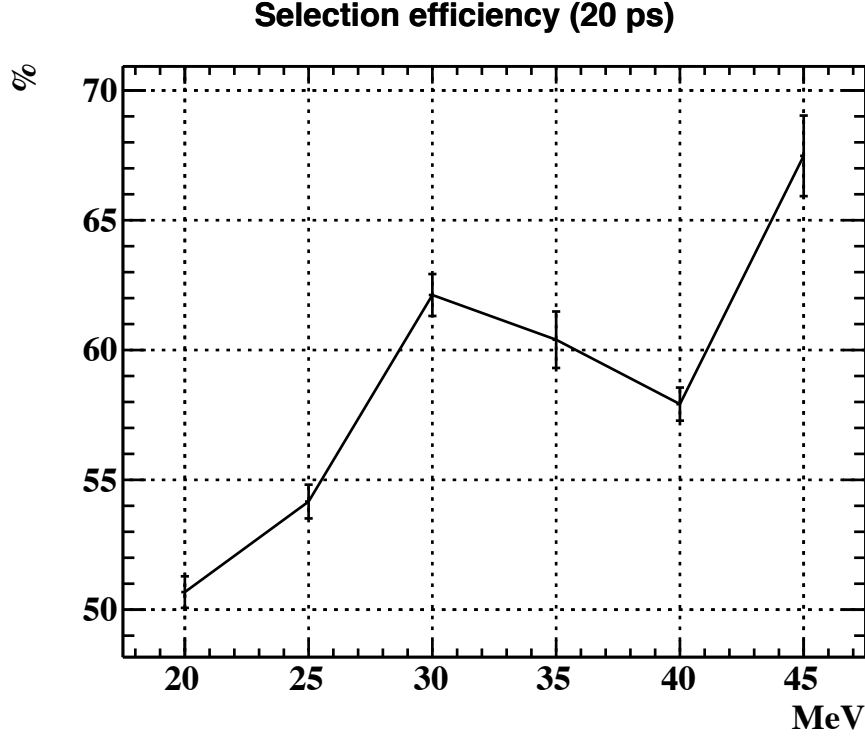
Figure 7.20: Selection efficiency (20 ps). The error bar only includes statistical uncertainty. Other systematic uncertainties are included in the estimation of the normalization factor (Section 7.6.9).

These values are weighted averaged efficiency using normalization of each year. Results of 5 ps and 40 ps are also summarized in Appendix B.

7.6.7   Selection efficiency

$$\epsilon_{\text{selection}} \equiv P(\text{selected}|e^+, 2\gamma \in \text{acceptance}, e^+ \text{triggered},$$
$$e^+ \text{reconstructed}, \gamma \text{triggered}, 2\gamma \text{reconstructed}, \text{DM-triggered})$$

The selection conditions determined in Section 7.4 are applied. The selection efficiency is estimated using the signal MC and summarized in Figure 7.20 for 20 ps. These values are weighted averaged efficiency using normalization of each year. No significant difference among lifetimes is observed and the results of 5 ps and 40 ps are also summarized in Appendix B. As discussed in Section 7.4, selection conditions depend on different mass assumptions; the conditions are optimized for each mass. Thus the mass dependence is not monotonic. In particular, for $m_X = 45$ MeV, the number of background events is smaller than others and looser selection conditions are applied compared with other masses, which gives the highest selection efficiency close to $\sim 70\%$.
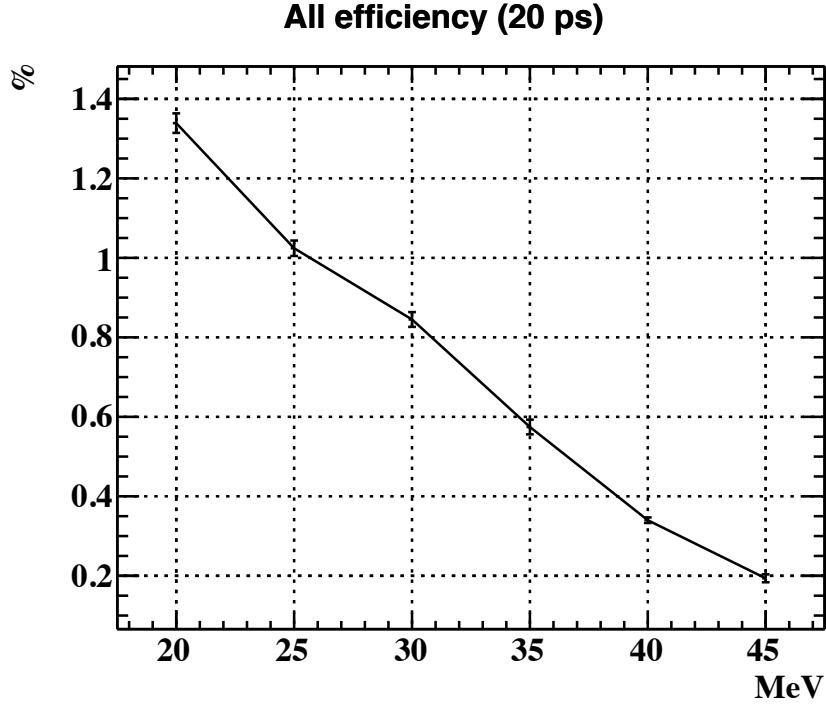
Figure 7.21: Overall efficiency (20 ps). The error bar only includes statistical uncertainty. Other systematic uncertainties are included in the estimation of the normalization factor (Section 7.6.9).

In the previous study [62], selection conditions were fixed regardless of m_X. The efficiency was ~50% at 20 MeV and it decreased as a function of mass down to ~20% at 45 MeV. As we discuss in Section 7.9, this (at most) ~ 3 times improvement of the selection efficiency gives the better single event sensitivity than expected from statistics.

7.6.8 Overall efficiency

The overall signal efficiency, which is defined by $\epsilon_\gamma \times \epsilon_{DM} \times \epsilon_{selection}$, is shown in Figure 7.21. This estimation is based on the signal MC. These values are weighted averaged efficiencies using normalization of each year. No large difference among lifetimes is observed and the results of 5 ps and 40 ps are also summarized in Appendix B. m_X dependence comes mainly from the gamma acceptance (Section 5.3.4) and the DM efficiency (Section 7.6.6).

7.6.9 Uncertainties on normalization

Uncertainties on the normalization factor are summarized in Table 7.9 for 20 ps. Systematic uncertainty of the signal efficiency is dominant. Statistical one comes from the MC statistics. In 2013, for $\tau = 20$ps, it is 9.3% for 20 MeV and 17% for 45 MeV. The systematic uncertainty is further factorized as shown in Table 7.10. Among them, γ detection efficiency (Section 5.3.4.5) and MC smearing are major components. Other systematic uncertainties including the direction match trigger efficiency and the selection efficiency are included in the MC smearing. In the reconstruction chapters (Chapter 4, Chapter 5), we estimate smearing parameters to reproduce data resolution using MC. To estimate the effect of its systematic uncertainty on the signal efficiency, we prepare several MC settings smeared by mean, mean + systematic uncertainty, and mean − systematic uncertainty of each smearing parameter. Then, we define difference of the signal efficiency from the mean as systematic uncertainty of the signal efficiency originating from MC smearing parameter.

In the signal estimation (Section 7.3, Section 7.8), the uncertainties (both systematic and stasitical ones) are assigned as systematic uncertainties of the normalization factor.

Table 7.9: Relative uncertainty (in 2013, 20 ps)

m_X (MeV)	20	25	30	35	40	45
Number of Michel events (Section 7.6.1)	0.43 % (stat)	0.56 % (stat)	0.64 % (stat)	0.90 % (stat)	1.3 % (stat)	2.7 % (stat)
Branching ratio of Michel	-	-	-	-	-	-
Fraction of Michel events (Section 7.6.2)	0.01 % (sys)	0.00 % (sys)	0.00 % (sys)	0.00 % (sys)	0.00 % (sys)	0.00 % (sys)
Prescale (Section 7.6.3)	-	-	-	-	-	-
Prescale correction (Section 7.6.3)	0.89 % (stat)	0.89 % (stat)	0.89 % (stat)	0.89 % (stat)	0.89 % (stat)	0.89 % (stat)
Relative positron efficiency correction (Section 7.6.4)	1.4 % (sys)	-	-	-	-	-
Missing turn correction (Section 7.6.4)	0.21 % (sys)	0.18 % (sys)	0.31 % (sys)	0.55 % (sys)	0.89 % (sys)	1.3 % (sys)
Signal efficiency	9.3 % (sys) 1.8 % (stat)	8.8 % (sys) 1.9 % (stat)	9.4 % (sys) 2.2 % (stat)	10 % (sys) 3.1 % (stat)	10 % (sys) 1.9 % (stat)	17 % (sys) 4.7 % (stat)
Total: systematic uncertainty (Table 7.10)	9.5 %	8.8 %	9.4 %	10 %	10 %	17 %
Total: statistical uncertainty	2.1 %	2.2 %	2.5 %	3.4 %	2.4 %	5.5 %

Table 7.10: Systematic uncertainties of signal efficiency (in 2013, 20 ps)

mass (MeV)	20	25	30	35	40	45
γ acceptance (Section 5.3.4.1)	1.3 %	2.0 %	3.4 %	2.9 %	5.4 %	1.9 %
γ trigger efficiency (Section 5.3.4.2)	0.98 %	0.32 %	0.26 %	0.52 %	1.4 %	3.2 %
γ detection efficiency (Section 5.3.4.5)	7.4 %	7.4 %	7.4 %	7.4 %	7.4 %	7.4 %
γ pileup inefficiency (Section 5.3.4.4)	2.8 %	2.8 %	2.8 %	2.8 %	2.8 %	2.8 %
MC smearing	4.8 %	3.4 %	3.8 %	5.3 %	3.3 %	14 %

S.E.S (20 ps)
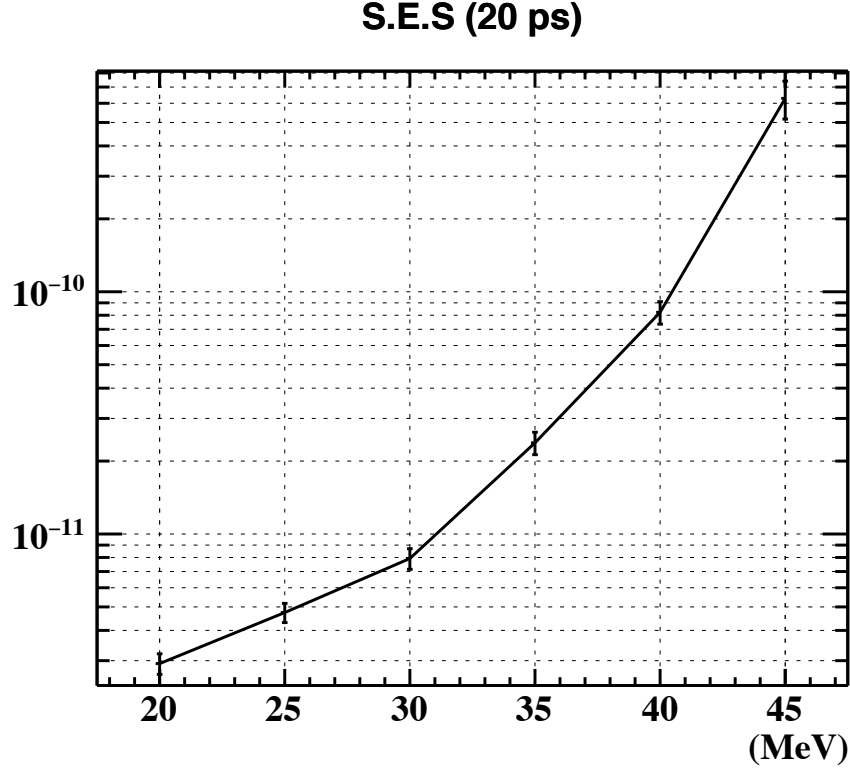
Figure 7.22: Single event sensitivity (20 ps).

