Doctoral Dissertation
博士論文

Search for resonances in light-by-light scattering in 14.6 fb^{-1} of pp collisions at $\sqrt{s}$ = 13 TeV

（重心系エネルギー 13 TeV の陽子-陽子衝突データ 14.6 fb^{-1} を用いた光子-光子散乱における共鳴の探索）

A Dissertation Submitted for the Degree of Doctor of Philosophy
December 2022
令和 4 年 12 月博士（理学）申請

Department of Physics, Graduate School of Science,
The University of Tokyo
東京大学大学院理学系研究科
物理学専攻

Gen Tateno
舘野　元

Abstract

This thesis reports a search for resonances in light-by-light scattering using 14.6 fb^{-1} of $\sqrt{s}$ = 13 TeV pp collision data collected by the ATLAS experiment at the LHC. The light-by-light scattering mediates a scattering of the beam protons. The scattered proton is detected by the ATLAS Forward Proton (AFP) spectrometer, while the ATLAS detector reconstructs the photons. At least one of the outgoing protons is required to match the kinematics of the final state diphoton to reduce the background events. Diphoton resonances in the mass range from 150 GeV to 1600 GeV are searched using signal and background distribution models with unbinned fitting. The maximum observed local significance is 2.51 at m_X = 454 GeV, and there is no excess from the standard model. Upper limits are placed on the production cross section as a function of the resonance mass. The results are interpreted as the upper limits of the ALP coupling constant to two photons. All the results are obtained assuming the branching ratio to two photons is 100%. The observed upper limits of the coupling constant are about 0.04 TeV^{-1} in the range m_X = 200 to 1000 GeV.

Contents

1 Introduction

1.1 Light-by-light scattering at the LHC

The studies of light-by-light scattering, a process in which two photons interact and turn back into two photons, have long attracted interest. This interaction is supposed not to occur in a vacuum in the classical electromagnetism. However, the quantum electrodynamics (QED), a part of the standard model (SM) of particle physics, two photons can interact with each other through a loop of charged fermions or $W^{\pm}$ bosons [1]. In addition to studies of this interaction, observation of the effects of new physics beyond the standard model (BSM) mediating light-by-light scattering has been aimed in many experiments. Light-by-light scattering has been measured in photon scattering in the Coulomb field of a nucleus (Delbrück scattering) [2–5] and in the photon-splitting process [6]. Both apply electromagnetic fields to observe light-by-light scattering. Detailed backgrounds and the other experiments concerning light-by-light scattering are described in Ref. [7, 8].

Currently, high-energy colliders are the only place where the Schwinger limit of 10^{18} V m^{-1} is surpassed [9]. At the Large Hadron Collider (LHC) [10], the electromagnetic fields between beam hadrons are strong enough to create fluxes of high-energy virtual photons. These photons can then interact with each other, enabling light-by-light scattering. In 2016, the ATLAS collaboration claimed the first direct evidence of high-energy light-by-light scattering without any additional electromagnetic field applied [11]. This experiment made use of Pb+Pb collisions, which are LHC beam collision experiments, to collide lead ion beams with each other. The electromagnetic fields around the colliding lead ion can be treated as quasi-real photons with a small virtuality [12–14], and the photons can collide with each other, as known as ultraperipheral collision (UPC) [15, 16]. Hence, the ATLAS observation means the LHC was used as a photon-photon ($\gamma\gamma$) collider. Fig. 1.1 shows the Feynman diagram of such an event. Light-by-light scattering can occur in collisions between charged particles in general. Further progress of light-by-light scattering measurement in the LHC is described in Sec. 1.3.2.

Historically, the modeling of photon production from a fast-moving charged particle was developed following a discussion by Fermi [17]. The electric field produced by the charged particle is predominantly transverse. It is, in association with the magnetic field surrounding it, described as a flux of photons with

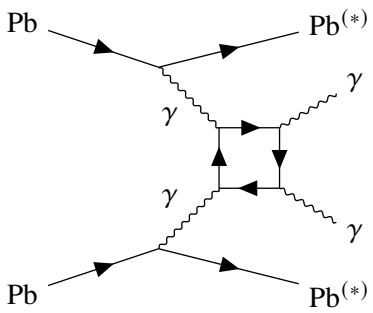


Figure 1.1: Feynman diagram of SM light-by-light scattering in Pb+Pb collisions. An electron is an example of a particle that forms a loop. The potential electromagnetic excitation of the lead ion is denoted by $^{(*)}$.

density $n(\omega)$, where ω is the energy equivalent to frequency. The cross section of the electromagnetic interaction of such photons with another particle was then approximated by the convolution of the photon flux with the relevant real photon production cross section. This model was simplified as Weizsäcker-Williams approximation (WWA) [13, 18] to enable easier calculation of $n(\omega)$ and the cross section which is given as

$$\mathrm{d}\sigma = \sigma_\gamma \mathrm{d}n,$$

where σ_γ is the cross section for absorption of photons with energy ω. The original WWA neglects photon virtuality. The story so far is detailed in Refs. [19, 20]. Likewise, the light-by-light scattering cross section in a collision between charged particles is given as

$$\mathrm{d}\sigma = \sigma_{\gamma\gamma} \mathrm{d}n_1 \mathrm{d}n_2,$$

where $\sigma_{\gamma\gamma}$ is the two-photons interaction cross section, and $\mathrm{d}n_i$ is the photon flux of charged particle i.

The WWA has been improved as equivalent photon approximation (EPA) [20] considering photon virtuality. The EPA is similar to the quantum chromodynamics (QCD) factorization theory which calculates the convolution of the parton distribution function (PDF) and the hard-process cross section to derive the cross section in a collider experiment. The photon distribution n_i is also defined as photon PDF and convoluted with the cross section of light-by-light scattering, an example of a hard process. In the EPA, the photon flux is given as a function of energy ω and virtuality Q^2 of the photon,

$$\mathrm{d}n_i = \frac{\alpha}{\pi}\frac{\mathrm{d}\omega_i}{\omega_i}\frac{\mathrm{d}Q_i^2}{Q_i^2}\left[\left(1-\frac{\omega_i}{E_i}\right)\left(1-\frac{(Q_{\min})_i^2}{Q_i^2}\right)(F_\mathrm{E})_i + \frac{\omega_i^2}{2E_i^2}(F_\mathrm{M})_i\right], \tag{1.1}$$

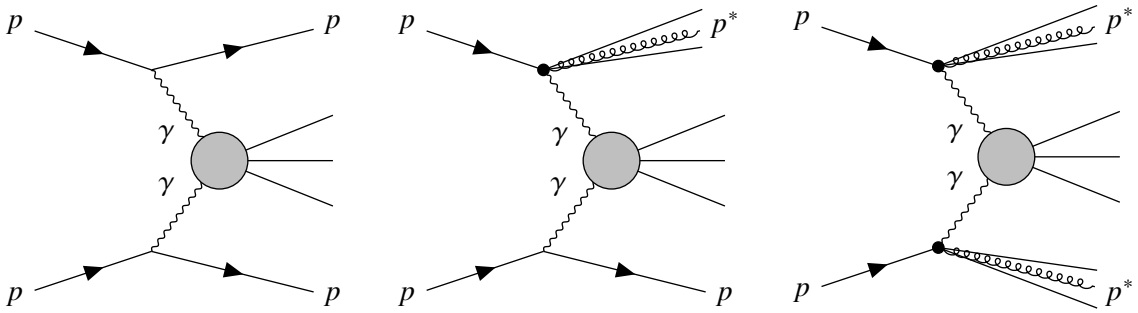
where α is the fine-structure constant, E is the incoming charged particle energy, $Q^2_{\min}$ is the minimum photon virtuality [20, 21]. The F_E and F_M depend on beam charged particle i. For example, if the particle i is electron, $(F_\mathrm{E})_i = (F_\mathrm{M})_i = 1$.

1.1.1 Light-by-light scattering in pp collisions

This thesis focuses on protons as the beam particles in LHC proton-proton (pp) collisions. Phenomenologically, photon production from incoming proton beams can be described as elastic or inelastic, which refers to how each photon is produced from the proton beam. Elastic production occurs as coherent photon radiation from the whole proton without disintegration. In inelastic production, the proton breaks up and induces hadronic activity, which is often called dissociative production. Inelastic production can occur in two regimes. For $Q^2 \lesssim 5$ GeV^2, the proton excites and subsequently decays into a low-multiplicity final state. For higher Q^2, the photon can be considered as being radiated off a parton inside the proton. Suppose both photon productions from beam protons colliding with each other are elastic. In that case, the two-photons interaction in the pp collision is called exclusive or elastic, referred to as EL in this thesis. On the other hand, an event in which one photon production is inelastic is called a single-dissociative (SD) event, and an event in which both photon productions are inelastic is called a double-dissociative (DD) event. The SD and DD events are collectively called dissociative events. Fig. 1.2 shows the conceptual Feynman diagram of EL, SD, and DD events in pp collisions.

In the pp collisions, the $Q^2_{\min}$ is approximated as

$$(Q_{\min})_i^2 = \frac{m_p^2\xi_i^2 + \left[\left(m_{p^{(*)}}\right)_i^2 - m_p^2\right]\xi_i}{1-\xi_i},$$



(a) EL (b) SD (c) DD

Figure 1.2: Feynman diagrams illustrating (a) EL (b) SD (c) DD photon-initiated event in pp collisions.

where $\xi = \omega/E$, m_p is proton mass, and $m_{p^{(*)}}$ is the invariant mass of the final state proton which can be either intact or dissociative. The $(F_{\mathrm{E}})_i$ and $(F_{\mathrm{M}})_i$ depend on whether the i-side proton remains intact or is dissociated. In the elastic case, they are given in terms of proton form factors G_{E} and G_{M},

$$(F_{\mathrm{E}})_i = \frac{4m_p^2 + (G_{\mathrm{E}})_i^2}{4m_p^2 + Q_i^2},$$
$$(F_{\mathrm{M}})_i = (G_{\mathrm{M}})_i^2 .$$

As a characteristic of proton, the G_{E} approximately satisfies

$$(G_{\mathrm{E}})_i^2 = \frac{(G_{\mathrm{M}})_i^2}{\mu_p^2} = \left(1 + \frac{Q_i^2}{0.71\ \mathrm{GeV}^2}\right)^{-4},$$

where the proton magnetic moment is taken to be $\mu_p^2 = 7.78$. In the inelastic case, the F factors are given in terms of proton structure function $F_2(x, Q^2)$,

$$(F_{\mathrm{E}})_i = \int \frac{F_2\left(x_i, Q_i^2\right)}{x_i}\mathrm{d}x_i,$$
$$(F_{\mathrm{M}})_i = \int \frac{F_2\left(x_i, Q_i^2\right)}{x_i^3}\mathrm{d}x_i,$$

where x_i is the fraction of the energy of the parton interacting with the photon over the proton energy and is approximated as

$$x_i = \frac{Q_i^2}{\left(m_{p^*}\right)_i^2},$$

where m_{p^*} is the invariant mass of the dissociated proton system, in the limit that the quark mass is negligible. The maximum photon virtuality Q^2_{max} is approximated as $1/r_p^2$ and typically assumed to be 2-4 GeV^2 [22], where r_p is the proton radius.

The distribution of photon flux n peaks at low Q because it is proportional to $1/Q^2$. This means the protons with low scattering angles are dominant. Especially in the elastic photon productions, the angle is typically as slight as $O(10)$ µrad. The photon fluxes are integrated over Q^2,

$$n_i(\omega_i) = \int_{Q^2_{\mathrm{min}}}^{Q^2_{\mathrm{max}}} \frac{\mathrm{d}n_i}{\mathrm{d}\omega_i \mathrm{d}Q_i^2}\mathrm{d}Q_i^2.$$

In the pp collisions with the center-of-mass energy $\sqrt{s}$, the luminosity of the $\gamma\gamma$ collisions as a function of center-of-mass energy of two-photons,

$$m_{\gamma\gamma} \equiv \sqrt{\hat{s}} \equiv \sqrt{s\xi_1\xi_2} = 2\sqrt{\omega_1\omega_2},$$

is calculated as a convolution of the photon spectra,

$$\frac{\mathrm{d}L_{\gamma\gamma}}{\mathrm{d}m_{\gamma\gamma}} = n_1(\omega_1) \otimes n_2(\omega_2).$$

The cross section of light-by-light scattering in pp collisions is then given by

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \int \frac{\mathrm{d}L_{\gamma\gamma}}{\mathrm{d}m_{\gamma\gamma}} \frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}\Omega} \mathrm{d}m_{\gamma\gamma},$$

where $\mathrm{d}\hat{\sigma}/\mathrm{d}\Omega$ is the cross section of light-by-light scattering as the elementary process [23].

1.1.2 Forward proton detectors

After EL or SD processes at the interaction point (IP), the intact protons remain part of the beam. They can be detected, tagged, and have their kinematics measured by the LHC's forward detectors. In the ATLAS experiment [24], the ATLAS Forward Proton (AFP) detectors [25] are used as such detectors. The AFP detectors are placed about 210 m away from the IP in the LHC beam pipe on both sides of the beam direction. The detail of AFP is described in Sec. 2.3. The AFP can measure proton fractional energy loss ξ for each side and can be used for analyses of $\gamma\gamma$ collisions tagging the events with the protons.

Fig. 1.3 shows the simulated $\mathrm{d}L_{\gamma\gamma}/\mathrm{d}m_{\gamma\gamma}$ as a function of $m_{\gamma\gamma}$ for each of Pb+Pb and pp collisions. The $\gamma\gamma$ luminosity in pp collisions at $\sqrt{s} = 14$ TeV is higher than the one in Pb+Pb collisions at $\sqrt{s} = 5.52$ TeV in the high-mass range. On the other hand, the luminosity in Pb+Pb collisions is enhanced in the lower mass range because it is proportional to Z^4, where Z is the atomic number of the ion. This is because an ion form factor in the Eq. 1.1 is proportional to Z^2. The luminosities for pp collisions with a requirement of proton tagging with forward detectors are also illustrated in the plot.

If the decay products from light-by-light scattering are detected in the ATLAS detector, the ξ can be calculated from their kinematics as well.[1] In the event selection, requiring the consistency of the measured ξ values between the ATLAS and AFP detectors for each side reduces the background events. It improves the sensitivity to the events of interest. The analysis in this thesis makes use of this feature as the primary strategy for event selection. Conventional $\gamma\gamma$ collision analyses specific to EL events without the forward proton detectors usually require less track multiplicity around the interaction vertex or low calorimeter activity instead.

The CMS-TOTEM Precision Proton Spectrometer (PPS) [27], managed by the CMS [28] and TOTEM [29] collaborations, is an AFP equivalent in the LHC. The PPS detectors are also around 200 m away from the IP in the CMS detector. It is proposed to add additional forward proton detectors at around 400 m positions to enhance the proton acceptance, as also discussed for the AFP (see Fig. 1.3) [30].

[1] Since AFP is also operated by the ATLAS , some papers regard AFP as part of the ATLAS detector and refer to the ATLAS detector as the central detector. However, this thesis distinguishes the ATLAS and AFP detectors.

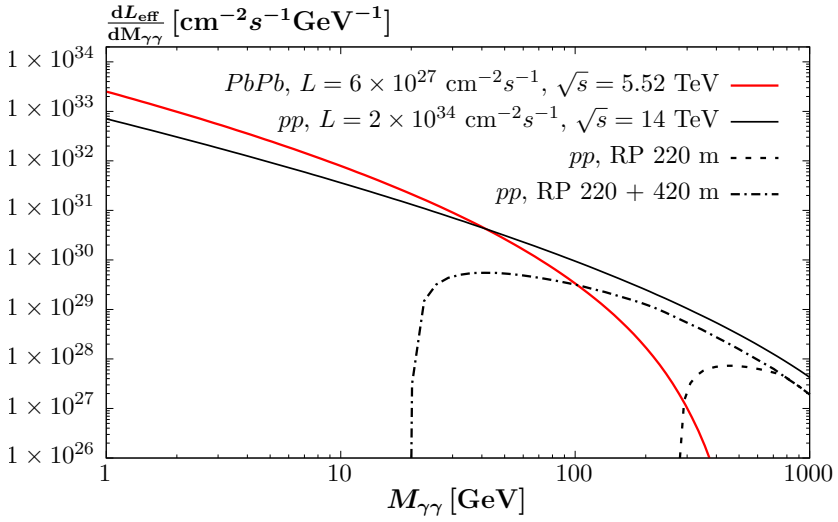


Figure 1.3: Effective $\gamma\gamma$ luminosity as a function of photon-fusion mass in Pb+Pb and pp collisions at the LHC [26]. In the pp case, the $\gamma\gamma$ luminosity is also shown with proton tagging at 220 m and 420 m away from the IP. The "RP" stands for the Roman Pots, containing the AFP detectors described in Sec. 2.3. Though the installation of forward proton detectors at around 420 m has been proposed, no experiment has been realized yet.

1.2 Light-by-light scattering mediated by new physics

The cross section of SM light-by-light scattering in pp collisions is too small to be observed in the current LHC. However, BSM particles can enhance the cross section of light-by-light scattering and make it observable. Although the cross section is still small, the proton tagging with the forward proton detectors enhances the sensitivity and enables the search for high-mass BSM particles mediating light-by-light scattering. Many BSM models predict the existence of new particles that couple with photons. The analysis of light-by-light scattering can probe such new physics. The analysis in this thesis assumes an axion-like particle (ALP), an extension of the axion, as a target signal process.

The QCD field has multiple vacuums mixing with each other and is naturally expected to violate CP symmetry. The CP-violating term in the QCD Lagrangian is

$$\mathcal{L}_\Theta = -\bar{\Theta}\frac{\alpha_s}{8\pi}G^{\mu\nu}\tilde{G}_{\mu\nu},$$
$$\tilde{G}^{\mu\nu} \equiv \frac{1}{2}\epsilon^{\mu\nu\lambda\rho}G_{\lambda\rho},$$

where $\bar{\Theta} \in [-\pi, \pi]$ is the CP-violating parameter, $G_{\mu\nu}$ is the color field strength tensor, and $\tilde{G}^{\mu\nu}$, with $\varepsilon^{0123} = 1$, is its dual [31]. However, experimentally, CP has been measured to be preserved so far with precision 10^{-10} at least [32, 33]. This unnatural fine-tuning is called strong CP problem [34, 35].

A field with a global U(1) symmetry spontaneously broken in the early universe is introduced to solve this problem. The symmetry is called Peccei-Quinn (PQ) symmetry, and the underlying U(1) is denoted $U(1)_{PQ}$. The spontaneous symmetry breaking (SSB) selects a single vacuum from the multiple vacuums. As a general property of global symmetry SSB, a massless Nambu-Goldstone boson (NGB) appears. If the global symmetry is also explicitly broken, most likely due to quantum effects, the NGB acquires mass and is called pseudo-NGB (PNGB). This PNGB is called axion when the symmetry of interest is $U(1)_{PQ}$ [36, 37], and its mass is denoted m_X. The axion is a pseudo-scalar particle with negative parity. The color anomaly

$$\mathcal{L}_{agg} = \Xi\frac{a}{f_a}\frac{\alpha_s}{8\pi}G^{\mu\nu}\tilde{G}_{\mu\nu}$$

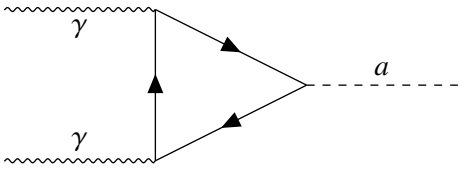


Figure 1.4: Feynman diagram of axion and two-photons coupling. Fermions mediate the interaction.

is induced by the axion field a, where f_a is the axion decay constant regarded as the PQ scale, and Ξ is a model-dependent anomaly coefficient. This term also represents the potential for the axion field whose minimum is at $a = \bar{\Theta} f_a/\Xi$, cancelling the CP-violating term and thus solving the strong CP problem.

The coupling of an axion to two photons is induced by mixing with π^0 [38–40],

$$\mathcal{L}_{a\gamma\gamma} = -\frac{1}{f} a F_{\mu\nu} \tilde{F}^{\mu\nu},$$

where f^{-1} is the coupling constant,[2] $F_{\mu\nu}$ is the electromagnetic field strength tensor, and $\tilde{F}^{\mu\nu}$ is its dual. Charged fermions mediate the coupling as shown in Fig. 1.4. The axion two-photon decay width is calculated as

$$\Gamma = \frac{m_X^3}{4\pi f^2}. \tag{1.2}$$

As a general property of PNGB, the coupling constant is proportional to axion mass. The quantitative relation depends on the PQ scale and the PQ charge of the fermions, which are model-dependent. However, motivated by the fact that axion has not been observed up to now, one may loosen a condition of axion and introduce ALPs whose mass and coupling constant are independent. There is no change in other properties, such as Lagrangian and decay width. The ALP field is also denoted a, and the Lagrangian is

$$\mathcal{L} = \frac{1}{2}\partial^\mu a \partial_\mu a - \frac{1}{2} m_X^2 a^2 - \frac{1}{f} a F^{\mu\nu} \tilde{F}_{\mu\nu}, \tag{1.3}$$

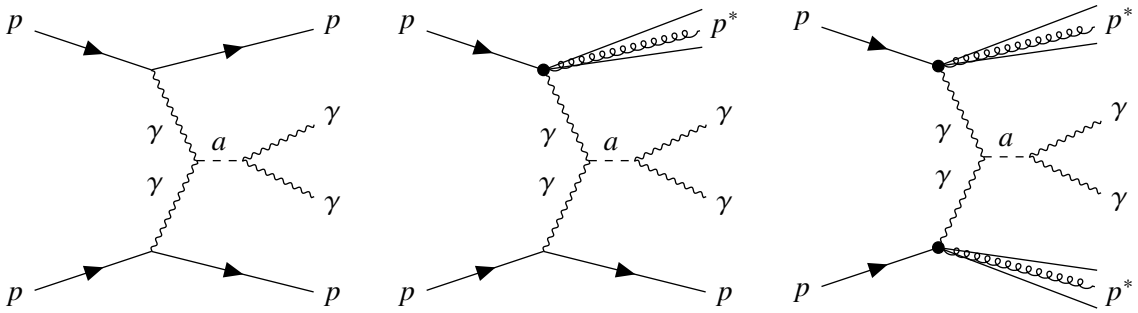
where the first and second terms are kinetic and mass terms, respectively.

Any BSM neutral spin-0 particle couples to fermions only via dimension-5 operators proportional to the fermion mass, while its dominant coupling to gauge bosons is via dimension-5 operators containing derivatives [23]. Thus, in energy regimes exceeding the top quark mass, ALPs are only accessible through their coupling to the gauge bosons and the Higgs boson. In this thesis, only photon coupling is considered. Fig. 1.5 shows the Feynman diagrams for the EL, SD, and DD processes of light-by-light scattering in pp collisions mediated by ALP. In this thesis, these processes are called signal events.

Using the $\gamma\gamma \to \gamma\gamma$ helicity amplitudes

$$\begin{aligned}
\mathcal{M}_{++++} &= -\frac{4}{f^2} \frac{\hat{s}^2}{\hat{s} - m_X^2}, \\
\mathcal{M}_{++--} &= \frac{4}{f^2} \left(\frac{\hat{s}^2}{\hat{s} - m_X^2} + \frac{\hat{t}^2}{\hat{t} - m_X^2} + \frac{\hat{u}^2}{\hat{u} - m_X^2} \right), \\
\mathcal{M}_{+++-} &= 0, \\
\mathcal{M}_{+-+-}(\hat{s}, \hat{t}, \hat{u}) &= \mathcal{M}_{++++}(\hat{u}, \hat{t}, \hat{s}), \\
\mathcal{M}_{+--+}(\hat{s}, \hat{t}, \hat{u}) &= \mathcal{M}_{++++}(\hat{t}, \hat{s}, \hat{u}),
\end{aligned}$$

[2] Alternatively, $g \equiv f^{-1}$ is often used as the coupling constant.



(a) EL (b) SD (c) DD

Figure 1.5: Feynman diagrams illustrating ALP a mediated light-by-light scattering. This process can take place in three distinct topologies: (a) exclusive $pp \to p(\gamma\gamma \to \gamma\gamma)p$, (b) single dissociative $pp \to p^*(\gamma\gamma \to \gamma\gamma)p$, and (c) double dissociative $pp \to p^*(\gamma\gamma \to \gamma\gamma)p^*$.

where $\hat{s}$, $\hat{t}$, and $\hat{u}$ are the Mandelstam variables of the diphoton system, the elementary process unpolarized cross section is given by

$$\frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}\Omega} = \frac{1}{128\pi^2\hat{s}}\left(|\mathcal{M}_{++++}|^2 + |\mathcal{M}_{+-+-}|^2 + |\mathcal{M}_{+--+}|^2 + |\mathcal{M}_{++--}|^2\right).$$

The calculation is discussed in Ref. [23] and is valid up to $m_{\gamma\gamma} \sim 4\pi f$. In the $m_{\gamma\gamma}$ mass range where ALP is produced resonantly, the elementary process cross section is proportional to the coupling constant squared,

$$\hat{\sigma} \propto f^{-2}\mathcal{B}_{a\to\gamma\gamma}, \tag{1.4}$$

where $\mathcal{B}_{a\to\gamma\gamma}$ is the branching ratio of ALP decaying to two photons assumed to be 100% in this analysis. A more detailed modeling is in Ref. [22].

The ALP model described above is used for signal Monte Carlo (MC) sample generation and is assumed to interpret the results in this thesis. However, there are other BSM models that can be examined using light-by-light scattering, such as warped extra dimensions [41], Kaluza-Klein gravitons [42], and composite Higgs [43]. There is also a model-independent Effective Field Theory (EFT) to represent BSM light-by-light scattering by the effective Lagrangian using dimension-8 operators,

$$\mathcal{L}_{4\gamma} = \zeta_1 F_{\mu\nu}F^{\mu\nu}F_{\rho\sigma}F^{\rho\sigma} + \zeta_2 F_{\mu\nu}F^{\nu\rho}F_{\rho\lambda}F^{\lambda\mu}, \tag{1.5}$$

where the coefficients $\zeta_{1,2}$ are the parameters and predicted by any BSM model [22]. The $\gamma\gamma \to \gamma\gamma$ cross section is calculated as

$$\frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}\Omega} = \frac{1}{16\pi^2\hat{s}}\left(\hat{s}^2 + \hat{t}^2 + \hat{s}\hat{t}\right)^2\left(48\zeta_1^2 + 40\zeta_1\zeta_2 + 11\zeta_2^2\right).$$

Although searches dedicated to resonant BSM signals such as the ones introduced in Sec. 1.3.1 can also study several different BSM models inclusively, the EFT model can take non-resonant BSM signals into account as well.

1.3 Previous researches

ALPs have been strongly constrained by multiple observations [44]. Many such results are model-independent and thus can be represented in the coupling constant or cross section versus mass plane of ALP. The parameter space is loosely constrained at high masses above the GeV scale, where searches are

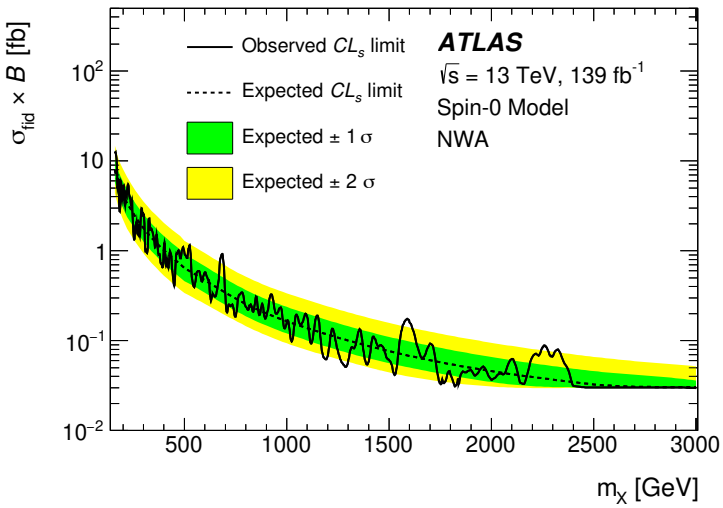


Figure 1.6: Exclusion limits of the fiducial cross section of resonant decay spin-0 BSM particles into two-photons in NWA as a function of their mass [52]. The analysis uses 139 fb^{-1} of pp collision data at $\sqrt{s}$ = 13 TeV. The region above the solid black line is excluded.

collider-based. Some studies explicitly searching for ALPs have been performed, but their mass ranges are lower than the research target in this thesis [45].

This section introduces three kinds of previous research whose methodologies are inherited in the analysis of this thesis. Each sort of analysis sets exclusion limits on a parameter space of BSM.

1.3.1 High-mass inclusive diphoton resonance searches

Some resonant BSM particles can decay into two photons. Requiring high-energy photons in the final state reduces background events, and the excellent energy resolutions of the electromagnetic (EM) calorimeters of the ATLAS and CMS detectors enable the diphoton mass ($m_{\gamma\gamma}$) distribution to have a narrow peak corresponding to the resonance.[3] The Higgs discovery made use of these features in the $H \to \gamma\gamma$ channel [46, 47]. After the discovery, high-mass diphoton resonance searches have been performed in the ATLAS [48–52] and CMS [53–57] experiments in $m_{\gamma\gamma}$ ranges beyond $O(100)$ GeV. Despite the excellent detector resolution, the peak width is dominated by the detector resolution rather than the natural width in some models. Especially the natural width is assumed to be 0 if it is smaller than a few GeV, which is called narrow-width approximation (NWA).

To use the narrow feature of the resonance peak effectively, the $m_{\gamma\gamma}$ distribution is modelled by a function for each hypothetical mass of the BSM particle of interest. On the other hand, the background $m_{\gamma\gamma}$ distribution has a smoothly-falling shape, which is also parametrized with another function. Finally, the sum of these signal and background functions, the signal-plus-background (s+b) function, is fitted for each BSM particle mass. This enhances the sensitivity to signal events compared to binning the distribution into a histogram where the binning optimization is not trivial.

The latest ATLAS result uses 139 fb^{-1} of pp collision data at $\sqrt{s}$ = 13 TeV [52]. The exclusion limits on the cross section of spin-0 particle resonance in NWA are shown in Fig. 1.6. The latest CMS journal publication uses of 35.9 fb^{-1} of pp collision data at $\sqrt{s}$ = 13 TeV [57]. Since these analyses search for BSM particles inclusively, the results can be interpreted as resonant ALPs with narrow widths.

[3] The ATLAS EM calorimeter also has an excellent angular resolution.

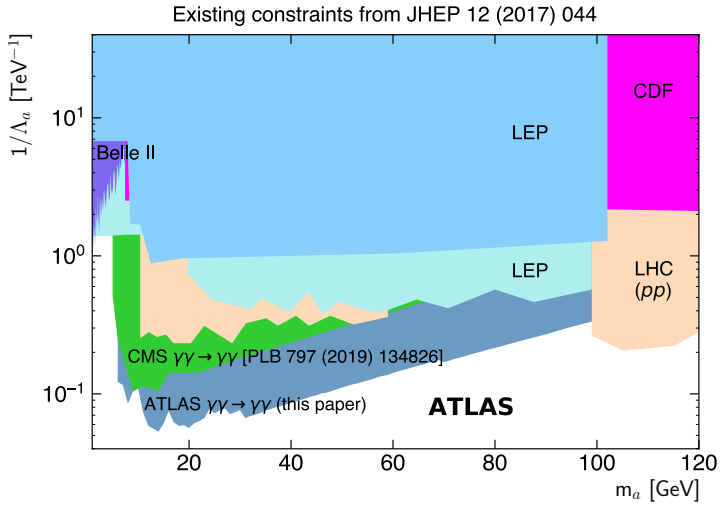


Figure 1.7: Statistically excluded regions in a plane of ALP coupling constant to two-photons and ALP mass [60]. The coupling constant is defined as $\Lambda_a^{-1} \equiv 4f^{-1}$ [62]. The ATLAS light-by-light scattering analysis result using 2.2 nb^{-1} Pb+Pb of collision data at $\sqrt{s} = 5.02$ TeV (indicated by "this paper") is compared with the other results derived from Refs. [44, 58, 63–65].

1.3.2 Light-by-light scattering observation in Pb+Pb collisions

After the reports of the strong evidence of light-by-light scattering [11, 58], the ATLAS achieved an initial direct observation of high-energy light-by-light scattering [59] in 2019 by performing an exclusive diphoton analysis in 1.73 nb^{-1} of Pb+Pb collision data at $\sqrt{s} = 5.02$ TeV. The interaction does not break the beam lead ions. Therefore, no QCD interaction occurs at the signal vertex. In the final state, two photons are exclusive in that no other objects are associated with the signal vertex. Thus, the final state diphoton is back-to-back in a plane orthogonal to the beam direction, and the event selection uses this characteristic.

The ATLAS analysis was updated using 2.2 nb^{-1} Pb+Pb collision data, adding the result of the ALP search [60, 61]. Fig. 1.7 shows the exclusion limits of ALP coupling constant to two-photons derived from Refs. [58, 60]. The limits are up to 100 GeV and do not overlap with the search region of the analysis in this thesis.

1.3.3 Analyses using the forward proton detectors

The ATLAS observed SM $\gamma\gamma \to \ell^+\ell^-$ process in 14.6 fb^{-1} pp collision data using the AFP for the first time [66]. Fig. 1.8 shows the diagram of the signal event in the analysis. This result serves as proof-of-principle to demonstrate that the AFP system and software can be used in conjunction with the ATLAS detector information to deliver physics results. Since the initial beam proton momenta are known, the proton energy loss fractions ξ_i can be calculated from the measured dilepton kinematics simultaneously. The events are required to have at least one proton with ξ measured by the AFPs close to the ξ calculated from the dilepton system. This is called diphoton-proton *matching*. The elementary $\gamma\gamma \to \ell^+\ell^-$ processes have already been measured by the ATLAS [67–69] and CMS [70, 71] collaborations without using the forward proton detectors in the past. Thus, the 14.6 fb^{-1} data analysis using the AFPs has roles not only as a verification of the SM but also as an anomaly search. In addition, the signal events are used for the calibration and performance evaluation of the AFPs. The CMS and TOTEM s also reported the observation of the same kind of events with proton tagging on one side of the outgoing protons [72].

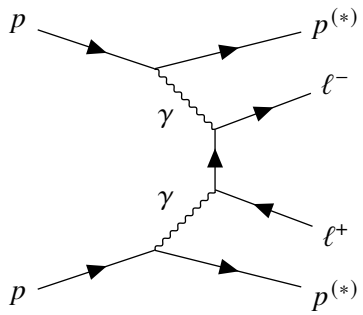


Figure 1.8: Feynman diagram of SM $\gamma\gamma \to \ell^+\ell^-$ process in pp collisions. Lepton ℓ is electron or muon. The potential proton dissociation is denoted by $^{(*)}$.

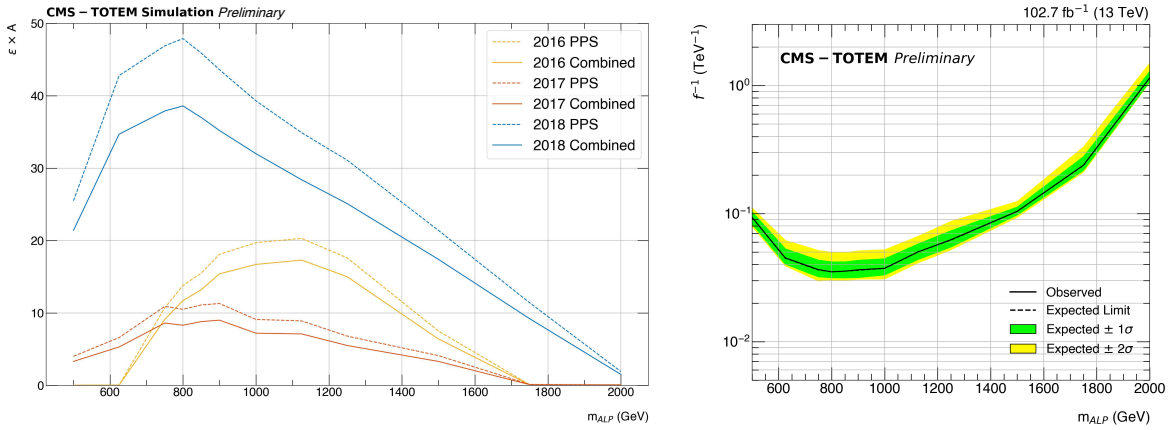
As another process searched for using the forward proton detectors, the CMS and TOTEM published the first result of a search for light-by-light scattering mediated by BSM particles in 9.4 fb^{-1} of pp collisions at $\sqrt{s}$ = 13 TeV using the PPS detectors [73]. The signal event of the analysis is the same as the one in the research in this thesis. However, the CMS and TOTEM analysis targets only EL processes requiring the kinematical matching of both sides of the outgoing intact protons. In this case, only electromagnetic coupling needs to be taken into consideration. In other words, this search is independent of the strong and weak force couplings that most other LHC searches rely on, enabling clean events analysis. Such a double-matching results in extreme background events reduction. Then, a counting experiment is performed to set exclusion limits on the parameter spaces of the BSM models. MC samples are used for the background events contribution estimation.

The CMS-TOTEM PPS analysis results were updated using the 102.7 fb^{-1} data [74]. Even this updated analysis observed only one event after all the event selection due to the strict double-matching requirement. Both the previous and the later analyses set exclusion limits on the ζ_1-ζ_2 plane described in Sec. 1.2. In addition, this analysis set the exclusion limits of the ALP coupling constant to two photons as a function of ALP mass. In contrast to the resonance searches with $m_{\gamma\gamma}$ distribution shape fitting described in Sec. 1.3.1, the number of events after the event selection are counted inclusively in terms of ALP mass. The signal efficiency, the rate of signal events passing the event selection, is estimated for each hypothetical ALP mass using signal MC samples. The exclusion limits are derived from the number of observed events and the signal efficiency for each ALP mass. The signal efficiency and the ALP exclusion limits are shown in Fig. 1.9. The single peak structure of the efficiency reflects the PPS geometrical acceptance and is propagated to the mass dependency of the limits.

1.4 Aim and overview of this work

This research aims to establish methods for searches for BSM light-by-light scattering and other similar research, inspired by the observation of light-by-light scattering in 2019 [59] first. Then, ALPs with mass ranging from 150 GeV to 1600 GeV mediating light-by-light scattering in the LHC pp collisions are chosen as the target in the analysis as the first example. This is the first time for the ATLAS to study light-by-light scattering in the pp collisions. The appearance of the new detectors, AFPs, was the decisive factor in starting this research. The analysis inherits the methodologies in Refs. [52, 60, 66] introduced in Sec 1.3.

The pp collision data taken in 2017 is used in this analysis. The ATLAS detector and the AFPs are used to detect the final state of two photons and protons of the signal events, respectively. This analysis is the second one using the AFP after the dilepton analysis [66], and the same datasets of 14.6 fb^{-1} are used.



(a) Signal efficiency (b) Exclusion limits

Figure 1.9: Results of a search for light-by-light scattering mediated by BSM particles using the CMS-TOTEM PPS detectors [74]. (a) Signal efficiency as a function of ALP mass estimated using the signal MC samples for each data-taking year. The dashed lines only consider the detector efficiency, and the solid lines include the other object and event selection efficiency. (b) The exclusion limits of ALP coupling constants to two photons as a function of ALP mass. The region above the solid black line is excluded.

The experimental setup is detailed in Sec. 2. The analysis in this thesis uses two kinds of samples, the data samples and the signal MC samples. The former was taken in the actual experiment in 2017 when AFPs became operational stably. Unfortunately, the 2018 data of the AFPs failed to be recorded correctly. The signal MC samples are used for modeling signal event distribution, evaluation of signal efficiency, and some tests for systematic uncertainty evaluation. The details of the samples are described in Sec. 3.

Events with a high-energy back-to-back diphoton system are selected as done in the Pb+Pb light-by-light scattering analysis [60]. The proton energy loss fractions ξ_i are calculated from the diphoton kinematics for each side i, assuming the initial photon virtuality is small enough. The formula is derived in Sec. 4. To clarify that the ξ_i is calculated from the diphoton, this is denoted $\xi^i_{\gamma\gamma}$. The ξ_i can also be measured by the AFP if the outgoing proton i remains intact and has momentum within the geometrical acceptance of the AFP. The ξ_i measured by the AFP is denoted ξ^i_{AFP}. Only the events with at least one outgoing proton within a ξ^i_{AFP} range in which the proton reconstruction efficiency and its uncertainty are measured are selected. The events are also required to have at least one ξ_i within the corresponding $\xi^i_{\gamma\gamma}$ range. In addition, the matching between $\xi^i_{\gamma\gamma}$ and ξ^i_{AFP} is required as done for each i as done in Refs. [66, 74]. The concepts of the matching are illustrated in Fig. 1.10. Finally, the events with matched pair of $\xi^i_{\gamma\gamma}$ and ξ^i_{AFP} for at least one i are selected. This is called OR selection in contrast to the selection used in the CMS-TOTEM PPS analysis [74] requiring the matchings for both protons i, which is called AND selection in this thesis (see Sec. 1.3.3). The event selection are detailed in Sec. 4.

The statistical model inherits one used in the high-mass diphoton resonance search [52]. The s+b fitting with the model enhances the sensitivity to resonant ALPs. The signal distribution is modelled in Sec. 5. The signal efficiency is also modelled as a part of signal modeling in this section. The ALP mass dependency of the efficiency reflects the AFP acceptance and differs from the CMS-TOTEM PPS case because of the difference between OR matching and AND matching. The unit of signal yield in the final results, signal strength, is also defined here. The background distribution is modelled in Sec. 6. For the modelling, a background sample is prepared in a fully data-driven way. This means no background MC sample is used, unlike usual analyses. Such data-driven background sample creation method inherits the dilepton analysis [66], but some points were improved in this analysis. The uncertainties of the signal and background models are considered in each section. For the background modeling uncertainty, the

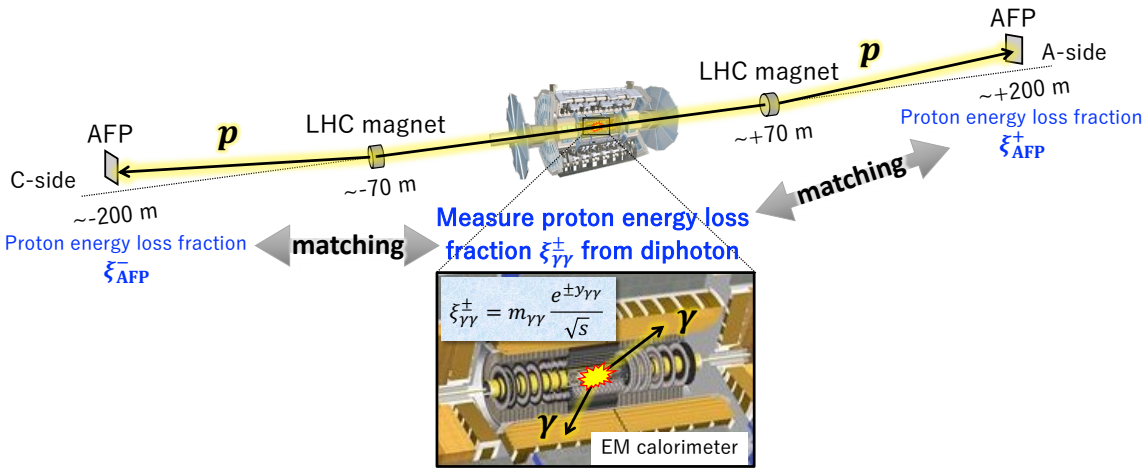


Figure 1.10: Conceptual illustration of matching between $\xi^{i}_{\gamma\gamma}$ and ξ^{i}_{AFP}. In the ATLAS detector illustrated at the center [75], the final state two-photons are detected by the EM calorimeter, and $\xi^{i}_{\gamma\gamma}$ is calculated. The outgoing proton is curved by the LHC magnets and enters the AFPs to have ξ^{i}_{AFP} measured.

data-driven background distributions are smoothed, and *spurious signals* are evaluated by s+b fitting to artificial distributions derived from the smoothing.