7.6.10 Single event sensitivity

Based on the factorization of the normalization factor k defined in Equation (7.15), we discussed all the components in the previous sections. The single event sensitivity (SES) is defined as a BR which gives 1 event observation. It is given by $1/k$. It is calculated for each m_X and τ. Figure 7.22 shows SES for $\tau = 20$ ps (these measured values correspond to $1/k_0$).

The signal efficiency is estimated using MC generated with an interval of 5 MeV while the signal search is performed with an interval of 1 MeV. Therefore, SES curve in Figure 7.22 is fitted with exponential of quadratic function [7] to get SES with a interval of 1 MeV by interpolation. Averaged uncertainties of edge points (5 MeV interval) are assigned to uncertainties of in-between points. In addition, systematic uncertainties coming from the fitting are added to the systematic uncertainties of interpolated points. Table 7.11 summarizes all the SESs after the interpolation. Note that listed SES of 20, 25, 30, 35, 40, 45 MeV are estimated values while those for the other masses in between are fitted values.

---

[7] $y = \exp(ax^2 + bx + c)$, where y is SES and x is m_X. This function is used to interpolate points with a small number of parameters.

Table 7.11: Single event sensitivity. The uncertainty includes both statistical and systematic ones. Interpolated points have larger uncertainties coming from the fitting.

MeV	5 ps		20 ps		40 ps	
	SES	uncertainty (%)	SES	uncertainty (%)	SES	uncertainty (%)
20	3.3×10^{-12}	9.4	2.9×10^{-12}	9.7	3.0×10^{-12}	9.5
21	3.6×10^{-12}	12.3	3.2×10^{-12}	12.2	3.3×10^{-12}	12.0
22	3.7×10^{-12}	11.8	3.3×10^{-12}	11.5	3.4×10^{-12}	11.3
23	4.0×10^{-12}	11.5	3.6×10^{-12}	11.1	3.7×10^{-12}	11.0
24	4.3×10^{-12}	11.4	3.9×10^{-12}	11.1	3.9×10^{-12}	10.9
25	5.4×10^{-12}	10.2	4.7×10^{-12}	9.1	4.8×10^{-12}	9.0
26	5.2×10^{-12}	11.8	4.7×10^{-12}	11.3	4.8×10^{-12}	11.2
27	5.8×10^{-12}	12.0	5.3×10^{-12}	11.6	5.3×10^{-12}	11.4
28	6.6×10^{-12}	12.2	6.1×10^{-12}	11.8	6.1×10^{-12}	11.7
29	7.7×10^{-12}	12.4	7.0×10^{-12}	12.0	7.0×10^{-12}	11.8
30	8.6×10^{-12}	9.8	7.9×10^{-12}	9.8	7.9×10^{-12}	9.8
31	1.1×10^{-11}	13.2	9.9×10^{-12}	12.8	9.8×10^{-12}	12.4
32	1.3×10^{-11}	13.2	1.2×10^{-11}	12.8	1.2×10^{-11}	12.4
33	1.6×10^{-11}	13.2	1.5×10^{-11}	12.7	1.4×10^{-11}	12.3
34	2.0×10^{-11}	13.1	1.8×10^{-11}	12.5	1.8×10^{-11}	12.2
35	2.6×10^{-11}	11.7	2.4×10^{-11}	10.6	2.3×10^{-11}	10.0
36	3.2×10^{-11}	13.0	2.9×10^{-11}	12.6	2.8×10^{-11}	11.9
37	4.1×10^{-11}	12.9	3.8×10^{-11}	12.5	3.7×10^{-11}	11.8
38	5.4×10^{-11}	12.8	5.0×10^{-11}	12.4	4.8×10^{-11}	11.7
39	7.2×10^{-11}	12.8	6.6×10^{-11}	12.5	6.4×10^{-11}	11.8
40	9.0×10^{-11}	10.3	8.2×10^{-11}	10.6	7.9×10^{-11}	9.8
41	1.3×10^{-10}	14.0	1.2×10^{-10}	16.2	1.2×10^{-10}	15.3
42	1.9×10^{-10}	14.3	1.7×10^{-10}	16.6	1.6×10^{-10}	15.7
43	2.7×10^{-10}	14.6	2.4×10^{-10}	17.2	2.3×10^{-10}	16.2
44	3.8×10^{-10}	15.0	3.4×10^{-10}	17.8	3.2×10^{-10}	16.7
45	6.3×10^{-10}	14.1	6.3×10^{-10}	17.9	5.6×10^{-10}	17.1

7.7 Statistical methods

Confidence interval The confidence interval of the number of signal is constructed based on the Feldman-Cousins approach [108]. We use a profile-likelihood ratio ordering [53] when building the interval. The test statistic λ_p is defined as

$$(7.19) \qquad \lambda_p(s) = \frac{\mathscr{L}(s, \hat{\hat{\theta}})}{\mathscr{L}(\hat{s}, \hat{\theta})},$$

where the likelihood function is defined in Equation (7.9), s is the number of signal and θ is a set of nuisance parameters (N_1, N_2, N_3, r). $\hat{s}$ and $\hat{\theta}$ are the values of s and θ which maximize the likelihood and $\hat{\hat{\theta}}$ is the value of θ which maximizes the likelihood with a fixed s. The interval is constructed using the distribution of the likelihood ratio of many pseudo experiments (toyMC) simulated with PDFs defined in Equation (7.9). As we discussed in Section 7.3, the uncertainty of the normalization factor estimated in Section 7.6.9 is incorporated as one of the nuisance parameters, r.

Significance Hypothesis testing is performed on null hypothesis with the same test statistic as Equation (7.19). After opening the blind box, we calculate null p-values and they are converted into the values of significance.

7.7.1 Expected sensitivity

[8] The sensitivity of the analysis is evaluated by a weighted average of the distribution of the branching ratio limits at 90% C.L. under a null signal hypothesis. The expected number of background events ($N_{\text{exp,BG}}$) is summarized in Table 7.3. Assuming a null signal hypothesis, the expected number of observed events ($N_{\text{exp,obs}}$) is 0 or 1. Therefore, we show the expected upper limit (= sensitivity) as weighted average of probabilities of these two cases; the sensitivity is given by internal division points between expected branching ratio limits for $N_{\text{exp,obs}} = 0$ and $N_{\text{exp,obs}} = 1$ with the following ratio:

$$\text{Poi}(\lambda = N_{\text{exp,BG}}, k = 0) : \text{Poi}(\lambda = N_{\text{exp,BG}}, k = 1) = 1 : N_{\text{exp,BG}},$$

where $\text{Poi}(\lambda, k) = \lambda^k e^{-\lambda}/k!$. Figure 7.23 shows sensitivity (in red). Upper (lower) bound of the blue band corresponds to 90% upper limit assuming the number of observed events to be 1 (0). In most cases, the sensitivity is close to the lower bound of the blue band.

---

[8]Sensitivity is usually given by the median of the expected number of events under the background only hypothesis. In our case, however, the expected number of events are 0 or 1 and the sensitivity is 0 for most of the masses. Thus, we define sensitivity as the mean of the expected number of events, not the median.
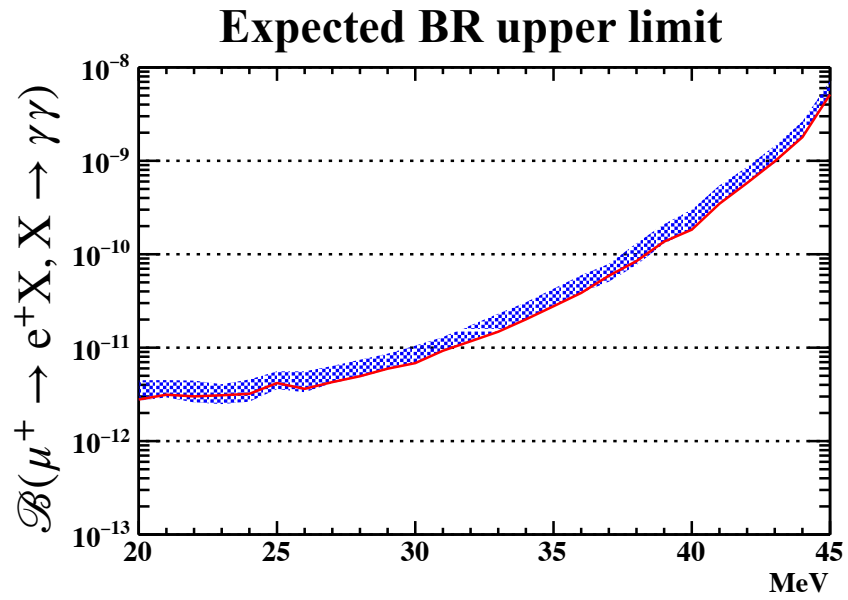
Expected BR upper limit

Figure 7.23: Expected branching ratio upper limit (90%C.L., red line). See text for the explanation of the blue band.

Table 7.12: The number of observed events in signal region (the most right column). The number of sideband events in Table 7.3 and best fit value of N_{BG} in the signal region are also listed together.

mass (MeV)	sideband			best fit N_{BG} in signal region	observed events in signal region
	$A(=A_1+A_2)$	$B(=B_1+B_2)$	$C(=C_1+C_2+C_3+C_4)$		
20	0	0	1	$0.048^{+0.228}_{-0.048}$	1
21	0	0	3	$0.139^{+0.184}_{-0.084}$	0
22	1	0	5	$0.278^{+0.275}_{-0.275}$	0
23	3	0	3	$0.516^{+0.362}_{-0.362}$	0
24	2	0	1	$0.415^{+0.321}_{-0.321}$	1
25	2	0	3	$0.344^{+0.330}_{-0.330}$	0
26	0	0	3	$0.143^{+0.186}_{-0.143}$	0
27	0	0	1	$0.048^{+0.173}_{-0.048}$	0
28	0	0	1	$0.046^{+0.171}_{-0.046}$	0
29	0	0	1	$0.048^{+0.229}_{-0.048}$	1
30	0	0	0	$0.000^{+0.167}_{-0.000}$	0
31	0	0	1	$0.046^{+0.171}_{-0.046}$	0
32	0	0	0	$0.000^{+0.167}_{-0.000}$	0
33	0	0	0	$0.000^{+0.206}_{-0.000}$	0
34	0	0	0	$0.000^{+0.286}_{-0.000}$	1
35	0	0	0	$0.000^{+0.286}_{-0.000}$	2
36	0	0	0	$0.000^{+0.286}_{-0.000}$	2
37	1	0	0	$0.400^{+0.450}_{-0.400}$	1
38	0	0	2	$0.156^{+0.230}_{-0.156}$	0
39	0	0	1	$0.078^{+0.216}_{-0.078}$	0
40	0	0	0	$0.000^{+0.206}_{-0.000}$	0
41	0	0	0	$0.000^{+0.206}_{-0.000}$	0
42	0	0	0	$0.000^{+0.206}_{-0.000}$	0
43	0	0	1	$0.078^{+0.216}_{-0.078}$	0
44	0	0	0	$0.000^{+0.206}_{-0.000}$	0
45	0	0	0	$0.000^{+0.206}_{-0.000}$	0

7.8 Results

7.8.1 Observed events in the signal region

In previous sections, we discussed how to obtain the signal number and we determined event selections using MC and sideband data. Thus, we open the blind box to count events after the selection. The number of events in the signal region is summarized in the most right column of Table 7.12. We observed non-zero events in some masses. Summing up observed events gives 9 events. 5 events of them are unique events. Its multiplicity is summarized in Table 7.13. For instance, the event of (run number, event number) = (103742, 1386) is appeared in the four adjacent mass bins. Reconstructed variables of observed events are summarized in Table C.1. Figure 7.24 shows event displays of LXe for the five events. No strange event is observed.

Table 7.13: Multiplicity of observed events

run number	event number	year	multiplicity	in which mass
103742	1386	2011	4	34, 35, 36, 37
109469	998	2011	1	29
156793	761	2012	1	24
161743	1102	2012	1	20
217733	683	2013	2	35, 36
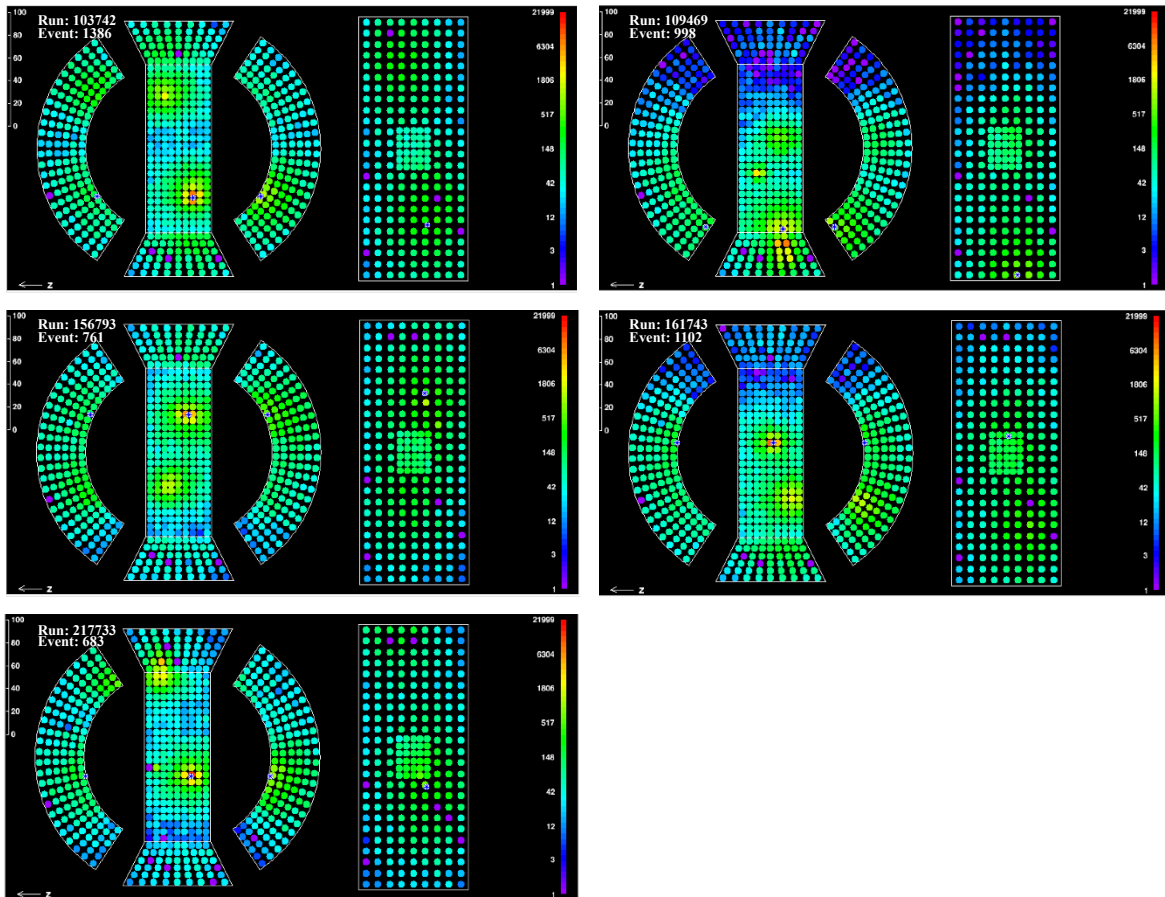
Figure 7.24: Event displays of observed events
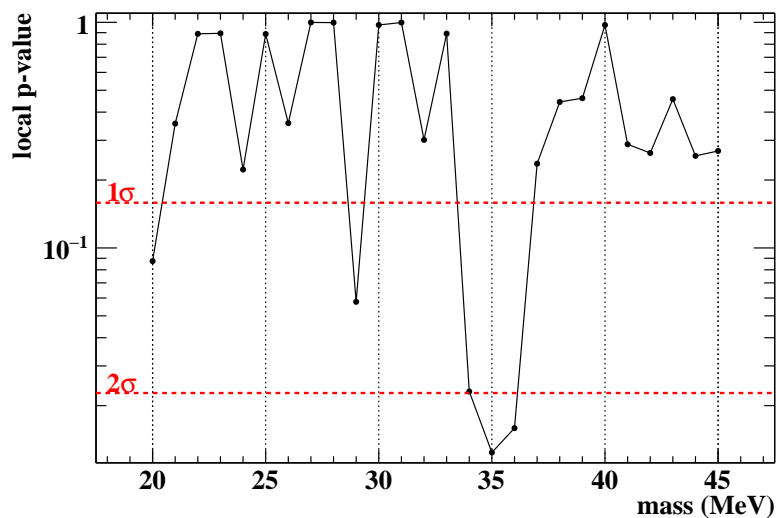
Figure 7.25: p-value under null signal hypothesis as a function of assumed masses.

7.8.2 p-value and significance

Under null signal hypothesis p-values are calculated for all mass points and the results are shown in Figure 7.25. Likelihood ratio is used as the test statistic (Equation (7.19)). We observed the lowest p-value of 0.012 at 35 MeV, which corresponds to 2.2 σ significance.

Since we do not know the signal mass, we need to take the look elsewhere effect into account to calculate the global significance. This effect is estimated according to Refs. [53, 109, 110]. For the local p-value of $p_{\text{local}} = 0.012$, its global one p_{global} is obtained to be $p_{\text{local}} = 0.10$, which corresponds to 1.3 σ.

7.8.3 Branching ratio limits

Figure 7.26 shows branching ratio limits (90%C.L.) obtained from this analysis (in blue) together with the sensitivity (Section 7.7.1) and upper limits of the other experiment [57] and the previous MEG study [62]. The upper bound of the blue region corresponds to the upper limits of the branching ratio of $\mu^+ \to e^+ X, X \to \gamma\gamma$. Since we observed small excess of events and uses the Feldman-Cousins approach to calculate the confidence interval of the signal number, some masses also have lower limits. The branching ratio upper limits of 10^{-11} (20 MeV) $- 10^{-9}$ (45 MeV) are set.

Figure 7.26: Branching ratio limits (90% C.L.) for $\tau = 20$ ps. The blue band are obtained from upper and lower limits in this analysis. The red dashed line is sensitivity calculated in Section 7.7.1. Yellow and green lines are results from the other experiment [57] and the previous MEG study [62]. The MEG2012 result is recalculated based on the Feldman-Cousins approach for fair comparison.

7.9 Discussion
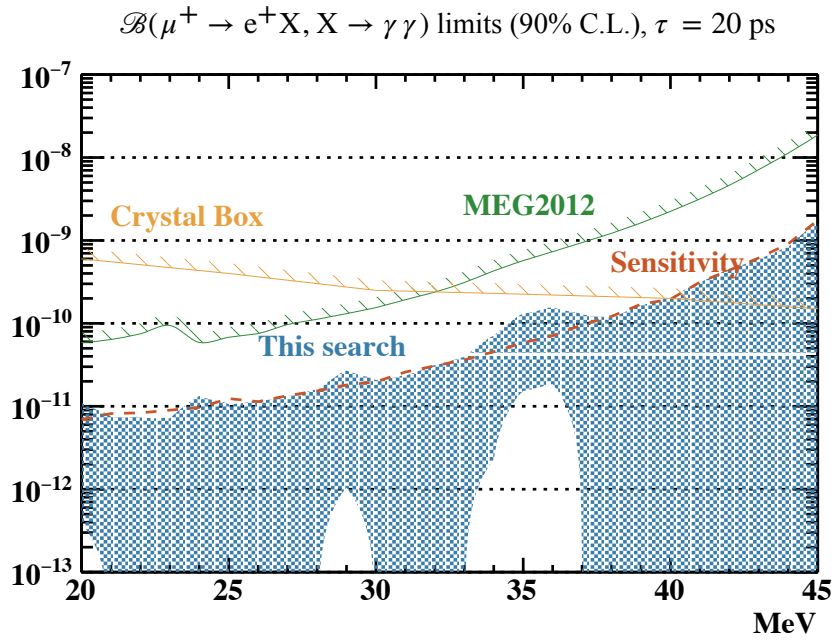
We observed excess events in some mass regions. The lowest p-value of 0.012 was observed at $m_X = 35$ MeV, which corresponds to 2.2 σ (local) and 1.3 σ (global) significance. Therefore, this excess is not statistically significant. We assumed decay length of X to be less than 1 cm and three lifetime settings (5, 20, and 40 ps) were used as an example. However, no significance difference was observed among these settings. If the signal exist below 5 ps, it should be observed in these settings because of the limitation of the vertex resolution (~1 cm). Therefore, we conclude that this null result holds for $\tau < 40$ ps.

We set limits on the branching ratio of $\mu^+ \to e^+X, X \to \gamma\gamma$. As shown in Figure 7.26, the branching ratio upper limits are improved by a factor of 4.4–13 depending on m_X from the previous analysis using the first two years of the MEG data [62]. Generally, the higher m_X is, the larger improvements become.

SES (Single event sensitivity) is improved by a factor of 5.6–13 depending on m_X. Statistics and positron analysis updates contribute at most ~ 5 times smaller SES compared with the previous analysis. This improvement mainly comes from the optimization of selection conditions depending on m_X. We set the criteria to optimize selection conditions and m_X dependent thresholds are determined in Section 7.4 while selection conditions were not optimized well in the previous analysis. Selection efficiency is improved by a factor of at most ~3 in 45 MeV.

Owing to the larger statistics and updates on analyses, the branching ratio upper limit has been pushed down to the level of $\mathscr{O}(10^{-11})$. Our results set more stringent upper limits comparing with not only the previous MEG studies [62], but also the upper limits converted from the Crystal Box experiment's results [57] in $m_X > 30$ MeV, which was not reached by the previous MEG studies.

Prospects We use the full datasets of the MEG experiment in this analysis. The beam intensity was forced to be reduced in order to suppress the accidental background and operate the detector stably. The upgraded experiment, the MEG II experiment is currently being prepared [7]. We plan to upgrade all the detectors to make the maximum use of the muon beam at the world's highest intensity ($7 \times 10^7 \mu$/s) at Paul Scherrer Institut in Switzerland. Experimental sensitivity of the MEx2G search is expected to be improved by one order of magnitude ($\mathscr{O}(10^{-12})$) using the MEG II datasets in all mass regions.

The higher mass has worse sensitivity in this analysis. This is mainly due to the 2γ acceptance and the direction match efficiency. The 2γ acceptance is determined by the geometry of liquid xenon gamma calorimeter and this part cannot be improved because the geometry is not changed so much in the upgrade. The direction match efficiency can be worsened in the upgrade if we consider only the $\mu^+ \to e^+\gamma$ search in the data-taking. γ position resolution is expected to be improved by a factor two, which enables us to tighten the direction match trigger condition. This means that non back-to-back events are more unlikely to be triggered, resulting in a worse trigger

efficiency for $\mu^+ \to e^+X, X \to \gamma\gamma$. However, the MEG II trigger development is underway and trigger efficiencies for higher mass can be improved by at most factor ~2 if a dedicated trigger for $\mu^+ \to e^+X, X \to \gamma\gamma$ is prepared. This will be a matter of discussion in the MEG II collaboration.