The systematic uncertainties originate from the preciseness of object detection, simulation, and reconstruction, for example. Such effects are summarized in Sec. 7, including the signal and background modeling uncertainties. The complete statistical modeling is discussed in Sec. 8. The systematic uncertainties are incorporated into the model as the constraints to the likelihood. Two kinds of statistical tests are performed using popular test statistics. One is for testing BSM, and the other is for setting the exclusion limits for the ALP parameter space. The results are shown in Sec. 9. Discussions about the interpretation of the results, comparison to the other analyses, expected improvement in the future analysis, and how the methods established in this research can be used in future analyses follow in the same section.

2 Experimental Setup

The experimental setup consists of the LHC, ATLAS detector, and AFPs. This section introduces them.

2.1 LHC

The LHC [10] is an energy-frontier hadron-hadron collider operated by the European Organization for Nuclear Research (CERN) that aims to search for new physics at the TeV scale directly. It is placed across the border of Switzerland and France in a tunnel with 27 km circumference and a depth of down to about 100 m from the ground level. It is designed to collide proton beams at $\sqrt{s} = 14$ TeV and an instantaneous luminosity of 10^{34} cm^{-2}s^{-1}. The LHC also accelerates heavy ions, but only pp collisions are discussed in this thesis. The LHC consists of straight sections and arc sections. The straight sections are used as the insertion regions (IRs) where the experimental apparatus are inserted into the LHC tunnel. The ATLAS detector is at an IR. To have the protons with $E = 7$ TeV remain in the LHC, the dipole magnets with 8.3 T are needed in the arc sections. The current LHC magnets cannot have a higher magnetic field because the NbTi superconducting wire induces quenching[1]. The beam proton loses its energy through synchrotron radiation, having it complemented by radiofrequency (RF) acceleration. The energy loss is proportional to γ^4, where γ is the beam Lorentz factor. For example, a beam proton with $E = 7$ TeV loses 7 keV per revolution (turn). During the operation from 2010 to 2012 (Run 1), the LHC performed pp collisions at $\sqrt{s} = 7$ TeV and 8 TeV. During the operation from 2015 to 2018 (Run 2), the LHC worked at $\sqrt{s} = 13$ TeV. The setups in 2017 when the data used in this research were taken are discussed in this section.

The LHC proton beams are transported in a bunch structure, and each beam bunch consists of approximately 10^{11} protons. The beam bunches orbit the two parallel beam pipes in opposite directions and cross at the IP of the ATLAS detector at a frequency of 40 MHz. The beams circling the LHC clockwise and counterclockwise are called beam 1 and beam 2, respectively. It takes about 90 µs per turn for a bunch, and its inverse (revolution frequency) is about 11 kHz. Since the pp collision cross section is about 0.1 b, the pp collision frequency is $O(1)$ GHz. Thus, the averaged number of pp collisions per bunch crossing is $O(1)$ GHz/40 MHz $\sim O(10)$. All these pp collisions, except for the events of interest, are called pileups. A bunch crossing or its timing referred from the detectors is called BC. Detected pileups are subdivided into the ones originating from the same BC (in-time pileups) and different BCs (out-of-time pileups). Although a BC is a set of pp collision events, it is often referred to as an event, especially in physics analyses.

The proton beams are accelerated in a series of other accelerators before being injected into the LHC. Fig. 2.1 illustrates the CERN accelerator complex. The origins of the protons are hydrogen anions. The initial injector, LINAC 2, accelerates them to 50 MeV. The anions are injected into the Proton Synchrotron Booster (PSB), and the electrons are removed from the anions using the carbon foils to obtain the proton beam. The protons are accelerated to 1.4 GeV in the PSB and injected into the Proton Synchrotron (PS) six times until the PS is filled. In the PS, the protons in the six bunches are accelerated to 26 GeV. In this process, 72 of the 40 MHz beam bunches are formed by gradually inducing higher RFs. This method,

[1] The critical surface limits the superconducting phase in the space of temperature, current density, and magnetic field. A too-strong magnetic field causes heating and hence violates the superconductivity. This is called quenching.

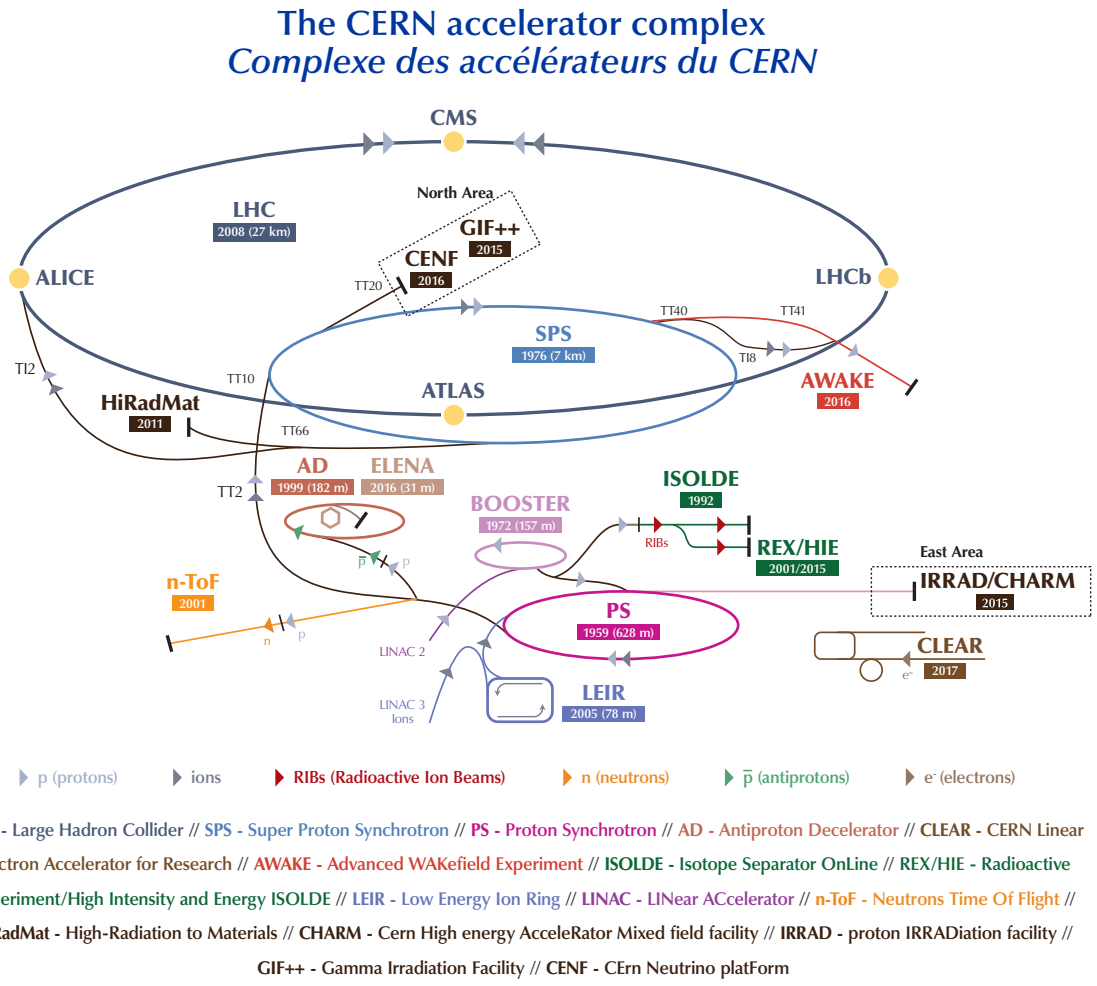


Figure 2.1: The complex chain of particle accelerators in CERN [77]. The dark blue line illustrates the LHC. The smaller machines are used in the chain to help boost the particles to their final energy and provide beams to a whole set of smaller experiments, which also aim to uncover the mysteries of the universe.

called multiple splitting, is illustrated in Ref. [76]. The bunches are then injected into the Super Proton Synchrotron (SPS) in 72×4 times at most until the SPS is filled. The protons are accelerated to 450 GeV with the SPS. Finally, the bunches are injected into the LHC until the LHC is filled. Then, the protons are accelerated to 6.5 TeV taking about 20 minutes for the Run 2 case.

The injections and extractions of the beam into and from the accelerators are controlled by the septum and kicker magnets. A septum magnet creates spaces with a strong magnetic field and without a magnetic field in a small region, enabling the injection or extraction while the orbiting bunches are unaffected. A kicker magnet is used to put the protons from the magnetic space into the orbit or extract the protons from the orbit to the magnetic space. The kicker magnets are magnetized only during the injection and extraction. Their magnetization and demagnetization take a finite time, depending on the beam energy. Thus, there exist empty bunches for that operation in the 40 MHz bunches.

The quadrupole magnets have a so-called FODO cell structure in conjunction with the dipole magnets to repeat the focusing and defocusing of the beam in the axes orthogonal to the beam direction. This is called betatron oscillation, and its period is tuned not to be an integral multiple of the period of the turn in order to stabilize the beam. Position $x(s)$ in one of the transverse axes and its derivative $x'(s)$ are defined for each position s in the accelerator. The geometric emittance $\epsilon = \sigma_x^2(s)/\beta_x(s)$ is the degree

of the deviation of the beam in the (x, x')-plane, where $\sigma_x^2(s)$ is the x-direction deviation of the beam, and $\beta_x(s)$ is determined by the accelerator setup and is called β-function. The normalized emittance of the beam is defined as $\epsilon_n = \beta\gamma\epsilon$, where β is velocity. This is a conserved quantity with respect to s and is 3.5 µm in the LHC. Thus, the beam width has a negative correlation with the beam energy. After the beam energy reaches 6.5 TeV, the final focusing system dynamically adjusts the β_x of the series of quadrupole magnets closest to the IP to squeeze the beam and maximize the luminosity at the IP. After the squeezing, $\beta^* \equiv \beta_x(\text{IP})$ is 0.3-0.4 m in the standard runs in 2017. The instantaneous luminosity $\mathcal{L}$ at the IP is calculated as

$$\mathcal{L} = \frac{n_b N_{b1} N_{b2} f_{rev}}{4\pi\sigma_x\sigma_y} F \approx \frac{n_b N_{b1} N_{b2} f_{rev}\gamma}{4\pi\epsilon_n\beta^*} F,$$
$$F = \left[1 + \left(\frac{\theta_c\sigma_z}{2d}\right)^2\right]^{-\frac{1}{2}},$$
$$d \approx \sqrt{\frac{\epsilon_n\beta^*}{\gamma}},$$

where $n_b \sim 2500$ is the number of bunches per revolution, N_{bi} is the number of particles per bunch in beam i, f_{rev} is the revolution frequency, $\sigma_{x,y}$ and $d \approx \sigma_x \approx \sigma_y$ are the transverse beam size at the IP,[2] F is the geometric luminosity reduction factor due to the crossing angle at the IP, θ_c is the beam crossing angle at the IP, and σ_z is the longitudinal length of the beam bunches at the IP [78]. θ_c is typically 100-150 µrad in the standard runs in 2017.

2.2 ATLAS detector

The ATLAS detector [24] at the LHC covers nearly the entire solid angle around the collision point. It consists of an inner tracking detector surrounded by a thin superconducting solenoid, EM and hadron calorimeters, and a muon spectrometer incorporating three large superconducting air-core toroidal magnets. Fig. 2.2 shows the image of the whole ATLAS detector. To specify a position in the detectors, ATLAS uses a right-handed coordinate system with its origin at the nominal IP in the center of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the center of the LHC ring, and the y-axis points upwards. The region with $z > 0$ is called A-side, and one with $z < 0$ is called C-side. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The r-direction or, more generally, a direction in the x-y plane is often referred to as the transverse direction, while the z-direction is often referred to as the longitudinal direction.The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$. The angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

Since the pp collision vertices are approximately at the origin, the angles of the generated particles or their systems are defined in the same manner. The rapidity y is defined as

$$y \equiv \operatorname{arctanh}\beta_z = \frac{1}{2}\ln\frac{E + p_z}{E - p_z},$$

where the velocity is $\boldsymbol{\beta} = (\beta_x, \beta_y, \beta_z)$, E is the energy, and the momentum is $\boldsymbol{p} = (p_x, p_y, p_z)$. The pseudorapidity is then defined as the massless limit of the rapidity,

$$\eta \equiv \frac{1}{2}\ln\frac{E + E\cos\theta}{E - E\cos\theta} = -\ln\tan\frac{\theta}{2}.$$

[2] The beam envelope is circular in the transverse plane at the IP.

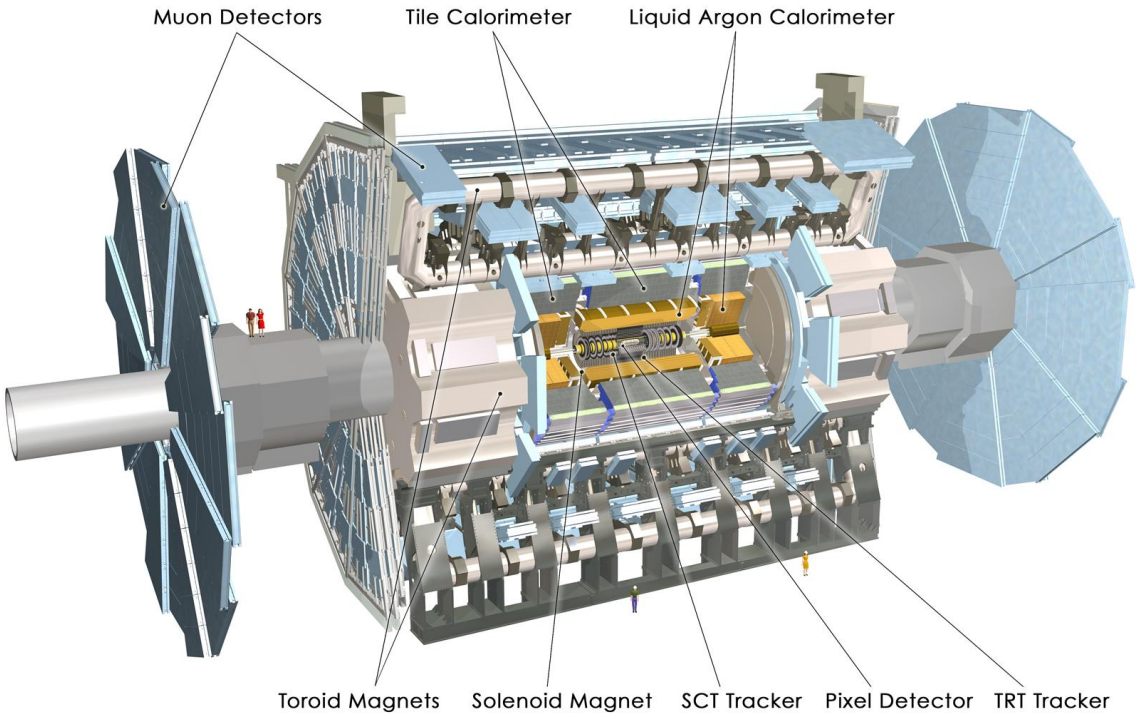


Figure 2.2: Computer generated image of the whole ATLAS detector [75]. People are also illustrated for comparison.

In addition, the transverse energy $E_{\mathrm{T}} \equiv E \sin\theta$ and transverse momentum $p_{\mathrm{T}} \equiv \sqrt{p_x^2 + p_y^2}$ are often used in the analysis.

The subdetectors are introduced in Sec. 2.2.1. Since photon reconstruction mainly uses the information from the EM calorimeters, Sec. 2.2.2 details them. Sec. 2.2.3 describes how the data from the subdetectors are collected and processed.

2.2.1 Subdetectors

The inner detectors cover $|\eta| < 2.5$ and track charged particles in the 2 T axial magnetic field. The innermost layer has the high granularity silicon pixel detector, among which the insertable B-layer (IBL) [79, 80] was installed before Run 2. The silicon pixel detector provides four measurements per track, and the IBL usually has the first hit. The next layer is the silicon microstrip tracker (SCT), where eight measurements are usually provided per track. The outer layer transition radiation tracker (TRT) based on straw drift tubes covers $|\eta| < 2.0$ and provides information for tracking and electron identification.

Outside the inner detector is the calorimeter system covering $|\eta| < 4.9$. The EM calorimeter consists of high-granularity lead/liquid-argon (LAr) calorimeters in the barrel and end cap, covering $|\eta| < 3.2$. The innermost LAr presampler covering $|\eta| < 1.8$ corrects the energy. The downstream hadron calorimeter consists of - the tile calorimeter with steel and scintillator covering $|\eta| < 1.7$, - the hadron endcap calorimeter with copper and LAr covering $1.7 < |\eta| < 3.2$, - the forward calorimeter with copper, tungsten and LAr covering $3.2 < |\eta| < 4.9$, and measures hadronic energy.

The outermost layer is the muon spectrometer (MS), which measures the muon momentum using the magnetic field generated by the superconducting air-core toroidal magnets. The MS consists of the monitored drift tubes for $|\eta| < 2.7$ and the cathode-strip chambers for the forward region for precision

measurement, the resistive-plate chambers in the barrel and the thin-gap chambers in the endcap and for triggering (see Sec. 2.2.3) in $|\eta| < 2.7$.

The Luminosity Cherenkov Integrating Detector-2 (LUCID-2) [81] is mainly placed ± 17 m away from the IP and 13 cm away from the beam pipe, covering $5.6 < |\eta| < 5.9$, and measures the beam luminosity using the pileup events BC by BC. Quartz is used as the Cherenkov radiation medium to detect the radiated light with the photomultiplier tubes (PMTs). ATLAS beam condition monitor (BCM) complements LUCID-2. The usual subdetectors, such as the EM and tile calorimeters, are also used for BC-integrated measurements based on quantities proportional to instantaneous luminosity. The integrated and instantaneous luminosities are measured by these detectors in conjunction with some other detectors and methods [82]. The luminosity-sensitive detectors are calibrated in the dedicated low-luminosity runs using van der Meer (vdM) method [83, 84]. The calibration is then transported to the physics data-taking regime at high luminosity. These procedures map detector responses to luminosity.

2.2.2 EM calorimeter

High-energy photons of interest lose most of their energy in the EM calorimeters. Electrons and photons are the primary targets for EM calorimeters and are called EM objects. The EM calorimeters are the sampling calorimeters which use LAr as the detection layer. Some hadron calorimeters on the endcap parts also use the LAr. The set of calorimeters using LAr as the detection layer is called LAr calorimeter. Fig. 2.3 shows the overview of the LAr calorimeter. The LAr calorimeter is subdivided into four parts, EM barrel (EMB), EM endcap (EMEC), hadronic endcap (HEC), and forward calorimeter (FCal). The EMB covers $|\eta| < 1.5$, and the EMEC covers $1.4 < |\eta| < 3.2$. The EMEC consists of the outer and inner wheels covering $1.4 < |\eta| < 2.5$ and $2.5 < |\eta| < 3.2$, respectively. The HEC and FCal cover higher $|\eta|$ regions. The photon angle used in this research is limited to $|\eta| < 2.4$.

The EMB has layers of lead, copper, LAr, etc., as shown in Fig 2.4. The layers have the *accordion* structure to geometrical uniformity keeping ϕ-symmetry. When a high-energy particle enters the detector, the EM shower is generated in the lead absorption layer, and the shower ionizes the argon in the LAr layer (LAr gap). The movement of the popped electrons is detected as an electric signal. The movement of the electrons is induced by a voltage of about 2 kV applied between a grounded stainless steel layer and a copper layer and is detected by a copper readout electrode placed between two copper anodes insulated by kapton. Since the Ar ions have negligible mobility compared to electrons, the voltage generated by their movement is negligible. The LAr gap is as thin as about 2 mm, so the current obtained by a single electron can be approximated as a rectangular pulse having the electron drift time as its length. The pulse height of the rectangular pulse does not depend on the electrons, and the drift time depends linearly on the uniform position of the ionization. Thus, the rectangular pulses are summed into a triangular pulse with a height proportional to the energy.

The EMB is celled in the r, η, and ϕ directions, as shown in Fig. 2.5. Each set of cells classified by the division in the r direction is called a layer. Typically, the 0th layer (presampler), first layer (front layer), second layer (middle layer), and fourth layer (back layer) are arranged in ascending order of r. The strip cells with high granularity in η in the front layer are designed to distinguish a usual photon and π^0 decaying into two photons. The second layer is thick, and EM objects drop most of their energy here. A larger unit of readout cell, Trigger Tower (TT), is used for the first-level (L1) triggering introduced in Sec. 2.2.3.

The EMEC is also ϕ-symmetric and has the accordion structure in a different direction from the EMB case. The mechanism to acquire the signals is the same as EMB, and the triangular pulses are generated. Similar cell division is used as the EMB, but there are many areas where all layers from the 0th to the third are not entirely arranged.

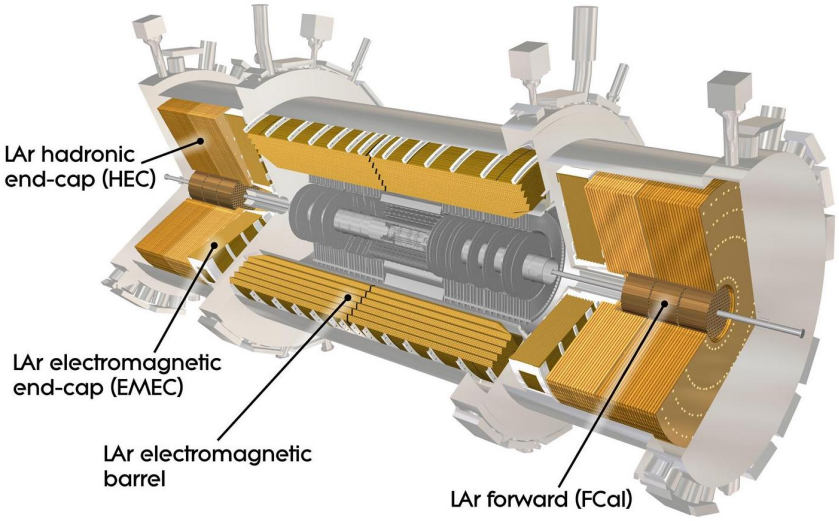


Figure 2.3: Computer generated image of the ATLAS LAr calorimeter [85].

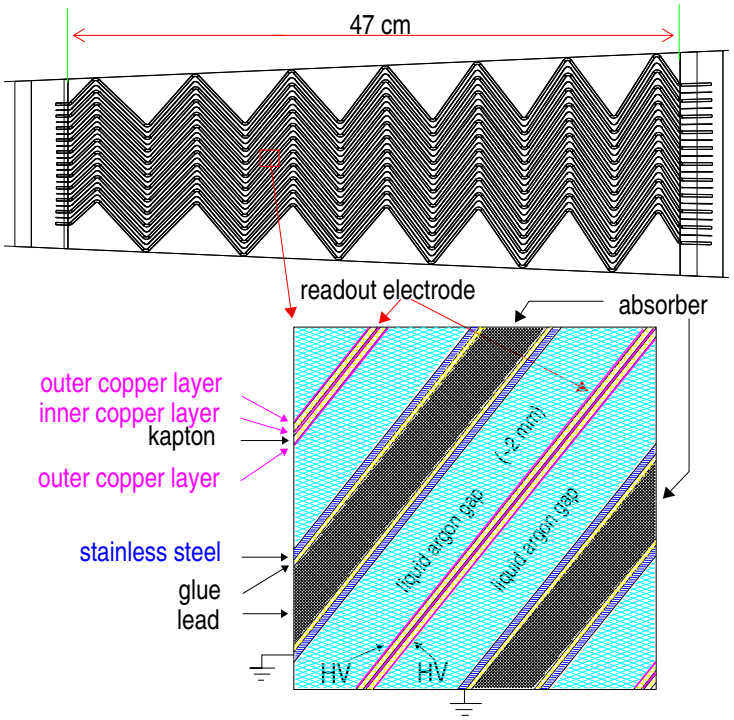


Figure 2.4: Structure of the LAr EMB shown in the ϕ-η plane [86].

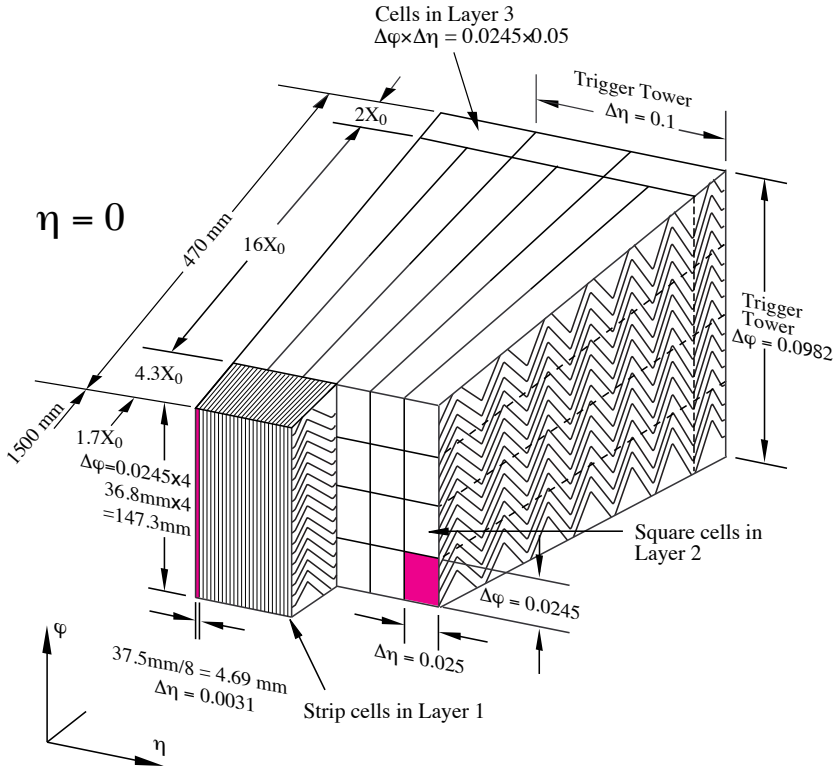


Figure 2.5: Cell division in the LAr EMB [24]. The front, middle, and back layers are illustrated. The radiation length is denoted X_0.

The functions of the LAr calorimeter are divided into signal acquisition for precise energy calculation and triggering. The Front-End boards (FEBs) are in the Front-End crates (FECs) mounted on the cryostat and read out and process the triangular pulses delivered through the feedthroughs. In the FEBs, a triangular pulse from a readout cell is amplified by the preamplifier and shaped into a bipolar pulse by the CR-$(RC)^2$ circuit.[3] The bipolar pulse is sampled at 40 MHz and temporarily stored in the Switched Capacitor Array (SCA). On the other hand, the triangular pulses are summed and shaped to create a bipolar TT signal in the Tower Builder Board (TBB), and the TT signal is sent to the L1 trigger system. The trigger system analyses the signals from all the cells in conjunction with the signals from other subdetectors and fires a trigger as an L1-accept (L1A) signal when the BC is judged to be interesting. Only the signals stored in the SCA from the BCs where L1A is issued are digitized by the ADC in the FEBs and sent to the backend Read Out Driver (ROD) through 80-100 m long fibers. The energy deposit in a cell is then calculated with the *optimal filtering* method[4] using the linearity of the pulse height with respect to the energy.

The cell-level energy computation of the LAr calorimeter is calibrated using test pulses output from the calibration boards in the FECs. The calibration pulses have shapes similar to the triangular pulses for each cell. They are sent to the LAr detector to collect the responses in a similar environment as the actual data-taking for physics analyses. The input pulse heights are associated with particle energy, and the correspondence is calibrated using the test beams [87, 88]. More precise calibration and object-level calibration are performed using pp collisions data. An example is to use $Z \rightarrow ee$ events whose final state kinematics are known.

[3] The typical length of the triangular pulses is 400 ns, which is much larger than the BC interval, 25 ns. Thus, if this pulse shape is used, the energy calculation is affected by the pileups from the previous BCs (out-of-time pileups). However, the bipolar pulse shape cancels this effect on average with the negative lobe with the same integration as the positive lobe.

[4] The optimal filtering algorism is designed to minimize the noise effect. This meets the demand for precise energy calculations.

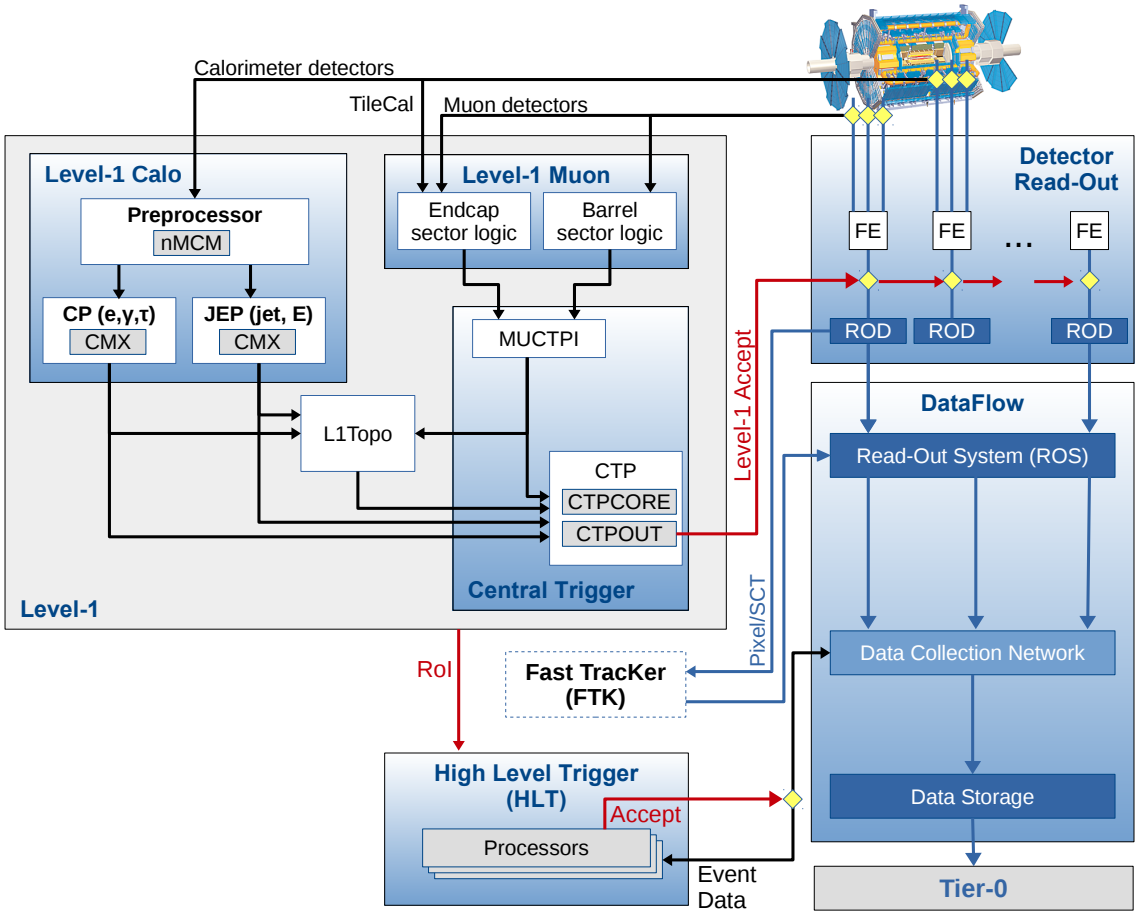


Figure 2.6: ATLAS TDAQ system in Run 2 focusing on the components of the L1 trigger system [89].

2.2.3 TDAQ system

The trigger and data acquisition (TDAQ) system selects the BCs to store the corresponding data. Since the BC frequency of 40 MHz is extremely high, not all BCs can be recorded to storage for offline analyses. Therefore, the trigger system illustrated in Fig. 2.6 is used to select interesting BCs. The detector signals are temporarily stored in the front-end (FE) electronics, and the same signals are also sent to the trigger readout to judge whether the BC is worth to be processed in the following steps. If the trigger is fired, the L1A signal makes the stored data sent to ROD and processed further. The procedures up to this point are exemplified with the LAr calorimeter case in Sec. 2.2.2. The L1 trigger system implemented in custom hardware is used for this triggering. It uses the information from the calorimeters and the muon detectors. The L1 trigger accepts events from the 40 MHz BCs at a rate below 100 kHz. In addition, it sends the information about the region of interest (RoI) for the BC to the following trigger system.

The high-level trigger (HLT) [89] then selects the BCs with algorithms implemented in software in order to record events to disk at about 1 kHz. It uses full information from the ATLAS detector, including the finely segmented calorimeter cells and the inner trackers in the region specified by the RoI signals.

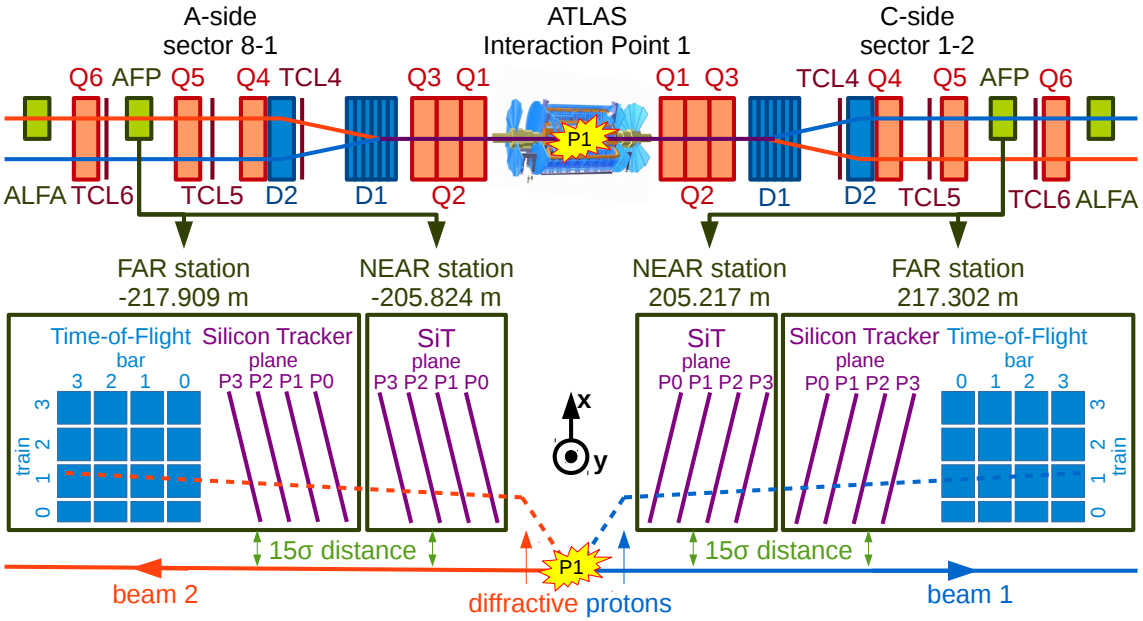


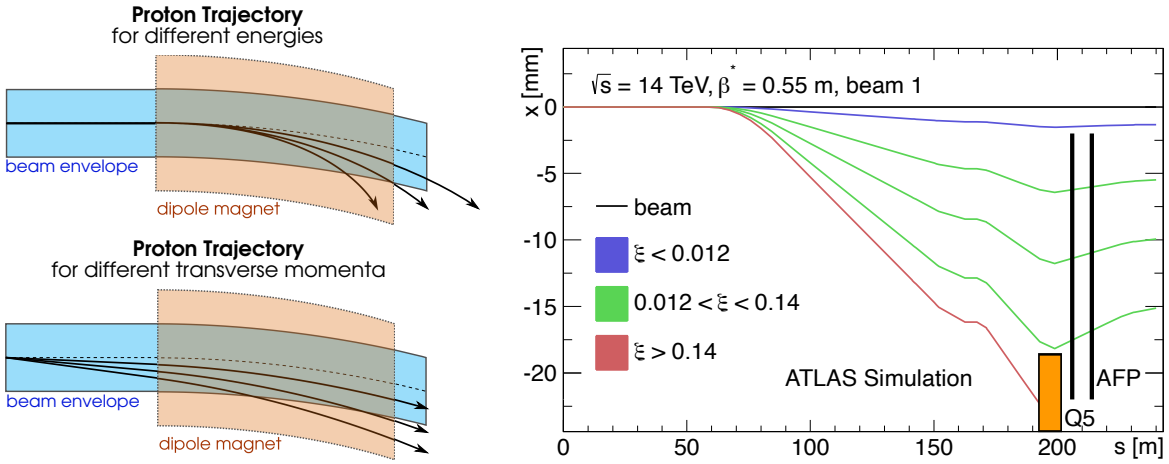
Figure 2.7: AFP system and the related LHC beam optics between the A-side and C-side forward regions [92]. D1 and D2 are the dipole magnets that bend the beam, while Q1-6 are the quadrupole magnets that focus it. The collimators TCL4-6 (Target Collimator Long) are also shown. The NEAR and FAR stations are indicated. The quantitative information about the AFPs uses their local coordinates.

2.3 AFP detectors

The AFP detectors [25] are proton spectrometers developed to study diffraction and search for new physics mainly. They reconstruct protons from small or zero-angle scattering at the ATLAS IP. A schematic of the detectors is presented in Fig. 2.7. The signal proton trajectories are mainly bent by the dipole magnets, which were installed to split the incoming and outgoing beams, beam 1 and beam 2, being merged near the IP in order for BCs for the original purpose. When a proton loses its energy in an event, its trajectory is bent by the magnets more significantly than the usual beam protons. Fig. 2.8 shows the typical proton trajectories with the AFPs. Such effects of the magnets allow the signal protons to reach the forward detectors inserted near the beam. In such forward regions, the protons are measured by the AFP and Absolute Luminosity for ATLAS (ALFA) [90, 91] detectors. Detailed descriptions and simulation results about the LHC beam optics and the related instruments are available in Ref. [78]. In the AFPs, the positions of the protons are measured with silicon trackers (SiTs). Therefore, the proton momenta and hence their fractional energy loss $\xi = \Delta E/E$ can be measured using the information of the positions and the magnetic fields at the LHC magnets. Fig. 2.9 shows the hit position in a SiT plane depending on the proton p_{T} and ξ as well as a distribution of track positions.

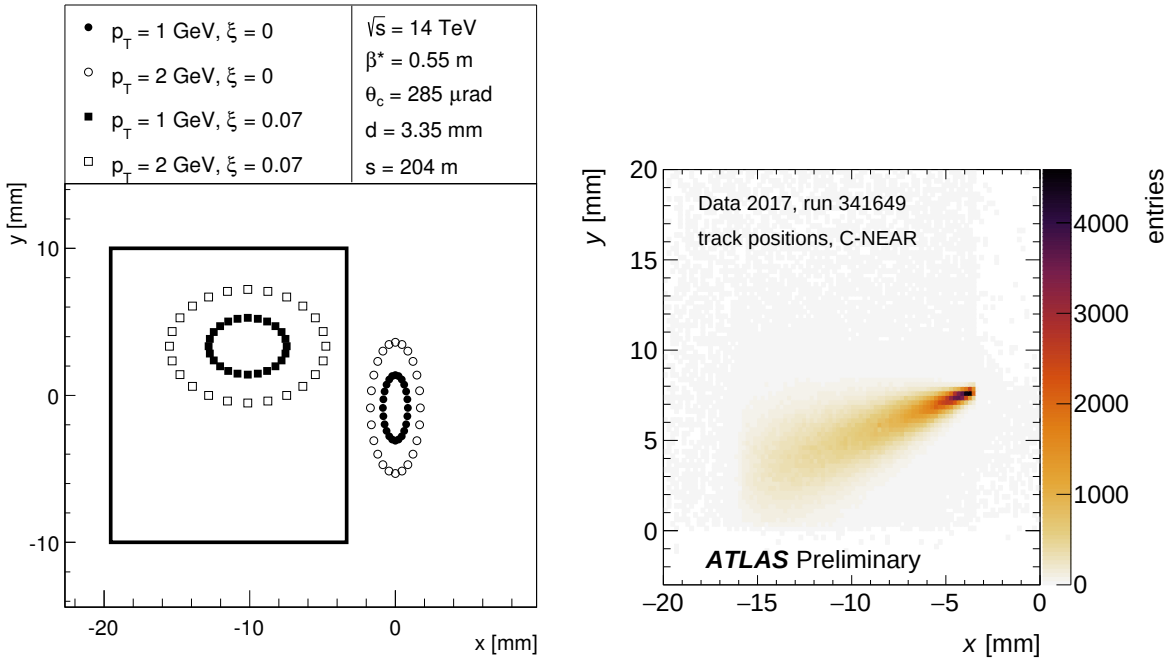
Each side of the AFP system is referred to as an arm. There are two tracking units per arm, referred to as the NEAR and FAR stations. These subdivisions of the AFP detectors are located at $z_{\mathrm{NEAR}} \approx \pm 205$ m and $z_{\mathrm{FAR}} \approx \pm 217$ m, respectively, from the IP. Having two stations allows the system to measure not only proton energy but also its p_{T}. In addition, the proton reconstruction efficiency can be measured using the two stations as described in Sec. 3.5. A detailed description of AFP and its testing and installation is available in Ref. [95].