All the resolutions will be improved by a factor of two, which will also improve the signal selection efficiency to some extent. There is no dedicated time calibration source for 2γ events in this analysis. Developments of dedicated analysis tools for 2γ from annihilation-in-flight positrons provide us with the new time calibration source. We may set tighter selection conditions for timing distributions, resulting in better selection efficiency.

A further update on the MEx2G search can be possible in the future CLFV experiments. The MEG II experiment and the Mu3e experiment are being prepared and these experiments will be CLFV experiments in the 2020s. There are several considerations of the next generation CLFV experiments [68, 111, 112]. Proposed detectors have full acceptance around the beam axis, unlike the MEG detector. In addition, the muon beam intensity at PSI will also be upgraded around 10^9 [113]. The expected sensitivity will be further pushed down by one order of magnitude, $\mathcal{O}(10^{-13})$.

CONCLUSION

We have searched for a lepton flavor violating muon decay mediated by a new light particle. The charged lepton flavor violation is one of the powerful tools to search for new physics beyond the standard model. On the other hand, light new physics has attracted a great deal of attention. In the analysis performed in this thesis, we combined these two different directions and have searched for the $\mu^+ \to e^+ X, X \to \gamma\gamma$ decay using the full datasets (2009–2013) of the MEG experiment.

No significant excess was found in the mass region of 20–45 MeV and lifetime below 40 ps for a new light particle. Thus, we set the most stringent upper limits of the branching ratio in the mass region of 20–40 MeV. In particular, the upper limits are pushed down to the level of $\mathscr{O}(10^{-11})$ for 20–30 MeV.

It is at most 60 times more stringent result than the bound converted from the previous experiment, the Crystal Box experiment [57]. Together with the previous analysis [62] using the first-two-year physics data of the MEG experiment, this is the first direct search of the $\mu^+ \to e^+ X, X \to \gamma\gamma$ decay in the world.

The upgrade of the MEG experiment, the MEG II experiment, is planned and the sensitivity is expected to be improved by one order of magnitude, resulting in $\mathscr{O}(10^{-12})$ in mass region of 20–30 MeV.

LINEAR ASSUMPTION

I n this appendix, we discuss the linearity assumption on the background distribution. We use the following relations in the background and signal estimation (Section 7.2 and Section 7.3).

$$(A.1) \qquad N_{\mathrm{A}}^{\mathrm{exp}} \quad = \quad N_2 \times \frac{y_{\mathrm{C1}} + y_{\mathrm{C2}}}{y_{\mathrm{B}}} + N_3 \times \frac{y_{\mathrm{C1}} + y_{\mathrm{C2}}}{y_{\mathrm{B}}}$$

$$(A.2) \qquad N_{\mathrm{B}}^{\mathrm{exp}} \quad = \quad N_1 \times \frac{x_{\mathrm{C1}} + x_{\mathrm{C2}}}{x_{\mathrm{A}}} + N_3 \times \frac{x_{\mathrm{C1}} + x_{\mathrm{C2}}}{x_{\mathrm{A}}} + N_2 \times f_{\mathrm{escape}}$$

$$(A.3) \qquad N_{\mathrm{C}}^{\mathrm{exp}} \quad = \quad N_3 \times \frac{y_{\mathrm{C1}} + y_{\mathrm{C2}}}{y_{\mathrm{B}}} \times \frac{x_{\mathrm{C1}} + x_{\mathrm{C2}}}{x_{\mathrm{A}}} + N_2 \times f_{\mathrm{escape}} \times \frac{y_{\mathrm{C1}} + y_{\mathrm{C2}}}{y_{\mathrm{B}}}$$

The same equations as Equation (7.2), (7.3), and (7.4) are listed again. The number of events in the sidebands can be calculated from N_1, N_2, N_3 assuming their linear distribution over time. Hereafter, we check this assumption.

We need to check the following three factors:

$$(A.4) \qquad \frac{x_{\mathrm{C1}} + x_{\mathrm{C2}}}{x_{\mathrm{A}}}, \frac{y_{\mathrm{C1}} + y_{\mathrm{C2}}}{y_{\mathrm{B}}}, f_{\mathrm{escape}}$$

Equation (A.1) means that the expected number of background events in the region A has type2 (N_2) and type3 (N_3) contributions. If we assume the background distribution is linear, we can scale the number of events with a factor of ratio between the box sizes. The scaling factors of $t_{\gamma\gamma}$ and $t_{\gamma_1 e}$ are $\frac{x_{\mathrm{C1}} + x_{\mathrm{C2}}}{x_{\mathrm{A}}}$ and $\frac{y_{\mathrm{C1}} + y_{\mathrm{C2}}}{y_{\mathrm{B}}}$, respectively.

On the other hand, the first two terms in Equation (A.2) means that the region B has type1 (N_1) and type3 (N_3) contributions. In addition, we need to consider escape events from the $t_{\gamma\gamma}$ peak. f_{escape} is defined as a fraction of events in $|t_{\gamma\gamma}| > 1\,\mathrm{ns}$ escaped from $|t_{\gamma\gamma}| \leq 1\,\mathrm{ns}$ whose origin is $t_{\gamma\gamma}$ peak.
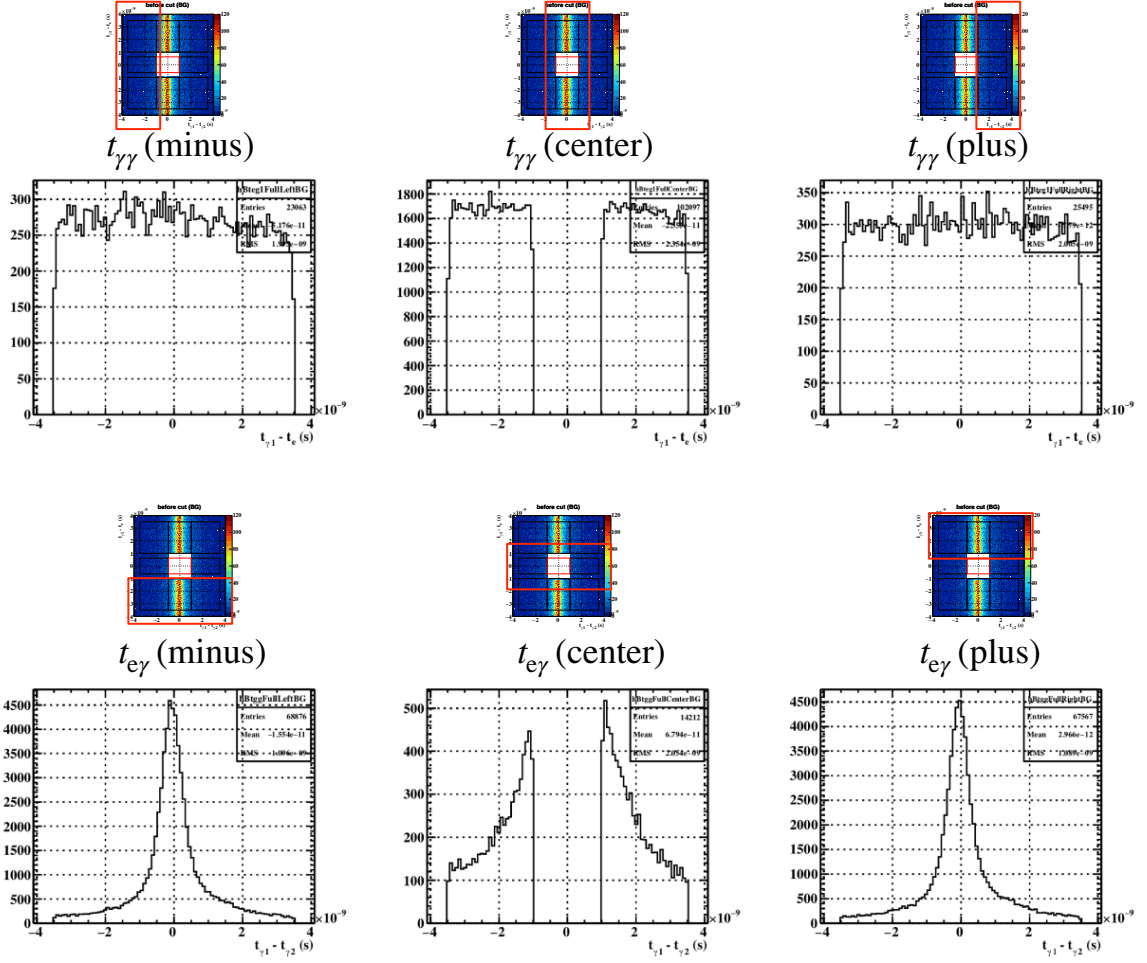
Figure A.1: Sideband distributions. The small figure with red box above the each plot shows which part of sideband data is used to make the corresponding distribution. Top left: $t_{e\gamma}$ distribution ($t_{\gamma\gamma} \leq -1$ns), top center: $t_{e\gamma}$ distribution ($|t_{\gamma\gamma}| < 1$ns), top right: $t_{e\gamma}$ distribution ($t_{\gamma\gamma} \geq 1$ns), bottom left: $t_{\gamma\gamma}$ distribution ($t_{e\gamma} \leq -1$ns), bottom center: $t_{\gamma\gamma}$ distribution ($|t_{e\gamma}| < 1$ns), bottom right: $t_{\gamma\gamma}$ distribution ($t_{e\gamma} \geq 1$ns)

Sideband distributions　Figure A.1 shows sideband distributions of BG events (data). Top three plots show that $t_{\gamma_1 e}$ distribution seems to be linear. Bottom three plots show that $t_{\gamma\gamma}$ floor distribution seems to be linear but escaped events from the $t_{\gamma\gamma}$ peak exist in $t_{\gamma\gamma} > 1$ns. Then, we discuss factors listed in Equation (A.4) using these distributions quantitatively.

Scaling factor: $\frac{y_{C1}+y_{C2}}{y_B}$　First, we discuss $\frac{y_{C1}+y_{C2}}{y_B}$. Figure A.2 is $t_{e\gamma}$ distribution ($t_{\gamma\gamma} \geq 1$ns). The left plot is $|t_{e\gamma}| \geq 1$ns and the right plot is $|t_{e\gamma}| < 1$ns. To check linearity of the distribution, a linear function (in red in the left plot of Figure A.2) is fitted to the left distribution and take difference between the fitted function and $t_{e\gamma}$ distribution in $|t_{e\gamma}| < 1$ns. The blue line in the right plot is the same one as the red line in the left one. Reduced χ^2 of the fitting is 1.25 (in the left
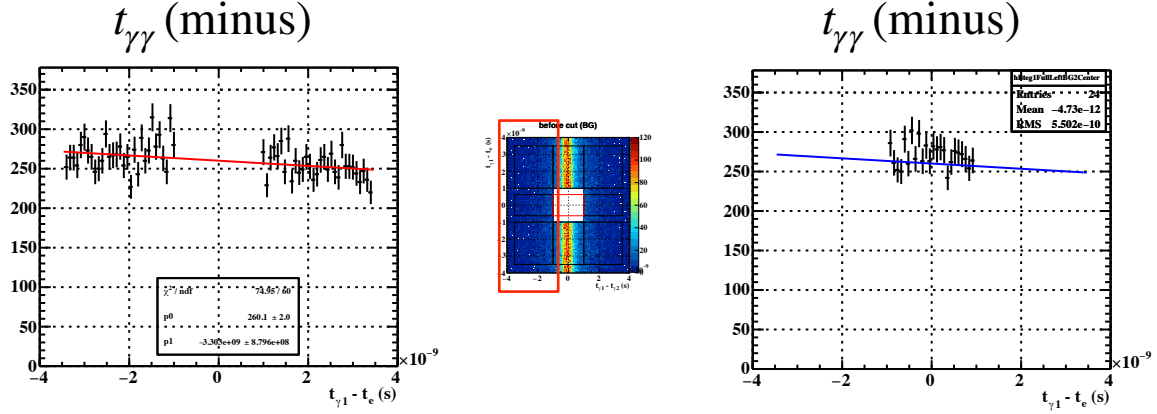
Figure A.2: $t_{e\gamma}$ distribution ($t_{\gamma\gamma} \leq -1$ns)
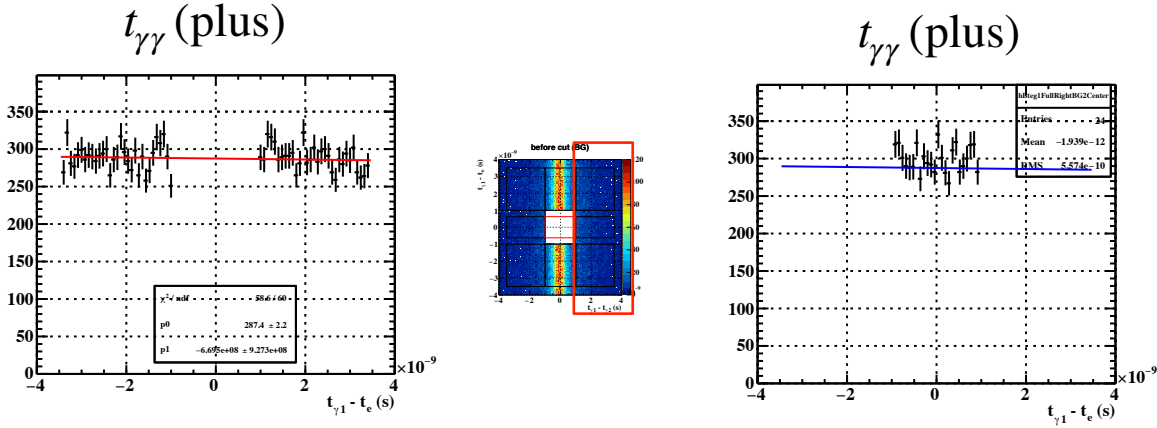
Figure A.3: $t_{e\gamma}$ distribution ($t_{\gamma\gamma} \geq 1$ns)

plot). The difference between number of events from the right distribution and integral of the fitted function (in the right plot) is 4.0%.

The same fitting is applied on $t_{e\gamma}$ distribution ($t_{\gamma\gamma} \geq 1$ns) as shown in Figure A.3. The reduced χ^2 of the fitting is 0.98 (in the left plot). The difference between number of events from the right distribution and integral of the fitted function (in the right plot) is 3.9%.

First, the reduced χ^2s are close to 1 and the background distribution is linear over time. Second, the difference between fitting function and the distribution is ~4% and it is statistical uncertainty is dominant. Therefore, systematic uncertainty of the scaling factor is negligible compared with statistical uncertainty when estimating signal number ($\sim \mathcal{O}(10)\%$) due to the fact that low statistics is expected after event selections. In conclusion, we can assume the background distribution is linear and the scaling factor $\frac{y_{C1} + y_{C2}}{y_B}$ can be used with no other corrections.

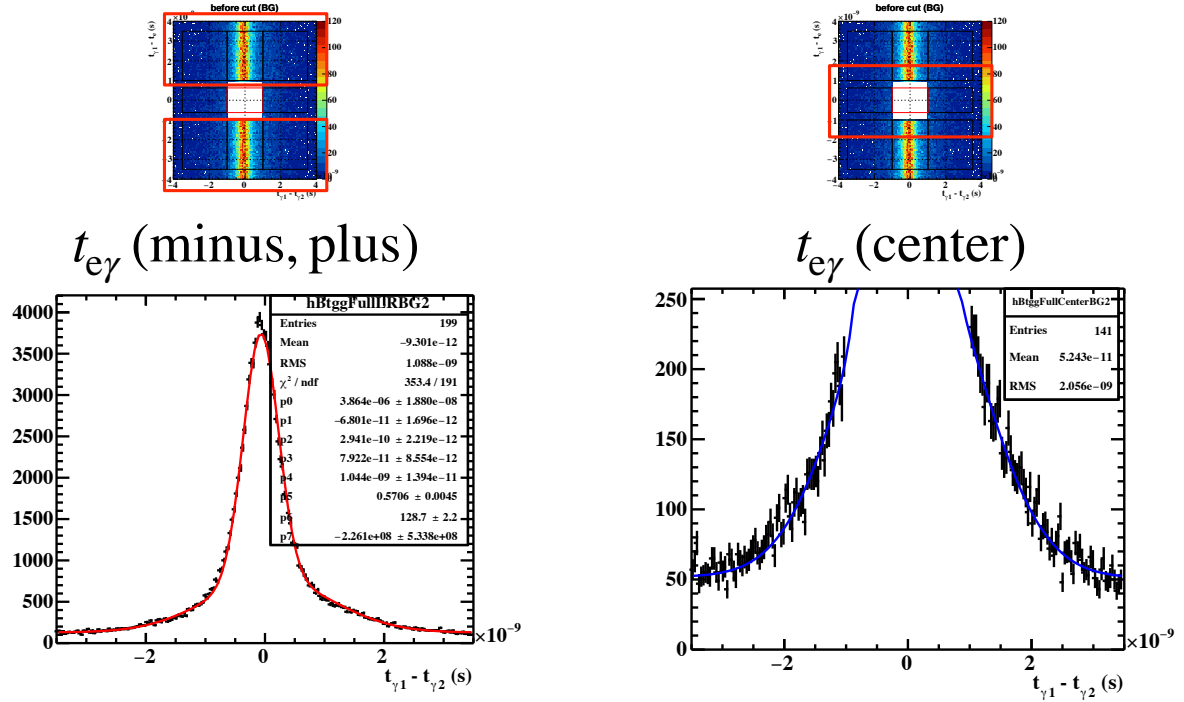Figure A.4: Left: $t_{\gamma\gamma}$ distribution ($|t_{e\gamma}| \geq 1$ns), right: $t_{\gamma\gamma}$ distribution ($|t_{e\gamma}| < 1$ns). See texts in detail. Note that the blue curve in the right figure is not a fitted curve of this histogram, it is estimated from the other region (in the left figure).

Scaling factor: $\frac{x_{C1}+x_{C2}}{x_A}$ Next we check if the $t_{\gamma\gamma}$ distribution (bottom plots in Figure A.1) is linear. However, it is difficult to pick up linear components solely and check the assumption because tail component from the $t_{\gamma\gamma}$ peak is also included in $t_\gamma > 1$ns as shown in the plots. Therefore we assume that the background (type 1) is linear and deviation from the linear function is included in the estimation of escape factor f_{escape}.

Escape factor: f_{escape} To check the linearity of the background distribution and estimate the escape factor, first, a double Gaussian is fitted to $t_{\gamma\gamma}$ distribution in $|t_{e\gamma}| \geq 1$ns as shown in the red curve in the left figure of Figure A.4. Then, it is scaled with the box size to the region of $|t_{e\gamma}| < 1$ns as shown in the blue curve in the right figure. A difference between the blue curve and the event distribution in the right figure include deviation from the assumption that $t_{\gamma\gamma}$ background is linear ($\frac{x_{C1}+x_{C2}}{x_A}$) and uncertainty of f_{escape}.

The difference is estimated to be 2.8%, which is statistical component is dominant. Therefore, systematic uncertainty coming from the two components listed above is less than 2.8%.

f_{escape} is estimated from the fitted double Gaussian function to be 17.1%. This is added to Equation (A.2) and (A.3) as a correction term. Its uncertainty is negligible compared with statistical uncertainty when estimating signal number ($\sim \mathcal{O}(10)\%$) due to the fact that low

statistics is expected after event selections.

Finally, concerning the scaling factor $\frac{x_{C1}+x_{C2}}{x_A}$, its systematic uncertainty is negligible because its effect is included in f_{escape} and is smaller than $17.1\% \times 2.8\%$.

Summary   In conclusion, we can assume the background distributions are linear and we use the factors ($\frac{x_{C1}+x_{C2}}{x_A}$, $\frac{y_{C1}+y_{C2}}{y_B}$, f_{escape}) as constant values in the background and signal estimation.

In this section, efficiency and S.E.S. plots of 5 ps and 40 ps are shown. Only 20 ps case is shown in the main section.

B.1 Gamma acceptance
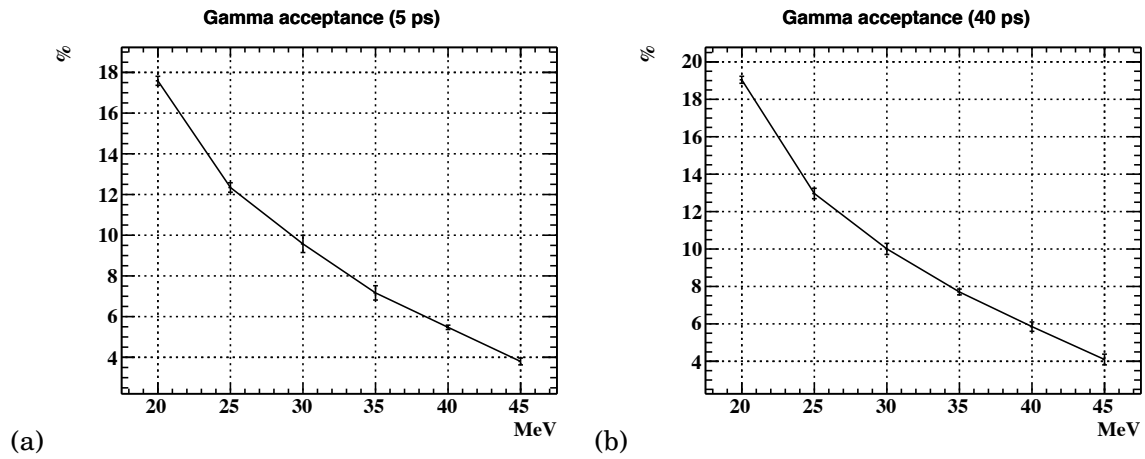
Figure B.1: Gamma acceptance of (a) 5 ps and (b) 40 ps.

B.2   Gamma efficiency
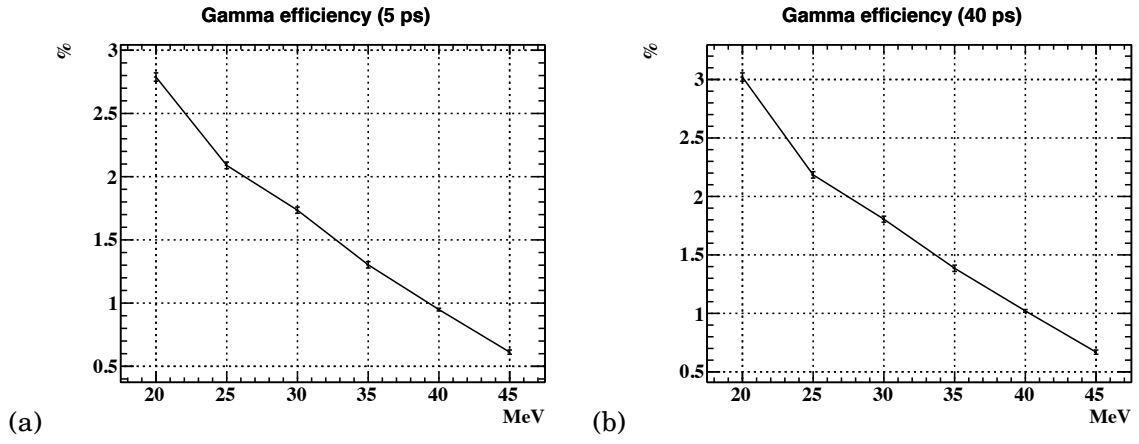
Figure B.2: Gamma efficiency of (a) 5 ps and (b) 40 ps.