Each station has four planes with 3D silicon pixel sensors (3D-Si) that form the SiTs to measure the protons' trajectories. These planes are based on the IBL tracker in the ATLAS detector and have 336×80 pixels with 50×250 μm^2 area each. The total area of one SiT plane is 1.7×2.0 cm^2, and it is 230 μm



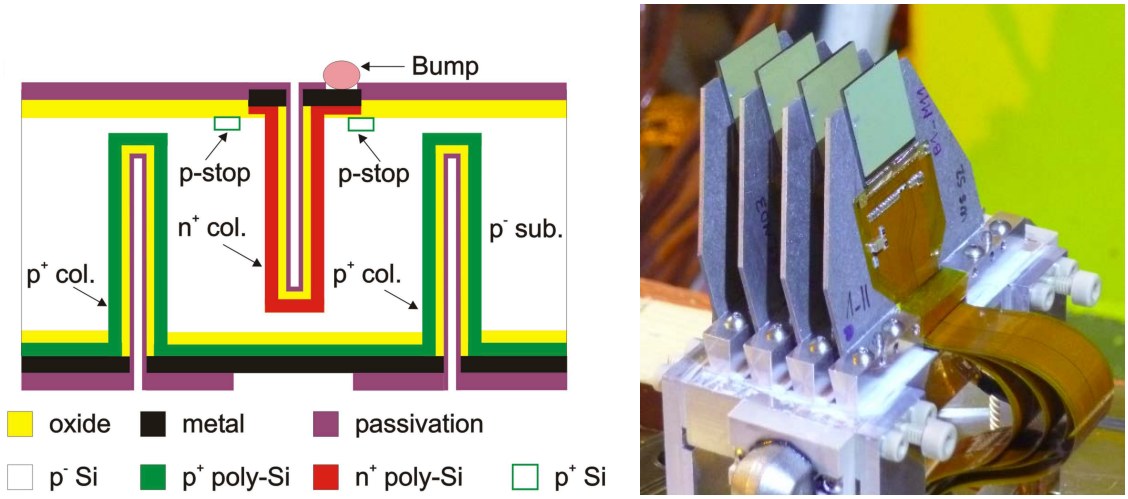
(a) Conceptual illustration (b) Proton trajectories

Figure 2.8: Proton trajectories from the ATLAS IP to the AFPs [93]. The horizontal axes correspond to the z-axis. (a) Conceptual illustration of the trajectories of the proton beam (blue) and its constituent proton being bent by an LHC dipole magnetic field (orange). The vertical axis is the x-axis. Comparisons for different energy and p_{T} values of the proton are illustrated. (b) Simulated proton trajectories in C-side for $\sqrt{s} = 14$ TeV and $\beta^* = 0.55$ m. The horizontal axis is the distance from the IP, and the vertical axis is the local coordinate of the AFPs corresponding to the difference of the ATLAS x-coordinates from the $\xi = 0$ case.



(a) p_{T} and ξ dependency (b) Distribution

Figure 2.9: Hit positions in an AFP SiT plane. (a) Simulated proton hit position for different assumptions of p_{T} and ξ [25]. The thick solid rectangle stands for the SiT plane. (b) Track positions of a SiT plane in the A-side NEAR station in a special run in 2017 [94].



(a) Pixel structure (b) SiT planes

Figure 2.10: (a) Partial structure of a pixel in the AFP 3D-Si sensor [97]. The protons of interest pass through the pixel from top to bottom. (b) The SiT planes in an AFP station [98]. The protons of interest pass through the planes from the left back to the right front. The 3D-Si sensor is on the right front side of each plane. The hit signal is read out by the FE-I4 on the left back side and sent outside the RP through the flexes seen in front.

thick. They are tilted 14° with respect to the x-axis to improve the resolution in the x-direction, to pass the resolution requirement of $\sigma_x = 10$ µm, $\sigma_y = 30$ µm. Clustering in each SiT plane uses the information of charge-weighted pixel hit distribution. Multiple pixel hits per proton improve the positional resolution rather than a single pixel hit. However, a tilt larger than 15° may cause inefficiency because of the columnar electrodes of the 3D-Si. The planes are placed only about 2 mm away from the beam center because the scattering angle of the protons of interest is shallow. Thus, they have slim edges (inactive regions) smaller than 200 µm. Fig. 2.10 (a) shows a partial structure of a pixel. Each pixel is the n^+-in-p type and consists of two n^+-junction columns from the front side and six p^+-ohmic columns from the back side. The isolation between n^+ columns at the surface is realized by the p-stop method, where p-type semiconductors are partially implanted between the electrodes [96]. Protons ionize the p-type bulk, and the electrodes collect the popped electrons.

Each 3D-Si plane is connected to a FE chip by SnAg bump-bonding. The FE-I4 is used as the FE chip. The signals are readout for each channel, amplified and shaped, followed by the discriminators. The processed pulses are triangular pulses with lengths proportional to the collected charge. Thus, the collected charge is measured using the time between exceeding and falling below the discriminator threshold, called time-over-threshold (ToT). The threshold is calibrated using the pulse injectors implemented in the FE-I4 chips. Typically, 2000 electrons are required to send the signal to the next stage. The ToT and the discriminator firing time are stored until an L1 trigger decision is taken as per the ATLAS TDAQ system described in Sec. 2.2.3. The FE-I4 can also send a trigger signal using the logical OR of all fired discriminators. However, any AFP trigger was not used in Run 2 and hence in this research. The FE-I4 is glued and wire-bonded to a flexible printed circuit (flex) to read out the signals. Fig. 2.10 (b) shows the whole picture of the SiT planes.

The FAR stations also house time-of-flight (ToF) detectors to measure the difference in the arrival times Δt of the two protons on the A-side and C-side. The z-position of the vertex is calculated by $c\Delta t/2$, where c is the light speed, and the timing resolution of 25 ps is propagated to the vertex position resolution of about 5 mm. Fig. 2.11 shows the position of a ToF detector relative to the SiT. Each ToF detector consists of 4×4 L-shaped Quartz (LQ)-bars placed at the Cherenkov angle to the beam [99]. Each LQ-bar consists of the radiator arm and light-guide arm transparently glued together at 90°. The Cherenkov light emitted

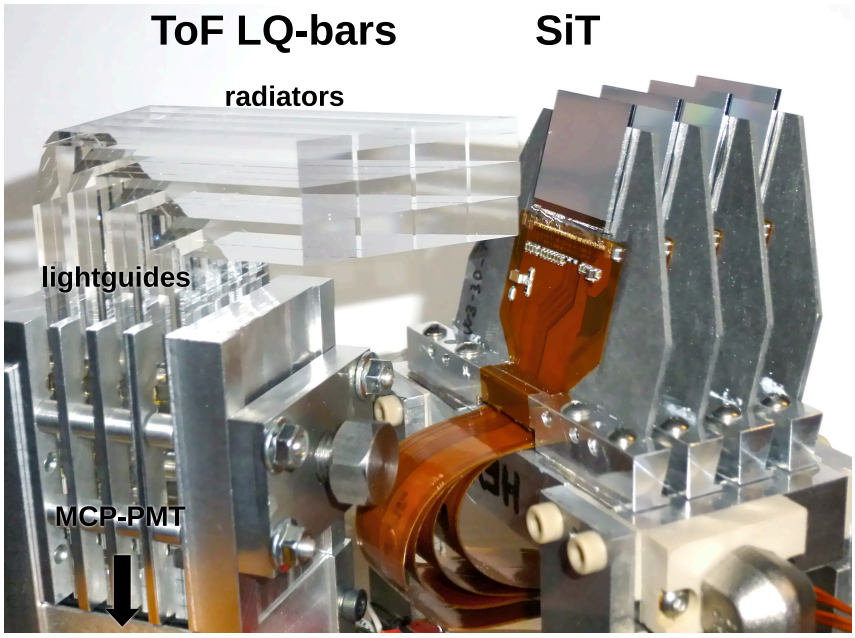


Figure 2.11: The ToF and SiT detectors in an AFP FAR station [99]. The protons of interest pass through the SiT planes and the ToF radiators in order from right back to left front.

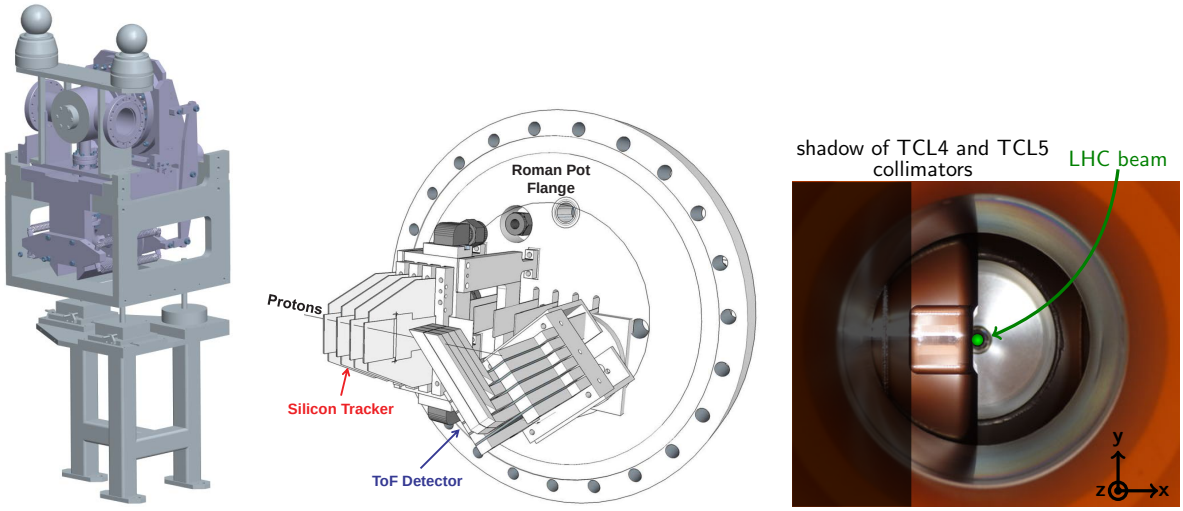
from a radiation arm is propagated to the micro-channel plate multi-anode PMT (MCP-PMT) through the light-guide arm. The photo-electrons created in the MCP-PMT induce pulses processed by a discriminator and high performance time-to-digital converter (HPTDC). The processed pulses are sampled in 1024 bins of about 25 ps. The ToF detectors can also send a trigger signal. However, this analysis does not consider the ToF detectors because of their low efficiency in 2017. The reason is discussed in Sec. 3.1.

The AFP and ALFA stations are in separate Roman Pots (RPs) [90]. The AFP detectors are inserted horizontally in the x-direction into the beam pipe by the RPs. The RP technology provides movable near-beam devices that bring the detectors within a few mm of the beam without affecting the accelerator vacuum. Whenever stable beams are declared, the RP systems insert the AFPs so that the SiT edge is about 2 mm from the beam center. The nominal beam-detector distance is 15σ, where σ is the standard deviation of the spatial beam distribution in the horizontal plane. Fig. 2.12 shows an AFP station with its contents to illustrate how the RP is used. The RP covers the AFP detectors with about 1 mm thick steel to protect them from the ultra-high vacuum in the beam pipe,[5] but the window where the protons enter the AFP and the floor separating the AFP and the beam have 300 µm thickness.

Data from the AFP are sent to the ATLAS TDAQ system through a 200 m long optical MTP fiber ribbon to be recorded [102]. The signals for the trigger, on the other hand, are sent to the system through two air-core coaxial cables corresponding to the NEAR and FAR stations for each arm. The cables propagate the signals within a time to comply with the L1 trigger latency. The L1A signals are sent back to the AFP detectors through another optical fiber. The recorded data from the main readout are used for AFP clustering, tracking, and proton reconstruction. The clusters are defined in each SiT plane using pixel-hit data. The tracks and the corresponding protons are then reconstructed using the cluster information. The reconstruction steps starting from the clustering, are conducted at a software level and described in Sec. 3.5.

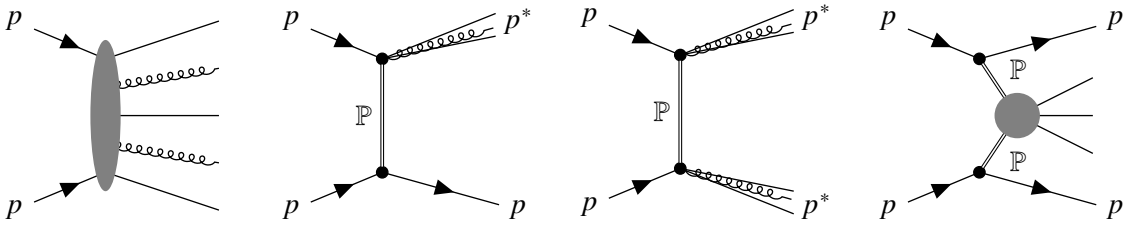
The primary source of the background hits is pileup protons. In general, the pileup events are subdivided into non-diffractive dissociation (ND), single-diffractive dissociation (SD), double-diffractive dissociation

[5] The inside of the RP is also highly vacuumed to prevent the RP material from being bent.



(a) Whole station (b) SiT and ToF detectors (c) Longitudinal view

Figure 2.12: AFP station [100, 101]. (a) Whole AFP station. The LHC beam penetrates the illustrated tube from the left back to the right front. The RP is inserted into the beam pipe from right back to left front. (b) RP flange with SiT and ToF detectors attached. This is built into the top of (a) at the same angle. (c) View from inside of the beam pipe to the RP inserted from left to right. The collimator shadow is illustrated on the left.



(a) ND (b) SD (c) DD (d) CD

Figure 2.13: Simplified Feynman diagrams of the pileup events in the pp collisions. (a) Non-diffractive dissociation. (b) Single-diffractive dissociation. (c) Double-diffractive dissociation. (d) Centra diffraction. $\mathbb{P}$ stands for pomeron.

(DD), and central diffraction (CD). Diffraction is a process where a strongly interacting color singlet object, pomeron, is exchanged, sometimes remaining the beam proton intact [25]. Fig. 2.13 shows the simplified diagrams of these processes. The ND is the most usual pileup with the largest cross section. The SD and DD pileups are diffractions where single and double protons dissociate, respectively. The CD pileup has two pomerons exchanged, called the double pomeron exchange (DPE) process. Detailed descriptions and a diagram about diffraction are available in Ref. [103], for example. The SD pileup events have the second dominant cross section of about 15% of the total cross section of the pileup events in the Run 2 pp collisions. They are the dominant pileup in this research requiring AFP proton tagging. Most tracks in Fig. 2.9 (b) are from the SD pileup events. The contributions from the out-of-time pileups are negligible.

Protons can induce showers, where many AFP tracks are reconstructed. The showering occurs when a proton interacts with the AFP detector, RP steel window and floor, or collimator for a higher ξ region.[6] The average number of reconstructed tracks is slightly larger in the FAR stations than in the NEAR stations, mainly due to the showering. If a proton showers in the NEAR stations, the proton does not reach the FAR station. The proton reconstruction in this research requires its associated tracks in both NEAR and FAR

[6] Which causes the most significant influence of the shower depends on the position in the plane.

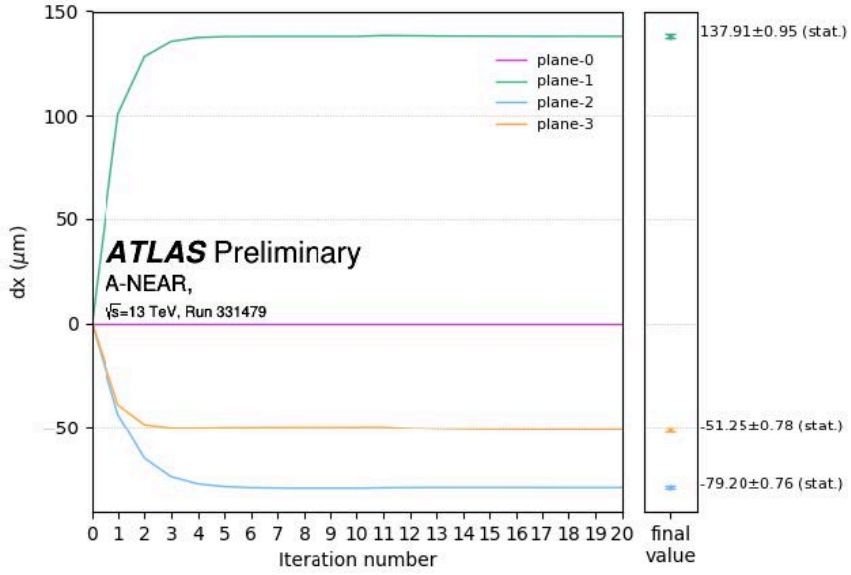


Figure 2.14: The alignment in the x-direction over 20 iterations of the algorithm for the A-side NEAR station [104], where the dx is the x-direction positional difference of the coordinate from the first plane (plane-0). Each line corresponds to a plane of the station.

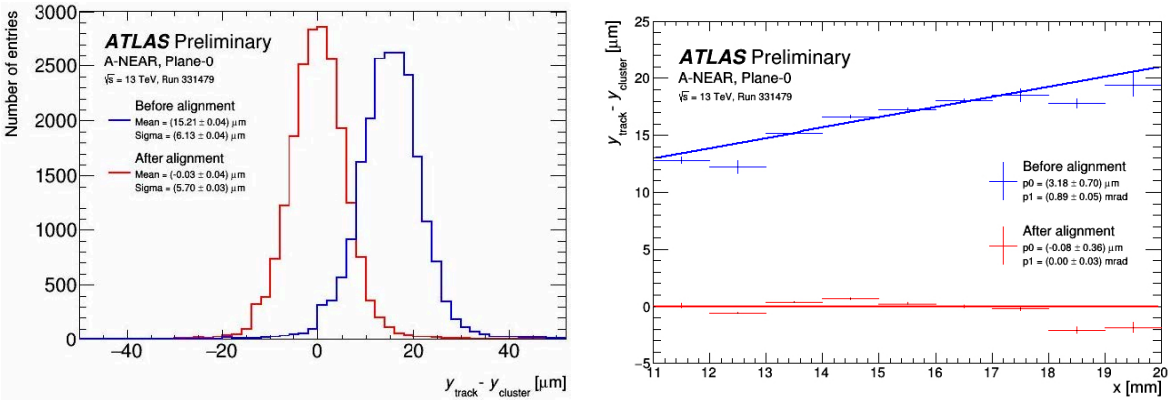
stations as described in Sec. 3.5, so the showering has the largest contribution to inefficiency of proton reconstruction.

ALFA also detects protons but is only used in special low-luminosity runs. Although ALFA data are used to measure exclusive hadrons with $O(1)$ GeV, SD dijets with $O(10)$ GeV, and scarcely DPE dijets, the ALP stays inaccessible by this detector due to very low integrated luminosity.

2.3.1 Calibration and performance evaluation

This section briefly describes the calibration and performance evaluation of the AFP SiTs. Although the performance study results are not published in any paper, some results and plots are available in Refs. [94, 104].

The alignment of the SiTs consists of local and global alignment. The local alignment is for the relative positions of the SiT planes in the RP in each station and is sometimes called inter-plane alignment. The offsets of z-coordinate and rotation in x-z and y-z planes are negligible. The local alignment starts by assuming perfect alignment and that all tracks are parallel to the beam-line. Only the data events with a single cluster per plane are used for this study, and the track is reconstructed by the nominal method described in Sec. 3.5. Then, the residuals between a track and its associated clusters are calculated, and the alignment parameters are derived from the average of the residuals and updated. This procedure is repeated 10 times first and is iterated 10 times further applying the selection of $\chi^2 < 2$ to remove the outliers. Fig. 2.14 shows how the relative position of other planes to a plane is determined in a station by local alignment. Fig. 2.15 (a) shows the comparison of the y-direction residual distribution between before and after the iteration in the first plane of the A-side NEAR station, while (b) shows the rotation alignment in the same manner in the same plane. The alignment parameter uncertainties are corrected by a run-by-run evaluation. The uncertainty of the local alignment is within 10 μm and is negligible in this analysis as described in Sec. 7.2.



(a) Residual distribution (b) x dependency

Figure 2.15: (a) Comparison of the y-direction residual distribution between before and after the alignment procedure in the first plane (plane-0) of the A-side NEAR station with the statistical uncertainties [104]. (b) Comparison of the x dependency of the y-direction residual representing the AFP rotation alignment.

The global alignment is for the x-direction distance of the SiT from the beam center in each station. The x-position of the SiT is calculated as

$$x(r,s) = x_{\text{pre-align}} + x_{\text{tracker}} - x_{\text{beam}}(s) + x_{\text{RP}}(r,s) + \delta x_{\text{corr}}(s),$$

where r and s stand for run and station, respectively. $x_{\text{pre-align}}$ is the coordinate before the global alignment, and $x_{\text{tracker}} = -0.5$ mm is the distance of the edge of the active region of the SiT from the outer side of the RP floor. x_{beam} is the beam position determined using the beam-based alignment (BBA) procedure [105] in dedicated runs, moving the collimators towards the beam until a sharp change in the rate is observed in Beam Loss Monitors (BLMs) [106]. The beam position is determined from the BBA run at the beginning of 2017. The Beam Position Monitor (BPM) continuously monitors the beam position [107], and the change over time is found to be less than 100 μm. x_{RP} is the inserted RP position.

Considering

$$x_{\text{AFP}}(r,s) = x_{\text{pre-align}} + x_{\text{tracker}} - x_{\text{beam}}(s) + x_{\text{RP}}(r,s)$$

as the preliminary alignment result, δx_{corr} is determined as the *in-situ* corrections using the exclusive dimuon data in the dilepton analysis [66]. From the proton energy loss fraction calculated by the dimuon system $\xi_{\mu\mu}$, the ξ_{AFP} and hence the x-position of the proton is estimated as $x_{\mu\mu}$ and compared with the detected position in the AFP x_{AFP}. Proton p_x is very small and almost negligible. The correction is then calculated as the mean of the difference,

$$\delta x_{\text{corr}} = \left\langle x_{\text{AFP}} - x_{\mu\mu} \right\rangle,$$

and is taken as the overall global alignment uncertainty. Fig. 2.16 shows the $x_{\text{AFP}} - x_{\mu\mu}$ distribution before and after the in-situ correction and with and without the background events subtraction. The dominant component of the background is called combinatorial background. It is estimated using the event mixing method from which the background estimation method used in this research originates as described in Sec. 6.1. This is subtracted from the sample to obtain the pure $x_{\text{AFP}} - x_{\mu\mu}$ distribution. The muon momentum scale and resolution are calibrated separately as a few percent [108]. Although Fig. 2.16 shows $|\delta x_{\text{corr}}|$ over 400 μm, the resulting global alignment uncertainty is determined as 300 μm considering the results of the other stations and testing various event selection criteria to be conservative enough. Finally, the x-direction inter-station alignment is performed for each arm.

The overall ξ_{AFP} resolution is measured in the dilepton analysis [66] using the dileptons as the probe. Fig. 2.17 shows the observed $\xi_{\text{AFP}} - \xi_{\ell\ell}$ distribution for each side, where $\xi_{\ell\ell}$ is the proton energy loss

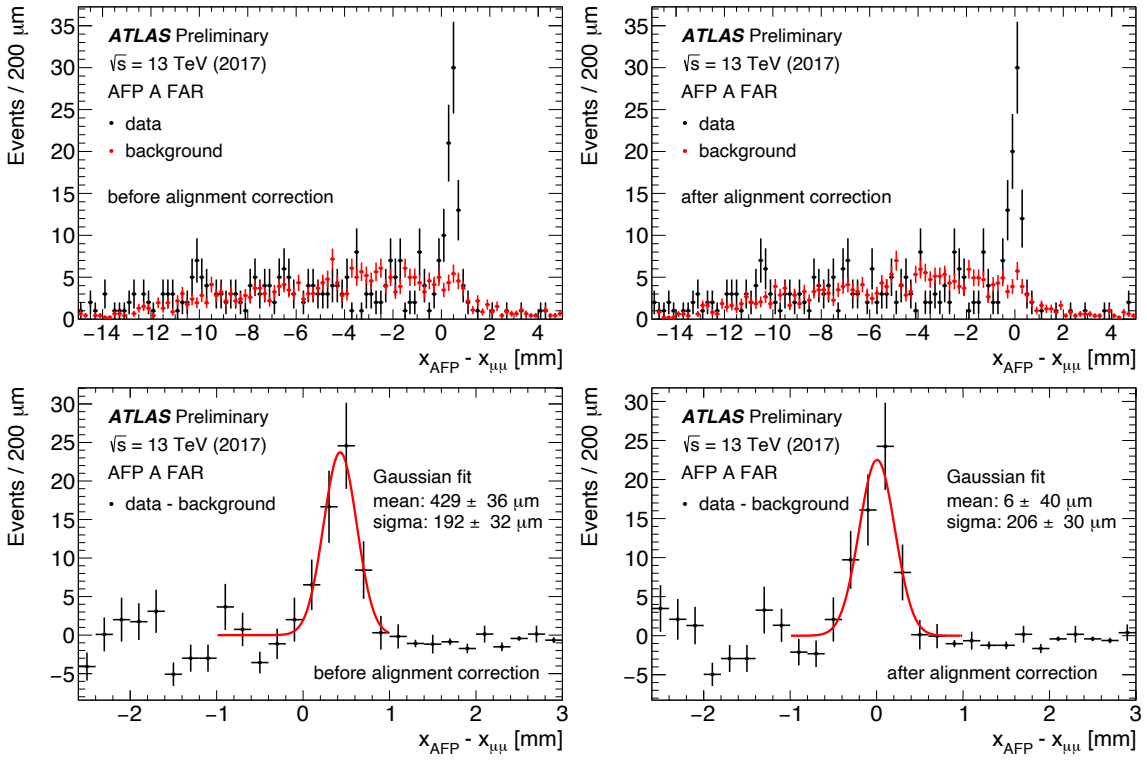


Figure 2.16: The $x_{\mathrm{AFP}} - x_{\mu\mu}$ distributions in the in-situ AFP alignment [109]. The left and right plots show the distributions before and after the alignment correction, respectively. In the upper plots, the estimated background distribution is overlaid. The background contributions are subtracted in the lower plots. The Gaussian fitting results are shown.

fraction calculated from the dilepton system. Although the $\xi_{\ell\ell}$ uncertainty affects the result, the effect is expected to be much smaller than the true ξ_{AFP} uncertainty. The relative uncertainty of most ξ value is determined to be about 10% as described in Sec. 3.5 in which the proton reconstruction efficiency and the acceptance are discussed.

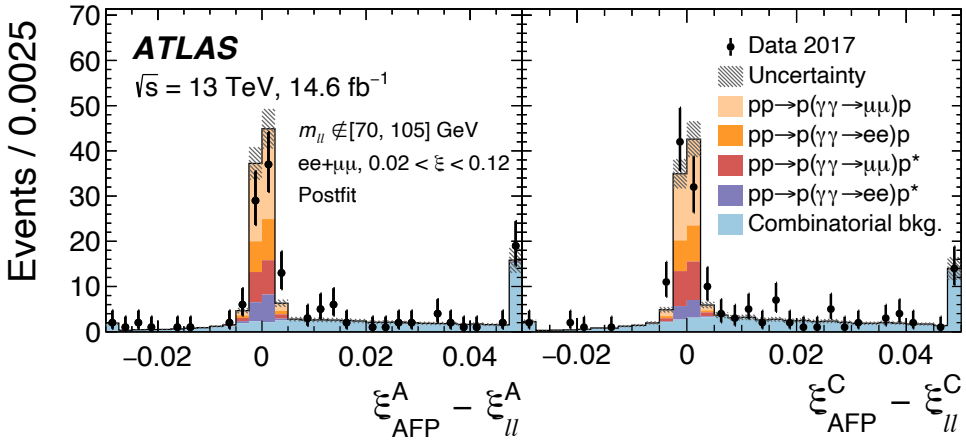


Figure 2.17: The $\xi_{\mathrm{AFP}} - \xi_{\ell\ell}$ distribution for each side observed in the dilepton analysis [66]. Dissociated proton is denoted by *.

3 Data and Monte Carlo Samples

This section details data and signal MC samples. The object reconstruction and event preselection using the object information in these samples is also discussed.

3.1 Data samples

This research uses the data of 14.6 fb^{-1} integrated luminosity collected in 2017 during the LHC Run 2 pp collisions, which is exactly the same as the one used in the dilepton analysis [66]. 2017 is the first year the AFP was installed on both sides and recorded data suitable for diphoton-proton matching. However, in 2018, the last year of Run 2, although data was recorded with AFP, it was synchronized with the incorrect BCs. Thus, there is no diphoton-proton correlation, and it is not considered in this analysis.

The AFPs were inserted for most of the 2017 runs. There were 213 insertions corresponding to 46 fb^{-1} recorded by ATLAS out of 50 fb^{-1} delivered by LHC. Among them, the data when the AFP data acquisition (DAQ) system was in operation corresponds to 132 insertions and 32 fb^{-1}. Fig. 3.1 shows the changes in integrated luminosities in 2017 corresponding to these criteria. The ATLAS global good run list (GRL) [110] applies data quality requirements and reduces the data to 26 fb^{-1}. A subset GRL dedicated to the analyses using the AFPs is then applied to reduce the data to 19.2 fb^{-1}. This GRL requires

- all AFP stations are inserted into their nominal data-taking positions,
- at least two SiT planes in each station have high voltage on,
- no DAQ problems associated with the Reconfigurable Cluster Element (RCE) data processing compute modules,
- no DAQ problems associated with the Readout Subsystem (ROS) S-link cables [102].

These requirements are imposed for each *lumiblock*, a subdivision of a run of typically 1 minute duration with fixed TDAQ conditions. A further tighter requirement is finally applied as the *nominal* GRL to select the 14.6 fb^{-1} data. This GRL requires at least three AFP SiT planes operational at high voltage for each station. The lumiblocks removed by this requirement are in the runs showing severe drops in AFP track reconstruction efficiency due to the A-side FAR station having fewer than three functional planes.

The data were recorded using a diphoton trigger (HLT) that required two EM clusters, clusters of EM detector responses indicating electron or photon, with transverse energy E_{T} above 35 and 25 GeV, respectively, both fulfilling photon identification criteria based on shower shapes in the EM calorimeter [111],[1] as done in the high-mass diphoton resonance search [52].

[1] The trigger is referred to as `g35_medium_g25_medium` in Ref. [111].

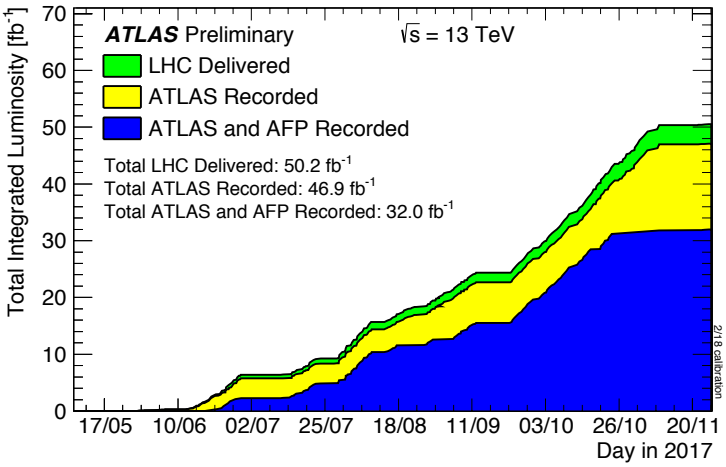


Figure 3.1: Integrated luminosity recorded by ATLAS and AFP in the LHC pp collisions in 2017 [104].

3.2 Signal MC samples

Simulated signal MC samples are used primarily to construct the model of signal $m_{\gamma\gamma}$ distribution and evaluate estimate the signal efficiency. The EL ALP signal MC samples were generated with SuperChic 4.02 [112–114] in June 2021, and the SD and DD samples were generated with SuperChic 4.14 [115] in October 2022. SuperChic is a generator for exclusive and photon-initiated production in proton and heavy ion collisions. It uses state-of-the-art formalism for cross section calculation using a *structure function approach*. This approach uses the proton structure function directly bypassing any explicit reference to photon PDF in Eq. 1.1 to reduce the uncertainty of cross section [116]. For any of the EL, SD, and DD signals, the cross section is calculated as

$$\sigma = \frac{1}{2s}\int \frac{\mathrm{d}\boldsymbol{p}_1'\mathrm{d}\boldsymbol{p}_2'\mathrm{d}\Gamma}{E_1'E_2'}\alpha\left(Q_1^2\right)\alpha\left(Q_2^2\right)\frac{\rho_1^{\mu\mu'}\rho_2^{\nu\nu'}\mathcal{M}^*_{\mu'\nu'}\mathcal{M}_{\mu\nu}}{Q_1^2Q_2^2}\delta^{(4)}\left(q_1+q_2-k_1-k_2\right),$$

where the subscripts 1 and 2 identify the initial proton, $\boldsymbol{p}$ and E are the momentum and energy of the outgoing hadronic system, α is the structure constant, q and k are the incoming and outgoing photon 4-momenta with $q^2 = -Q^2$, and $\mathrm{d}\Gamma = \prod_{j=1}^{2}\mathrm{d}\boldsymbol{k}_j/2E_j(2\pi)^3$ is the standard phase space volume. s is the Mandelstam variable, where $\sqrt{s} = 13$ TeV. The photon density matrix

$$\rho_i^{\alpha\beta} = \int \mathrm{d}\left(m_{p^{(*)}}\right)_i^2\left[A_i^{\alpha\beta}F_1\left(x_{B,i},Q_i^2\right) + B_i^{\alpha\beta}F_2\left(x_{B,i},Q_i^2\right)\right],$$

where A, B, and $x_{B,i}$ are functions of photon and hadronic system kinematics, has the proton structure functions $F_{1,2}$, which are precisely determined from previous experiments [31]. $\mathcal{M}$ corresponds to the $\gamma\gamma \to a \to \gamma\gamma$ amplitude, where only the s-channel is considered because the contributions of t and u-channels are expected to be extremely small. The cross section is multiplied by the *soft survival factor* S^2, which is the probability of no additional soft inelastic interactions between the intact protons. If such an interaction occurs, there is a color flow between the protons, and hence both A and C-sides protons would no longer remain intact. Fig. 3.2 shows the soft survival factor for the dilepton events illustrated in Fig. 1.8. For the EL case, the survival probability is higher than the other processes because the impact parameter is large since the photon is quasi-real. Similar values and diphoton kinematics dependency are expected for the ALP signals.

The ALP mass m_X, coupling constant $g = 4f^{-1}$, and event type (EL, SD, or DD) are specified in SuperChic for each sample along with some other configuration. In this research, signal samples ranging

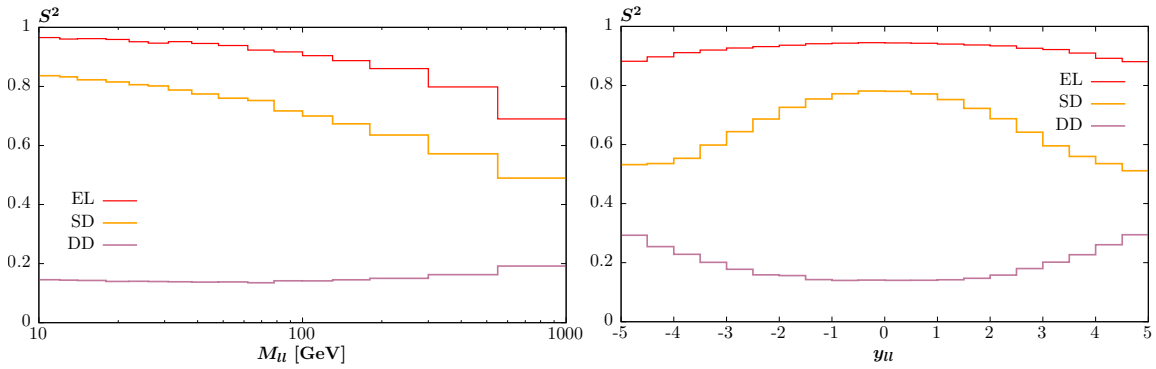


Figure 3.2: Soft survival factor calculated for $\gamma\gamma \to \ell\ell$ events in the pp collisions at $\sqrt{s}$ = 13 TeV [114]. The diphoton mass dependency (left) and dilepton rapidity dependency (right) are illustrated for the EL, SD, and DD processes.

Table 3.1: Summary of the used set of SuperChic parameters with their descriptions and assigned values.

Description	Value
Number of events	10000
Center-of-mass energy	$\sqrt{s}$ = 13 TeV
Collision type	pp collisions
Coupling constant	$f^{-1} = 0.05$ TeV^{-1}
Diphoton rapidity range	$y_{\gamma\gamma} \in [-2.4, 2.4]$
Final state photon p_{T} range	$p_{\mathrm{T}} \in [20, \infty)$ GeV
Final state photon η range	$\eta \in [-2.4, 2.4]$

from 200 GeV to 1600 GeV were generated with fixed coupling constant $f^{-1} = 0.05$ TeV^{-1}. The major configuration commonly used for the signal samples is listed in Table 3.1. The computed generation cross sections are considered in Sec. 5.2. The generated objects at this stage before the detector simulation are referred to as the *truth* objects.

After the event generation by SuperChic, Pythia 8.307 [117] is used for the SD and DD signals to hadronize the dissociative system. SuperChic removes the final state intact protons of the SD signals from its output which is the input to Pythia because Pythia reports a violation of the momentum conservation law and does not work. Instead, Pythia automatically adds additional QCD objects not to violate the conservation laws due to the lack of intact proton and hadronizes them as well. However, these additional impossible QCD objects are boosted enough not to contaminate the ATLAS detector. This was confirmed by comparing the production vertices of the two final state truth photons with the first and the second highest energy and seeing if they are identical for each event. On the other hand, they can contaminate AFP and sometimes hadronize into a proton and behave as a fake intact proton. Therefore, such QCD objects were removed at the truth level, and an intact proton was added artificially, having its momentum calculated from the initial proton and photon.

The initial photon p_{T} on the intact side of the SD samples is exactly given as 0. Since the initial beam proton p_{T} is also assumed to be 0, this means calculated intact proton p_{T} is 0 as well. This simplified assumption was used because it was confirmed to have negligible effects on the final results by testing the case where p_x and p_y are independently smeared by Gaussian distribution with the width taken from the EL sample intact protons. Fig. 3.3 shows the p_x and p_y width of intact protons in the EL samples for each m_X. To be conservative, 0.23 GeV was used for the smearing test.

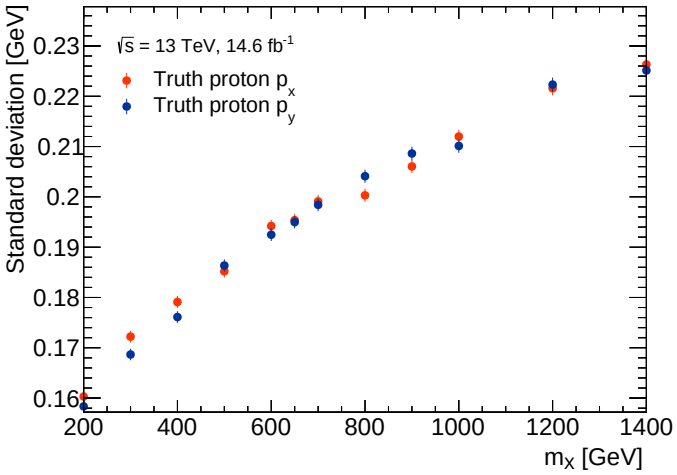


Figure 3.3: Width of proton p_x and p_y distributions of the EL signal samples for each ALP mass.

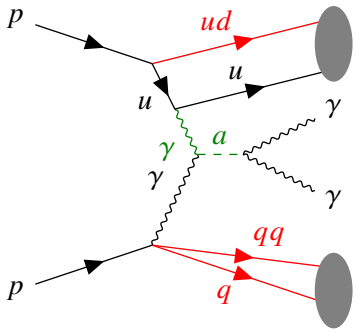


Figure 3.4: Feynman diagram illustrating an SD signal event processed in PYTHIA. The inelastic photon production is modelled as $u \to u\gamma$. The particles indicated by the black lines are input to PYTHIA with their momenta calculated in SUPERCHIC. The grey ovals mean hadronization. The particles indicated by the green lines are not input to PYTHIA but considered only in SUPERCHIC. The particles indicated by the red lines are automatically added by PYTHIA. The ud object interacts with the final state u-quark to hadronize. The QCD objects $q + qq$ replace the intact proton that should have been one of the input particles to PYTHIA.

It is assumed that there is an underlying $u \to u\gamma$ process in inelastic photon production of the SD and DD signal events in order for PYTHIA to work, even in the lower photon virtuality Q^2 region where the proton is likely to excite and dissociate without $q \to q\gamma$ process. This does not cause any problem because the SD and DD signal events scarcely have inelastic photons with $Q^2 < 5$ GeV^2. Fig. 3.4 shows the diagram of an SD signal processed in PYTHIA.

The effect of multiple interactions in the same and neighboring bunch crossings (pileup) was modelled by overlaying the simulated signal event with inelastic pp events generated with PYTHIA 8.210 [118] using the NNPDF2.3LO set of parton distribution functions [119] and the A3 set of tuned parameters [120], in association with EVTGEN 1.2.0 [121]. The MC events were weighted to reproduce the distribution of the actual number of interactions per BC (μ) observed in the data. In contrast, the pileup protons in the AFP detectors are neglected in the nominal signal MC samples. Effects of the pileup protons on the result were estimated as the systematic uncertainty separately and found to have a small effect as expected as described in Sec. 7.2.

All simulated signal events were processed using a fast simulation of the ATLAS detector [122], where the full simulation[2] of the calorimeter is replaced with a parameterization of the calorimeter response [124].

[2]The full simulation of the ATLAS detector [122] is based on GEANT4 [123]. This research does not use the full simulation.

The response of the AFP spectrometer is modelled by another fast simulation, where the Gaussian smearing is applied to track positions based on the AFP spatial resolution. Each proton is selected or discarded according to the reconstruction efficiency described in Sec. 3.5. All simulated events were reconstructed with the same reconstruction algorithms as those used for data.

3.3 Track and primary vertex reconstruction

The energy of charged particles is deposited as hits on different layers in the ATLAS inner-detector system. Although the charged particle tracks reconstructed using the hits are not directly used for the event selection in this research, they are used for reconstructing the other objects. The tracking algorithm consists of three steps.

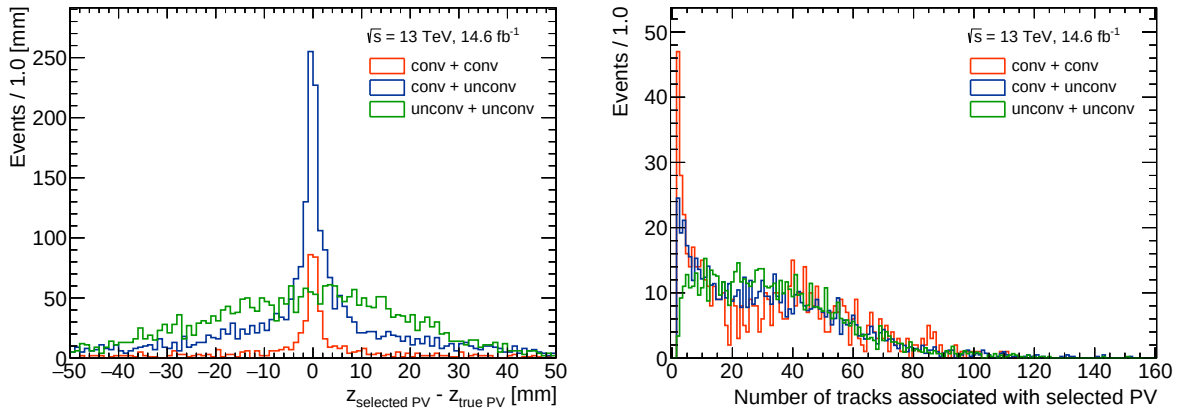
1. Neighboring silicon-sensor hits are merged into a cluster. Then, a position in the (x, y, z) space, called space-point, is defined for each cluster using the connected component analysis (CCA) technique [125].
2. A seed-track is defined as a set of three or more space-points [126].
3. Each seed-track is extrapolated to the TRT hits using the Kalman filter [127] and incorporates the space-points on the trajectory to construct a track object. Momentum is calculated for each track.