B.3   Direction match efficiency
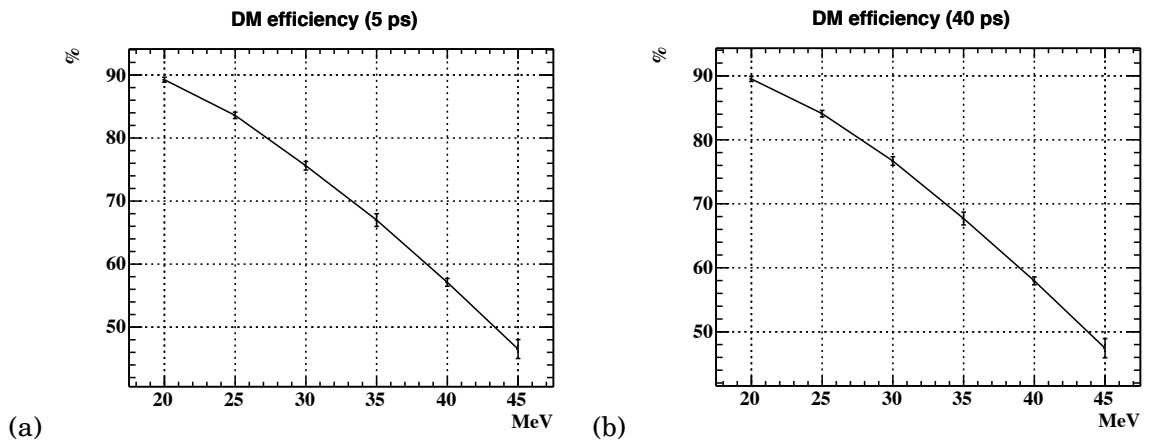
Figure B.3: DM efficiency of (a) 5 ps and (b) 40 ps.

B.4 Selection efficiency
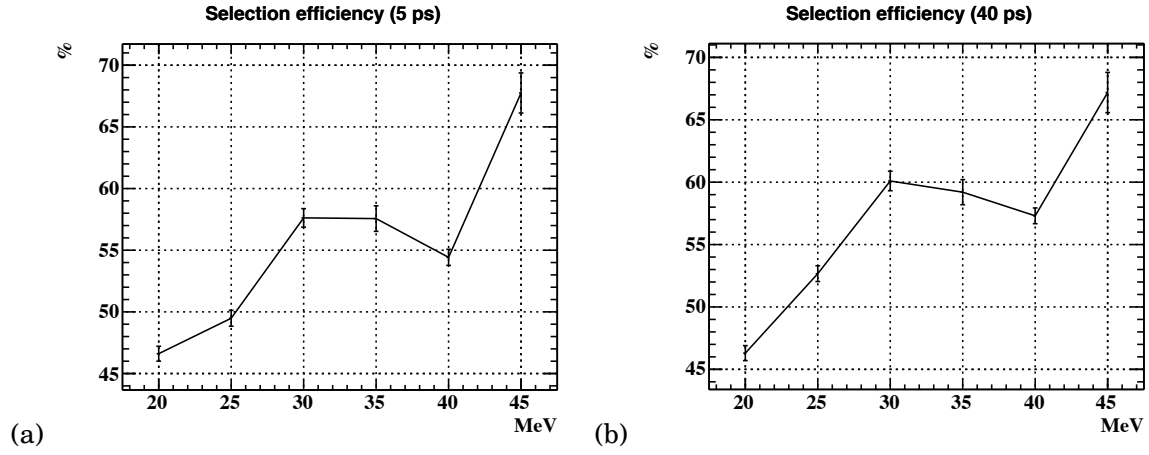
Figure B.4: Selection efficiency of (a) 5 ps and (b) 40 ps.

B.5 Signal efficiency
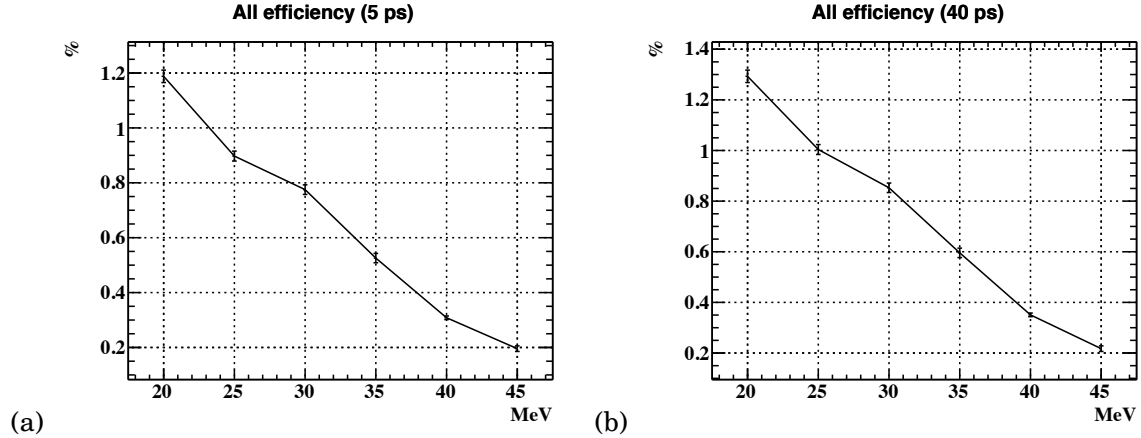
Figure B.5: Overall signal efficiency of (a) 5 ps and (b) 40 ps.

B.6 Single event sensitivity

(a)

(b)
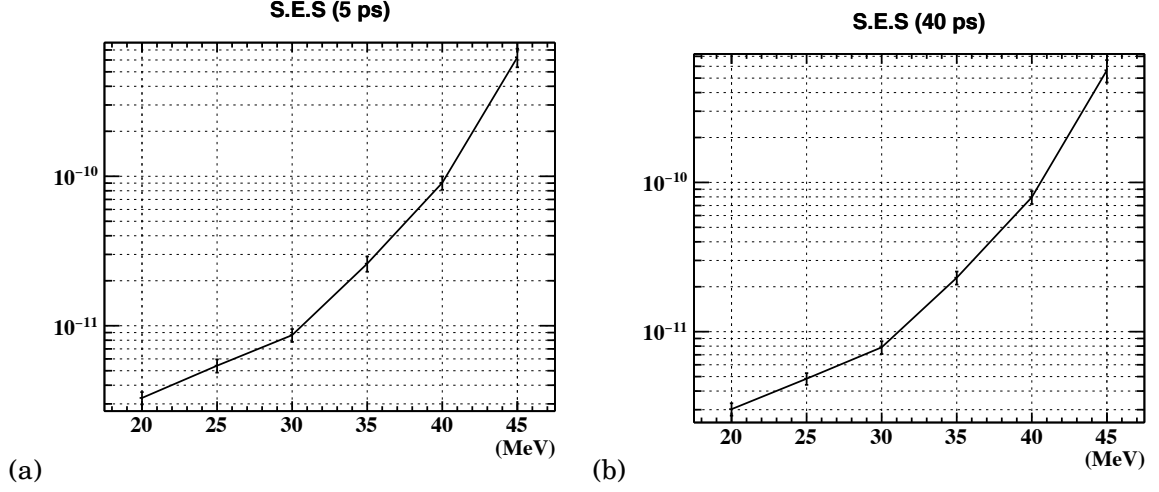
Figure B.6: Single event sensitivity of (a) 5 ps and (b) 40 ps.

B.7 Branching ratio limits

(a)

(b)

Figure B.7: Branching ratio limits of (a) 5 ps and (b) 40 ps. The blue band are obtained upper and lower limits in this analysis.

C

FURTHER INFORMATION OF THE EXCESS

C.1 Reconstructed variables of excess events

Table C.1 shows reconstructed variables of the observed 5 events.

Table C.1: Reconstructed variables of excess events

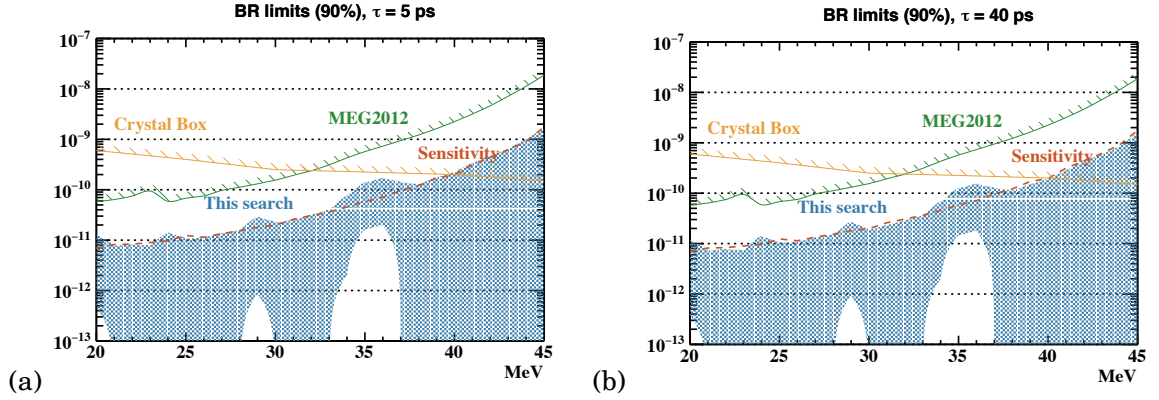| mass (MeV)	20	24	29	34	35	35	36	36	37
run number	161743	156793	109469	103742	103742	217733	103742	217733	103742
event number	1102	761	998	1386	1386	683	1386	683	1386
year	2012	2012	2011	2011	2011	2013	2011	2013	2011
$t_{e\gamma_1}$ (ns)	-5.36×10^{-2}	3.79×10^{-3}	2.54×10^{-1}	4.15×10^{-1}	4.11×10^{-1}	1.86×10^{-1}	3.93×10^{-1}	1.62×10^{-1}	4.00×10^{-1}
$t_{\gamma\gamma}$ (ns)	5.96×10^{-1}	8.63×10^{-1}	-7.78×10^{-1}	-1.53×10^{-1}	-1.51×10^{-1}	2.98×10^{-1}	-1.4×10^{-1}	3.14×10^{-1}	-1.45×10^{-1}
u_1 (cm)	-18.5	-8.14	-11	-12	-12	-11.3	-12	-11.3	-12
v_1 (cm)	-39.3	33.2	-70.7	-41.9	-41.9	-16.5	-41.9	-16.5	-41.9
w_1 (cm)	10.7	3.68	8.96	3.56	3.56	1.75	3.56	1.75	3.56
u_2 (cm)	-1.83	7.78	-8.76	12.2	12.2	14.5	12.2	14.5	12.2
v_2 (cm)	8.99	-29.5	8.54	48.1	48.1	70.3	48.1	70.3	48.1
w_2 (cm)	2.11	9.16	14.2	12.5	12.5	6.58	12.5	6.58	12.5
uv distance (cm)	51.3	64.9	79.6	93.5	93.5	90.8	93.5	90.8	93.5
E_{γ_1} (MeV)	30.1	34.3	38.4	35.6	35.6	35.5	35.6	35.5	35.6
E_{γ_2} (MeV)	20.9	18.3	15.0	21.8	21.8	17.7	21.8	17.7	21.8
E_{e^+} (MeV)	50.3	50.6	47.9	47.3	47.3	46.1	47.3	46.1	47.3
momentum sum (MeV)	3.19	3.65	2.38	1.75	1.80	3.63	2.09	4.11	1.96
vertex χ^2	4.9	3.88	6.28	5.61	2.52	5.4	3.95	6.38	7.14
sum of energy (MeV)	101	103	101	105	105	99.3	105	99.3	105

INDEX OF ABBREVIATIONS

[1]   G. Aad *et al.*, Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, Phys. Lett. Sect. B Nucl. Elem. Part. High-Energy Phys. 716 (1) (2012) 1–29, `arXiv:1207.7214`, `doi:10.1016/j.physletb.2012.08.020`.