During the above steps, some cleaning procedures, such as overlap removal of the multiple seed-tracks and ambiguity resolution of the track candidates to suppress the fake tracks, are taken. The track reconstruction is detailed in Ref. [126].

The primary vertices are reconstructed using the reconstructed tracks. Each primary vertex is defined as a spacial position attributed to tracks requiring some criteria [128]. The two reconstructed and selected photon objects with the highest p_{T} (leading and sub-leading photons) and additional information from the tracking systems are used to identify the diphoton production vertex from the candidate primary vertices [129]. This is called *photon pointing* method. Although primary vertex information is not used in this research, the analysis framework automatically excludes the events with the selected primary vertex associated with less than two tracks.[3] However, the EL signal event has no charged particle in the final state in most cases, while the final state photons are sometimes converted to electron pairs due to interactions with the detector materials, leaving tracks. A photon splitting into an electron pair is referred to as a converted photon, while a photon without tracks is called an unconverted photon. Still, the automatic selection efficiency is almost 100% because the photon pointing selects a pileup vertex with multiple associated tracks. Fig. 3.5 shows the primary vertex-related distributions of an EL signal MC sample. When the leading two photons are unconverted, the photon pointing selects a pileup vertex with many associated tracks.

Diphoton production vertex reconstruction methods without using track information have been developed. These methods, such as *non-pointing* [130, 131] and *calo-vertexing* [132] use only diphoton information to reconstruct a single diphoton production vertex with about 15 mm spacial resolution. It is aimed at a search for long-lived particle decaying into a final state with diphoton at a displaced vertex from the primary vertex. However, this method could also be efficiently used in the analysis of this research. This is discussed in Sec. 9.2.2. Still, such a method was not integrated into the analysis framework of this research because there is no motivation to impose additional event selection since the integrated luminosity of the nominal data is as small as 14.6 fb^{-1}.

[3] This is imposed for usual analyses for high event quality.



(a) Longitudinal positional difference (b) Number of tracks

Figure 3.5: The primary vertex-related distributions of the EL signal MC sample for $m_X = 400$ GeV. No event selection described in Sec. 4 is applied. Each color corresponds to the combination of the conversion type of the two leading photons. (a) The distribution of the longitudinal positional difference between the selected primary vertex and truth signal primary vertex. (b) The distribution of the number of tracks associated with the selected primary vertex.

3.4 Photon reconstruction

The photon reconstruction and selection inherit the high-mass diphoton resonance search [52]. Photons are reconstructed from calorimeter signals using a dynamical, topological cell clustering-based algorithm along with the electron reconstruction. The reconstruction, calibration, identification, and isolation of photons and their performances are detailed in Ref. [133]. Fig. 3.6 shows the algorithm flow of the reconstruction. The topological clusters (topo-clusters) [134, 135] are defined first. Then, the superclusters are determined based on them, and finally, the photon objects are identified and isolated.

A topo-cluster is defined from proto-clusters. The formation of each proto-cluster begins with finding a calorimeter cell with $|\varsigma^{\mathrm{EM}}_{\mathrm{cell}}| \geq 4$, where $\varsigma^{\mathrm{EM}}_{\mathrm{cell}}$ is the ratio of energy deposit at the EM scale[4] to expected noise. Neighboring cells with $|\varsigma^{\mathrm{EM}}_{\mathrm{cell}}| \geq 2$ are then selected to form a cluster including the initial cell. Each neighboring cell becomes a seed cell in the next iteration, recursively collecting each of its neighbors in the proto-cluster. Next, all neighboring cells with positive energy are incorporated. Topo-clusters are then formed by merging and separating within each proto-cluster and between the proto-clusters. Some preselections are applied to the energy in each topo-cluster only from the EM calorimeter cells and its ratio to the total energy in the topo-cluster to define an EM topo-cluster. Topo-clusters are matched to tracks or conversion vertices using the information on the positions and energy of the clusters and the positions and momenta of the tracks. Two-track conversion vertices are reconstructed from two opposite-charge tracks forming a vertex consistent with that of a massless particle, while single-track vertices require no hits in the innermost sensitive layers.

A supercluster is then formed by merging some topo-clusters. Each seed topo-cluster is selected based on E_{T} required to be high. Satellite topo-clusters are combined with the seed to form a supercluster, as illustrated in Fig. 3.7. For both electrons and photons, a topo-cluster is considered a satellite topo-cluster if it is in the $\Delta\eta \times \Delta\phi = 0.075 \times 0.125$ window around the seed topo-cluster. For converted photons, in addition, a topo-cluster is added as a satellite if itself or its best-matched track shares a conversion vertex with the seed topo-cluster. For electrons, on the other hand, a wider window is set as the acceptance region

[4] The EM scale is the basic signal scale accounting correctly for the energy deposited in the calorimeter by EM showers.

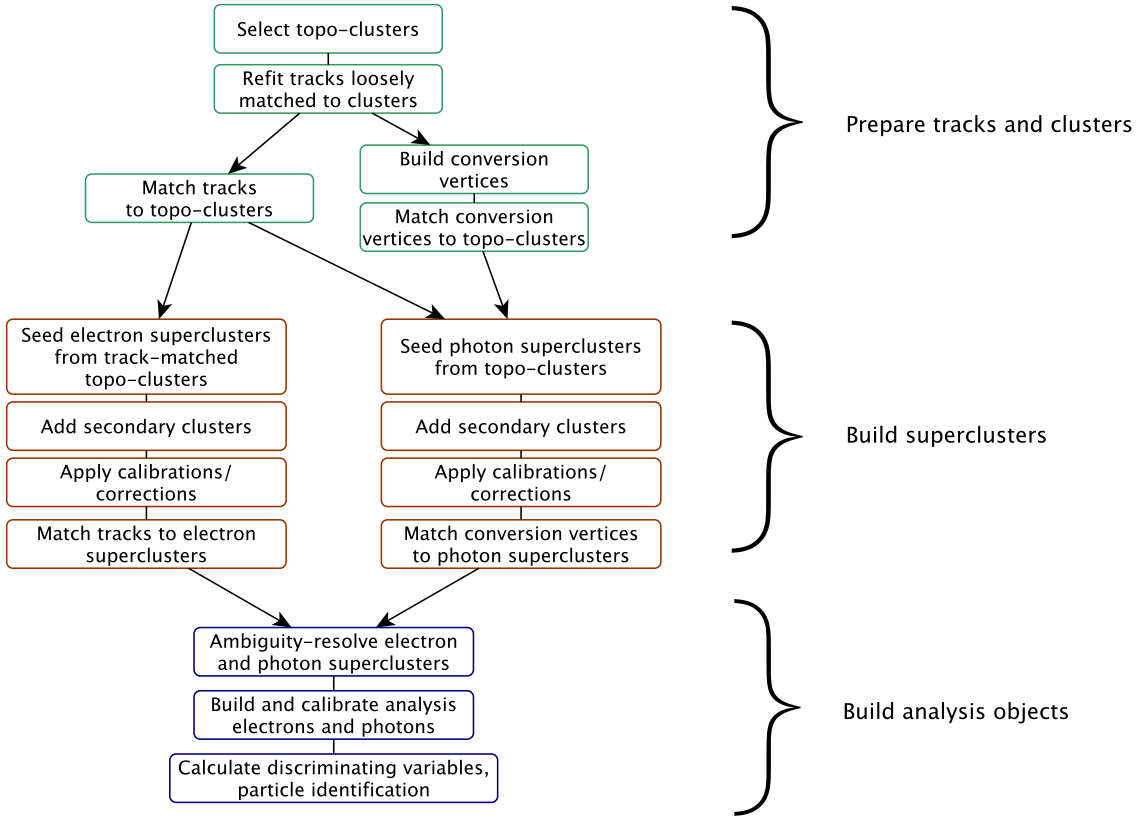


Figure 3.6: Algorithm flow of the photon and electron reconstruction [133].

to form a supercluster.[5] After the superclustering, conversion vertices are matched to photon superclusters, and tracks are matched to electron superclusters. An unconverted photon is defined as a cluster matched to neither an electron track nor a conversion vertex. About 20% of photons at low $|\eta|$ convert in the inner detectors, and up to about 65% convert at $|\eta| \approx 2.3$.

The superclustering is performed for photons and electrons independently, so a given seed cluster can produce both an electron and a photon. This ambiguity is resolved using a dedicated algorithm. Then, superclusters' energy scale and resolution are calibrated using $Z \to ee$ events considering pileup dependency. The energy scale correction is validated using $Z \to \ell\ell\gamma$ events for photons. Subsequently, shower shape and other discriminating variables are calculated for electron and photon identification. Photons and electrons are defined as the superclusters having these variables.

Photon candidates are classified based on cut-based identification requirements imposed on the calorimeter shower shape variables. The primary purpose is to discriminate photons from hadronic jets. Based on the levels of signal acceptance and background rejection of these cuts, several photon identification working points are prepared. The definition of the variables and working points can be found in Ref. [133]. This research uses the *tight* working point. The *loose* working point is based on the shower shape in the middle layer of the EM calorimeter and the energy deposited in the hadronic calorimeter. The tight working point additionally uses information from the finely-segmented front layer of the calorimeter. It is separately optimized for unconverted and converted photons to account for the generally broader lateral shower profile

[5] This is an advantage of topo-cluster over the sliding-window algorithm used previously instead of the topo-clustering in the ATLAS experiment [134, 136, 137]. The sliding-window algorithm uses a fixed-size window to define a cluster. Topo-clustering is superior to sliding-window in that dynamic clusters change in size as needed to recover energy from bremsstrahlung photons or electrons from photon conversions.

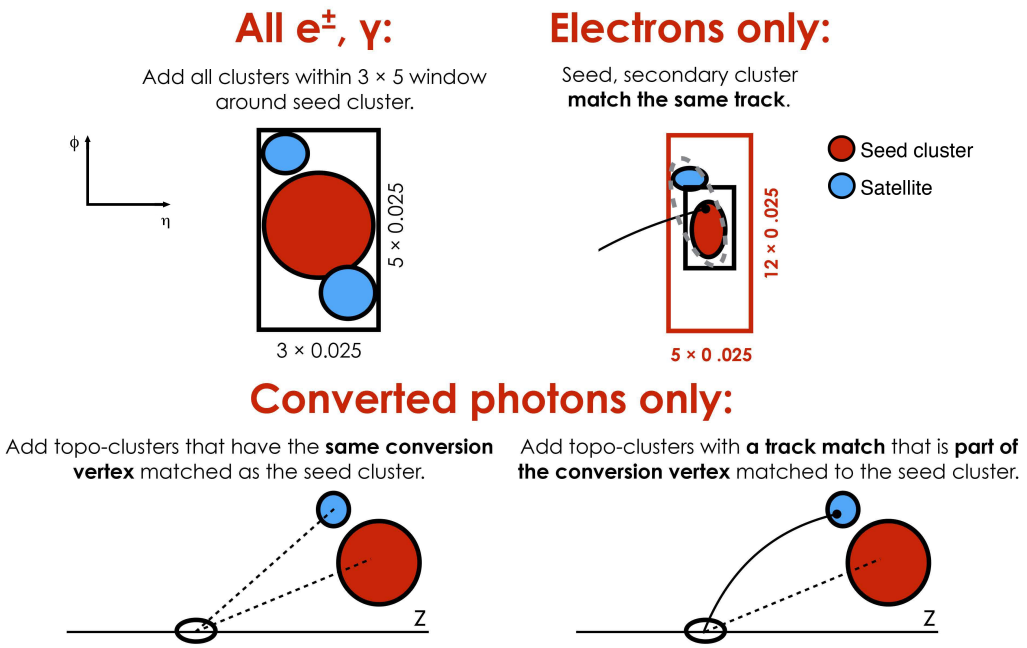


Figure 3.7: Superclustering algorithm for photons and electrons [133]. Seed clusters are shown in red, and satellite clusters in blue.

of the latter. The optimization is performed in sub-ranges of photon η and E_{T} using TMVA [138]. Using the p_{T}-dependent tight identification cuts, photons with $p_{\mathrm{T}} > 25$ GeV are expected to have a reconstruction and identification efficiency of greater than 80% that reaches a plateau of about 90-95% at 40-50 GeV.

Photon candidates are also required to be isolated to prevent faking from jets further, using criteria based on the calorimeter as implemented in the TightCaloOnly photon isolation working point defined in Ref. [133]. Other working points based on both the inner detector and calorimeter are used in the usual analyses. Any of such working points assume that the reconstruction of the primary vertex is successful and imposes conditions using the variables of the primary vertex. This research does not use such a working point because the primary vertex of the EL signal is hardly reconstructed correctly, as described in Sec. 3.3. The isolation variable is defined as the sum of E_{T} of the topo-clusters with positive energy reconstructed in the calorimeter around each photon candidate in a cone of radius $\Delta R = 0.4$. This energy calculation requires subtracting the contributions from the photon itself in the core window $\Delta\eta \times \Delta\phi = 0.125 \times 0.175$, correcting for the leakage of the photon energy from the core window, and finally using an event-by-event energy subtraction based on the jet area method to remove contributions from the underlying event and pileup interactions [139]. The total calorimeter isolation energy is required to be smaller than $0.022E_{\mathrm{T}} + 2.45$ GeV.

The photon object selection requires the photon candidates with $p_{\mathrm{T}} > 40$ GeV and $|\eta| < 2.37$, excluding the barrel-to-endcap transition regions of the calorimeter, $1.37 < |\eta| < 1.52$. An event preselection requires at least two photons fulfilling all the criteria described above.

3.5 Proton reconstruction

Protons are reconstructed from their associated AFP tracks, and the tracking begins with clustering. A cluster is defined from neighboring pixel hits in the x-direction in a 3D-Si sensor of the AFP detector. A usual cluster has one or two pixel hits, while a cluster sometimes has more than three pixel hits due

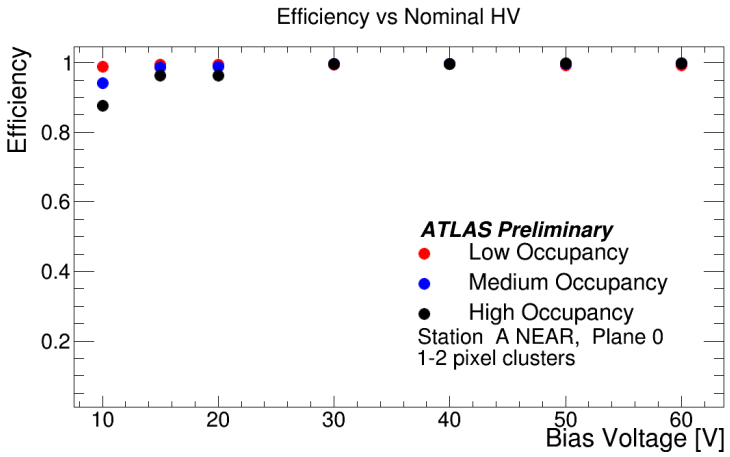


Figure 3.8: Clustering efficiency as a function of bias voltage [104]. Each color corresponds to an occupancy level.

to showering. The cluster (x, y) position is determined from the charge-weighted distribution of the component hits. A track is then defined by searching for a set of clusters in the same station aligned parallel to the beam-line. Deviations of the transverse cluster positions from one another within 500 µm are allowed. The track (x, y) position is determined by a linear regression fit, assuming the same uncertainties on each cluster. Clusters from at least two different planes are required in this research. Each station was found to have about 0.02 reconstructed tracks on average per pp collision.

The clustering efficiency for each SiT plane is measured using clusters in the other planes in the same station sharing the same track. The clustering rate was calculated within a window of one pixel. The clustering efficiency depends on the applied bias voltage in the 3D-Si sensors. Fig. 3.8 shows the dependency in the first plane of the A-side FAR station. The data are divided into three occupancy levels using the rate of traversing tracks as a proxy. Deterioration over time was observed during the data-taking due to the ageing effects originating from radiation damage. Therefore, the bias voltage was tuned after this observation, and the efficiency recovered. However, a SiT plane in the A-side FAR station was always non-operational during the data-taking in 2017 due to a problem with the electrical connections on the high-voltage (HV) lines.

Finally, a proton is reconstructed, searching for a pair of NEAR and FAR station tracks within 2 mm of transverse distance. If multiple such pairs are in the same arm, all possibilities are taken forward. The kinematics of the proton with 4-momentum (E, p_x, p_y, p_z) is reconstructed by mapping the track positions in the two stations, x_{NEAR} and x_{FAR}, to E and $\theta_{x,0} = p_x/p_z$ assuming the positions of the SiTs and the magnetic fields of the LHC magnet lattice are known. The E and $\theta_{x,0}$ are obtained by solving the equations

$$x \equiv \frac{x_{\mathrm{NEAR}} + x_{\mathrm{FAR}}}{2} = c_{1,x}(E) + c_{2,x}(E)\theta_{x,0},$$
$$\theta_x \equiv \frac{x_{\mathrm{FAR}} - x_{\mathrm{NEAR}}}{z_{\mathrm{FAR}} - z_{\mathrm{NEAR}}} = c'_{1,x}(E) + c'_{2,x}(E)\theta_{x,0},$$

where x_{NEAR} (x_{FAR}) and z_{NEAR} (z_{FAR}) is the x and z-coordinate of the track in the NEAR (FAR) station, respectively, and $c^{(\prime)}_{i,x}$ are polynomials with respect to E whose coefficients are determined by a proton transport simulation [140]. Additional terms,

$$c^{(\prime)}_{3,x}(E)x_0 + c^{(\prime)}_{4,x}(E)z_0 + c^{(\prime)}_{5,x}(E)z_0\theta_{x,0},$$

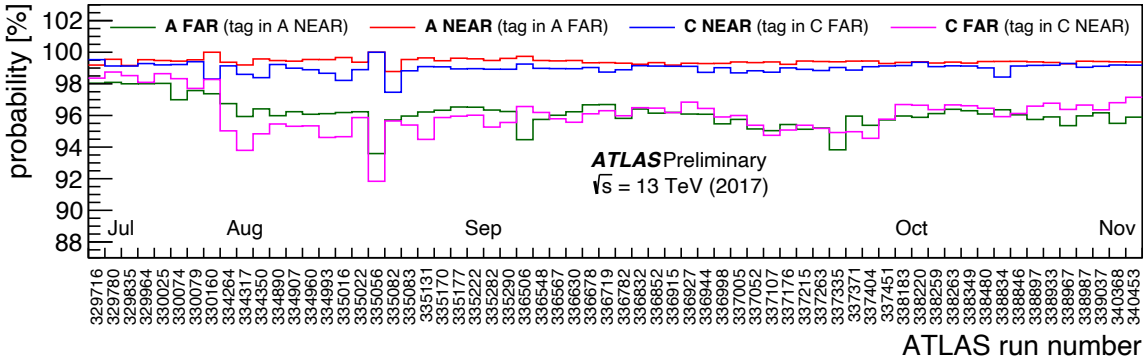


Figure 3.9: Tracking efficiency for each probe station for each run in the proton reconstruction efficiency measurement [94].

where (x_0, y_0, z_0) is the vertex position, were considered originally, but they are found to be negligible. The solved E is interpreted as the proton energy loss fraction,

$$\xi_{\mathrm{AFP}} = 1 - \frac{E}{E_{\mathrm{beam}}},$$

where $E_{\mathrm{beam}} = 6.5$ TeV is the initial beam proton energy, and $\theta_{x,0}$ is interpreted to p_x. p_y is not calculated in this research but is expected to be calculated in the future. The uncertainty of the assumed magnetic fields in the beam optics in the transport simulation has the dominant effect on proton ξ_{AFP} uncertainty, leading to about 10% for most ξ_{AFP} values.

Proton reconstruction efficiency was measured using a tag-and-probe method. It begins with the tracking efficiency measurement for each station using events where there is exactly one reconstructed track (tag) in one station of an arm and calculating the probability that a proton is reconstructed along with the existence of a reconstructed track (probe) in the other station in the same arm. The distance of x-position of the track in the probe station from the one in the tag station is required to be less than 2 mm as per the nominal analysis. Fig. 3.9 shows the tracking efficiency for each probe station for each run in 2017. The slight difference of the Far stations behaviours in the earlier runs is due to the larger distance between the RP and the beam, causing fewer showering. The Near stations have 1% detection inefficiency, while the Far stations have 5% inefficiency, where the additional 4% is due to showering. When reconstructing a proton using the Near and Far stations, assuming the tracking efficiency of Near and Far stations are assumed to be uncorrelated, the maximum contribution of the detection inefficiency is $1\% \times 2 = 2\%$, and the contribution of showering is from 4% to $4\% \times 2 = 8\%$, i.e. $6 \pm 2\%$. Thus, the total proton reconstruction efficiency was then determined as the $100 - (2 + 6 \pm 2) = 92 \pm 2\%$. In general, multiple protons are reconstructed for each side per bunch crossing. Though the typical number of reconstructed protons ranges from 0 to 2, there are sometimes more than 10 per BC. The primary cause is the showering and coincident hits from multiple pileup events.

The AFP acceptance is studied based on the proton reconstruction efficiency. The proton efficiency of $92 \pm 2\%$ is guaranteed only for $x < -3.5$ mm in the AFP local coordinate and $\xi_{\mathrm{AFP}} \in [0.035, 0.08]$. The Far stations still have high efficiency in the range $\xi_{\mathrm{AFP}} \in [0.02, 0.035]$, but the Near station does not cover the range, and hence the efficiency cannot be measured by the tag-and-probe method. Although the efficiency is expected to be still high for both stations for $\xi_{\mathrm{AFP}} \in [0.08, 0.12]$,[6] there were a small number of events in the range, and the efficiency could not be derived sufficiently. As object selection in this research, only the protons with an Near station seed track with $x < -3.5$ mm is selected.[7] Table 3.2

[6] The existence of collimator determines the 0.12.

[7] Selection of $\xi_{\mathrm{AFP}} \in [0.035, 0.08]$ is applied in the event selection as described in Sec. 4.3.

Table 3.2: Photon and proton selections and the event preselection using the photons.

	Selection
Photon	Identification (Tight) Isolation (TightCaloOnly) $p_{\mathrm{T}} > 40$ GeV $\lvert\eta\rvert \in [0, 2.37] \setminus [1.37, 1.52]$
Proton	$x_{\mathrm{NEAR}} < -3.5$ mm
Event	At least two selected photons

summarizes the photon and proton selections and the event selection using the photons. Cut on ξ_{AFP} is imposed in the event selection described in Sec. 4.

The data from AFP ToF detectors are not used in this research as described in Sec. 2.3. This is due to very low efficiency. The cause has not been entirely determined, but there are two main possibilities.

- The quantum efficiency decreased due to the material degradation of PMT. Since high HV was applied at the commissioning time, the quantum efficiency dropped at that point.
- The HV was lowered too much. Since a large number of photoelectrons were observed during the commissioning, the HV was adjusted to prevent a decrease in quantum efficiency. However, there was no online monitoring of efficiency in Run 2, so it was unclear how much to lower the HV to keep sufficient efficiency.

4 Event Selection

This section describes event selection, introduced in Sec. 1.4, in detail. The selection flow is summarized in Fig. 4.1. The acoplanarity selection and $\xi_{\gamma\gamma}$ selection is the requirement for diphoton kinematics. The ξ_{AFP} is the requirement for proton kinematics. The *matching* is the requirement that the $\xi_{\gamma\gamma}$ and ξ_{AFP} are close. The selection is applied for each side from the $\xi_{\gamma\gamma}$ selection. Thus, there are two sets of events after the matching selection. A set operation finally selects the events from the two samples.

4.1 Acoplanarity selection

First, the diphoton is constructed from the leading and sub-leading photons described in Sec. 3.4. The acoplanarity of the diphoton is defined as

$$A_{\phi}^{\gamma\gamma} \equiv 1 - \frac{|\Delta\phi_{\gamma\gamma}|}{\pi}, \tag{4.1}$$

where $\Delta\phi_{\gamma\gamma} \in (-\pi, \pi]$ is the difference of ϕ between the two photons. The acoplanarity is interpreted as the degree of the diphoton not to be back-to-back in the x-y plane. Figs. 4.2 and 4.3 show the acoplanarity distributions for the data and ALP signals after the whole preselection described in Sec. 3. The EL signals have exclusive final state diphoton, and the momentum conservation ensures that the two photons have the same p_{T} and the opposite direction.[1] On the other hand, the SD and DD signals are not back-to-back because the proton dissociations produce other objects in the final state. Still, all the signal samples have a peak around 0, as well as the data sample. Only the event with diphoton acoplanarity less than 0.01 is selected.

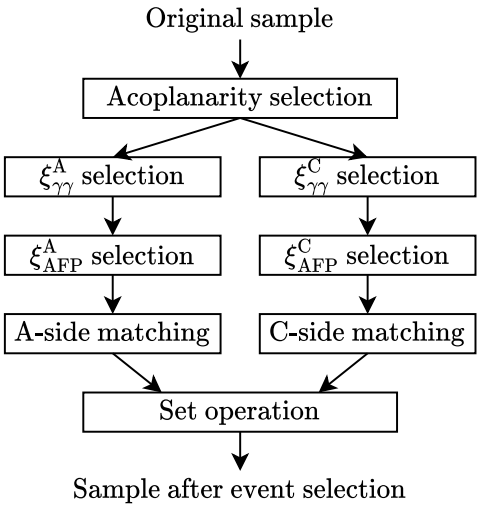


Figure 4.1: Event selection flow.

[1] The relative p_{T} difference distributions for the data and signal MC samples are shown in Figs. A.1 and A.2. The p_{T} balance characteristic could be used for an additional event selection. However, it is not used in this analysis because too much background reduction makes the background estimation difficult.

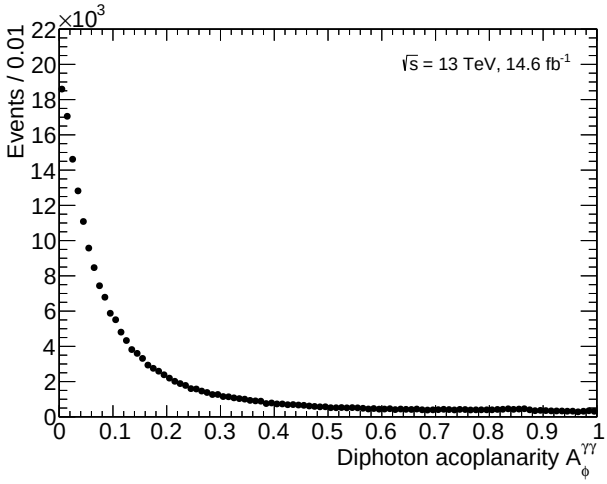
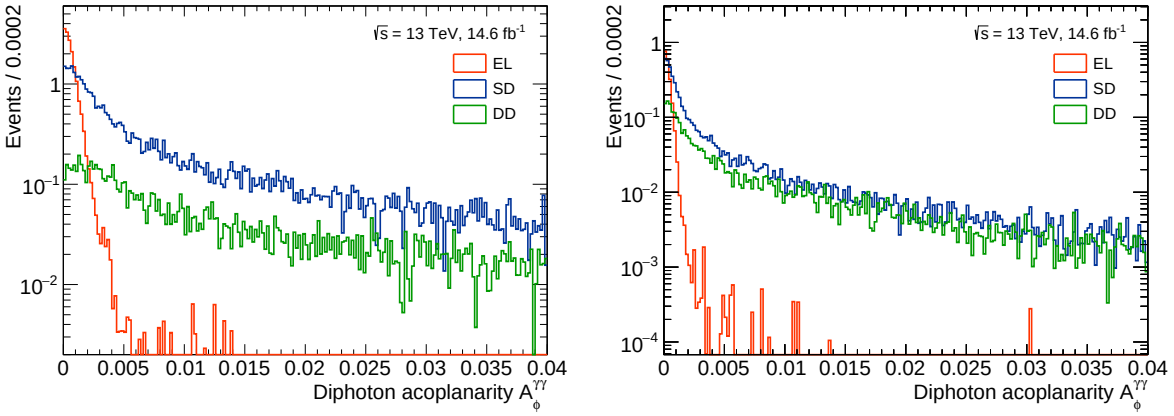


Figure 4.2: Diphoton acoplanarity distribution of the data after the preselection.



(a) $m_X = 300$ GeV (b) $m_X = 1200$ GeV

Figure 4.3: Diphoton acoplanarity distribution of the ALP signals after the preselection. The different colors stand for the EL, SD, and SD signals.

4.2 $\xi_{\gamma\gamma}$ selection

Through the conservation of 4-momentum for a given beam center-of-mass energy $\sqrt{s}$, the $\xi^{\pm}$ can be determined from the photon system kinematics, where $\xi^{+} \equiv \xi^{\mathrm{A}}$ and $\xi^{-} \equiv \xi^{\mathrm{C}}$ are the energy loss fraction of proton flying towards A-side and C-side, respectively. From the definition of the proton energy loss fraction, the energy of initial photons is $\xi^{\pm}\sqrt{s}/2$. Assuming the initial photons move almost along the beam direction, the 4-momenta of the initial photons are

$$\left(\xi^{\pm}\frac{\sqrt{s}}{2}, 0, 0, \pm\xi^{\pm}\frac{\sqrt{s}}{2}\right),$$

where the first element is energy, and the last is longitudinal momentum. The ALP 4-momenta is the sum of them,

$$\frac{\sqrt{s}}{2}\left(\xi^{+}+\xi^{-}, 0, 0, \xi^{+}-\xi^{-}\right) = \left(E_{\gamma\gamma}, p_x^{\gamma\gamma}, p_y^{\gamma\gamma}, p_z^{\gamma\gamma}\right).$$

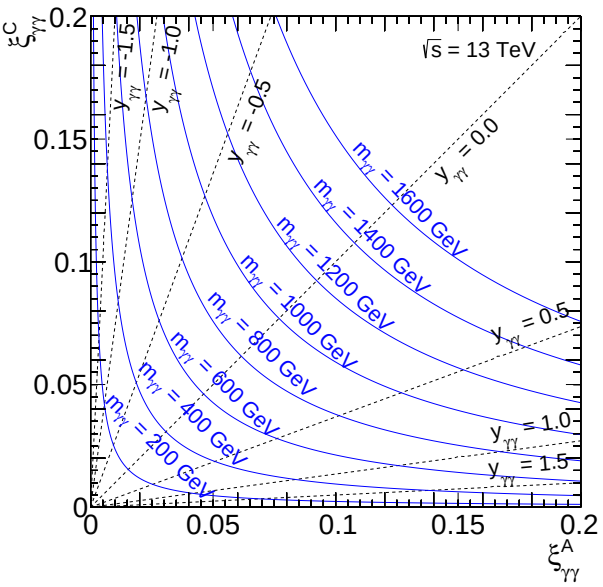


Figure 4.4: The correspondence between $(\xi^{\mathrm{A}}_{\gamma\gamma}, \xi^{\mathrm{C}}_{\gamma\gamma})$ and $(m_{\gamma\gamma}, y_{\gamma\gamma})$ coordinates.

As shown on the right-hand side, this is equivalent to the 4-momenta of the final state diphoton system. The diphoton mass is

$$m_{\gamma\gamma} = \sqrt{E^2_{\gamma\gamma} - (p^{\gamma\gamma}_x)^2 - (p^{\gamma\gamma}_y)^2 - (p^{\gamma\gamma}_z)^2} = \sqrt{s\xi^+\xi^-},$$

thus,

$$\frac{m_{\gamma\gamma}}{\sqrt{s}} = \sqrt{\xi^+\xi^-}. \tag{4.2}$$

The diphoton rapidity is

$$y_{\gamma\gamma} = \frac{1}{2}\log\frac{E_{\gamma\gamma} + p^{\gamma\gamma}_z}{E_{\gamma\gamma} - p^{\gamma\gamma}_z} = \frac{1}{2}\log\frac{\xi^+}{\xi^-},$$

thus,

$$e^{y_{\gamma\gamma}} = \sqrt{\frac{\xi^+}{\xi^-}}, \quad e^{-y_{\gamma\gamma}} = \sqrt{\frac{\xi^-}{\xi^+}}.$$

Therefore,

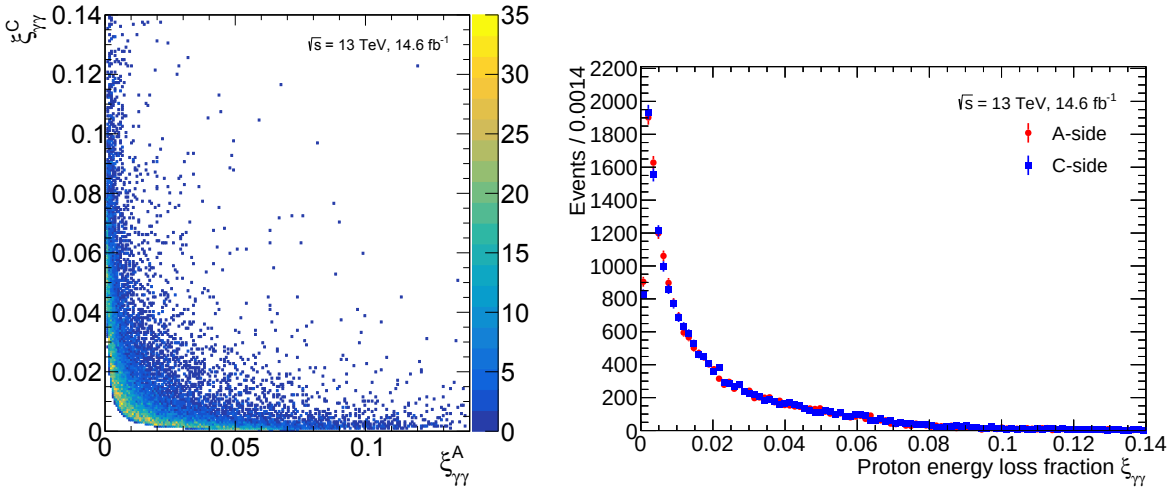
$$\xi^\pm = \frac{m_{\gamma\gamma}}{\sqrt{s}} e^{\pm y_{\gamma\gamma}}. \tag{4.3}$$

This is denoted by $\xi^\pm_{\gamma\gamma}$ or $\xi^{\mathrm{A(C)}}_{\gamma\gamma}$. The assumption on the above calculations is not always good approximations. Especially, SD or DD events have non-zero $p^{\gamma\gamma}_x$ or $p^{\gamma\gamma}_y$ in general. Still, this definition of the $\xi_{\gamma\gamma}$ is used for the event selection uniformly. The correspondence of $(\xi^{\mathrm{A}}_{\gamma\gamma}, \xi^{\mathrm{C}}_{\gamma\gamma})$ and $(m_{\gamma\gamma}, y_{\gamma\gamma})$ coordinates is shown in Fig. 4.4. The signal events for ALP mass m_X tend to concentrate on a mass contour $m_{\gamma\gamma} = m_X$.

Figs. 4.5 and 4.6 show the $\xi^\pm_{\gamma\gamma}$ distributions for the data and ALP signals after the acoplanarity selection. The data are distributed in the low $m_{\gamma\gamma}$ region. In contrast, high-mass signal events are distributed in relatively high ξ regions as expected by Fig. 4.4.

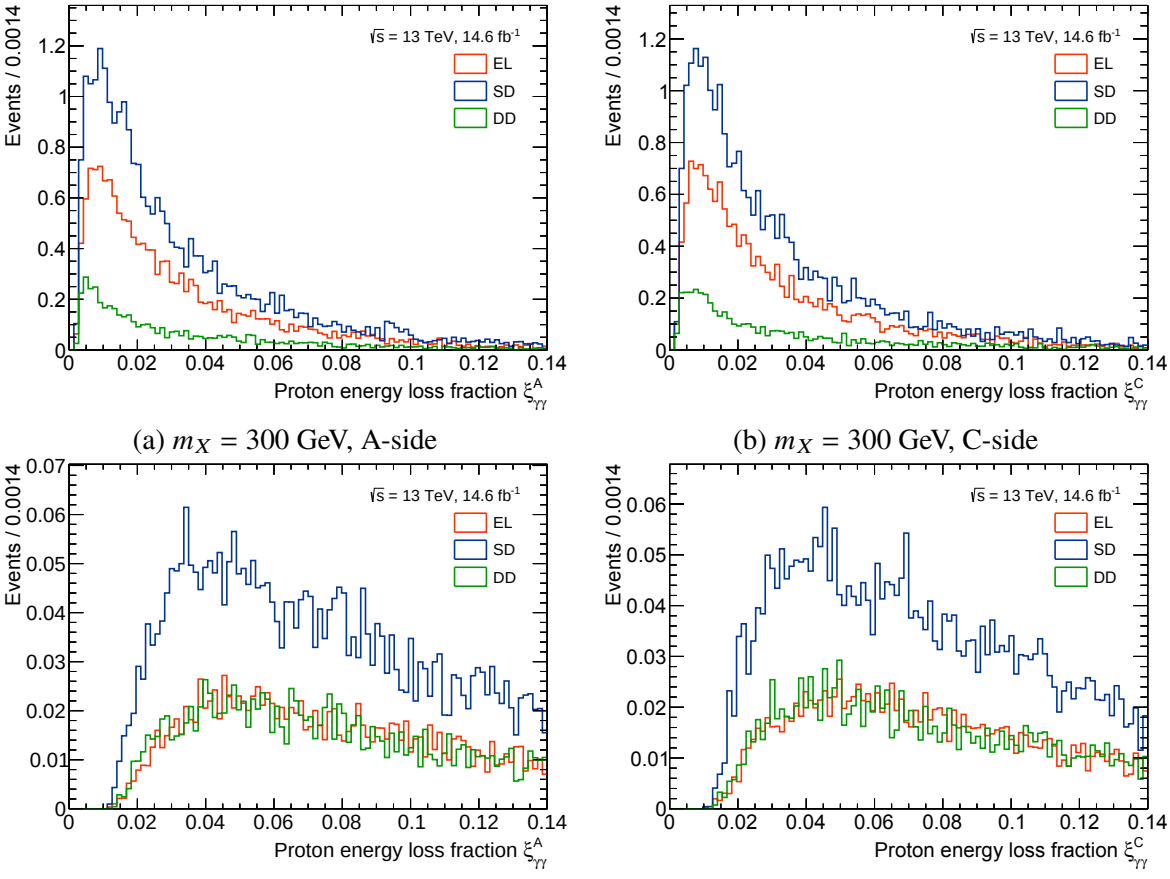
The event selection is

$$\xi^\pm_{\gamma\gamma} \in \left[\frac{0.031}{1.1}, \frac{0.084}{0.9}\right].$$



(a) $\xi^{+}_{\gamma\gamma}$ vs $\xi^{-}_{\gamma\gamma}$ (b) $\xi^{\pm}_{\gamma\gamma}$

Figure 4.5: $\xi^{\pm}_{\gamma\gamma}$ distributions of the data after the acoplanarity selection.



(a) $m_X = 300$ GeV, A-side (b) $m_X = 300$ GeV, C-side

(c) $m_X = 1200$ GeV, A-side (d) $m_X = 1200$ GeV, C-side

Figure 4.6: $\xi^{\pm}_{\gamma\gamma}$ distributions of the signals after the acoplanarity selection. The EL, SD, and DD signal distributions are illustrated separately.

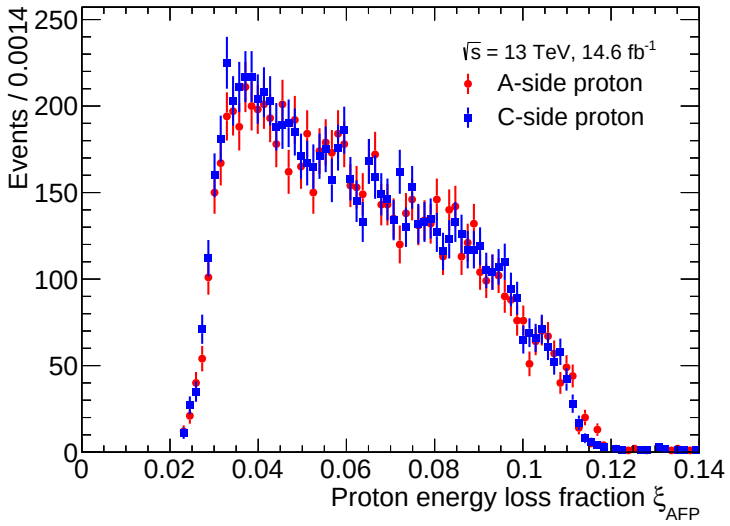


Figure 4.7: $\xi^{\pm}_{\mathrm{AFP}}$ distributions of the data after the acoplanarity selection. Each distribution has only entries for events in which reconstructed protons are present on the corresponding arm. The shape is almost the same as before the acoplanarity selection because the dominant combinatorial background is not correlated to the diphoton system as described in Sec. 6.