[2]   S. Chatrchyan *et al.*, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, Phys. Lett. Sect. B Nucl. Elem. Part. High-Energy Phys. 716 (1) (2012) 30–61, `arXiv:1207.7235`, `doi:10.1016/j.physletb.2012.08.021`.

[3]   The Nobel Prize in Physics 2013.
available at `https://www.nobelprize.org/prizes/physics/2013/summary/`(accessed on 2019-11-29) .

[4]   The Nobel Prize in Physics 2015.
available at `https://www.nobelprize.org/prizes/physics/2015/summary/`(accessed on 2019-11-29) .

[5]   Mori, T. and Ootani, W., Flavour violating muon decays (2014).
`doi:10.1016/j.ppnp.2014.09.001`, `doi:10.1016/j.ppnp.2014.09.001`.

[6]   A. M. Baldini *et al.*, Search for the lepton flavour violating decay $\mu^+ \to e^+\gamma$ with the full dataset of the MEG experiment, Eur. Phys. J. C 76 (8) (2016) 434, `doi:10.1140/epjc/s10052-016-4271-x`.

[7]   A. M. Baldini *et al.*, The design of the MEG II experiment: MEG II Collaboration, Eur. Phys. J. C 78 (5) (may 2018), `doi:10.1140/epjc/s10052-018-5845-6`.

[8]   A. Blondel *et al.*, Research Proposal for an Experiment to Search for the Decay $\mu \to$ eee (jan 2013), `arXiv:1301.6113`.

[9]   L. Bartoszek *et al.*, Mu2e Technical Design Report (jan 2015), `arXiv:1501.05241`.

[10] The COMET Collaboration *et al.*, COMET Phase-I Technical Design Report (dec 2018), `arXiv:1812.09018`.

149

[11]   M. Aaboud, Search for long-lived, massive particles in events with displaced vertices and missing transverse momentum in ffiffi s p = 13 TeV pp collisions with the ATLAS detector, Phys. Rev. D 97 (2018), `doi:10.1103/PhysRevD.97.052012`.

[12]   M. Aaboud *et al.*, Search for long-lived charginos based on a disappearing-track signature in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector, J. High Energy Phys. 2018 (6) (jun 2018), `arXiv:1712.02118`, `doi:10.1007/JHEP06(2018)022`.

[13]   M. Aaboud *et al.*, Search for heavy charged long-lived particles in proton–proton collisions at $\sqrt{s}$=13TeV using an ionisation measurement with the ATLAS detector, Phys. Lett. Sect. B Nucl. Elem. Part. High-Energy Phys. 788 (2019) 96–116, `doi:10.1016/j.physletb.2018.10.055`.

[14]   G. Aad *et al.*, The ATLAS experiment at the CERN large hadron collider, J. Instrum. 3 (8) (2008), `doi:10.1088/1748-0221/3/08/S08003`.

[15]   S. Chatrchyan *et al.*, The CMS experiment at the CERN LHC, J. Instrum. 3 (8) (2008), `doi:10.1088/1748-0221/3/08/S08004`.

[16]   H. Lubatti *et al.*, MATHUSLA: A Detector Proposal to Explore the Lifetime Frontier at the HL-LHC (jan 2019), `arXiv:1901.04040`.

[17]   A. Haas *et al.*, Looking for milli-charged particles with a new experiment at the LHC (oct 2014), `arXiv:1410.6816`, `doi:10.1016/j.physletb.2015.04.062`.

[18]   A. Ball *et al.*, A Letter of Intent to Install a milli-charged Particle Detector at LHC P5 (jul 2016), `arXiv:1607.04669`.

[19]   J. H. Yoo, The milliQan Experiment: Search for milli-charged Particles at the LHC (oct 2018), `arXiv:1810.06733`.

[20]   G. Aielli *et al.*, Expression of Interest for the CODEX-b Detector (nov 2019), `arXiv:1911.00481`.

[21]   W. Bonivento *et al.*, Proposal to Search for Heavy Neutral Leptons at the SPS (oct 2013), `arXiv:1310.1762`.

[22]   SHiP Collaboration *et al.*, A facility to Search for Hidden Particles (SHiP) at the CERN SPS (apr 2015), `arXiv:1504.04956`.

[23]   S. Alekhin *et al.*, A facility to search for hidden particles at the CERN SPS: The SHiP physics case, Reports Prog. Phys. 79 (12) (oct 2016), `arXiv:1504.04855`, `doi:10.1088/0034-4885/79/12/124201`.

[24] FASER Collaboration *et al*., FASER's Physics Reach for Long-Lived Particles (nov 2018), `arXiv:1811.12522`, `doi:10.1103/PhysRevD.99.095011`.

[25] FASER Collaboration *et al*., Letter of Intent for FASER: ForwArd Search ExpeRiment at the LHC (nov 2018), `arXiv:1811.10243`.

[26] FASER Collaboration *et al*., Technical Proposal for FASER: ForwArd Search ExpeRiment at the LHC (dec 2018), `arXiv:1812.09139`.

[27] FASER Collaboration *et al*., FASER: ForwArd Search ExpeRiment at the LHC (jan 2019), `arXiv:1901.04468`.

[28] A. J. Krasznahorkay *et al*., Observation of Anomalous Internal Pair Creation in Be 8: A Possible Indication of a Light, Neutral Boson, Phys. Rev. Lett. 116 (4) (jan 2016), `arXiv:1504.01527`, `doi:10.1103/PhysRevLett.116.042501`.

[29] A. J. Krasznahorkay *et al*., New evidence supporting the existence of the hypothetic X17 particle (oct 2019), `arXiv:1910.10459`.

[30] E. van Beveren, S. Coito, G. Rupp, On the existence of a superlight scalar boson, EPJ Web Conf. 95 (2015) 02007, `doi:10.1051/epjconf/20159502007`.

[31] K. Abraamyan *et al*., Check of the structure in photon pairs spectra at the invariant mass of about $38\,\mathrm{MeV}/c^2$, EPJ Web Conf. 204 (2019) 08004, `doi:10.1051/epjconf/201920408004`.

[32] Shinohara, S., Search for the rare decay $K_L \to \pi^0 \nu^- \nu$ at J-PARC KOTO experiment. available at `https://indico.cern.ch/event/769729/contributions/3510939/attachments/1904988/3145907/KAON2019_shinohara_upload.pdf`(accessed on 2019-11-30) .

[33] T. Kitahara *et al*., New physics implications of recent search for $K_L \to \pi^0 \nu^- \nu$ at KOTO`arXiv:1909.11111v2`.

[34] D. Egana-Ugrinovic, S. Homiller, P. Meade, Light Scalars and the KOTO Anomaly (nov 2019), `arXiv:1911.10203`.

[35] S. Gardner, X. Yan, LIght scalars with lepton number to solve the $(g-2)_e$ anomaly (jul 2019), `arXiv:1907.12571`.

[36] M. Bauer *et al*., Axion-like particles, lepton-flavor violation and a new explanation of a_μ and a_e (jul 2019), `arXiv:1908.00008`.

[37] R. D. Peccei, H. R. Quinn, CP conservation in the presence of pseudoparticles, Phys. Rev. Lett. 38 (25) (1977) 1440–1443, `doi:10.1103/PhysRevLett.38.1440`.

[38] S. Weinberg, A new light boson?, Phys. Rev. Lett. 40 (4) (1978) 223–226, `doi:10.1103/PhysRevLett.40.223`.

[39] F. Wilczek, Problem of strong P and T invariance in the presence of instantons, Phys. Rev. Lett. 40 (5) (1978) 279–282, `doi:10.1103/PhysRevLett.40.279`.

[40] C. Cornella, P. Paradisi, O. Sumensari, Hunting for ALPs with Lepton Flavor Violation, Tech. rep. (2019), `arXiv:1911.06279v1`.

[41] Y. Chikashige, R. N. Mohapatra, R. D. Peccei, Are there real goldstone bosons associated with broken lepton number?, Phys. Lett. B 98 (4) (1981) 265–268, `doi:10.1016/0370-2693(81)90011-3`.

[42] G. B. Gelmini, M. Roncadelli, Left-handed neutrino mass scale and spontaneously broken lepton number, Phys. Lett. B 99 (5) (1981) 411–415, `doi:10.1016/0370-2693(81)90559-1`.

[43] D. B. Reiss, Can the family group be a global symmetry?, Phys. Lett. B 115 (3) (1982) 217–220, `doi:10.1016/0370-2693(82)90647-5`.

[44] F. Wilczek, Axions and family symmetry breaking, Phys. Rev. Lett. 49 (21) (1982) 1549–1552, `doi:10.1103/PhysRevLett.49.1549`.

[45] Z. Berezhiani, M. Khlopov, Physical and astrophysical consequences of breaking of the symmetry of families. (In Russian), Sov.J.Nucl.Phys. 51 (1990) 1479–1491.

[46] J. Jaeckel, A family of WISPy dark matter candidates, Phys. Lett. Sect. B Nucl. Elem. Part. High-Energy Phys. 732 (2014) 1–7, `arXiv:1311.0880`, `doi:10.1016/j.physletb.2014.03.005`.

[47] K. Tsumura, L. Velasco-Sevilla, Phenomenology of flavon fields at the LHC, Phys. Rev. D - Part. Fields, Gravit. Cosmol. 81 (3) (feb 2010), `doi:10.1103/PhysRevD.81.036012`.

[48] M. Bauer, T. Schell, T. Plehn, Hunting the flavon, Phys. Rev. D 94 (5) (sep 2016), `arXiv:1603.06950`, `doi:10.1103/PhysRevD.94.056003`.

[49] Y. Ema *et al.*, Flaxion: a minimal extension to solve puzzles in the standard model, J. High Energy Phys. 2017 (1) (jan 2017), `arXiv:1612.05492`, `doi:10.1007/JHEP01(2017)096`.

[50] O. Davidi *et al.*, The hierarchion, a relaxion addressing the Standard Model's hierarchies, J. High Energy Phys. 2018 (8) (aug 2018), `arXiv:1806.08791`, `doi:10.1007/JHEP08(2018)153`.

[51] Y. Hochberg *et al.*, Mechanism for thermal relic dark matter of strongly interacting massive particles, Phys. Rev. Lett. 113 (17) (oct 2014), `doi:10.1103/PhysRevLett.113.171301`.

[52] Y. Hochberg *et al.*, Model for Thermal Relic Dark Matter of Strongly Interacting Massive Particles, Phys. Rev. Lett. 115 (2) (jul 2015), `doi:10.1103/PhysRevLett.115.021301`.

[53] M. Tanabashi *et al.*, Review of Particle Physics, Phys. Rev. D 98 (3) (aug 2018), `doi:10.1103/PhysRevD.98.030001`.

[54] C. D. Anderson, S. H. Neddermeyer, Cloud chamber observations of cosmic rays at 4300 meters elevation and near sea-level, Phys. Rev. 50 (4) (1936) 263–271, `doi:10.1103/PhysRev.50.263`.

[55] S. J. Freedman *et al.*, Limits on neutrino oscillations from $\bar{\mu}e$ appearance, Phys. Rev. D 47 (3) (1993) 811–829, `doi:10.1103/PhysRevD.47.811`.

[56] U. Bellgardt *et al.*, Search for the decay $\mu^+ \to e^+ e^+ e^-$, Nucl. Physics, Sect. B 299 (1) (1988) 1–6, `doi:10.1016/0550-3213(88)90462-2`.

[57] R. D. Bolton *et al.*, Search for rare muon decays with the Crystal Box detector, Phys. Rev. D 38 (7) (1988) 2077–2101, `doi:10.1103/PhysRevD.38.2077`.

[58] W. Bertl *et al.*, A search for μ-e conversion in muonic gold, Eur. Phys. J. C 47 (2) (2006) 337–346, `doi:10.1140/epjc/s2006-02582-x`.