This selection is applied for each side, and the condition is equivalent to

$$0.035 - \left(0.004 + 0.1\xi^{\pm}_{\gamma\gamma}\right) \leq \xi^{\pm}_{\gamma\gamma} \leq 0.08 + \left(0.004 + 0.1\xi^{\pm}_{\gamma\gamma}\right),$$

where 0.035 and 0.08 are the ξ_{AFP} selection thresholds as described in Sec. 4.3, and the $0.004 + 0.1\xi^{\pm}_{\gamma\gamma}$ is the maximum allowed absolute difference between $\xi^{\pm}_{\gamma\gamma}$ and $\xi^{\pm}_{\mathrm{AFP}}$ in the matching selection described in Sec. 4.4, which is introduced to improve the signal and background efficiency.[2]

4.3 ξ_{AFP} selection

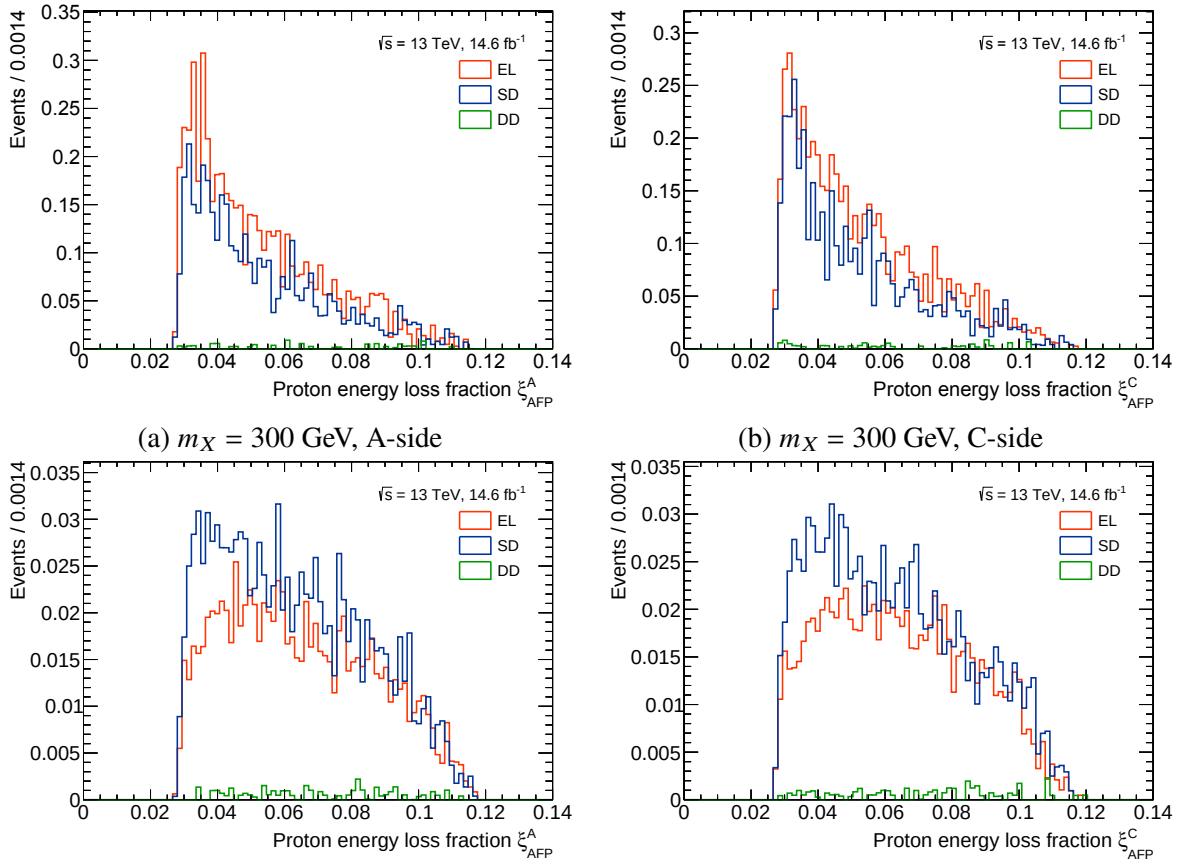
From this stage, protons are used for the event selection. First, for each side, if there are multiple reconstructed protons, a unique proton with energy loss fraction ξ_{AFP} closest to $\xi_{\gamma\gamma}$ is selected.[3] Mapping from $(\xi^{\mathrm{A}}_{\mathrm{AFP}}, \xi^{\mathrm{C}}_{\mathrm{AFP}})$ to $(m_{\gamma\gamma}, y_{\gamma\gamma})$ is only possible when protons are reconstructed on both sides.

Figs. 4.7 and 4.8 show the $\xi^{\pm}_{\mathrm{AFP}}$ distributions for the data and ALP signals after the acoplanarity selection. In contrast to the $\xi_{\gamma\gamma}$ distributions in Sec. 4.2, the distributions have entries only in the acceptable region of the AFP detectors. High-mass signal events are distributed in relatively high ξ region as expected by Fig. 4.4. The contribution of the DD signals is minimal because each distribution has only entries for events in which reconstructed protons are present on the corresponding arm, while the DD signals have no intact protons in their final states. The small existing DD contribution is due to the events that the dissociative proton systems are hadronized back to protons and coincidentally enter the AFP detectors.

As the event selection, the $\xi^{\pm}_{\mathrm{AFP}}$ of the selected proton is required to be in the range $[0.035, 0.08]$ in which the proton reconstruction efficiency is evaluated with uncertainty as described in Sec 3.5. This selection is applied for each side.

[2] Too small background efficiency would make background modeling difficult.

[3] This should not be done at the object selection stage in Sec. 3.5 because the background sample creation described in Sec. 6 would be biased. In the signal MC samples, an efficiency of $92 \pm 2\%$ that may be wrong is applied to a proton with $\xi_{\mathrm{AFP}} \notin [0.035, 0.08]$ (see Sec. 3.5), but events with such a proton closest to the diphoton in terms of ξ are discarded by the ξ_{AFP} selection for the side of interest.



(a) $m_X = 300$ GeV, A-side

(b) $m_X = 300$ GeV, C-side

(c) $m_X = 1200$ GeV, A-side

(d) $m_X = 1200$ GeV, C-side

Figure 4.8: $\xi^{\pm}_{\mathrm{AFP}}$ distributions of the signals after the acoplanarity selection. The EL, SD, and DD signal distributions are illustrated separately. Each distribution has only entries for events in which reconstructed protons are present on the corresponding arm.

4.4 Diphoton-proton matching

Diphoton-proton matching is then applied by comparing the $\xi_{\gamma\gamma}$ value with the ξ_{AFP} for each side as described in Sec. 1.4. The differences

$$\Delta\xi^{+} = \Delta\xi^{\mathrm{A}} \equiv \xi^{+}_{\mathrm{AFP}} - \xi^{+}_{\gamma\gamma},$$
$$\Delta\xi^{-} = \Delta\xi^{\mathrm{C}} \equiv \xi^{-}_{\mathrm{AFP}} - \xi^{-}_{\gamma\gamma}$$

are used for the selection. Figs. 4.9 and 4.10 show the $\Delta\xi^{\pm}$ distributions of the data and signal MC samples after the acoplanarity selecion. From the widths of the signal distributions, the bare selection condition is determined as $|\Delta\xi^{\pm}| < 0.004$. However, if the ξ_{AFP} relative uncertainty of about 10% described in Sec. 3.5 is considered, this threshold is found to be unacceptably tight. The dominant source of the 10% is the uncertainty of beam optics assumed in the proton reconstruction. Fig. 4.11 shows the $\Delta\xi^{\pm}$ distributions of the signal events where the beam optics configuration in the simulation is redefined as the nominal value added or subtracted by its uncertainty. The distributions represent the nominal case, upper-shifted case, and lower-shifted case. Fig. 4.12 shows their dependency on $\xi_{\gamma\gamma}$. The bare condition fails more than 80% events for the systematically shifted samples. Some $\Delta\xi^{\pm}$ plots for other systematic variations are in Appendix C.

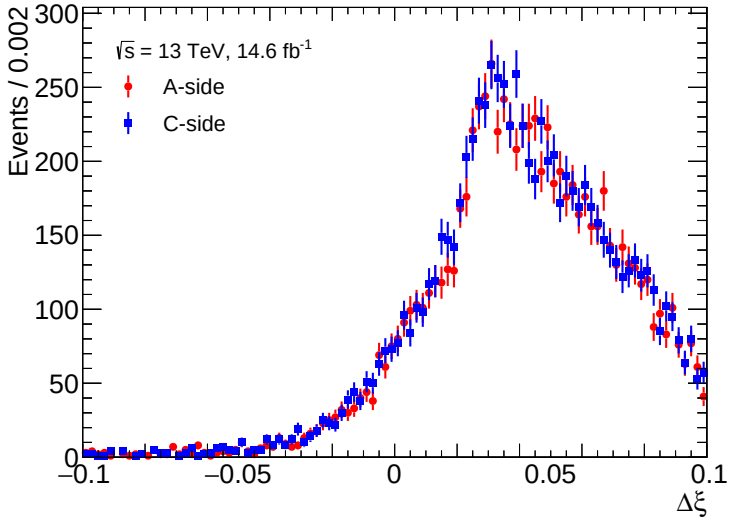
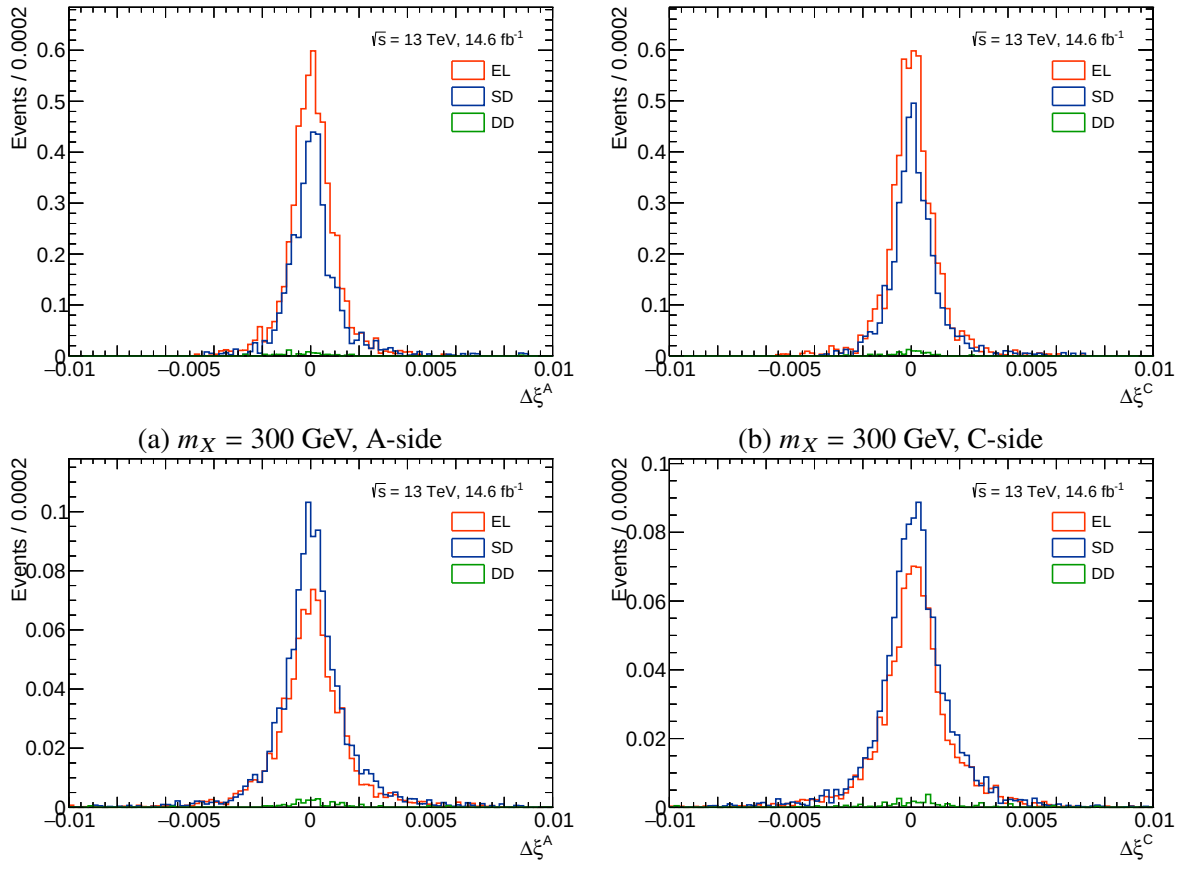


Figure 4.9: $\Delta\xi^{\pm}$ distributions of the data after the acoplanarity selection. Each distribution has only entries for events in which reconstructed protons are present on the corresponding arm.



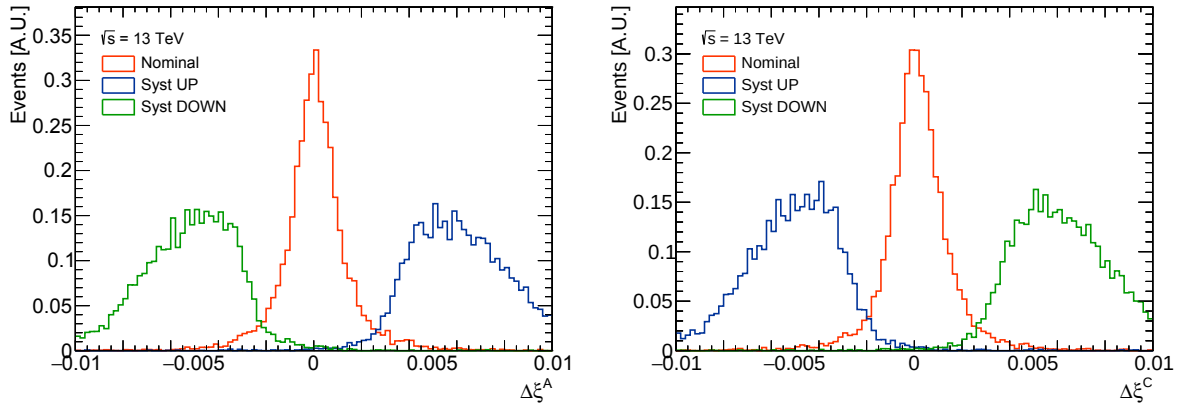
(a) m_X = 300 GeV, A-side

(b) m_X = 300 GeV, C-side

(c) m_X = 1200 GeV, A-side

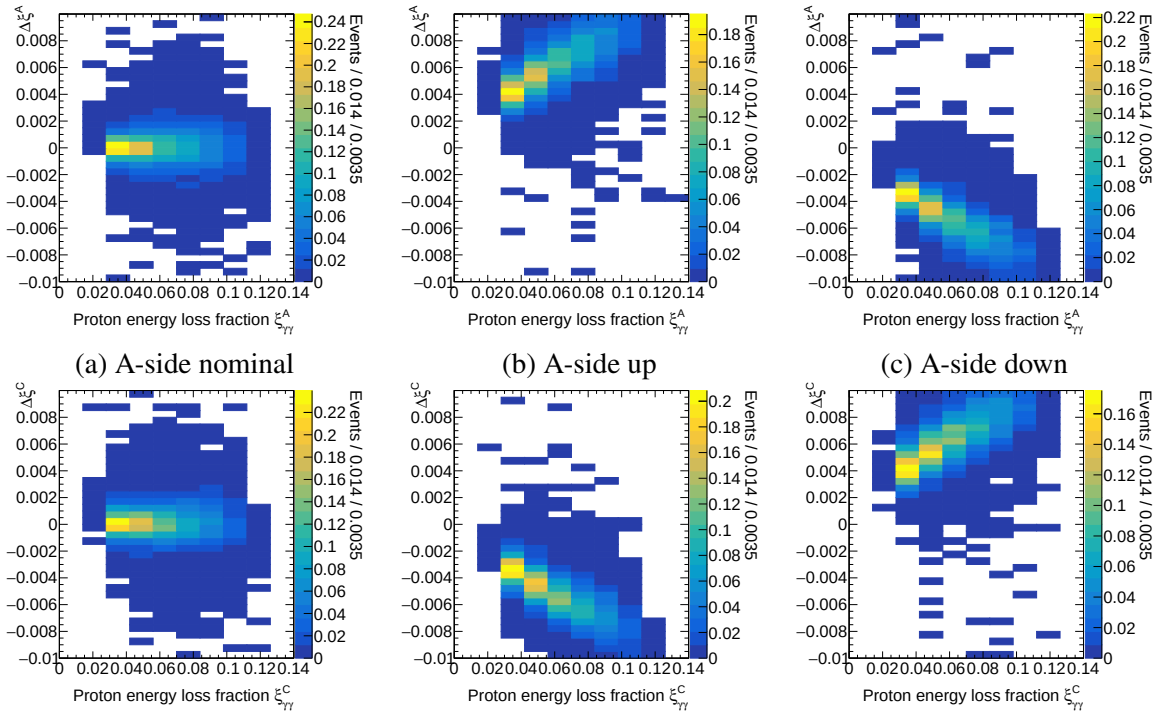
(d) m_X = 1200 GeV, C-side

Figure 4.10: $\Delta\xi^{\pm}$ distributions of the signals after the acoplanarity selection. The EL, SD, and DD signal distributions are illustrated separately. Each distribution has only entries for events in which reconstructed protons are present on the corresponding arm.



(a) A-side (b) C-side

Figure 4.11: $\Delta\xi$ distributions of the $m_X = 1000$ GeV simulated signal obtained with nominal and systematically varied beam optics configuration. The distributions are normalized by the integrals.



(a) A-side nominal (b) A-side up (c) A-side down

(d) C-side nominal (e) C-side up (f) C-side down

Figure 4.12: $\Delta\xi$ vs $\xi_{\gamma\gamma}$ 2D distributions of the $m_X = 1000$ GeV simulated signal obtained with nominal and systematically varied beam optics configuration.

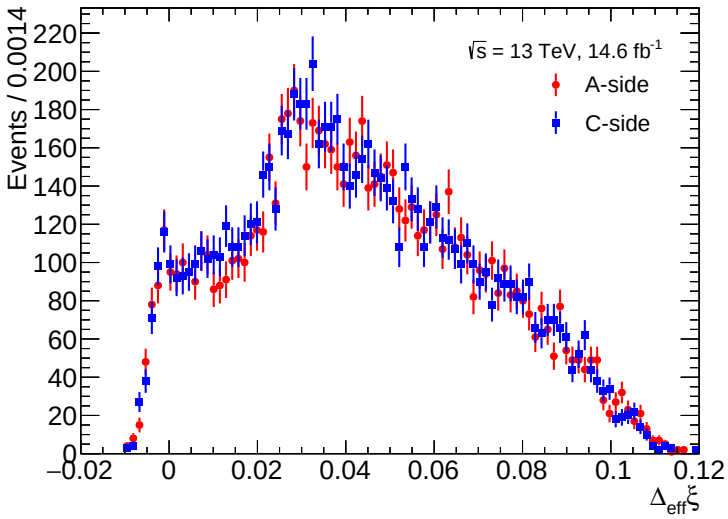


Figure 4.13: $\Delta_{\text{eff}}\xi^{\pm}$ distributions of the data after the acoplanarity selection. Each distribution has only entries for events in which reconstructed protons are present on the corresponding arm.

Thus, the matching condition is modified as

$$|\Delta\xi^{\pm}| < 0.004 + 0.1\xi^{\pm}_{\gamma\gamma}$$

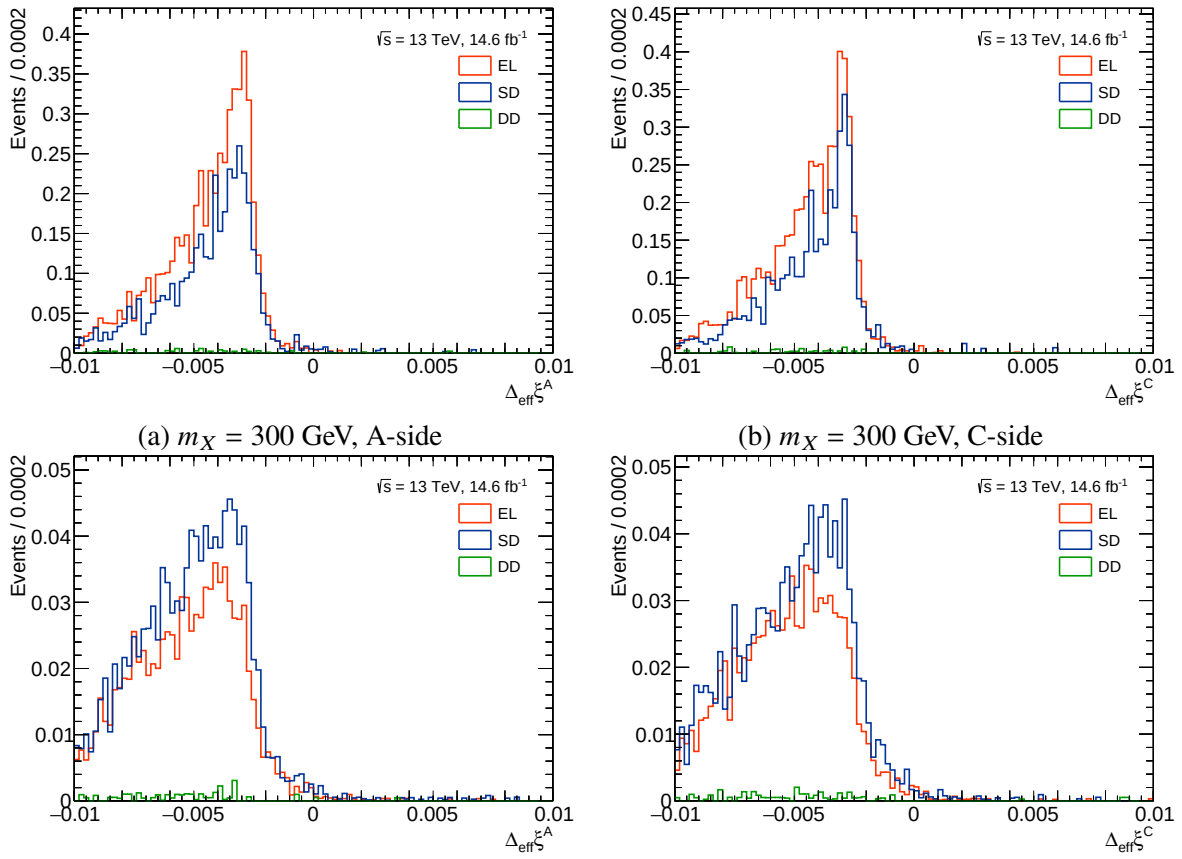
with the *correction* term to consider the 10% uncertainty. This condition is rewritten as

$$\Delta_{\text{eff}}\xi^{\pm} \equiv |\Delta\xi^{\pm}| - 0.1\xi^{\pm}_{\gamma\gamma} < 0.004$$

to make the threshold constant. Figs. 4.13 and 4.14 show the $\Delta_{\text{eff}}\xi$ distributions for the data and the ALP signal MC samples, respectively. This modified condition can cause the selection of events whose systematic uncertainty is not evaluated enough, but the effects of such events are evaluated in Appendix D and found to be small. The diphoton mass distribution of the data after the matching selection had been blinded to prevent potential psychological bias until the statistical modeling and the evaluation of systematic uncertainty were fixed. This is called the blinding strategy and is widely adopted in search experiments.

4.5 Set operation

There are two sets of events corresponding to the A-side and C-side per initial sample after the event selection up to the matching, because the $\xi_{\gamma\gamma}$ selection, ξ_{AFP} selection, and the matching are performed side-by-side as shown in Fig. 4.1. In this research, their union is taken as the sample after the whole event selection. In contrast, the CMS-TOTEM PPS analysis takes the intersection of them [74]. The former is referred to as the OR selection, while the latter is referred to as the AND selection, as described in Sec. 1.4. One could use another set operation between the two samples, such as the XOR or NOR conditions to combine the results with one another. Still, this research only uses the OR selection for simplicity. Fig. 4.1 is recreated as Fig. 4.15 with the quantitative information about the selection condition. Fig. 4.16 show the cut flow of the data for each of the A-side selection, C-side selection, AND selection, and OR selection, where the data after the A(C)-side selection is the sample after the side-by-side selection which would have been input to the set operation. No data event is observed after the AND matching selection. 441 data events remain after the whole selection with the OR condition and requiring $m_{\gamma\gamma}$ to be in the search range $[150, 1600]$ GeV. Table 4.1 quantifies the cut flow for the OR condition. The signal selection cut flow for the OR selection is summarized in Tables 4.2 and 4.3.



(a) m_X = 300 GeV, A-side

(b) m_X = 300 GeV, C-side

(c) m_X = 1200 GeV, A-side

(d) m_X = 1200 GeV, C-side

Figure 4.14: $\Delta_{\text{eff}}\xi^{\pm}$ distributions of the signals after the acoplanarity selection. The EL, SD, and DD signal distributions are illustrated separately. Each distribution has only entries for events in which reconstructed protons are present on the corresponding arm.

Table 4.1: The cut flow of the data for 14.6 fb^{-1} of pp collision at $\sqrt{s}$ = 13 TeV. The values for $\xi_{\gamma\gamma}$ and ξ_{AFP} are the numbers of events when the set operation is performed without the latter selection.

Cut variable	**Events**	**Efficiency**
Preselection	201,723	
$A_{\phi}^{\gamma\gamma}$	18,597	9%
$\xi_{\gamma\gamma}$	8,626	46%
ξ_{AFP}	2,624	14%
$\Delta\xi_{\text{eff}}$	936	36%
$m_{\gamma\gamma} \in [150, 1600]$ GeV	441	

The mass dependency of the selection efficiency is determined mainly at the $\xi_{\gamma\gamma}$ selection stage. Fig. 4.17 (a) shows the estimated efficiency of this selection on the data events for each $m_{\gamma\gamma}$ for the OR selection and the other conditions. A toy MC sample is made by generating $m_{\gamma\gamma}$ and $y_{\gamma\gamma}$ values uniformly, weighting them by the data-driven $y_{\gamma\gamma}$ distribution as Fig. 4.18, and the $\xi_{\gamma\gamma}$ selection is applied. It is compared with (b), the correspondence between $(\xi_{\gamma\gamma}^{\text{A}}, \xi_{\gamma\gamma}^{\text{C}})$ and $(m_{\gamma\gamma}, y_{\gamma\gamma})$ coordinates with the yellow OR selection range. The OR selection efficiency corresponds to the length of the diphoton mass contour line that overlaps the yellow region, weighted by the diphoton rapidity distribution obtained from data assuming almost all data events are background. Sec. 4.5.1 models the selection efficiency of the background. The signal selection efficiency is modelled in Sec. 5.2.

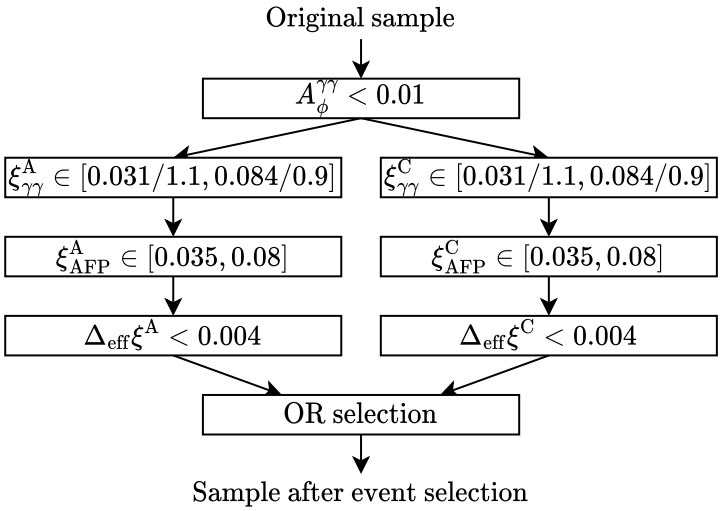


Figure 4.15: Event selection flow with quantitative selection criteria.

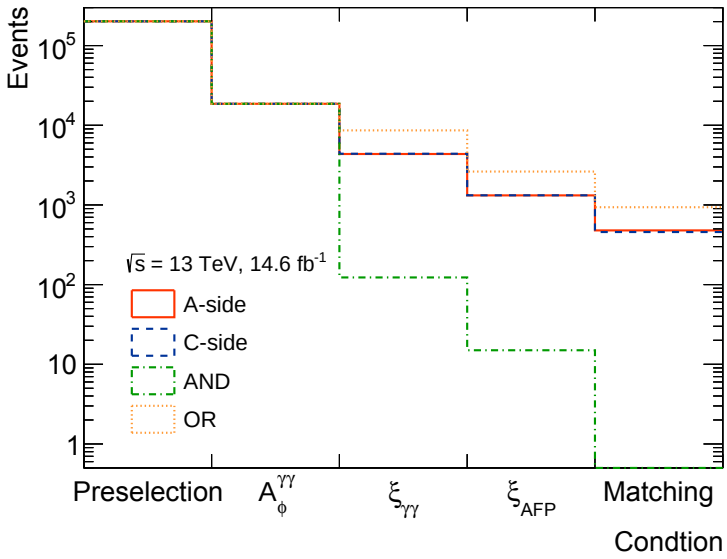


Figure 4.16: The cut flow of the data after the preselection. The values for $\xi_{\gamma\gamma}$ and ξ_{AFP} are the numbers of events when the set operation is performed without the latter selection.

Table 4.2: Signal MC sample cut flow for ALP mass from 200 GeV to 650 GeV for 14.6 fb^{-1} of pp collision at $\sqrt{s}$ = 13 TeV. The values for $\xi_{\gamma\gamma}$ and ξ_{AFP} are the numbers of events when the set operation is performed without the latter selection.

200 GeV								
Cut variable	EL		SD		DD		Total	
	Events	Efficiency	Events	Efficiency	Events	Efficiency	Events	Efficiency
Preselection	21.49		67.69		16.64		105.82	
$A_{\phi}^{\gamma\gamma}$	21.47	100%	31.65	47%	4.70	28%	57.82	55%
$\xi_{\gamma\gamma}$	11.55	54%	16.69	53%	2.60	55%	30.84	53%
ξ_{AFP}	6.98	60%	4.53	27%	0.10	4%	11.61	38%
$\Delta\xi_{\mathrm{eff}}$	6.98	100%	4.50	99%	0.10	96%	11.58	100%

300 GeV								
Cut variable	EL		SD		DD		Total	
	Events	Efficiency	Events	Efficiency	Events	Efficiency	Events	Efficiency
Preselection	16.54		48.77		13.41		78.72	
$A_{\phi}^{\gamma\gamma}$	16.51	100%	26.56	54%	4.94	37%	48.01	61%
$\xi_{\gamma\gamma}$	11.73	71%	18.63	70%	3.27	66%	33.63	70%
ξ_{AFP}	7.34	63%	5.17	28%	0.08	2%	12.59	37%
$\Delta\xi_{\mathrm{eff}}$	7.34	100%	5.15	100%	0.08	97%	12.57	100%

400 GeV								
Cut variable	EL		SD		DD		Total	
	Events	Efficiency	Events	Efficiency	Events	Efficiency	Events	Efficiency
Preselection	12.81		35.93		10.89		59.63	
$A_{\phi}^{\gamma\gamma}$	12.79	100%	22.14	62%	4.82	44%	39.75	67%
$\xi_{\gamma\gamma}$	9.66	76%	16.58	75%	3.31	69%	29.56	74%
ξ_{AFP}	6.69	69%	5.26	32%	0.13	4%	12.09	41%
$\Delta\xi_{\mathrm{eff}}$	6.69	100%	5.24	99%	0.11	83%	12.04	100%

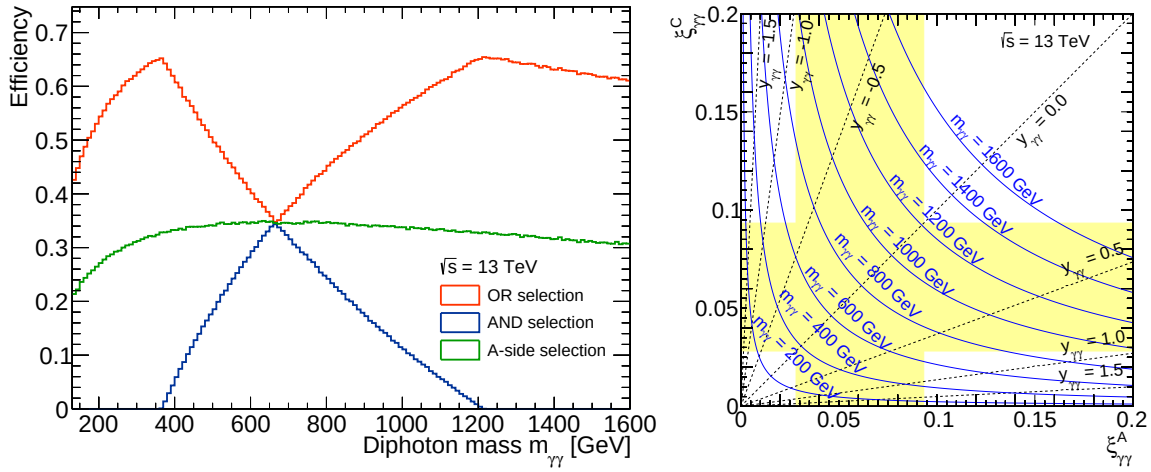
500 GeV								
Cut variable	EL		SD		DD		Total	
	Events	Efficiency	Events	Efficiency	Events	Efficiency	Events	Efficiency
Preselection	9.80		26.89		8.88		45.58	
$A_{\phi}^{\gamma\gamma}$	9.79	100%	17.62	66%	4.34	49%	31.76	70%
$\xi_{\gamma\gamma}$	6.51	66%	11.54	65%	2.52	58%	20.57	65%
ξ_{AFP}	5.13	79%	4.83	42%	0.09	4%	10.05	49%
$\Delta\xi_{\mathrm{eff}}$	5.13	100%	4.83	100%	0.08	83%	10.03	100%

600 GeV								
Cut variable	EL		SD		DD		Total	
	Events	Efficiency	Events	Efficiency	Events	Efficiency	Events	Efficiency
Preselection	7.82		20.97		7.61		36.40	
$A_{\phi}^{\gamma\gamma}$	7.81	100%	14.51	69%	4.10	54%	26.43	73%
$\xi_{\gamma\gamma}$	4.35	56%	7.97	55%	2.04	50%	14.37	54%
ξ_{AFP}	3.30	76%	4.31	54%	0.10	5%	7.72	54%
$\Delta\xi_{\mathrm{eff}}$	3.30	100%	4.31	100%	0.09	92%	7.70	100%

650 GeV								
Cut variable	EL		SD		DD		Total	
	Events	Efficiency	Events	Efficiency	Events	Efficiency	Events	Efficiency
Preselection	6.91		18.78		6.91		32.60	
$A_{\phi}^{\gamma\gamma}$	6.90	100%	13.32	71%	3.92	57%	24.13	74%
$\xi_{\gamma\gamma}$	3.58	52%	6.69	50%	1.83	47%	12.09	50%
ξ_{AFP}	2.67	75%	4.06	61%	0.10	6%	6.84	57%
$\Delta\xi_{\mathrm{eff}}$	2.67	100%	4.06	100%	0.09	91%	6.83	100%

Table 4.3: Signal MC sample cut flow for ALP mass from 700 GeV to 1400 GeV for 14.6 fb^{-1} of pp collision at $\sqrt{s}$ = 13 TeV. The values for $\xi_{\gamma\gamma}$ and ξ_{AFP} are the numbers of events when the set operation is performed without the latter selection.

Cut variable	700 GeV EL		SD		DD		Total	
	Events	Efficiency	Events	Efficiency	Events	Efficiency	Events	Efficiency
Preselection	6.13		16.33		6.27		28.74	
$A_\phi^{\gamma\gamma}$	6.12	100%	11.74	72%	3.65	58%	21.52	75%
$\xi_{\gamma\gamma}$	3.30	54%	6.34	54%	1.76	48%	11.40	53%
ξ_{AFP}	2.25	68%	3.78	60%	0.08	4%	6.10	54%
$\Delta\xi_{\mathrm{eff}}$	2.25	100%	3.77	100%	0.07	92%	6.09	100%
Cut variable	800 GeV EL		SD		DD		Total	
	Events	Efficiency	Events	Efficiency	Events	Efficiency	Events	Efficiency
Preselection	4.82		12.98		5.28		23.08	
$A_\phi^{\gamma\gamma}$	4.82	100%	9.68	75%	3.27	62%	17.78	77%
$\xi_{\gamma\gamma}$	3.14	65%	6.10	63%	1.98	61%	11.23	63%
ξ_{AFP}	2.27	72%	3.33	55%	0.07	3%	5.67	50%
$\Delta\xi_{\mathrm{eff}}$	2.27	100%	3.32	100%	0.06	92%	5.65	100%
Cut variable	900 GeV EL		SD		DD		Total	
	Events	Efficiency	Events	Efficiency	Events	Efficiency	Events	Efficiency
Preselection	3.78		10.22		4.39		18.39	
$A_\phi^{\gamma\gamma}$	3.77	100%	7.79	76%	2.91	66%	14.48	79%
$\xi_{\gamma\gamma}$	2.81	74%	5.49	70%	2.04	70%	10.33	71%
ξ_{AFP}	2.11	75%	2.70	49%	0.07	3%	4.88	47%
$\Delta\xi_{\mathrm{eff}}$	2.11	100%	2.69	100%	0.06	91%	4.86	100%
Cut variable	1000 GeV EL		SD		DD		Total	
	Events	Efficiency	Events	Efficiency	Events	Efficiency	Events	Efficiency
Preselection	3.00		8.32		3.73		15.05	
$A_\phi^{\gamma\gamma}$	3.00	100%	6.52	78%	2.55	68%	12.07	80%
$\xi_{\gamma\gamma}$	2.43	81%	5.05	77%	2.01	79%	9.49	79%
ξ_{AFP}	1.88	77%	2.30	45%	0.05	2%	4.22	44%
$\Delta\xi_{\mathrm{eff}}$	1.88	100%	2.29	100%	0.04	86%	4.21	100%
Cut variable	1200 GeV EL		SD		DD		Total	
	Events	Efficiency	Events	Efficiency	Events	Efficiency	Events	Efficiency
Preselection	1.95		5.49		2.69		10.13	
$A_\phi^{\gamma\gamma}$	1.95	100%	4.47	81%	1.92	72%	8.34	82%
$\xi_{\gamma\gamma}$	1.77	91%	3.91	88%	1.72	89%	7.40	89%
ξ_{AFP}	1.20	68%	1.51	39%	0.03	2%	2.75	37%
$\Delta\xi_{\mathrm{eff}}$	1.20	100%	1.51	99%	0.03	92%	2.74	100%
Cut variable	1400 GeV EL		SD		DD		Total	
	Events	Efficiency	Events	Efficiency	Events	Efficiency	Events	Efficiency
Preselection	1.29		3.77		1.97		7.03	
$A_\phi^{\gamma\gamma}$	1.29	100%	3.13	83%	1.47	74%	5.89	84%
$\xi_{\gamma\gamma}$	1.03	80%	2.51	80%	1.19	81%	4.72	80%
ξ_{AFP}	0.69	67%	0.92	37%	0.02	2%	1.64	35%
$\Delta\xi_{\mathrm{eff}}$	0.69	100%	0.92	100%	0.02	94%	1.63	100%



(a) Efficiency (b) $m_{\gamma\gamma}$ contours

Figure 4.17: (a) The estimated $\xi_{\gamma\gamma}$ selection efficiency of the data for the logical OR (red) and AND (blue) requirements of the A-side and the C-side, as well as the A-side (green) requirement, depending on diphoton mass $m_{\gamma\gamma}$. (b) $m_{\gamma\gamma}$ contours in a $\xi^A_{\gamma\gamma}$-$\xi^C_{\gamma\gamma}$ plane. The yellow region corresponds to the $\xi_{\gamma\gamma}$ OR selection.

4.5.1 Background selection efficiency

The background selection efficiency after the acoplanarity cut is modelled as a function of diphoton mass $m_{\gamma\gamma}$ for better understanding and use in Sec. 6.4.3. The efficiency is computed numerically using some analytical calculations with data-driven constants. The $\xi_{\gamma\gamma}$ selection efficiency $\varepsilon_{\gamma\gamma}$ is calculated as the integral of the probability density function of diphoton rapidity,

$$\varepsilon_{\gamma\gamma}(x_{\gamma\gamma}) = C \int_{I(x_{\gamma\gamma})} f(y_{\gamma\gamma}) \mathrm{d}y_{\gamma\gamma}, \tag{4.4}$$

where $x_{\gamma\gamma} \equiv m_{\gamma\gamma}/\sqrt{s}$, C is a constant, I is the interval of integration, and

$$f(y_{\gamma\gamma}) = \begin{cases} \dfrac{\sum_{i=0}^{4} p_i y_{\gamma\gamma}^{2i}}{\int_{-y_{\mathrm{end}}}^{y_{\mathrm{end}}} \left(\sum_{i=0}^{4} p_i y_{\gamma\gamma}^{2i} \right) \mathrm{d}y_{\gamma\gamma}} & \text{for } |y_{\gamma\gamma}| \leq y_{\mathrm{end}} \\ 0 & \text{for } |y_{\gamma\gamma}| > y_{\mathrm{end}} \end{cases},$$

where y_{end} is defined as the smallest $y_{\gamma\gamma} > 0$ such that $\sum_{i=0}^{4} p_i y_{\gamma\gamma}^{2i} = 0$. The coefficients p_i are derived by fitting the data distribution as Fig. 4.18 and summarized in Table B.2. The $m_{\gamma\gamma}$ dependency of $f(y_{\gamma\gamma})$ is ignored because it is not large.

The interval $I(x_{\gamma\gamma})$ is defined for the OR, AND, and A-side selection. The calculation result of the C-side

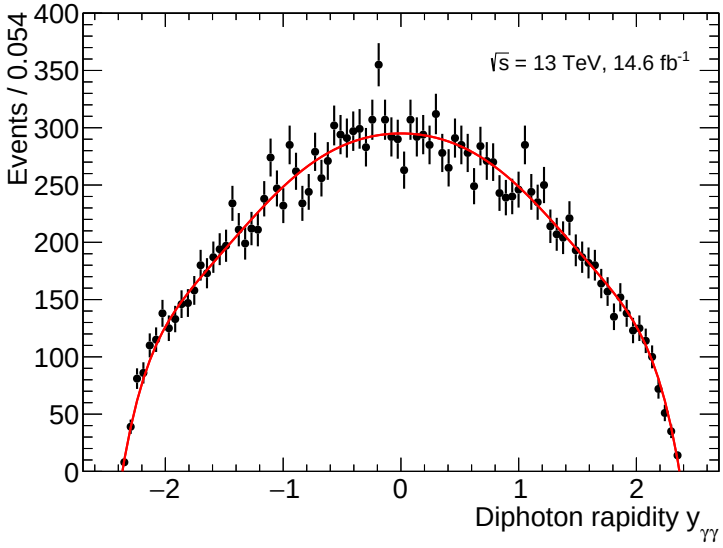


Figure 4.18: Diphoton rapidity distribution of the data after the acoplanarity selection with the fit result of a polynomial function.

Table 4.4: Intervals of integration for each $x_{\gamma\gamma} = m_{\gamma\gamma}/\sqrt{s}$ range. $\bar{\xi}_{\gamma\gamma}$ is defined as $\sqrt{\xi^{\gamma\gamma}_{\mathrm{min}}\xi^{\gamma\gamma}_{\mathrm{max}}}$.

Condition	C	$x_{\gamma\gamma} \in [0, \xi^{\gamma\gamma}_{\mathrm{min}})$	$x_{\gamma\gamma} \in [\xi^{\gamma\gamma}_{\mathrm{min}}, \bar{\xi}_{\gamma\gamma})$	$x_{\gamma\gamma} \in [\bar{\xi}_{\gamma\gamma}, \xi^{\gamma\gamma}_{\mathrm{max}})$	$x_{\gamma\gamma} \in [\xi^{\gamma\gamma}_{\mathrm{max}}, \infty)$
OR	2	$[y^{\mathrm{A}}_{\mathrm{min}}, y^{\mathrm{A}}_{\mathrm{max}}]$	$[0, y^{\mathrm{A}}_{\mathrm{max}}]$	$[0, y^{\mathrm{C}}_{\mathrm{min}}]$	$[y^{\mathrm{C}}_{\mathrm{max}}, y^{\mathrm{C}}_{\mathrm{min}}]$
AND	2	$\varnothing$	$[0, y^{\mathrm{C}}_{\mathrm{min}}]$	$[0, y^{\mathrm{A}}_{\mathrm{max}}]$	$\varnothing$
A-side	1	$[y^{\mathrm{A}}_{\mathrm{min}}, y^{\mathrm{A}}_{\mathrm{max}}]$	$[y^{\mathrm{A}}_{\mathrm{min}}, y^{\mathrm{A}}_{\mathrm{max}}]$	$[y^{\mathrm{A}}_{\mathrm{min}}, y^{\mathrm{A}}_{\mathrm{max}}]$	$[y^{\mathrm{A}}_{\mathrm{min}}, y^{\mathrm{A}}_{\mathrm{max}}]$

selection is assumed to be the same as the A-side. The end-points

$$
\begin{aligned}
y^{\mathrm{A}}_{\mathrm{min}} &= \ln \frac{\xi^{\gamma\gamma}_{\mathrm{min}}}{x_{\gamma\gamma}}, \\
y^{\mathrm{A}}_{\mathrm{max}} &= \ln \frac{\xi^{\gamma\gamma}_{\mathrm{max}}}{x_{\gamma\gamma}}, \\
y^{\mathrm{C}}_{\mathrm{min}} &= -\ln \frac{\xi^{\gamma\gamma}_{\mathrm{min}}}{x_{\gamma\gamma}}, \\
y^{\mathrm{C}}_{\mathrm{max}} &= -\ln \frac{\xi^{\gamma\gamma}_{\mathrm{max}}}{x_{\gamma\gamma}}.
\end{aligned}
$$

are defined for a given $x_{\gamma\gamma}$ using Eq. 4.3 beforehand, where $\xi^{\gamma\gamma}_{\mathrm{min}} = 0.031/1.1$, and $\xi^{\gamma\gamma}_{\mathrm{max}} = 0.084/0.9$. Table 4.4 shows C and the intervals for each $x_{\gamma\gamma}$ range. The integrations for $C = 2$ make use of the fact that the integrand is an even function. Fig. 4.19 shows the $m_{\gamma\gamma}$ dependency of the $\xi_{\gamma\gamma}$ selection efficiency computed numerically. This matches with Fig. 4.17 (a), validating the equation. Fig. A.3 shows the data-driven evaluation for comparison.

The $m_{\gamma\gamma}$ dependency of the background matching selection efficiency, including the ξ_{AFP} selection, is also calculated for each of OR, AND, XOR, and NOR conditions. First of all, for a pair of diphoton and proton for a single side, the probability that the diphoton is matched with the proton is calculated as

$$
\varepsilon^{1p}_{\mathrm{match}}(\xi_{\gamma\gamma}) = \int_{\xi_{\gamma\gamma}-\Delta\xi}^{\xi_{\gamma\gamma}+\Delta\xi} g(\xi_{\mathrm{AFP}}) \mathrm{d}\xi_{\mathrm{AFP}},
$$

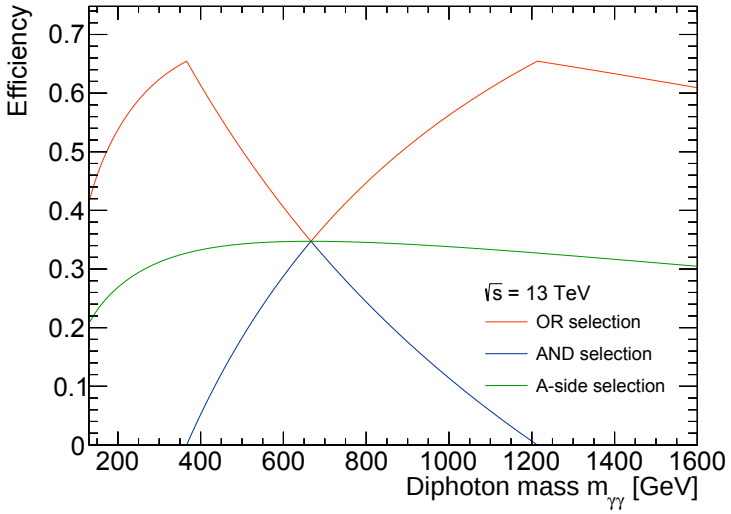


Figure 4.19: $m_{\gamma\gamma}$ dependency of the $\xi_{\gamma\gamma}$ cut efficiency computed numerically for each set operation condition.

where $\Delta\xi = 0.004 + 0.1\xi_{\gamma\gamma}$. $g(\xi_{\mathrm{AFP}})$ is the probability density function of ξ_{AFP} normalized in $\xi_{\mathrm{AFP}} \in [\xi^{\mathrm{AFP}}_{\mathrm{min}}, \xi^{\mathrm{AFP}}_{\mathrm{max}}]$, where $\xi^{\mathrm{AFP}}_{\mathrm{min}} = 0.035$, and $\xi^{\mathrm{AFP}}_{\mathrm{max}} = 0.08$, and is parametrized as

$$g(\xi_{\mathrm{AFP}}) = \begin{cases} \dfrac{g_{\mathrm{stem}}(\xi_{\mathrm{AFP}})}{\displaystyle\int_{\xi^{\mathrm{AFP}}_{\mathrm{min}}}^{\xi^{\mathrm{AFP}}_{\mathrm{max}}} g_{\mathrm{stem}}(\xi_{\mathrm{AFP}})\mathrm{d}\xi_{\mathrm{AFP}}} & \text{for } \xi_{\mathrm{AFP}} \in [\xi^{\mathrm{AFP}}_{\mathrm{min}}, \xi^{\mathrm{AFP}}_{\mathrm{max}}] \\ 0 & \text{for } \xi_{\mathrm{AFP}} \notin [\xi^{\mathrm{AFP}}_{\mathrm{min}}, \xi^{\mathrm{AFP}}_{\mathrm{max}}] \end{cases},$$

$$g_{\mathrm{stem}}(\xi_{\mathrm{AFP}}) = q_0 + \frac{q_1 \,\mathrm{erfc}\left(-\dfrac{\xi_{\mathrm{AFP}} - q_3}{q_4}\right)}{\xi_{\mathrm{AFP}} + q_2}.$$

The parameters q_i are derived by fitting the data distribution as Fig. 4.20 so that the fit result well represent the distribution in $[\xi^{\mathrm{AFP}}_{\mathrm{min}}, \xi^{\mathrm{AFP}}_{\mathrm{max}}]$, assuming that the distributions for both sides are the same. The parameter values are summarized in Table B.2. Since the thresholds of the $\xi_{\gamma\gamma}$ selection are set to $\xi^{\gamma\gamma}_{\mathrm{min}} = \xi^{\mathrm{AFP}}_{\mathrm{min}} - \Delta\xi$ and $\xi^{\gamma\gamma}_{\mathrm{max}} = \xi^{\mathrm{AFP}}_{\mathrm{max}} + \Delta\xi$, the possibility that the matching is performed on a particular side even though the $\xi_{\gamma\gamma}$ selection is not passed on the side is naturally out of consideration.

The matching efficiency $\tilde{\varepsilon}_{\mathrm{match}}(\xi^{\mathrm{A}}_{\gamma\gamma}, \xi^{\mathrm{C}}_{\gamma\gamma})$ is calculated for each condition,

$$\tilde{\varepsilon}^{\mathrm{OR}}_{\mathrm{match}}(\xi^{\mathrm{A}}_{\gamma\gamma}, \xi^{\mathrm{C}}_{\gamma\gamma}) = 1 - \left(1 - \varepsilon^{1p}_{\mathrm{match}}(\xi^{\mathrm{A}}_{\gamma\gamma})\right)^{\overline{N}} \left(1 - \varepsilon^{1p}_{\mathrm{match}}(\xi^{\mathrm{C}}_{\gamma\gamma})\right)^{\overline{N}},$$

$$\tilde{\varepsilon}^{\mathrm{AND}}_{\mathrm{match}}(\xi^{\mathrm{A}}_{\gamma\gamma}, \xi^{\mathrm{C}}_{\gamma\gamma}) = 1 - \left(1 - \varepsilon^{1p}_{\mathrm{match}}(\xi^{\mathrm{A}}_{\gamma\gamma})\varepsilon^{1p}_{\mathrm{match}}(\xi^{\mathrm{C}}_{\gamma\gamma})\right)^{\overline{N}},$$

$$\tilde{\varepsilon}^{\mathrm{A}}_{\mathrm{match}}(\xi^{\mathrm{A}}_{\gamma\gamma}, \xi^{\mathrm{C}}_{\gamma\gamma}) = 1 - \left(1 - \varepsilon^{1p}_{\mathrm{match}}(\xi^{\mathrm{A}}_{\gamma\gamma})\right)^{\overline{N}},$$

where the averaged number of reconstructed protons per event per side derived from the data is $\overline{N} = 0.74$. For any condition, the matching efficiency $\varepsilon_{\mathrm{match}}$ can be calculated as a function of $x_{\gamma\gamma}$ and $y_{\gamma\gamma}$,

$$\varepsilon_{\mathrm{match}}(x_{\gamma\gamma}, y_{\gamma\gamma}) = \tilde{\varepsilon}_{\mathrm{match}}(x_{\gamma\gamma}e^{y_{\gamma\gamma}}, x_{\gamma\gamma}e^{-y_{\gamma\gamma}}).$$

The matching efficiency for a fixed $x_{\gamma\gamma}$ is then calculated as the weighted average of $\varepsilon_{\mathrm{match}}(x_{\gamma\gamma}, y_{\gamma\gamma})$,

$$\varepsilon_{\mathrm{match}}(x_{\gamma\gamma}) = \frac{\displaystyle\int_{I(x_{\gamma\gamma})} \varepsilon_{\mathrm{match}}(x_{\gamma\gamma}, y_{\gamma\gamma}) f(y_{\gamma\gamma})\mathrm{d}y_{\gamma\gamma}}{\displaystyle\int_{I(x_{\gamma\gamma})} f(y_{\gamma\gamma})\mathrm{d}y_{\gamma\gamma}}.$$

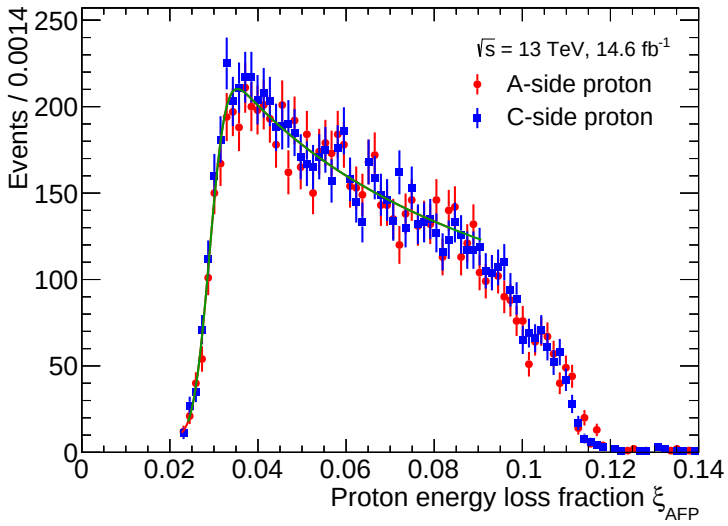
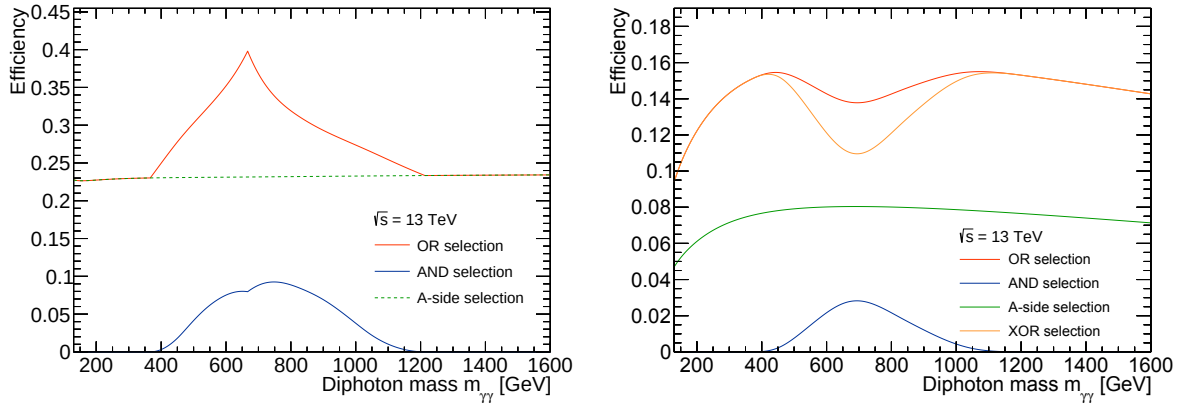


Figure 4.20: ξ_{AFP} distribution of the data with the fitting result. Distributions before the $\xi_{\gamma\gamma}$ selection are used because it is assumed that background events have no correlation between diphoton and proton.



(a) Matching efficiency

(b) Overall efficiency

Figure 4.21: $m_{\gamma\gamma}$ dependency of the matching and overall selection efficiency computed numerically for each condition.

This calculation holds when the AND, OR, or A(C)-side selection is applied. In such a case, the overall background selection efficiency after the acoplanarity selection is

$$\varepsilon_{\mathrm{all}}(x_{\gamma\gamma}) = \varepsilon_{\gamma\gamma}(x_{\gamma\gamma})\varepsilon_{\mathrm{match}}(x_{\gamma\gamma}) = C\int_{I(x_{\gamma\gamma})} \varepsilon_{\mathrm{match}}(x_{\gamma\gamma}, y_{\gamma\gamma}) f(y_{\gamma\gamma})\mathrm{d}y_{\gamma\gamma}. \tag{4.5}$$

The efficiency for XOR and NOR conditions can be calculated as

$$\varepsilon_{\mathrm{all}}^{\mathrm{XOR}}(x_{\gamma\gamma}) = \varepsilon_{\mathrm{all}}^{\mathrm{OR}}(x_{\gamma\gamma}) - \varepsilon_{\mathrm{all}}^{\mathrm{AND}}(x_{\gamma\gamma}),$$
$$\varepsilon_{\mathrm{all}}^{\mathrm{NOR}}(x_{\gamma\gamma}) = 1 - \varepsilon_{\mathrm{all}}^{\mathrm{OR}}(x_{\gamma\gamma}).$$

Fig. 4.21 shows the $m_{\gamma\gamma}$ dependency of the matching and overall selection efficiency computed numerically. Fig. A.4 shows the data-driven evaluation of them for comparison. The efficiency is not compatible, while the dependency shape is similar to each other. Since only the model shape affects the background estimation, this difference is not taken seriously.

5 Signal Modeling

5.1 Signal fitting function

This section describes the signal modeling strategy. In this research, signal modeling assumes that the variance of diphoton mass ($m_{\gamma\gamma}$) distribution of reconstructed signal events is dominated by the resolution of the ATLAS detector, and the ALP natural width is negligible. This assumption is called NWA, as described in Sec. 1.3.1. Under this assumption, the signal $m_{\gamma\gamma}$ distribution is modelled by a double-sided Crystal Ball (DSCB) function [141, 142],

$$
\begin{aligned}
f_X\left(m_{\gamma\gamma}; \boldsymbol{x}_X\left(m_X\right)\right) &= N \cdot \begin{cases} e^{-0.5t^2} & \text{if } -\alpha_{\mathrm{L}} \leq t \leq \alpha_{\mathrm{R}} \\ \dfrac{e^{-0.5\alpha_{\mathrm{L}}^2}}{\left[\dfrac{\alpha_{\mathrm{L}}}{n_{\mathrm{L}}}\left(\dfrac{n_{\mathrm{L}}}{\alpha_{\mathrm{L}}} - \alpha_{\mathrm{L}} - t\right)\right]^{n_{\mathrm{L}}}} & \text{if } t < -\alpha_{\mathrm{L}} \\ \dfrac{e^{-0.5\alpha_{\mathrm{R}}^2}}{\left[\dfrac{\alpha_{\mathrm{R}}}{n_{\mathrm{R}}}\left(\dfrac{n_{\mathrm{R}}}{\alpha_{\mathrm{R}}} - \alpha_{\mathrm{R}} + t\right)\right]^{n_{\mathrm{R}}}} & \text{if } t > \alpha_{\mathrm{R}} \end{cases}, \\
\boldsymbol{x}_X\left(m_X\right) &\equiv \left\{\mu_{\mathrm{CB}}(m_X), \sigma_{\mathrm{CB}}(m_X), \alpha_{\mathrm{L}}, \alpha_{\mathrm{R}}, n_{\mathrm{L}}, n_{\mathrm{R}}\right\}, \\
t &\equiv \frac{m_{\gamma\gamma} - \mu_{\mathrm{CB}}(m_X)}{\sigma_{\mathrm{CB}}(m_X)},
\end{aligned}
\tag{5.1}
$$

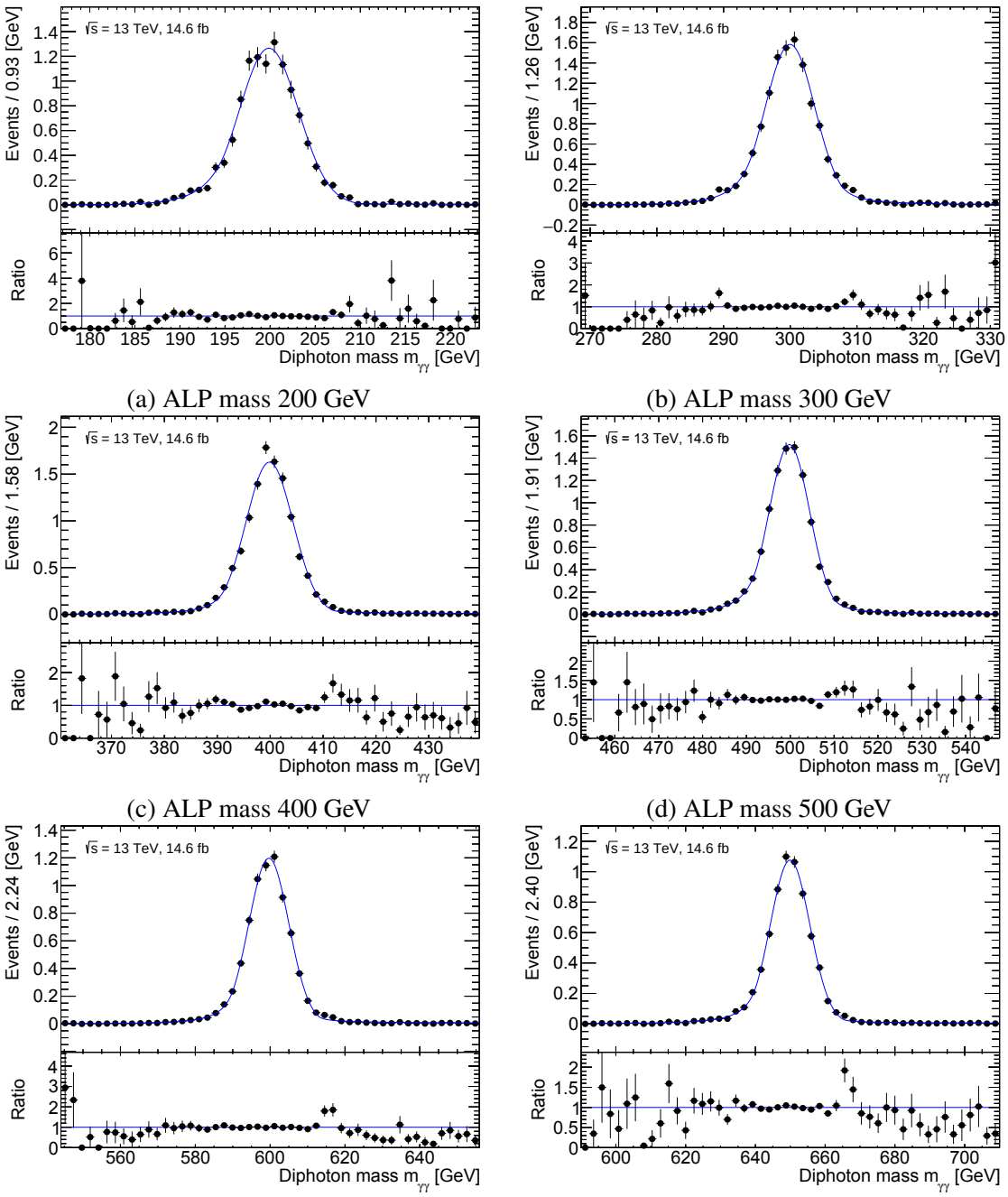
as per the high-mass diphoton resonance search [52]. μ_{CB} and σ_{CB}, depending on hypothetical ALP mass m_X, represent the peak and width of the Gaussian core of the function, N is a normalization parameter, α_{L} (α_{R}) is the position of the junction between the Gaussian and power law on the low (high) mass side in units of t, and n_{L} (n_{R}) is the exponent of this power law.[1] This function and its derivative are continuous for all t and converge to 0 for $t \to \infty$ with $n_{\mathrm{L,R}} > 0$.

The parameters of the DSCB are determined in a multi-step procedure. First, the $m_{\gamma\gamma}$ distribution of the signal MC sample for each mass point m_X after the whole event selection is fitted with the DSCB function, yielding a set of DSCB parameters. The fits are shown in Figs. 5.1 and 5.2. Next, linear or constant functions,

$$
\begin{aligned}
\mu_{\mathrm{CB}}(m_X) &= m_X + \Delta m(m_X) = m_X + p_{\Delta m0} + p_{\Delta m1} m_X, \\
\sigma_{\mathrm{CB}}(m_X) &= p_{\sigma_{\mathrm{CB}}0} + p_{\sigma_{\mathrm{CB}}1} m_X, \\
\alpha_{\mathrm{L}} &= p_{\alpha_{\mathrm{L}}}, \ \alpha_{\mathrm{R}} = p_{\alpha_{\mathrm{R}}} \\
n_{\mathrm{L}} &= p_{n_{\mathrm{L}}}, \ n_{\mathrm{R}} = p_{n_{\mathrm{R}}}
\end{aligned}
$$

are fit to the m_X dependency of each parameter to parametrize the mass-dependent evolution, where the constants p are the parameters. The parametrization is shown in Fig. 5.3 and is summarized in Table 5.1. The parameterization models the resonance peak for arbitrary hypothetical ALP mass. Each uncertainty in the table is considered a systematic uncertainty of the corresponding parameter as described in Sec. 8.1.

[1] Although $\alpha_{\mathrm{L,R}}$ and $n_{\mathrm{L,R}}$ are also expected to depend on m_X, the dependency is ignored because the effect is negligible.



(a) ALP mass 200 GeV

(b) ALP mass 300 GeV

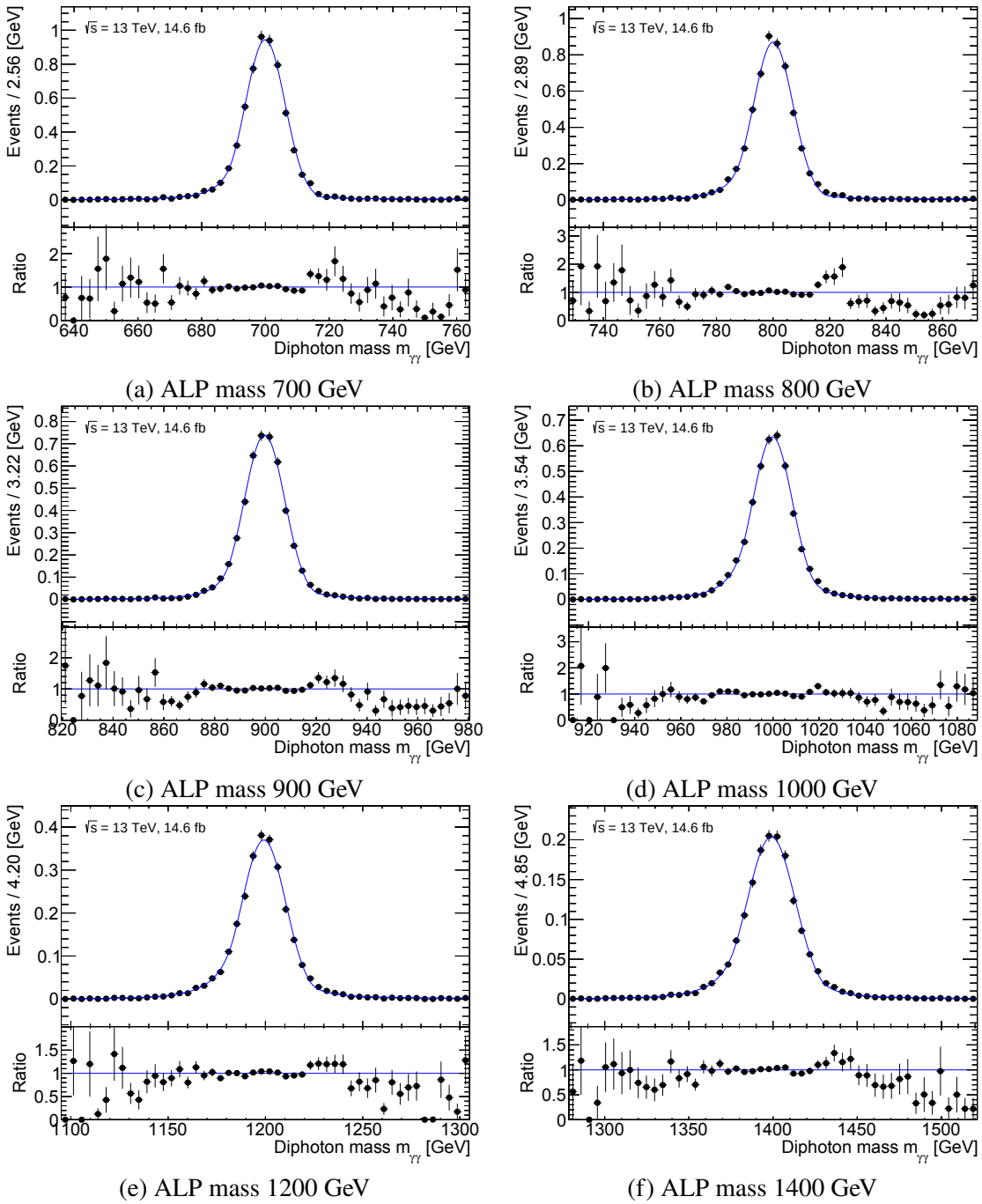
(c) ALP mass 400 GeV

(d) ALP mass 500 GeV

(e) ALP mass 600 GeV

(f) ALP mass 650 GeV

Figure 5.1: Fitting results of signal MC samples after the whole event selection using the DSCB function, shown here for different signal masses ranging from 200 GeV to 650 GeV. The bottom panel of each plot shows the ratio of MC sample distribution compared with the fit function.



(a) ALP mass 700 GeV

(b) ALP mass 800 GeV

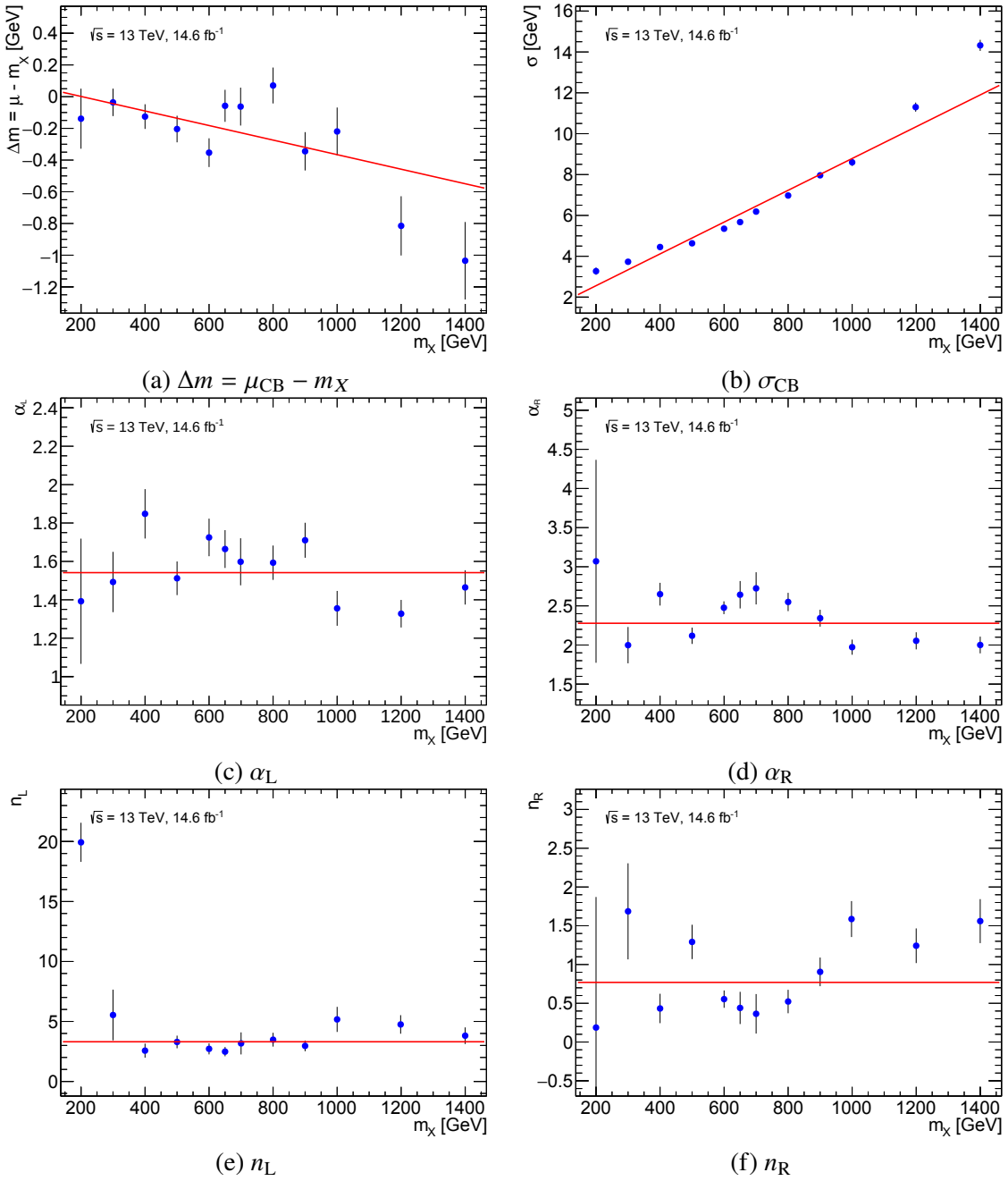
(c) ALP mass 900 GeV

(d) ALP mass 1000 GeV

(e) ALP mass 1200 GeV

(f) ALP mass 1400 GeV

Figure 5.2: Fitting results of signal MC samples after the whole event selection using the DSCB function, shown here for different signal masses ranging from 700 GeV to 1400 GeV. The bottom panel of each plot shows the ratio of MC sample distribution compared with the fit function.



(a) $\Delta m = \mu_{\mathrm{CB}} - m_X$ (b) σ_{CB}

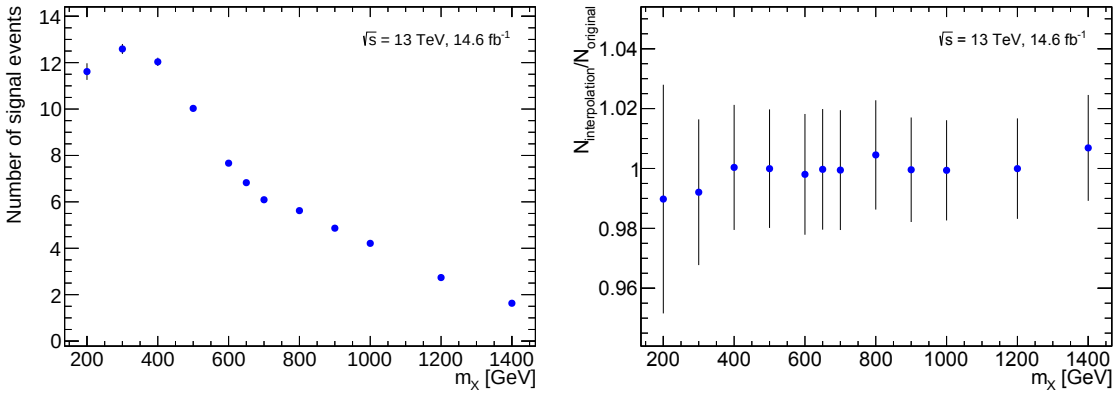
(c) α_{L} (d) α_{R}

(e) n_{L} (f) n_{R}

Figure 5.3: The mass dependence of the DSCB parameters. The linear or constant functions (red lines) are obtained from a fit to the DSCB fit results of each mass point (blue points).

Table 5.1: Signal model parameters. Parameters modelled by linear function have the subscripts followed by the number i, which means the coefficient of the i-th order.

Parameter	Value	Uncertainty
$p_{\Delta m0}$	9.33×10^{-2}	$\delta_{\Delta m0} \equiv 8.18 \times 10^{-2}$
$p_{\Delta m1}$	-4.59×10^{-4}	$\delta_{\Delta m1} \equiv 1.26 \times 10^{-4}$
$p_{\sigma_{\mathrm{CB}}0}$	1.00	$\delta_{\sigma_{\mathrm{CB}}0} \equiv 8.94 \times 10^{-2}$
$p_{\sigma_{\mathrm{CB}}1}$	7.78×10^{-3}	$\delta_{\sigma_{\mathrm{CB}}1} \equiv 1.39 \times 10^{-4}$
$p_{\alpha_{\mathrm{L}}}$	1.54	$\delta_{\alpha_{\mathrm{L}}} \equiv 2.89 \times 10^{-2}$
$p_{\alpha_{\mathrm{R}}}$	2.28	$\delta_{\alpha_r} \equiv 3.53 \times 10^{-2}$
$p_{n_{\mathrm{L}}}$	3.31	$\delta_{n_{\mathrm{L}}} \equiv 1.75 \times 10^{-1}$
$p_{n_{\mathrm{R}}}$	7.69×10^{-1}	$\delta_{n_{\mathrm{R}}} \equiv 5.84 \times 10^{-2}$



(a) Signal yields of original fittings

(b) Relative difference of signal yields

Figure 5.4: (a) The signal yields of the original fits. The non-smooth behaviour at around 650 GeV originates from the $\xi_{\gamma\gamma}$ cut efficiency. (b) The signal yields of fittings using the fixed parameters relative to (a) with maximum errors.

The parametrized model does not entirely match the original fit results in Fig. 5.3. Therefore, the effect of the difference on the resulting signal yield is studied by comparing the original fit with the fit using the DSCB function whose parameters are fixed to the interpolation results. Fig. 5.4 (a) shows the signal yields of original fittings. (b) shows the signal yields of fittings using the fixed parameters relative to (a). The difference is slight for any ALP mass m_X.

5.1.1 Signal modeling uncertainty

One of the systematic uncertainties of the signal model is evaluated. The systematic uncertainty of signal modeling is evaluated for each hypothetical ALP mass, comparing the resulting signal yield with the input number of events in the signal MC sample. The input signal events are weighted variously to check if the resulting yield follows the weight accordingly, following a linear relation. The yield and number of events are unitized by *signal strength*, which is the number of events (normalized to 14.6 fb^{-1}) of the nominal signal MC samples for the corresponding ALP mass after the whole event selection. Fig. 5.5 shows the results of the signal injection test performed by fitting the DSCB function to the signal MC distribution for each ALP mass hypothesis. The parameters of the DSCB function, except the signal yield, are fixed to the interpolated values to distinguish the sources of the systematic uncertainty. The input and output signal

Table 5.2: Signal modeling uncertainty for each ALP mass.

ALP mass [GeV]	δ_{smodel}
200	−0.0034
300	−0.0017
400	−0.0006
500	−0.0006
600	−0.0027
650	−0.0045
700	−0.0062
800	−0.0022
900	−0.0013
1000	−0.0008
1200	−0.0010
1400	−0.0212

strengths are compared for each ALP mass hypothesis. The fitting result of a linear function

$$\mu_{\text{out}}(\mu_{\text{in}}) = (1 + \delta_{\text{smodel}})\mu_{\text{in}}$$

is represented by the red lines, where δ_{smodel} is the parameter. Table 5.2 lists the fitting results of the linear function. The maximum deviation from 1 is taken as the signal modeling uncertainty,

$$\delta_{\text{smodel}} \simeq 2.1\%.$$

5.2 Signal efficiency and yield

The signal yield $N_X(m_X)$ in the results of this analysis is defined as the sum of EL, SD, and DD signal yields,

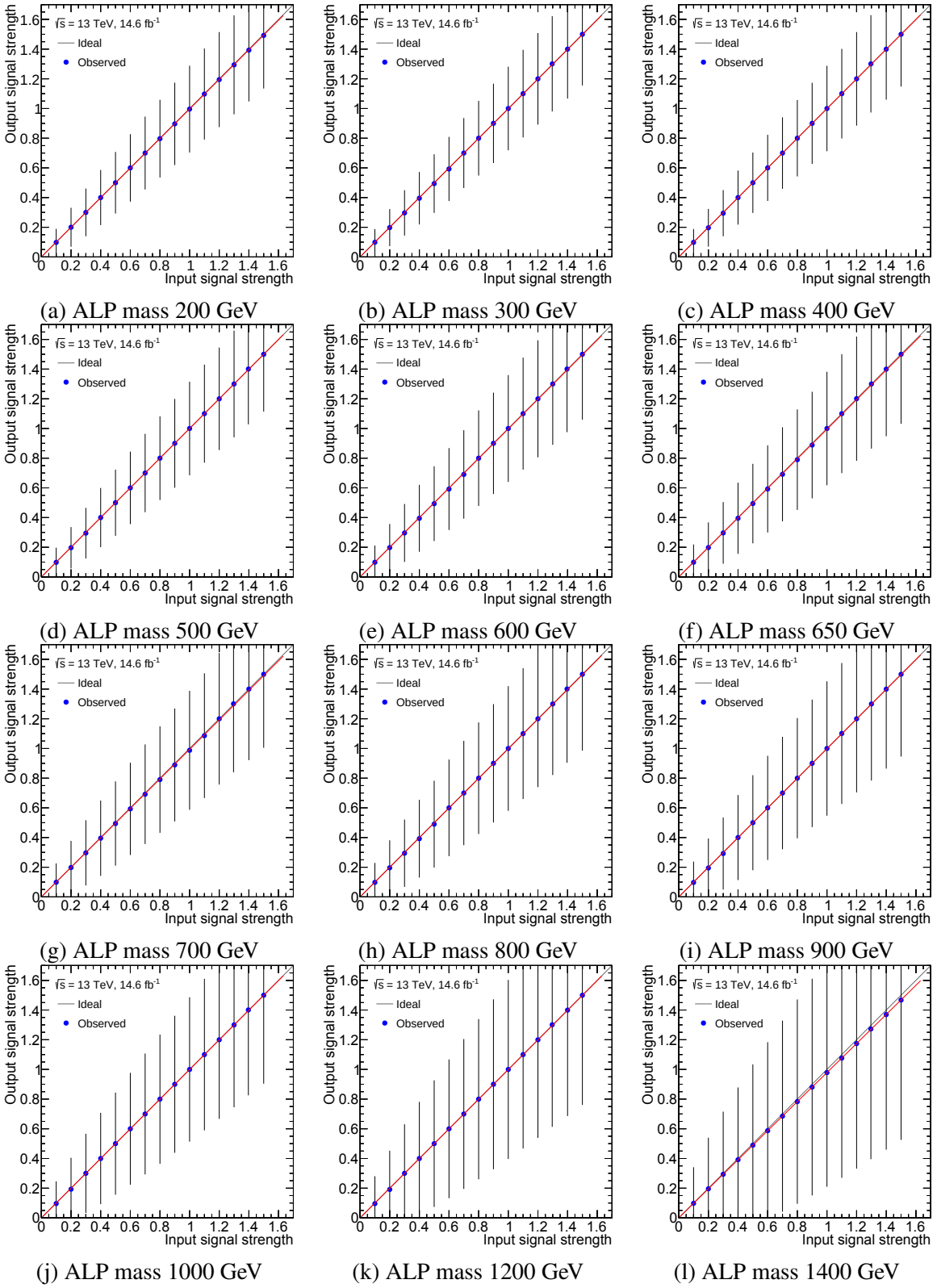
$$N_X(m_X) = \sum_{i \in \{\text{EL,SD,DD}\}} N_i(m_X),$$
$$N_i(m_X) = L_{\text{int}}\sigma_i(m_X)\varepsilon_i(m_X),$$

where $L_{\text{int}} = 14.6\ \text{fb}^{-1}$ is the integrated luminosity, N_i is the signal yield, σ_i is the signal cross section in the phase space limited to the region specified in Table 3.1, and ε_i is the selection efficiency, for each event type i.

5.2.1 Efficiency

The efficiency ε_i is parametrized as a function of ALP mass m_X with the evaluation of uncertainties to get results for arbitrary ALP mass. It is calculated as the fraction of the number of remaining events after the whole event selection over the total events for each ALP signal MC sample for each event type i,

$$\varepsilon^i_{m_X} = \frac{N^i_{m_X}}{L_{\text{int}}\sigma^i_{m_X}},$$



(a) ALP mass 200 GeV (b) ALP mass 300 GeV (c) ALP mass 400 GeV

(d) ALP mass 500 GeV (e) ALP mass 600 GeV (f) ALP mass 650 GeV

(g) ALP mass 700 GeV (h) ALP mass 800 GeV (i) ALP mass 900 GeV

(j) ALP mass 1000 GeV (k) ALP mass 1200 GeV (l) ALP mass 1400 GeV

Figure 5.5: The signal injection test results with statistical uncertainties for each ALP mass hypothesis. The input and output signal strengths are compared using the red fitted lines. The error bars represent the statistical uncertainty for 14.6 fb^{-1}.

Table 5.3: Signal efficiency parametrization.