[59] T. Goldman *et al.*, Light-boson emission in the decay of the μ^+, Phys. Rev. D 36 (5) (1987) 1543–1546, `doi:10.1103/PhysRevD.36.1543`.

[60] R. Bayes *et al.*, Search for two body muon decay signals, Phys. Rev. D - Part. Fields, Gravit. Cosmol. 91 (5) (mar 2015), `arXiv:1409.0638`, `doi:10.1103/PhysRevD.91.052020`.

[61] R. Eichler *et al.*, Limits for short-lived neutral particles emitted in μ^+ or π^+ decay, Phys. Lett. B 175 (1) (1986) 101–104, `doi:10.1016/0370-2693(86)90339-4`.

[62] Natori, Hiroaki, Search for a Lepton Flavor Violating Muon Decay Mediated by a New Light Neutral Particle with the MEG Detector (2012).
available at `http://inspirehep.net/record/1252114/?ln=ja`.

[63] Y. Kuno, Y. Okada, Muon decay and physics beyond the standard model, Rev. Mod. Phys. 73 (1) (2001) 151–202, `arXiv:9909265`, `doi:10.1103/RevModPhys.73.151`.

[64] T. Kinoshita, A. Sirlin, Radiative Corrections to Fermi Interactions, Phys. Rev. 113 (6) (1959) 1652–1660, `doi:10.1103/PhysRev.113.1652`.

[65] M. J. Dolan *et al.*, Revised constraints and Belle II sensitivity for visible and invisible axion-like particles, J. High Energy Phys. 2017 (12) (dec 2017), `arXiv:1709.00009`, `doi:10.1007/JHEP12(2017)094`.

[66] D. L. Anderson, C. D. Carone, M. Sher, Probing the light pseudoscalar window, Phys. Rev. D - Part. Fields, Gravit. Cosmol. 67 (11) (2003), `doi:10.1103/PhysRevD.67.115013`.

[67] Y. S. Liu, D. McKeen, G. A. Miller, Electrophobic Scalar Boson and Muonic Puzzles, Phys. Rev. Lett. 117 (10) (sep 2016), `arXiv:1605.04612`, `doi:10.1103/PhysRevLett.117.101801`.

[68] A.-K. Perrevoort, R.-K.-U. Heidelberg, F. für Physik, Sensitivity Studies on New Physics in the Mu3e Experiment and Development of Firmware for the Front-End of the Mu3e Pixel Detector, Ph.D. thesis.

[69] J. Heeck, W. Rodejohann, Lepton flavor violation with displaced vertices, Phys. Lett. B 776 (2018) 385–390, `doi:10.1016/J.PHYSLETB.2017.11.067`.

[70] Ulrich, Y., Physics of fundamental Symmetries and Interactions - PSI2019 (20-25 October 2019): High-precision QED prediction for low-energy lepton experiments · Indico. available at `https://indico.psi.ch/event/6857/contributions/19673/(accessed on 2019-11-25)`.

[71] G. Pruna, A. Signer, Y. Ulrich, Fully differential NLO predictions for the radiative decay of muons and taus, Phys. Lett. B 772 (2017) 452–458, `doi:10.1016/J.PHYSLETB.2017.07.008`.

[72] W. J. Marciano, T. Mori, J. M. Roney, Charged Lepton Flavor Violation Experiments, Annu. Rev. Nucl. Part. Sci. 58 (1) (2008) 315–341, `doi:10.1146/annurev.nucl.58.110707.171126`.

[73] J. Adam *et al.*, The MEG detector for $\mu^+ \to e^+ \gamma$ decay search, Eur. Phys. J. C 73 (4) (2013) 2365, `doi:10.1140/epjc/s10052-013-2365-2`.

[74] A. E. Pifer, T. Bowen, K. R. Kendall, A high stopping density $\mu+$ beam, Nucl. Instruments Methods 135 (1) (1976) 39–46, `doi:10.1016/0029-554X(76)90823-5`.

[75] A. M. Baldini *et al.*, Muon polarization in the MEG experiment: Predictions and measurements, Eur. Phys. J. C 76 (4) (2016) 223, `doi:10.1140/epjc/s10052-016-4047-3`.

[76] W. Ootani *et al.*, Development of a thin-wall superconducting magnet for the positron spectrometer in the MEG experiment, in: IEEE Trans. Appl. Supercond., Vol. 14, 2004, pp. 568–571, `doi:10.1109/TASC.2004.829721`.

[77] M. Hildebrandt, The drift chamber system of the MEG experiment, in: Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip., Vol. 623, 2010, pp. 111–113, `doi:10.1016/j.nima.2010.02.165`.

[78] M. De Gerone *et al.*, The MEG timing counter calibration and performance, Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip. 638 (1) (2011) 41–46, `doi:10.1016/j.nima.2011.02.044`.

[79] M. De Gerone *et al.*, Development and commissioning of the timing counter for the MEG experiment, IEEE Trans. Nucl. Sci. 59 (2) (2012) 379–388, `doi:10.1109/TNS.2012.2187311`.

[80] A. Baldini *et al.*, Absorption of scintillation light in a 100 l liquid xenon γ-ray detector and expected detector performance, Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip. 545 (3) (2005) 753–764, `arXiv:0407033`, `doi:10.1016/j.nima.2005.02.029`.

[81] R. Sawada, Performance of liquid xenon gamma ray detector for MEG, in: Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip., Vol. 623, 2010, pp. 258–260, `doi:10.1016/j.nima.2010.02.214`.

[82] XCOM: Photon Cross Sections Database | NIST.
available at `https://www.nist.gov/pml/xcom-photon-cross-sections-database`(accessed on 2019-12-26) .

[83] S. Kubota, M. Hishida, J. Raun, Evidence for a triplet state of the self-trapped exciton states in liquid argon, krypton and xenon, J. Phys. C Solid State Phys. 11 (12) (1978) 2645–2651, `doi:10.1088/0022-3719/11/12/024`.

[84] T. Doke, Fundamental Properties of Liquid Argon, Krypton and Xenon as Radiation Detector Media, Portugal.Phys. 12 (1981) 9–48.

[85] A. Hitachi *et al.*, Effect of ionization density on the time dependence of luminescence from liquid argon and xenon, Physical Review B 27 (9) (1983) 5279–5285, `doi:10.1103/PhysRevB.27.5279`.

[86] P. D. Thesis, B. Yusuke Uchiyama, Analysis of the First MEG Physics Data to Search for the Decay $\mu^+ \to e^+\gamma$, Ph.D. thesis (2009).

[87] S. Ritt, R. Dinapoli, U. Hartmann, Application of the DRS chip for fast waveform digitizing, in: Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip., Vol. 623, 2010, pp. 486–488, `doi:10.1016/j.nima.2010.03.045`.

[88] J. Adam *et al.*, Calibration and monitoring of the MEG experiment by a proton beam from a CockcroftWalton accelerator, Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip. 641 (1) (2011) 19–32, `doi:10.1016/j.nima.2011.03.048`.

[89] A. Papa, The calibration methods for the MEG experiment, Ph.D. thesis (2009). URL `https://meg.web.psi.ch/docs/theses/Angela.pdf`

[90] J. Spuller *et al.*, A remeasurement of the Panofsky ratio, Phys. Lett. B 67 (4) (1977) 479–482, `doi:10.1016/0370-2693(77)90449-X`.

[91] Y. Kuno, Y. Okada, Proposed $\mu \to e\gamma$ search with polarized muons, Phys. Rev. Lett. 77 (3) (1996) 434–437, `doi:10.1103/PhysRevLett.77.434`.

[92] R. Brun *et al.*, GEANT 3 : user's guide Geant 3.10, Geant 3.11, Tech. rep. (1987).

[93] Garfield.

[94] P. W. Cattaneo, R. Sawada, MEG simulation and analysis software, in: IEEE Nucl. Sci. Symp. Conf. Rec., 2010, pp. 1209–1213, `doi:10.1109/NSSMIC.2010.5873960`.

[95] P. W. Cattaneo *et al.*, The architecture of MEG simulation and analysis software, Eur. Phys. J. Plus 126 (7) (2011) 1–12, `doi:10.1140/epjp/i2011-11060-6`.

[96] D. Kaneko, The final result of $\mu^+ \to e^+\gamma$ search with the MEG experiment, Ph.D. thesis (2016).

[97] Y. Fujii, Search for the Lepton Flavor Violating Muon Decay $\mu^+ \to e^+\gamma$ with a Sensitivity below 10^{-12} in the MEG Experiment, Ph.D. thesis (2013).

[98] P. Billoir, Track fitting with multiple scattering: A new method, Nucl. Instruments Methods Phys. Res. 225 (2) (1984) 352–366, `doi:10.1016/0167-5087(84)90274-6`.

[99] R. Frühwirth, Application of Kalman filtering to track and vertex fitting, Nucl. Inst. Methods Phys. Res. A 262 (2-3) (1987) 444–450, `doi:10.1016/0168-9002(87)90887-4`.

[100] A. Fontana *et al.*, Use of GEANE for tracking in virtual Monte Carlo, in: J. Phys. Conf. Ser., Vol. 119, Institute of Physics Publishing, 2008, `doi:10.1088/1742-6596/119/3/032018`.

[101] S. Ogawa, Development of liquid xenon gamma-ray detector for MEG II experiment with highly granular readout realized by VUV sensitive large area MPPC, Ph.D. thesis (2016).

[102] ROOT: TSpectrum2 Class Reference.
available at `https://root.cern.ch/doc/master/classTSpectrum2.html`(accessed on 2019-11-12) .

[103] Y. Nishimura, A Search for the Decay $\mu^+ \rightarrow e^+\gamma$ Using a High-Resolution Liquid Xenon Gamma-Ray Detector, Ph.D. thesis (2010).

[104] A. Read, Linear interpolation of histograms, Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip. 425 (1-2) (1999) 357–360, `doi:10.1016/S0168-9002(98)01347-3`.

[105] G. Punzi, Sensitivity of searches for new signals and its optimization (aug 2003), `arXiv:0308063`.

[106] ROOT: TKDE Class Reference.
available at `https://root.cern.ch/doc/master/classTKDE.html`(accessed on 2019-11-13) .

[107] K. Cranmer, Kernel estimation in high-energy physics, Comput. Phys. Commun. 136 (3) (2001) 198–207, `arXiv:0011057`, `doi:10.1016/S0010-4655(00)00243-5`.

[108] G. J. Feldman, R. D. Cousins, Unified approach to the classical statistical analysis of small signals, Phys. Rev. D - Part. Fields, Gravit. Cosmol. 57 (7) (1998) 3873–3889, `arXiv:9711021`, `doi:10.1103/PhysRevD.57.3873`.

[109] E. Gross, O. Vitells, Trial factors for the look elsewhere effect in high energy physics, Eur. Phys. J. C 70 (1) (2010) 525–530, `arXiv:1005.1891`, `doi:10.1140/epjc/s10052-010-1470-8`.

[110] R. B. Davies, Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternatives, Biometrika 74 (1) (1987) 33, `doi:10.2307/2336019`.

[111] C.-h. Cheng, B. Echenard, D. G. Hitlin, The next generation of $\mu \rightarrow e\gamma$ and $\mu \rightarrow 3e$ CLFV search experiments (sep 2013), `arXiv:1309.7679`.

[112] G. Cavoto *et al.*, The quest for $\mu \rightarrow e\gamma$ and its experimental limiting factors at future high intensity muon beams, Eur. Phys. J. C 78 (1) (jan 2018), `arXiv:1811.12324`, `doi:10.1140/epjc/s10052-017-5444-y`.

[113] F. Berg *et al.*, Target studies for surface muon production, Phys. Rev. Accel. Beams 19 (2) (2016), `doi:10.1103/PhysRevAccelBeams.19.024701`.