Parameter	i = EL	i = SD	i = DD
p_0^i	-1.66×10^{-1}	-1.89×10^{-2}	-1.56×10^{-3}
p_1^i	2.22×10^{-3}	2.30×10^{-4}	2.31×10^{-5}
p_2^i	-2.66×10^{-6}	3.34×10^{-7}	-1.14×10^{-8}
p_3^i	1.49×10^{-9}	-4.20×10^{-10}	
p_4^i	-3.50×10^{-13}	1.05×10^{-13}	
p_5^i	4.00×10^{-1}		
p_6^i	6.94×10^{2}		
p_7^i	1.23×10^{2}		

where $\varepsilon^i_{m_X}$, $N^i_{m_X}$, and $\sigma^i_{m_X}$ are the efficiency, signal yield, and cross section. The $\sigma^i_{m_X}$ is output from SuperChic and known for each ALP mass. Since the signal MC samples are generated at discrete ALP mass points m_X, these efficiencies are discrete functions of m_X at this point. The mass dependency of signal efficiency is parametrized, fitting the functions

$$\varepsilon_{\mathrm{EL}}(m_X) = \left(\sum_{i=0}^{4} p_i^{\mathrm{EL}} m_X^i\right) \times \left(1 - G(m_X; p_5^{\mathrm{EL}}, p_6^{\mathrm{EL}}, p_7^{\mathrm{EL}})\right),$$

$$\varepsilon_{\mathrm{SD}}(m_X) = \sum_{i=0}^{4} p_i^{\mathrm{SD}} m_X^i,$$

$$\varepsilon_{\mathrm{DD}}(m_X) = \sum_{i=0}^{2} p_i^{\mathrm{DD}} m_X^i,$$

where

$$G(m_X; a, b, c) = a \exp\left[-\frac{1}{2}\left(\frac{m_X - b}{c}\right)^2\right] \tag{5.2}$$

is the Gaussian function. The parameters are listed in Table 5.3. Fig. 5.6 shows the computed efficiency values, and the functions are overlaid (orange). The envelopes are also illustrated to cover the difference between the computed efficiency and the functions. To derive the envelopes, the linear (blue) and cubic (cubic) interpolations are performed, and the upper (cyan) and lower (pink) envelopes are set to cover these interpolation results. The parametrization of the envelopes are

$$\delta^{\mathrm{up}}_{\varepsilon\mathrm{EL}} = 0.05 + 0.4e^{-0.05(m_X - 130\ \mathrm{GeV})} \qquad \delta^{\mathrm{down}}_{\varepsilon\mathrm{EL}} = 0.05 \tag{5.3}$$

$$\delta^{\mathrm{up}}_{\varepsilon\mathrm{SD}} = 0.04 + 0.4e^{-0.045(m_X - 130\ \mathrm{GeV})} \qquad \delta^{\mathrm{down}}_{\varepsilon\mathrm{SD}} = 0.04 \tag{5.4}$$

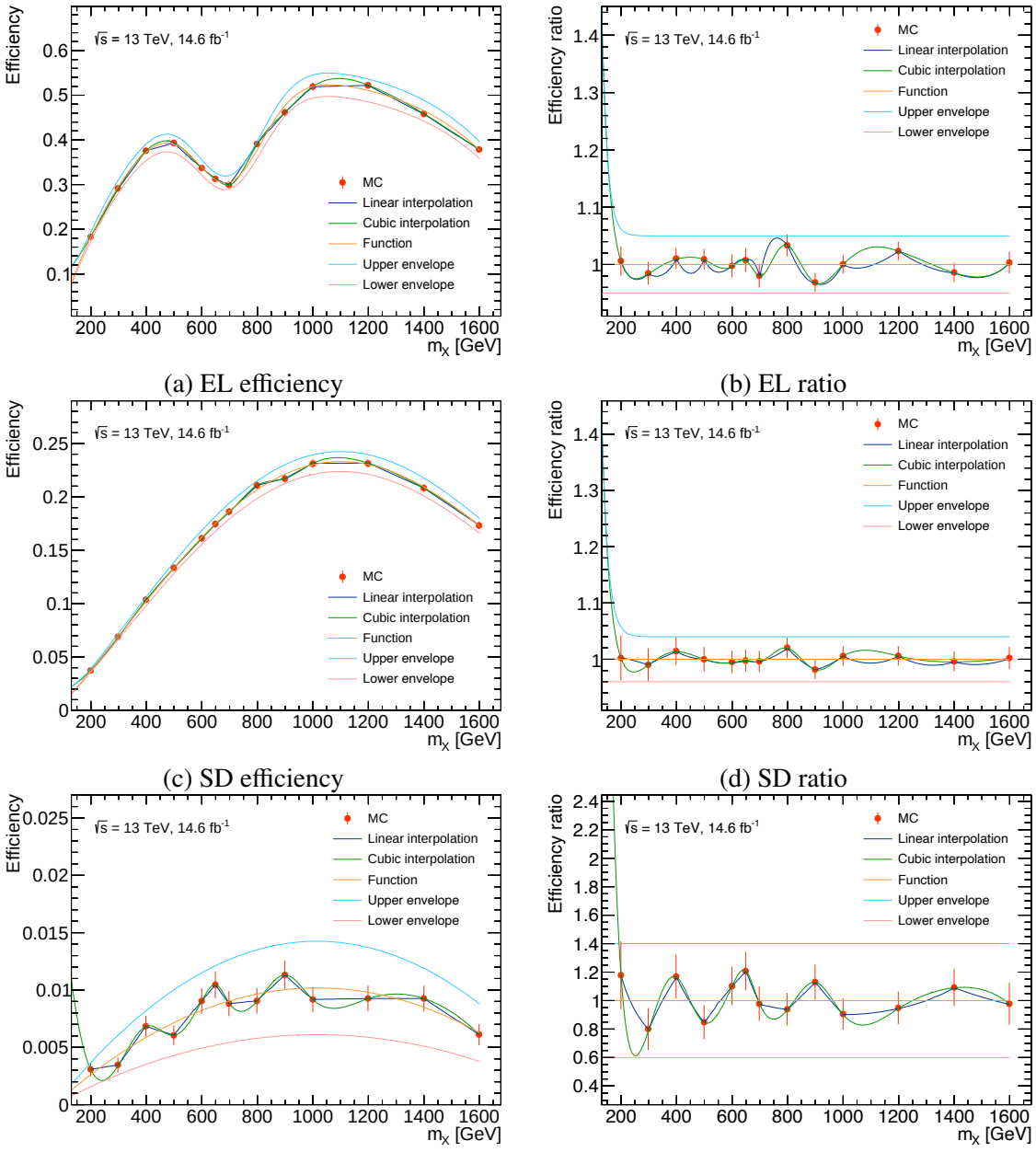
$$\delta^{\mathrm{up}}_{\varepsilon\mathrm{DD}} = 0.4 \qquad \delta^{\mathrm{down}}_{\varepsilon\mathrm{DD}} = 0.4 \tag{5.5}$$

in terms of the ratio to the nominal parametrization. They are used as the systematic uncertainty as described in Sec. 8.1.

5.2.2 Cross section and yield

The signal yield for the fixed ALP coupling constant $f^{-1} = 0.05\ \mathrm{TeV}^{-1}$, assumed when generating the signal MC samples, called *standard* coupling constant, is also parametrized. Such signal yield is referred to as the standard signal yield N_{std}, and the signal strength is redefined as the quantity μ satisfying

$$N_X(m_X) = \mu(m_X) N_{\mathrm{std}}(m_X), \tag{5.6}$$



(a) EL efficiency (b) EL ratio

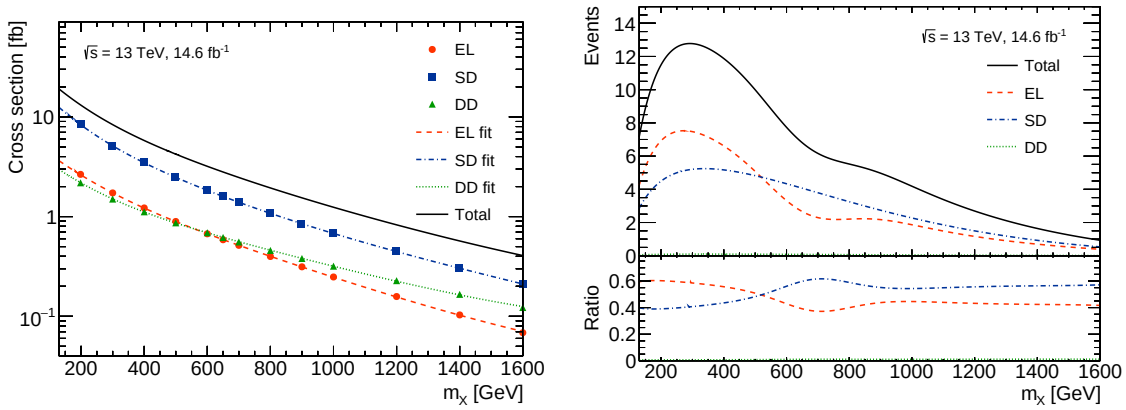
(c) SD efficiency (d) SD ratio

(e) DD efficiency (f) DD ratio

Figure 5.6: (a)(c)(e) Signal efficiency derived from the signal MC samples. The orange line is its parametrization. The envelopes are also shown. The cyan and magenta lines represent the envelopes. (b)(d)(f) The ratio of the signal efficiency taken with respect to the nominal efficiency modeling function.

Table 5.4: Signal cross section parametrization in fb unit for coupling constant $f^{-1} = 0.05$ TeV.

Parameter	i = EL	i = SD	i = DD
p_0^i	2.87983000	9.85015	2.04608
p_1^i	−0.00256053	−0.00417	−0.00191
p_2^i	4.60166000	19.75920	3.90991
p_3^i	−0.00838870	−0.01123	−0.00809
p_4^i	0.02282670	2.65697	0.00471
p_5^i	−0.00000639	−0.00162	0.00115



(a) Standard cross section

(b) Standard yield

Figure 5.7: (a) Signal cross sections and their parametrization for coupling constant $f^{-1} = 0.05$ TeV. (b) Signal yield parametrization for the same coupling constant. The lower panel shows the ratio of the number of events for each event type compared to the total number of events.

where N_X is the signal yield of interest. The standard signal yield is calculated as

$$N_{\text{std}}(m_X) = \sum_{i \in \{\text{EL,SD,DD}\}} N^i_{\text{std}}(m_X),$$
$$N^i_{\text{std}}(m_X) = L_{\text{int}} \sigma^i_{\text{std}}(m_X) \varepsilon_i(m_X),$$

where $\sigma^i_{\text{std}}(m_X)$ is the cross section corresponding to the fixed coupling constant, and $\varepsilon_i(m_X)$ is parametrized as described in Sec. 5.2.1. The cross sections are parametrized as

$$\sigma^i_{\text{std}}(m_X) = p^i_0 e^{p^i_1 m_X} + p^i_2 e^{p^i_3 m_X} + p^i_4 e^{p^i_5 m_X},$$

where the parameters are listed in Table 5.4. The sum

$$\sigma_{\text{std}}(m_X) = \sum_{i \in \{\text{EL,SD,DD}\}} \sigma^i_{\text{std}}(m_X)$$

σ_{std} is called standard cross section. Now that both cross section and efficiency are parametrized to be continuous, signal yield is also parametrized as the product of them and L_{int}. Fig. 5.7 shows the parametrization of σ^i_{std}, N^i_{std}, σ_{std}, and N_{std}. The ratio of signal yield between EL and SD is compatible with the estimations from the SuperChic developers illustrated in Fig. 5.8 for comparison, roughly supporting the implementation of the event selection in this research, although the selection criteria are slightly different.

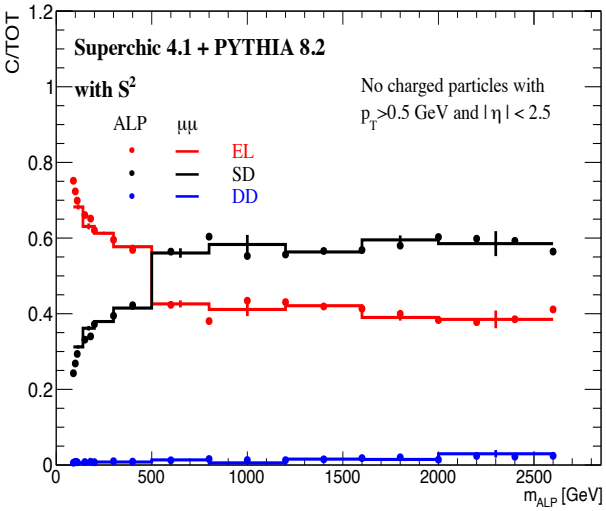


Figure 5.8: The ratio of the number of signal events for coupling constant $f = 0.05$ TeV^{-1} for each event type compared to the total number of events [115]. SuperChic and Pythia are used, and no detector simulation is performed. The object and event selection similar to this research is applied at the truth level. The dimuon events illustrated in Fig. 1.8 are also studied for comparison. The S^2 stands for the soft survival factor described in Sec. 3.2.

6 Background Modeling

In this section, the modeling of the diphoton mass distribution of the background events is described with systematic uncertainty. The dominant background arises from the coincident matching between the diphoton and proton objects from different pp collision vertices. Such events are referred to as *combinatorial* backgrounds, while all non-combinatorial background events are called *single-vertex* backgrounds. The sources of the diphoton objects in the combinatorial backgrounds are, for example, inclusive diphoton production and fake jets. On the other hand, the dominant source of the background protons in this research using the OR selection (see Sec. 4.5) is the single-diffraction as described in Sec. 2.3.

A background sample is prepared by a fully data-driven method to use as a template to determine the background model in Sec. 6.1. The background model is constructed in the same section. The validation of the data-driven method is described in Sec. 6.2. It is also proved that the single-vertex backgrounds are negligible. Signal contamination in the background sample is considered a systematic variation of the template in Sec. 6.3, and the systematic uncertainty of the background model is determined using the nominal and varied templates in Sec. 6.4.

6.1 Background sample creation and background modeling

A fully data-driven method is used to make a background template inheriting the procedure in the dilepton analysis [66] as described in Sec. 1.4. First, the nominal data sample after the object selection and the event preselection described in Sec. 3 is prepared. The amount of combinatorial background depends on instantaneous luminosity and hence the measured number of pp collisions μ in the BC, as shown in Fig. 6.1, which represents the μ dependency of the selection efficiency after the acoplanarity selection. Therefore, the preselected events are divided into the bins of the μ in the BC from which each event originates. Fig. 6.2 shows the number of events in each bin. The bin width is 1, and almost all events are in the bins with more than 100 entries, while the exceptional events are ignored in the background estimation.

In each bin, the events are processed as illustrated in Fig. 6.3. The sample has a set of proton objects for each event because the unique proton selection for each side has not been performed yet at this stage. The colorful boxes indicate the proton sets. The procedure is as follows.

1. The event selection up to the $\xi_{\gamma\gamma}$ selection with the OR selection is applied to the data sample in each bin. Since only the diphoton object information is used in this selection, this is referred to as the diphoton selection. The set of selected events necessarily contains the events that can pass the whole event selection with the OR selection.[1]
2. The set of protons in each event is then replaced by the sets from the other events before the diphoton selection to create multiple events with the same object information except for protons. Since the

[1] The diphoton selection can be skipped because the whole selection is applied after all, but it is applied in order to reduce the amount of computation in the following processes.

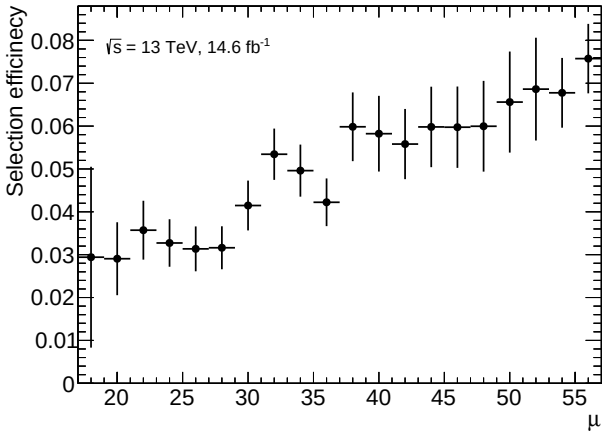


Figure 6.1: The ratio of the number of data events after the whole event selection to the selection up to the acoplanarity selection in each coarse μ bin. The statistical uncertainty is also illustrated.

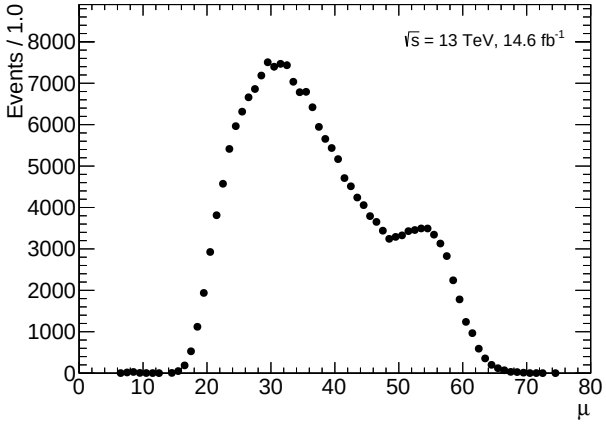


Figure 6.2: Distribution of the number of pp collisions μ in the BC from which each data event originates.

assigned set of protons is from a different BC for each created event, the diphoton and protons are not correlated. This procedure is referred to as the decorrelation by multiple reassignments.

3. The whole selection is applied to these many events. Only the events with diphotons and protons coincidentally matched remain, embodying the definition of combinatorial background.
4. Since a set of events originating from the same preselected event has the same diphoton mass value, such events are unified as one event with the weight calculated as the fraction of the selected events in the set.

Finally, the unified events are merged with respect to the μ bins to almost finalize the background sample. The difference in the method from the dilepton analysis [66] is that they mixed the proton sets without binning the events with respect to μ and reassigned only 100 proton sets to each event.

Although the backgrounds in data are expected to come from almost the combinatorial background, there could be non-negligible contamination of single-vertex backgrounds and signal events. However, the decorrelation method suppresses such contamination because the matching selection eliminates most of them. The suppression of signal contamination is quantified in Sec. 6.3. One could just shift the sets of protons to the next events one by one to decorrelate the diphotons and protons, but the original data size is regarded to be too small to use for the background modeling. Fig. 6.4 compares the $m_{\gamma\gamma}$ distribution of such a simple background sample and the nominal fully data-driven background sample after the whole event selection. The nominal background template is much smoother than the former one because events

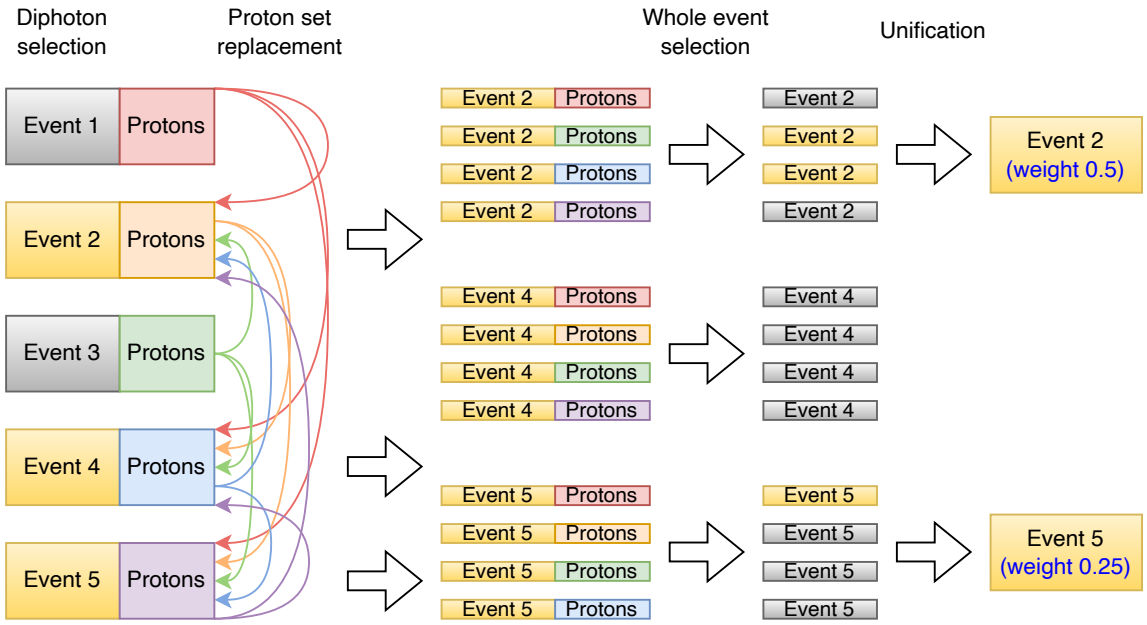


Figure 6.3: Algorithm flow of background sample creation. A μ bin with five preselected events is illustrated for simplicity. The yellow and grey boxes stand for the selected and not selected events, respectively. Each event has a set of proton objects represented by colorful boxes.

which should have been excluded by the matching selection in the simpler method remain with weights less than 1 in the nominal method. The distribution shape uncertainty is also evaluated as

$$\sqrt{\sum_i w_i^2}$$

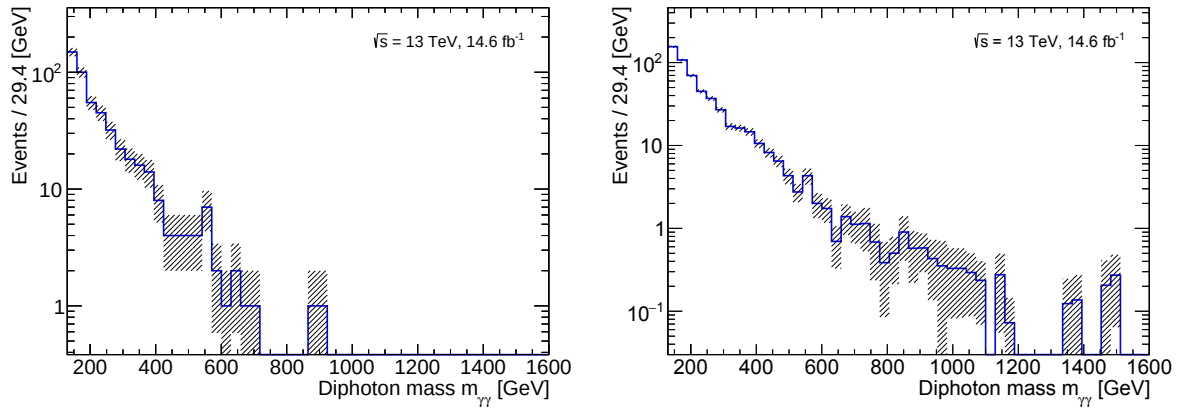
for each $m_{\gamma\gamma}$ bin, where w_i is the weight of the event with index i. The uncertainty is smaller than the former case, whose shape uncertainty is identical to the statistical uncertainty, the square root of the bin content. In the diphoton mass range from 150 GeV to 1600 GeV, the data after the diphoton selection has 3977 events, and the number of events is reduced to 441 after the whole event selection. However, the data-driven background sample, even after the whole selection, holds the information of 3964 events out of the 3977 events. The higher the integrated luminosity and the tighter the selection after the diphoton selection, the more powerful this background sample creation method becomes.

The weight of each event in the nominal background sample is slightly modified finally. Since the amount of combinatorial background and μ are positively correlated, for example, if a proton set from an event in a BC with a higher luminosity is assigned to a diphoton from an event in a BC with a lower luminosity, the coincident matching rate becomes overestimated. Each event in the background sample is weighted by $\mu_{\gamma\gamma}/\mu_p$ to consider the effect of the slight difference of the μ between the event from which the diphoton originates, $\mu_{\gamma\gamma}$, and the event from which the protons originate, μ_p, assuming the linearity between the amount of combinatorial background and μ. This weight has uncertainty within the μ bin. The μ dependency and the systematic uncertainty evaluation of the weight $\mu_{\gamma\gamma}/\mu_p$ are detailed in Appendix E, and the effect of the systematic uncertainty is found to be negligible.

The function form

$$f_{\mathrm{b}}\left(x;\{a,\alpha_0\}\right) \equiv f_0\left(x;\{a,\alpha_0\}\right) = \left(1-x^{1/3}\right)^a \cdot x^{\alpha_0}, \tag{6.1}$$

is used for the background function fitting in the mass range [150, 1600] GeV. $x = m_{\gamma\gamma}/\sqrt{s}$ is a transformation that improves the numerical stability of the fit. This function form is an element of a family of



(a) Single-shifted background sample (b) Nominal background sample

Figure 6.4: The $m_{\gamma\gamma}$ distribution of the background samples after the whole event selection with the shape uncertainty. (a) Background sample where the proton set in each event of the data is shifted to the next event. (b) The nominal background sample.

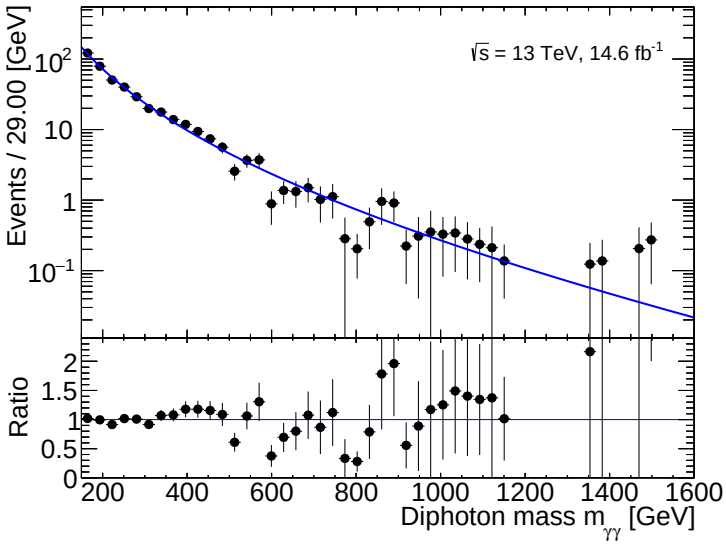


Figure 6.5: Comparison of $m_{\gamma\gamma}$ distribution between the fully data-driven background after the whole selection with shape uncertainty (black) and function fitted to it (blue) in the diphoton mass range [150, 1600] GeV.

functions f_k traditionally used in the high-mass diphoton resonance searches [50, 52] to describe the $m_{\gamma\gamma}$ spectrum, where

$$f_k\left(x;\{a,\alpha_i \mid i \in \mathbb{Z} \cup [0,k]\}\right) = \left(1 - x^{1/3}\right)^a \cdot x^{\sum_{i=0}^{k} \alpha_i (\log x)^i}.$$

Although the latest high-mass diphoton resonance search [52] used f_1, this research uses the simpler f_0 because the background modeling uncertainty is expected to be dominated by the statistical uncertainty considering the low luminosity, 14.6 fb^{-1}. The parameters a and α_0 are free parameters in this research as described in Sec. 8.1. The fit result is shown in Fig. 6.5. The best-fit parameters for the nominal background template are $a = 15.9$ and $\alpha_0 = -0.828$, which are used as the initial values in the latter analyses.

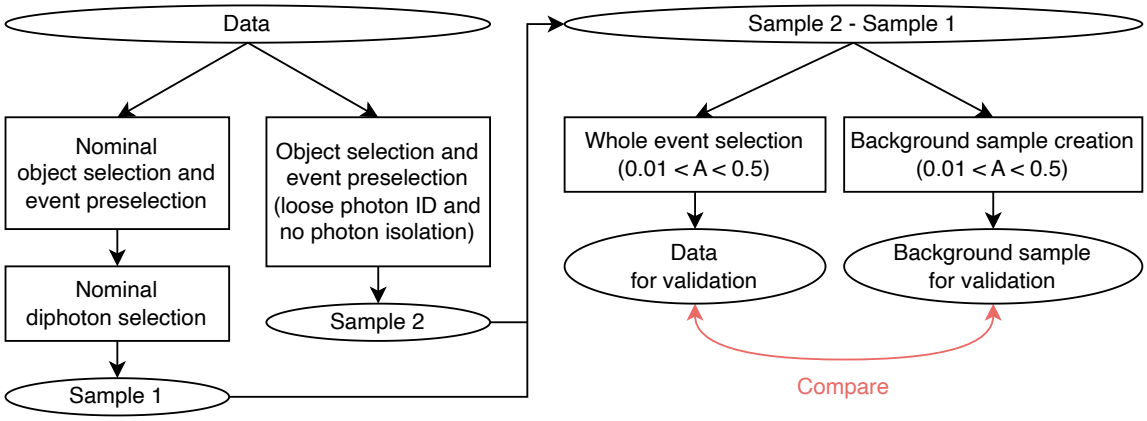


Figure 6.6: The algorithm flow of the data-driven validation of the fully data-driven background creation method in the acoplanarity range $[0.01, 0.5]$.

6.2 Validation

The data-driven background sample creation method described in Sec. 6.1 above is validated by comparing data and the background sample in validation regions. The agreement on the comparison supports that the combinatorial background is dominant. Sec. 6.2.1 defines the validation regions and compares the data and background sample. Furthermore, Sec. 6.2.2 examines some possible single-vertex backgrounds using MC samples to see if there is a non-negligible contribution.

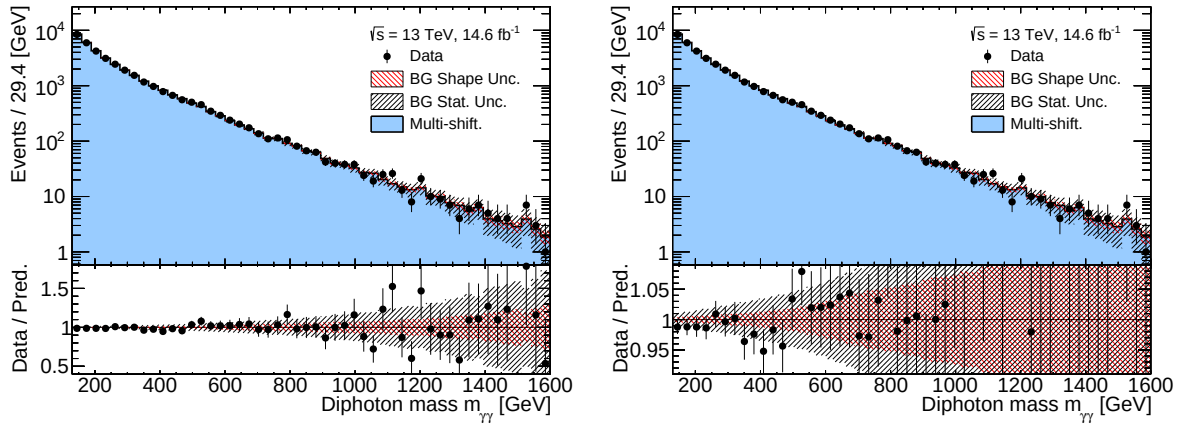
6.2.1 Data-driven validation of sample and dominance of combinatorial background

The background sample is compared with the data in terms of the $m_{\gamma\gamma}$ distribution in a background-rich region to validate the background sample creation method. Fig. 6.6 summarizes the procedure. The *loose* photon identification working point is used, and the photon isolation requirement is removed ("Sample 2"). Events that are identical to the nominal data events that pass the event selection up to the $\xi_{\gamma\gamma}$ selection ("Sample 1") are excluded from the validation samples. The acoplanarity range is changed to $0.01 < A_{\phi}^{\gamma\gamma} < 0.5$. The full nominal selection, except these criteria, is applied to the samples to create the data sample for validation for comparison. On the other hand, the background sample is made with the same selection by the same method as the nominal one. Fig. 6.7 shows the comparison between the data and the background samples in the validation region. The χ^2 test and the Kolmogorov-Smirnov test are performed for these distributions. The resulting reduced χ^2 and Kolmogorov test probability are 0.649 and 0.859, respectively. These results show the good compatibility of the background sample with the data and validate the background sample creation method.

Further tests are performed to validate the dominance of combinatorial background in a similar data-driven way. The background sample is compared with the data in another validation region closer to the signal region. A validation region is defined by the selection partially different from the nominal one in that

- Only one of the leading and the sub-leading photons is required to pass the tight ID criteria, and the other photon is required not to pass the tight identification and pass the loose identification,
- $0.01 < A_{\phi}^{\gamma\gamma} < 0.5$,
- $m_{\gamma\gamma} \in [130, 1600]$ GeV.

Since the first point makes the sample mutually exclusive from the nominal data, there is no need to take any difference set as illustrated in Fig. 6.6. The process is as simple as replacing the "Sample 2 - Sample 1"



(a) Unzoomed (b) Zoomed

Figure 6.7: Comparison between the data and background sample distributions in the background-rich region. (a) Unzoomed and (b) zoomed versions of the lower panels are represented for better visibility. The shape uncertainty and statistical uncertainty of the background sample are also illustrated. The shape uncertainty is smaller than the statistical uncertainty as described in Sec. 6.1.

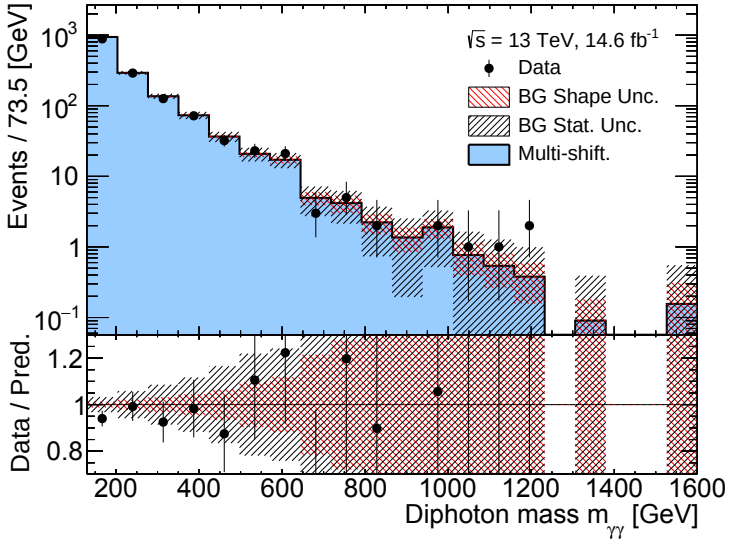


Figure 6.8: Comparison of $m_{\gamma\gamma}$ distributions between the data and background sample in the validation region with the non-back-to-back requirement and only one tight photon.

in the figure with the sample made following the first condition. Fig. 6.8 shows the $m_{\gamma\gamma}$ distribution of the data and background sample in the validation region. The χ^2 test is performed for these distributions. The resulting reduced χ^2 is 0.628.[2] Fig. 6.9 compares the acoplanarity distribution, where the χ^2 test and the Kolmogorov-Smirnov test are performed. The resulting reduced χ^2 and Kolmogorov test probability are 0.917 and 0.969, respectively. There is no positive excess of the data from the background sample. This means there is no significant single-vertex background near the signal region. The number of background and data events are 1534.3 and 1465, respectively. They are compatible within the 2σ uncertainty assuming the Poisson statistics.

[2]The Kolmogorov-Smirnov test is not performed because the number of events is insufficient in some bins.

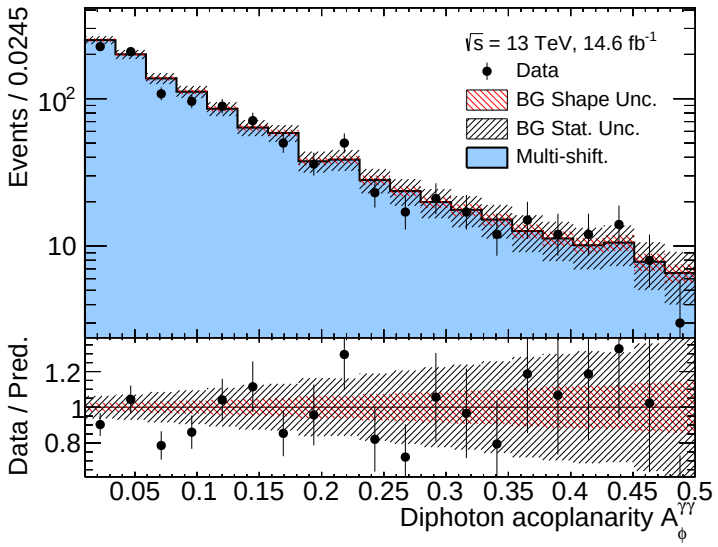


Figure 6.9: Comparison of acoplanarity distribution between the data and background sample in the validation region with the non-back-to-back requirement and only one tight photon.

6.2.2 MC-based validation of dominance of combinatorial background

An event with a diphoton and protons from the same vertex in their final state is called a single-vertex background event. Since the diphoton and protons are correlated, such an event is not a combinatorial background. If there is non-negligible single-vertex background, especially having a resonance-like peak in $m_{\gamma\gamma}$ distribution, the background modeling would be biased. In this case, the distribution of such events needs to be added to the background template. In addition, the contribution of such events to the background sample itself should be evaluated since the background sample is fully data-driven. This section evaluates the effect of single-vertex background and whether it is negligible. The MC samples of the single-vertex background considered possibly dominant are generated and analyzed at the truth level. Two generators are used to create different kinds of MC samples.

Test with SuperChic

SuperChic 4.13 is used to generate the MC events. Selection similar to the nominal one is applied to the final state of two central particles regarding them as the two photons even if they are not actually photons. The "photon" objects are required to satisfy

- $|\eta| \in [0, 1.37] \cup [1.52, 2.37]$,
- $p_{\mathrm{T}} > 40$ GeV,
- $A_{\phi}^{\gamma\gamma} < 0.01$,
- $\xi_{\gamma\gamma}^{\mathrm{A}} \in [0.031/1.1, 0.084/0.9] \vee \xi_{\gamma\gamma}^{\mathrm{C}} \in [0.031/1.1, 0.084/0.9]$,

as per the nominal selection. The following events are considered as possible single-vertex backgrounds.

- Dilepton production events analyzed in the dilepton analysis [66], $pp \to p(\gamma\gamma \to \ell^+\ell^-)p^{(*)}$. The EL and SD processes are generated separately. 100,000 events are generated for each channel. For the latter case, the matching efficiency of the side to which the dissociative proton directs is set as 0. Likewise, it is assumed that the DD contribution is negligible.

  - $pp \to p(\gamma\gamma \to e^+e^-)p^{(*)}$. The electron to photon misidentification rate is assumed to be 2.5% for $|\eta| < 0.6$, 5.5% for $|\eta| < 1.37$, and 16% for the other η region considering Ref. [136]. The product of the rates of the two particles is used as the event weight.
  - $pp \to p(\gamma\gamma \to \mu^+\mu^-)p^{(*)}$. Since the misidentification rate should be much lower than the dielectron case, this kind of event is neglected.
  - $pp \to p(\gamma\gamma \to \tau^+\tau^-)p^{(*)}$. The tauon-to-photon misidentification rate is calculated as the product of tauon to neutral pion (π^0) decay branching ratio, π^0-to-photon misidentification rate, and photon isolation efficiency for π^0. The product of the rates of the two particles is used as the event weight.

- SM light-by-light scattering, $pp \to p(\gamma\gamma \to \gamma\gamma)p^{(*)}$. 100,000 events are generated. Since the SD process of this kind of event is not implemented in this generator, SD contribution is just assumed to be the same as EL. Therefore, the event weight is 2.
- Central exclusive production (CEP). In this research, the CEP is defined as gluon-initiated events without proton dissociation. The detailed definition of CEP is in Ref. [112]. Since the SD process of this kind of event is not implemented in this generator, SD contribution is just assumed to be the same as EL. Therefore, the event weight is 2.
  - $pp \to p(gg \to \gamma\gamma)p^{(*)}$. 100,000 events are generated. There is no additional event weight. Ref. [23] supports that this event is negligible.
  - $pp \to p(gg \to q\bar{q})p^{(*)}$. 50000 events are generated. The quark-to-photon misidentification rate, photon identification efficiency multiplied by isolation efficiency, is assumed to be 0.02%. Thus, the event weight is 4×10^{-8}.
  - $pp \to p(gg \to gg)p^{(*)}$. 50000 events are generated. The same misidentification rate as the quark case is assumed. The event weight is 4×10^{-8}.

Fig. 6.10 shows the $m_{\gamma\gamma}$ stack distribution of the above single-vertex background events. Its ratio compared to the combinatorial background estimated with the nominal background sample is shown in Fig. 6.11. Since the contribution is less than the 1% of the combinatorial background, any of these single-vertex backgrounds are neglected.

Test with PYTHIA

PYTHIA 8.306 is used to generate the usual pileup events. Although they do not have intact protons in their final state, the partons could be hadronized into protons. The generated events are as follows.

- SD pileup with diphoton final state, $q\bar{q} \to \gamma\gamma$ or $gg \to \gamma\gamma$. 100,000 events are generated. The nominal selection is applied to the photons and protons. Only 0.17 events remain after the event selection, including the matching, indicating this kind of event and the double dissociative diffraction assumed to be rare is negligible.
- SD pileup with dijet final state, $gg \to jj$. Two photons or jets with the highest p_{T} are used to define diphoton, where the jet is selected probabilistically, assuming the product of jet-to-photon misidentification rate and photon isolation efficiency is 0.01. After generating 5 billion events, no event remained after the nominal selection. The upper limit of the number of events after the event selection is 0.74, calculated simply as the number of events at 14.6 fb^{-1} inferred from the generation cross section divided by 5 billion. This is negligible, and the DD pileup assumed to be rare is also considered negligible.

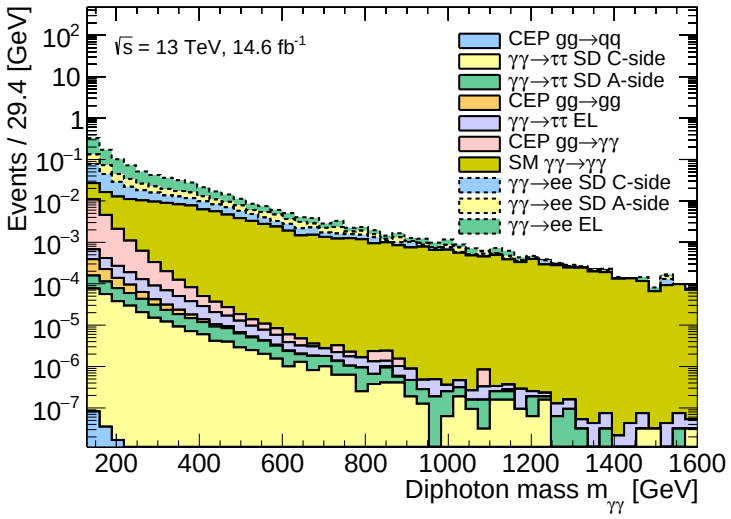


Figure 6.10: $m_{\gamma\gamma}$ stack distribution of the single-vertex background events. Each kind of event is normalized to 14.6 fb^{-1}. The legends are in ascending order in terms of the weighted integral. "SD A(C)-side" means A(C)-side proton is dissociative.

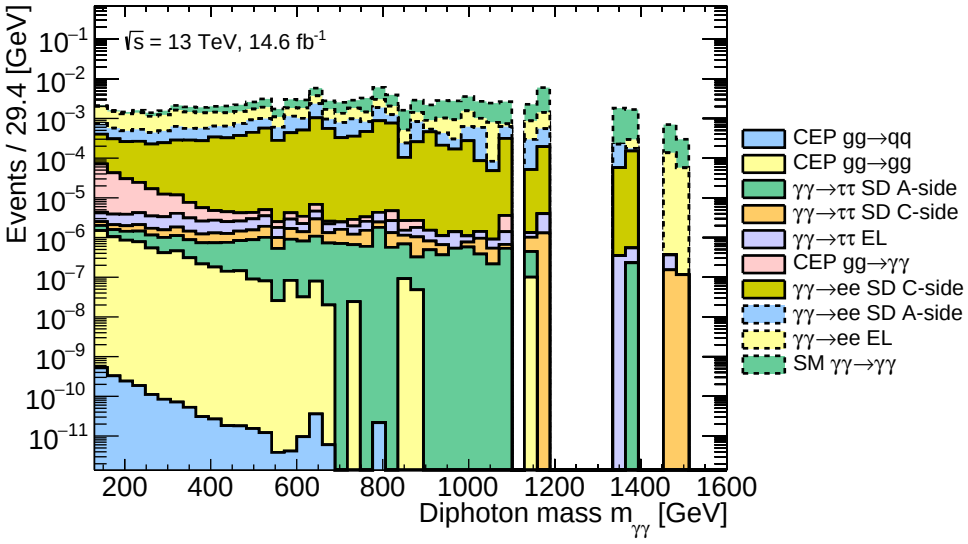
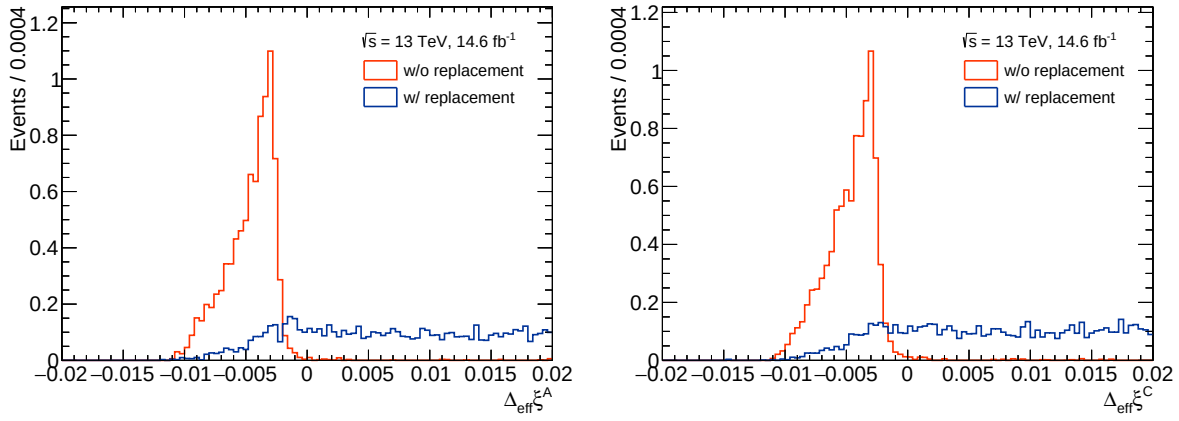


Figure 6.11: $m_{\gamma\gamma}$ stack distribution of the single-vertex background events divided by the combinatorial background distribution taken from the nominal background sample. The legends are in ascending order in terms of the weighted integral. "SD A(C)-side" means A(C)-side proton is dissociative.



(a) A-side (b) C-side

Figure 6.12: $\Delta_{\text{eff}}\xi$ distribution for the signal samples for ALPs with mass 400 GeV and coupling constant $f^{-1} = 0.05\ \text{TeV}^{-1}$ after the acoplanarity selection.

- ND pileup with dijet production. The procedure after the generation is the same as the SD $\rightarrow jj$ case. After generating 50 billion events, no event survives the nominal selection. The upper limit of the number of events after the event selection is 9.33. This number is large, but it is neglected considering that this kind of background does not produce any resonance-like peaks in the $m_{\gamma\gamma}$ distribution and that the number of events contributes only a few % compared to the combinatorial background.

Table B.1 shows each process's configurations in PYTHIA.

6.3 Background template systematic uncertainty

If there are signal events in the data, the fully data-driven background sample is contaminated by such events. This section evaluates the contributions of such events to the background template. The effect of signal contamination is estimated using the nominal signal MC samples with coupling constant $f^{-1} = 0.05\ \text{TeV}^{-1}$, which is regarded to have been excluded by the high-mass diphoton resonance search [52] (see Sec. 1.3.1). The proton sets in the signal samples are replaced by proton sets randomly extracted from the nominal data event-by-event to reproduce the data-driven proton sets reassignment in the nominal background sample creation method. These samples are called *mixed signal* samples. Fig. 6.12 compares $\Delta_{\text{eff}}\xi$ distribution after the acoplanarity selection for the 400 GeV signal samples with and without the replacement for each side. When the replacement decorrelates diphotons and protons, the peak around $\Delta_{\text{eff}}\xi = -0.003$ is flattened out. This leads to much fewer selected events after the matching selection, $\Delta_{\text{eff}}\xi < 0.004$, and is why the nominal background sample creation method suppresses the signal contamination.

However, a small number of signal events can contaminate the background sample even after the matching selection. Fig. 6.13 compares diphoton mass distribution after the whole selection for the 400 GeV signal samples with and without the proton sets replacement. A small resonant peak remains even with the replacement and after the matching selection. A merge of the mixed signal sample with the nominal background sample is created for each ALP mass. Fig. 6.14 shows the $m_{\gamma\gamma}$ distributions of such merged samples with the nominal distribution for comparison.

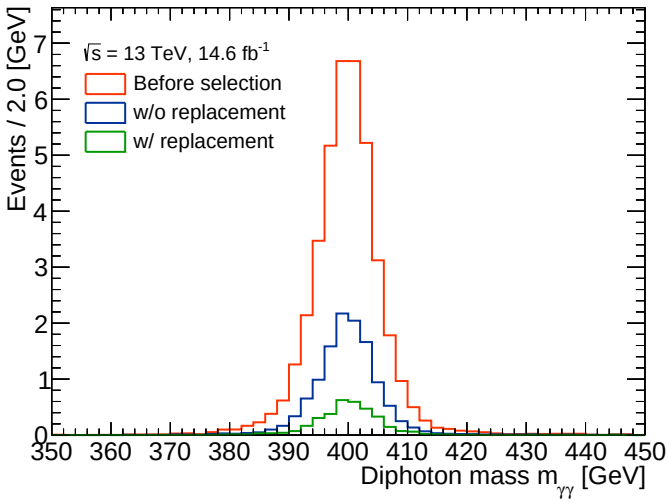


Figure 6.13: Diphoton mass distribution for the signal samples for ALPs with mass 400 GeV and coupling constant $f^{-1} = 0.05$ TeV^{-1}. The distributions with and without the proton set replacement after the whole event selection are compared. The distribution just after the acoplanarity selection is illustrated as the sample before the event selection using the proton information.

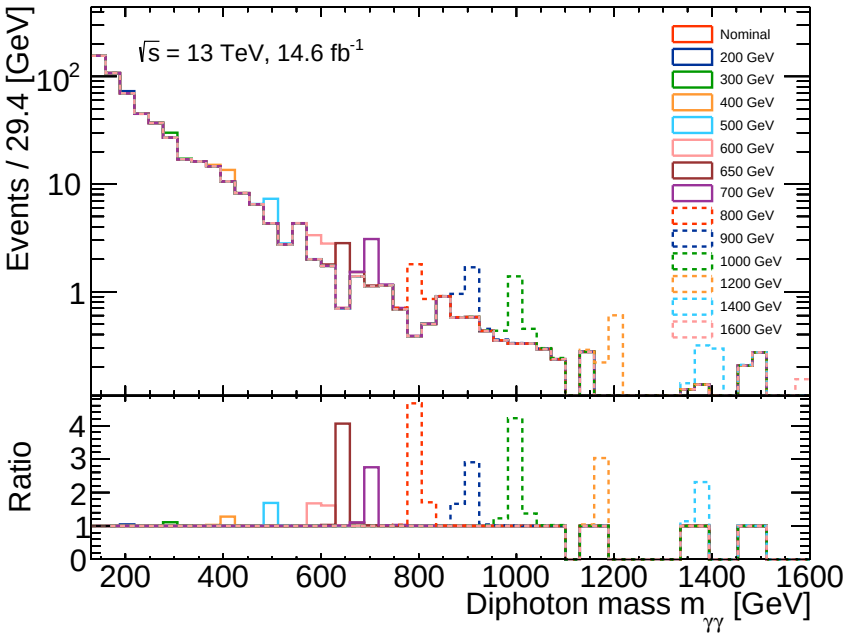


Figure 6.14: Diphoton mass distributions of the samples made by combining the background and mixed signal samples. The whole selection is applied. Each legend stands for the mass of the ALP, whose events are added to the background sample. The ALP coupling constant $f^{-1} = 0.05$ TeV^{-1} is assumed.

6.4 Background modeling uncertainty

In this section, the systematic uncertainty of the background model, Eq. 6.1, is evaluated using the background templates in Fig. 6.14. First, the templates are smoothed in Sec. 6.4.1 in order to remove the statistical uncertainty and focus on the systematic uncertainty. Then, in Sec. 6.4.2, the inflexibility of the background model is evaluated as the signal yield uncertainty originating from the background modeling, assuming the smoothed lines are the true background distribution. The evaluated systematic uncertainty is then modified considering the bias of this assumption in Sec. 6.4.3.

6.4.1 Functional Decomposition smoothing

The statistical uncertainty is desired to be removed when evaluating the systematic uncertainty of the background modeling so as not to overestimate the latter. The data-driven Functional Decomposition (FD) method [143] is used to smooth the background distribution to remove the statistical uncertainty. The FD method is originally implemented for resonance searches and generally takes multiple assumptions about the resonance peak shapes. However, if no signal shape is assumed, the process becomes much simpler and behaves as a background distribution smoother. The background template from the diphoton mass $m_0 = 130$ GeV is input to the tool, and the FD method expresses the background distribution using a combination of complete orthogonal functions,

$$\tilde{f}(z) = \sum_{n=0}^{\infty} c_n \tilde{E}_n(z),$$

where c_n is the real coefficient, and

$$\begin{aligned}
z &= \left(\frac{m_{\gamma\gamma} - m_0}{\lambda}\right)^{\alpha}, \\
\tilde{E}_n(z) &= \sum_{m=0}^{n} d_{nm} \tilde{F}_m(z), \\
\tilde{F}_n(z) &= \sqrt{2} e^{-nz}, \\
d_{nm} &= \sqrt{n}(-1)^{n+m} \left(\frac{2m}{n+m}\right) \prod_{i=1}^{m-1} \frac{m+i}{m-i} \prod_{i=m+1}^{n} \frac{i+m}{i-m}, \\
\left\langle \tilde{E}_n, \tilde{E}_m \right\rangle &= \delta_{nm}, \\
\left\langle \tilde{a}, \tilde{b} \right\rangle &= \int_0^{\infty} \mathrm{d}z \tilde{a}(z) \tilde{b}(z).
\end{aligned}$$

The hyperparameters λ and α are positive scalars. λ is the scale parameter similar to the bandwidth in the usual kernel density estimation and has the same dimension as mass, while α is dimensionless. The exponential family is used for the bases because the simple exponential is an *entropy-maximizing* distribution, i.e. most delocalized, and they are suitable for representing the smooth-falling background distribution. Although the background is modelled in the mass range [150, 1600] GeV, the x_0 is set to 130 GeV with the 20 GeV margin because the FD results have an oscillating behaviour around the left end. The coefficients $\{d_{nm}\}$ for the orthogonalization can be calculated recurrently and used to derive recurrence relations among $\{\tilde{E}_n\}$, which are used for the computation of $\tilde{E}_n$.

The input sample is processed as the sequence of $m_{\gamma\gamma}$ values in $m_{\gamma\gamma} \in [130, 1600]$ GeV, $\{m_{\gamma\gamma,i}\}$, and they are converted to $\{z_i\}$. The conventional approach to obtain the coefficients $\{c_n\}$ is the maximum-likelihood

method, but the FD uses another method for better performance as follows. Each coefficient is computed as

$$c_n = \frac{1}{M}\sum_{i=1}^{M}\tilde{E}_n\left(z_i\right),$$

where M is the number of events, and this is referred to as the decomposition of the sample. The expression of the sample is truncated to

$$\tilde{\Omega}(z) = \sum_{n=0}^{\mathcal{N}} c_n \tilde{E}_n(z),$$

which corresponds to the distribution smoothing. The $\mathcal{N}$, the degree to which the FD describes the input sample in detail, is another hyperparameter. Too large $\mathcal{N}$ causes a large modeling bias, i.e. overfitting, while if $\mathcal{N}$ is too small, the variance will increase.

The hyperparameters λ, α, and $\mathcal{N}$ are optimized so that

$$D_{\mathrm{KL}}\left(\tilde{f}\|\tilde{\Omega}\right) + D_{\mathrm{KL}}\left(\tilde{\Omega}\|\tilde{\mathcal{P}}\right)$$

is minimized, where

$$D_{\mathrm{KL}}\left(\tilde{a}\|\tilde{b}\right) \equiv \int_{-\infty}^{\infty}\tilde{a}(z)\log\left(\frac{\tilde{a}(z)}{\tilde{b}(z)}\right)\mathrm{d}z$$

is the Kullback-Leibler (KL) divergence, which is a measure of the difference between the two distributions given as its arguments. The first term requires the FD result to express the sample well. $\tilde{\mathcal{P}} = \sqrt{2}e^{-z}$ is a simple exponential function, so the second term suppresses the overfitting. The maximum allowed $\mathcal{N}$ is set to 8 in this research. Finally, the obtained $\tilde{\Omega}(z)$ is transformed to a function of $m_{\gamma\gamma}$,

$$\Omega(m_{\gamma\gamma}) = \sum_{n=0}^{\mathcal{N}} c_n \left(\frac{\mathrm{d}z}{\mathrm{d}m_{\gamma\gamma}}\right)\tilde{E}_n(z).$$

Fig. 6.15 shows this FD smoothing result of the nominal background sample, and Fig. 6.16 shows all the FD results for the background templates in Fig. 6.14. The resulting hyperparameters for the nominal case are $\lambda = 106$, $\alpha = 0.445$, and $\mathcal{N} = 6$.

6.4.2 Spurious signal evaluation

The inflexibility of the background model, Eq. 6.1, is evaluated using the smoothed lines, assuming the nominal FD smoothing result is the true background distribution. This method inherits the previous diphoton analyses [52, 129, 144]. The signal-contaminated cases are also tested to see the effect of the potential bias of the fully data-driven background sample creation method due to the signal contamination. Samples with a large number of events weighted to have the same integral as the smoothed lines, so-called Asimov samples, are created from the lines one by one. A primitive MC method is used to generate the samples. Each of randomly generated $m_{\gamma\gamma}$ values following a uniform distribution in the mass range $[150, 1600]$ GeV is selected if $\Omega(m_{\gamma\gamma})$ is smaller than a value randomly generated following another uniform distribution in $[0, \Omega(150\ \mathrm{GeV})]$. The event generation is repeated until 20 million $m_{\gamma\gamma}$ values are selected. The signal-plus-background (s+b) function,

$$N_X f_X(m_{\gamma\gamma}; \mu_{\mathrm{CB}}, \sigma_{\mathrm{CB}}, n_{\mathrm{R}}, n_{\mathrm{L}}, \alpha_{\mathrm{R}}, \alpha_{\mathrm{L}}) + N_{\mathrm{b}} f_{\mathrm{b}}(m_{\gamma\gamma}; a, \alpha_0) \tag{6.2}$$

is then fitted to the Asimov data for each hypothetical ALP mass m_X, where the signal function f_X and background function f_{b} are defined in Eqs. 5.1 and 6.1, respectively. The N_X and N_{b} are the signal and background yield, respectively. The signal function parameters are fixed to the values in Table 5.1 in order

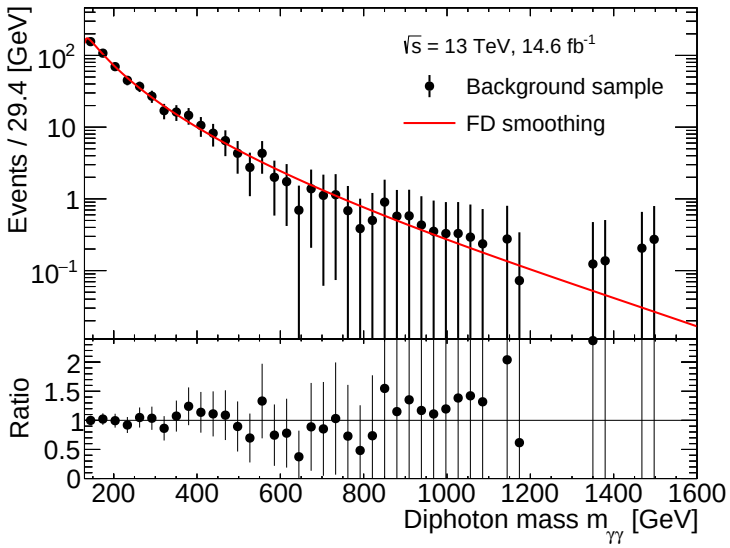


Figure 6.15: Comparison of $m_{\gamma\gamma}$ distribution between the nominal background sample with statistical uncertainty (black) and the FD smoothing result (red).

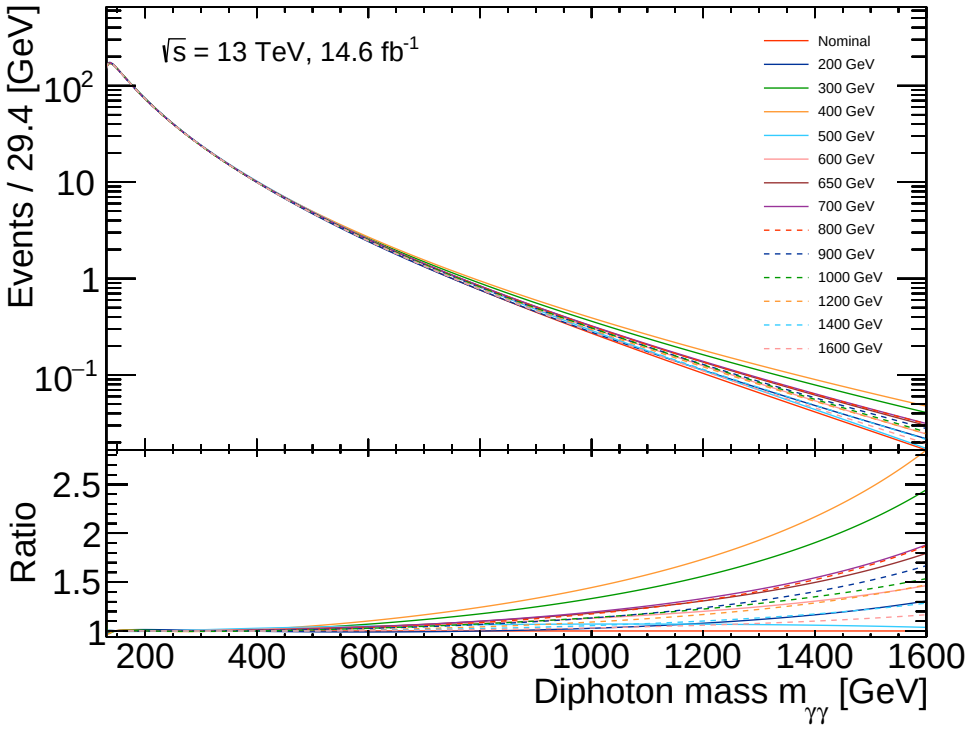
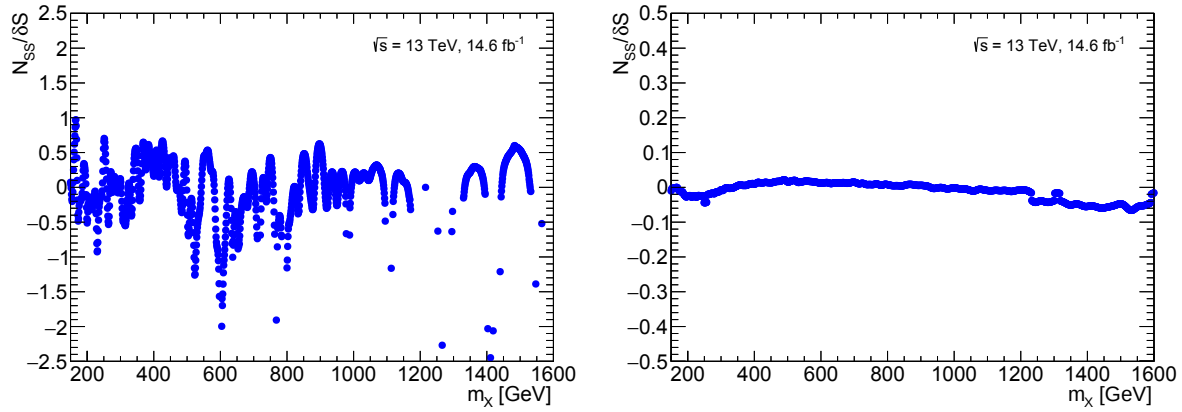


Figure 6.16: FD smoothing results of the background templates for the nominal and the signal-contaminated cases. Each legend stands for the mass of the ALP, whose events are added to the background sample. The ALP coupling constant $f^{-1} = 0.05$ TeV^{-1} is assumed.



(a) Nominal background sample (b) FD driven Asimov data

Figure 6.17: m_X dependency of SS relative to the statistical uncertainty determined using the s+b function and (a) the nominal background sample (b) the FD driven Asimov data.

Table 6.1: Spurious signal envelope parametrization.

Parameter	Value
p_0	0.019
p_1	1.2
p_2	−0.0054

to distinguish the sources of systematic uncertainties, while the background function parameters are set to be variable. Since it is assumed that the FD smooth line is the true background distribution, the fitting result of N_X should be exactly 0. However, it can actually be non-zero because the background distribution shape model has only two free parameters. This incompleteness of the background model is expressed as the non-zero N_X values, called *spurious signal* (SS) and denoted N_{SS}, for each ALP mass hypothesis.

Fig. 6.17 (b) shows the SS evaluation results relative to its uncertainty δS. Traditionally, $|N_{\mathrm{SS}}|/\delta S < 50\%$ is set as the threshold for considering the background model to be flexible.[3] The SS result for the nominal background sample illustrated in Fig. 6.17 (a) does not pass the criterion because the statistical fluctuation deteriorate the result. On the other hand, the result after the smoothing (b) is much better than (a) and pass the test. The unstable fitting behaviour of the negative N_{SS} results does not cause any problem in this research because the test statistics do not use such values as described in Sec. 8

An envelope of $|N_{\mathrm{SS}}|$ is taken as the bare background modeling systematic uncertainty as a function of m_X. Fig. 6.18 shows the $|N_{\mathrm{SS}}|$ and $|N_{\mathrm{SS}}|/\delta S$ for all the Asimov samples as well as the envelope, which is parametrized as a function of ALP mass. The parametrization is obtained from a fit of a function

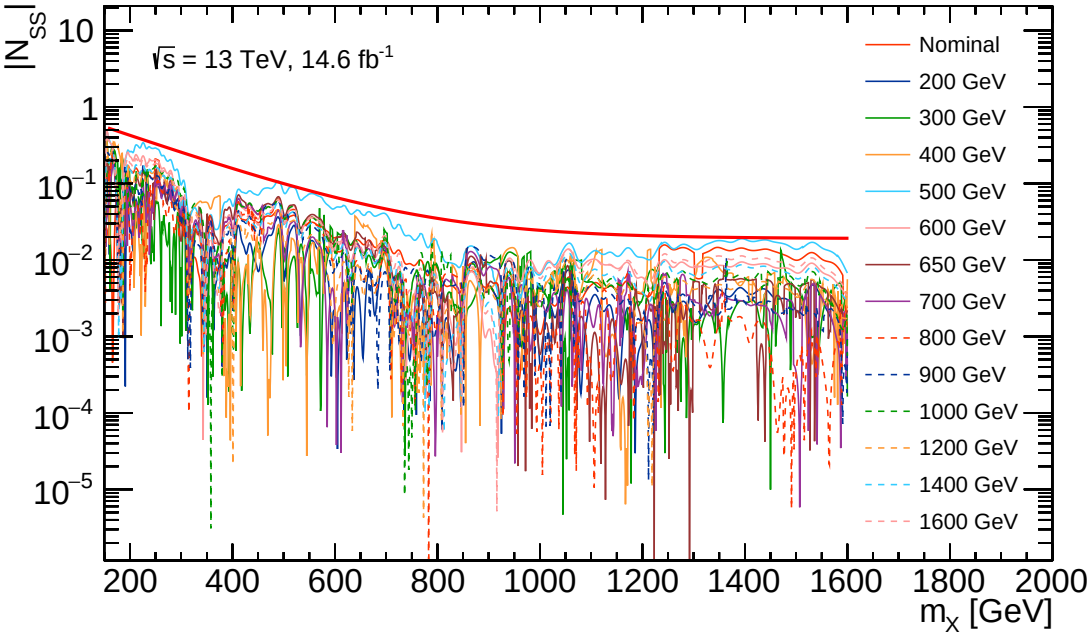
$$\delta_{\mathrm{SS}}(m_X) = p_0 + p_1 e^{p_2 m_X}, \tag{6.3}$$

to the local maxima, where the parameters are listed in Table 6.1.

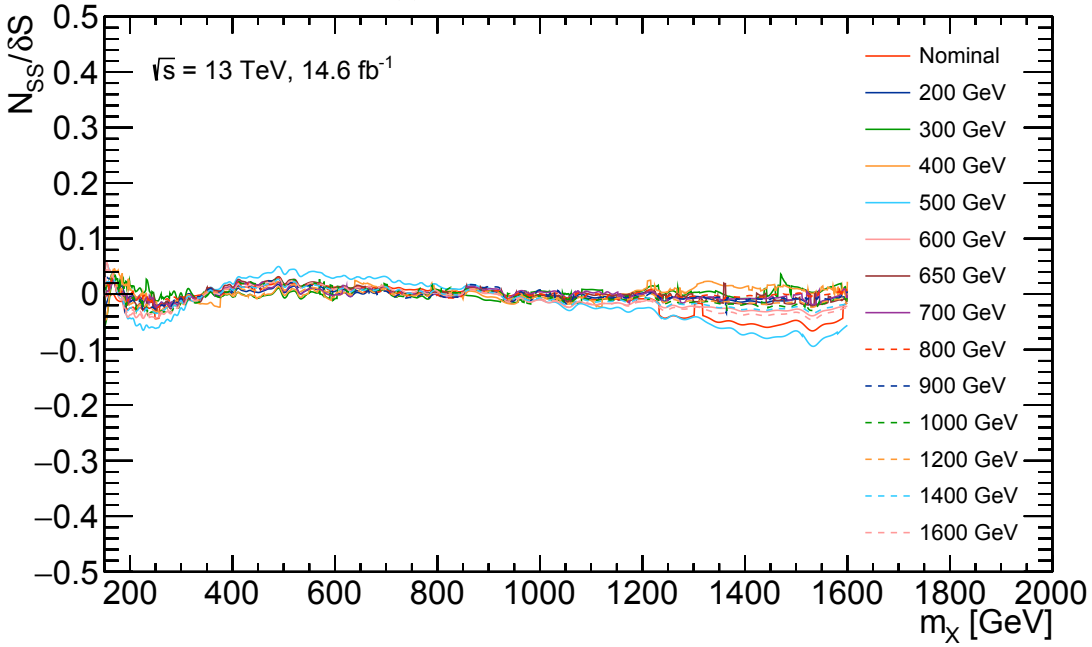
6.4.3 Bias of Functional Decomposition

The systematic uncertainty evaluation with the SS test assumes that the smooth distribution created by the FD method is identical to the true background distribution. However, obviously, this assumption

[3] Since no assumption is made on the signal production cross section, there is no strict requirement on $|N_{\mathrm{SS}}|/\delta S$.



(a) Absolute number of SS



(b) Relative number of SS

Figure 6.18: (a) Absolute number of SS as functions of m_X evaluated for the Asimov samples. The thick red line is the result of a fit of an analytical function to the local maxima of the graphs. (b) Number of SS relative to the statistical uncertainty obtained from the Asimov samples.

should not be completely correct. Therefore, the background modeling systematic uncertainty, Eq. 6.3, is modified considering such bias evaluated using some toy MC samples. First, the following three $m_{\gamma\gamma}$ distribution functions are prepared in the range [130, 1600] GeV.

1. The background function, Eq. 6.1. The parameters are determined by fitting the function to the nominal data just after the acoplanarity selection.
2. The background function 1 multiplied by the $\xi_{\gamma\gamma}$ OR selection efficiency, Eq. 4.4. This has a non-smooth shape illustrated in Fig. 4.19 and hence is an extreme case to be conservative in the bias evaluation.
3. The background function 1 multiplied by the $\xi_{\gamma\gamma}$ and matching OR selection efficiency, Eq. 4.5. This function is more realistic than the function 2.

Assuming that each of these distribution functions is the true background distribution, the difference in the SS evaluation results between the cases where the distribution is smoothed and not smoothed by the FD method is investigated. A similar approach is taken in Ref. [144].

Fig. 6.19 shows the procedure as well as the nominal SS evaluation procedure. For each of the three distribution functions, there are two different paths as follows.

- The case without the FD smoothing is checked. An Asimov sample with 20 million events and integral normalized to the nominal background sample in [150, 1600] GeV is created from the assumed distribution function. The generation method is the same as described in Sec. 6.4.2. The SS is evaluated for the sample as per the nominal SS evaluation.
- The case with the FD smoothing is checked. 100 toy samples with the same numbers of events as the nominal background sample are generated in [130, 1600] GeV from the same distribution function with the same method as the Asimov sample creation. These toys are also normalized to the nominal background sample in terms of the integral. The toy samples are then smoothed by the FD method, and an Asimov sample is created from each smooth line in [150, 1600] GeV. Subsequently, the SS evaluation is performed for each Asimov sample. The average of the 100 SS results is then calculated.

The results of the two evaluations are illustrated in order in Figs. 6.20 and 6.21.[4] The SS result of the function 2 without the FD smoothing has a non-smoothed shape as expected, while such behaviour is not visible for the FD smoothing case.

The two results are compared with each other for each initial distribution function. Fig. 6.22 shows the difference between these results in terms of the number of SS relative to the statistical uncertainty. An analytical function

$$\delta_{\mathrm{FD}}(m_X) = \sum_{i=0}^{5} p_i m_X^i$$

is used as the envelope. The coefficients are listed in table 6.2. The absolute signal yield uncertainty originating from the background modeling is then modified as

$$\delta_{\mathrm{BG}}(m_X) = \delta_{\mathrm{SS}}(m_X)\,(1 + \delta_{\mathrm{FD}}(m_X))\,, \tag{6.4}$$

where δ_{SS} is taken from Eq. 6.3 and Table 6.1.

[4] Some mass points are removed because the fit did not converge, but this does not cause any problem in this research as long as the fit results for the nominal data converge since the background modeling uncertainty has nothing to do with local convergence problems.

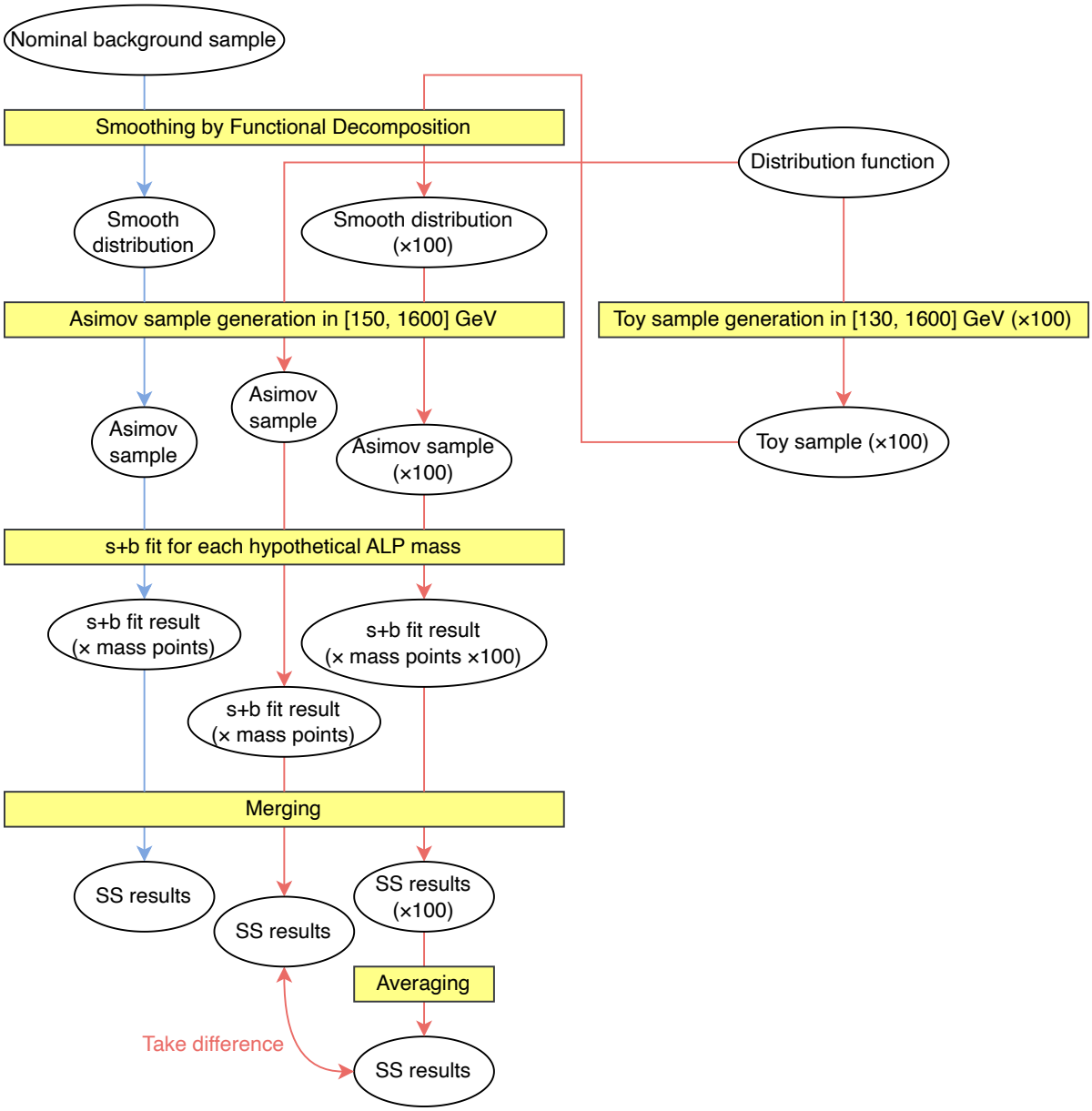
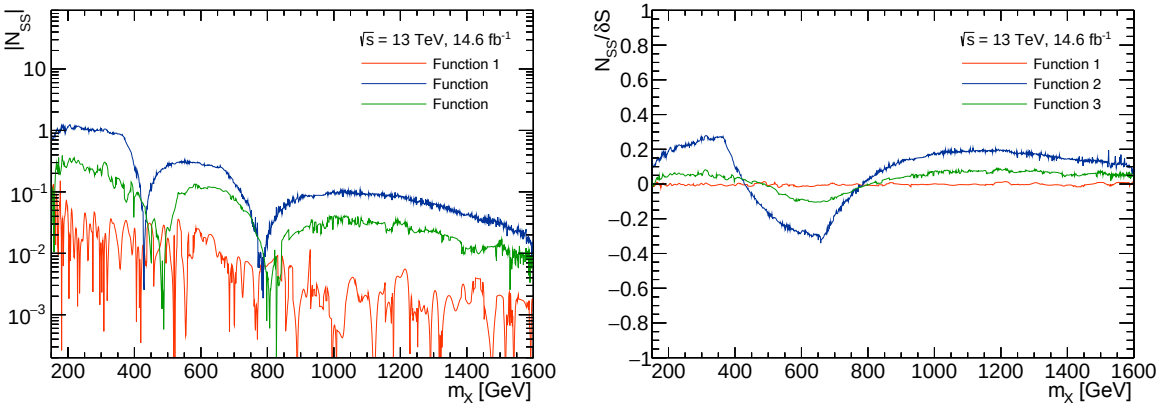


Figure 6.19: The procedures of the nominal SS evaluation (blue) and the FD smoothing bias evaluation (red). Ellipses represent objects, and rectangles represent processes. The FD bias study procedure is repeated for three different background functions.

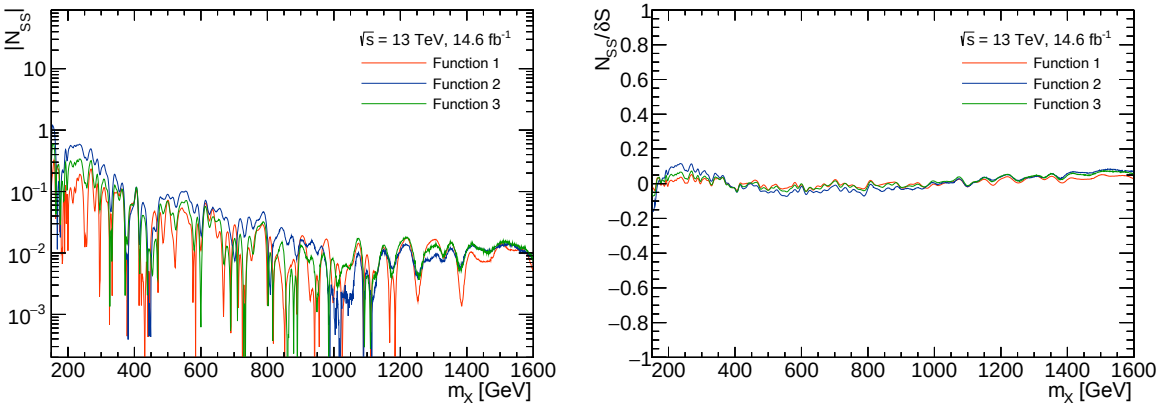
Table 6.2: FD bias parameters.

Parameter	Value
p_0	0.460
p_1	−0.00185
p_2	5.64×10^{-06}
p_3	-7.05×10^{-09}
p_4	3.81×10^{-12}
p_5	-7.54×10^{-16}



(a) Absolute number of SS (b) Relative number of SS

Figure 6.20: SS evaluation results for the FD bias evaluation derived without the FD smoothing. (a) Absolute number of SS as a function of m_X. (b) SS relative to the statistical uncertainty.



(a) Absolute number of SS (b) Relative number of SS

Figure 6.21: SS evaluation results for the FD bias evaluation derived with the FD smoothing. (a) Absolute number of SS as a function of m_X. (b) SS relative to the statistical uncertainty.

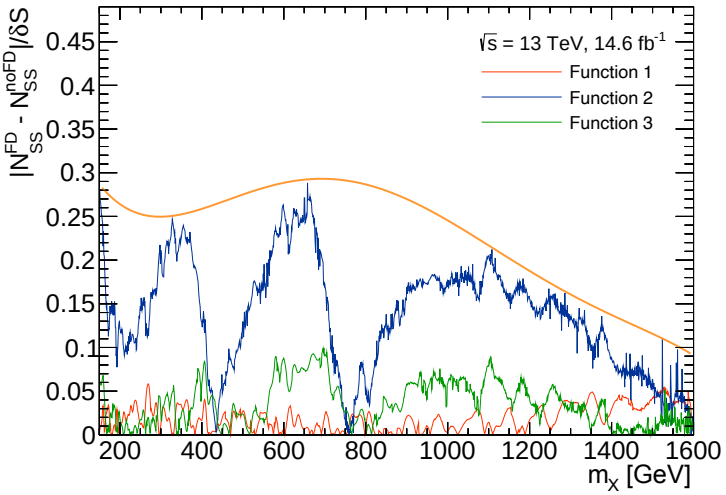


Figure 6.22: FD smoothing bias defined as the difference in SS evaluation results between the with and without FD smoothing cases in terms of the number of SS relative to the statistical uncertainties. The thick orange line is the envelope.

7 Systematic Uncertainty

Various sources of systematic uncertainty in this research are considered, quantified, and summarized in this section. They are categorized into photon object-related, proton object-related, fitting-related, and theoretical uncertainty, and described in Sec. 7.1, 7.2, 7.3, and 7.4, respectively.

To evaluate the effect of each systematic uncertainty, basically, the value of a parameter that is supposed to have uncertainty is shifted up and down by the amount of its uncertainty (referred to as shifting in this research), and the signal MC samples are recreated. The relative difference in each of

- mean of $m_{\gamma\gamma}$ distribution,
- standard deviation of $m_{\gamma\gamma}$ distribution,
- number of events normalized to 14.6 fb^{-1}

from the nominal signal MC sample are evaluated for each ALP mass m_X after the whole selection. In most cases, the maximum relative difference considering all the ALP mass points is taken as the systematic uncertainty for the corresponding source. The uncertainty is propagated to μ_{CB}, σ_{CB} (see Eq. 5.1), and signal yield, respectively. Non-negligible contributions among all the systematic uncertainties are summarized in Table 7.1 with the quantitative information, while negligible sources are not listed.

The following two uncertainties, which do not belong to any aforementioned category, are also considered.

- Integrated luminosity uncertainty δ_{lumi}. For the 2017 dataset, the uncertainty is 2.4% [82]. This is directly used as the relative uncertainty of signal yield.
- Proton beam energy uncertainty. The 6.5 TeV beam energy in 2017 has 0.1% uncertainty [145], which is negligible.

7.1 Photon object-related uncertainty

The following sources are considered for the photon object-related systematic uncertainty.

- Photon energy scale uncertainty originating from the photon energy calibration described in Sec. 3.4. After the shifting, the resulting distributions are compared with the nominal sample in Fig. 7.1 for m_X = 1000 GeV. The mean values of the modified $m_{\gamma\gamma}$ distributions subtracted by m_X are plotted as a function of m_X and compared to those of the nominal $m_{\gamma\gamma}$ distributions in Fig. 7.2. The maximum relative difference is regarded as the uncertainty of μ_{CB}.
- Photon energy resolution uncertainty originating from the photon energy calibration described in Sec. 3.4. This primarily affects the width of $m_{\gamma\gamma}$ distribution. The shift is applied as the change of the degree of photon kinematics smearing in the simulation. The resulting distribution after the shifting is compared with the nominal distribution in Fig. 7.3. The standard deviations of the nominal and systematically varied $m_{\gamma\gamma}$ distributions are plotted as a function of m_X in Fig. 7.4. The maximum relative difference is regarded as the uncertainty of σ_{CB}.
- Photon-related efficiency uncertainty. This primarily affects the signal yield. The MC-to-data correction scale factor of event weight is varied to consider this uncertainty.

Table 7.1: Summary of the non-negligible systematic uncertainty considered in this research.

Symbol	Systematic uncertainty description	Value
	$m_{\gamma\gamma}$ scale uncertainty	
δ_{scale}	Photon energy scale	$\pm 0.1\%$
	$m_{\gamma\gamma}$ resolution uncertainty	
$\delta_{\sigma_{\text{scale}}}$	Photon energy scale	$^{+1.6}_{-2.1}\%$
δ_{res}	Photon energy resolution	$^{+14.1}_{-4.8}\%$
$\delta_{\sigma_{\text{PRW}}}$	Pileup reweighting	$^{+2.1}_{-1.4}\%$
	Signal yield uncertainty	
δ_{lumi}	Integrated luminosity	$\pm 2.4\%$
δ_{ID}	Photon identification efficiency	$^{+1.6}_{-1.5}\%$
δ_{iso}	Photon isolation efficiency	$\pm 1.9\%$
δ_{PRW}	Pileup reweighting	$^{+2.7}_{-2.6}\%$
δ_{optics}	Beam optics	$^{+0.8}_{-3.4}\%$
δ_{glob}	AFP global alignment	$^{+10.0}_{-8.6}\%$
$\delta_{\varepsilon_{\text{AFP}}}$	Proton reconstruction efficiency	$^{+3.0}_{-2.2}\%$
δ_{Nclust}	Required number of AFP clusters	$^{+0}_{-6.6}\%$
δ_{smear}	AFP track smearing	$^{+2.1}_{-1.5}\%$
$\delta_{0.1\xi_{\gamma\gamma}}$	Selection correction	$\pm 2.2\%$
δ_{smodel}	Signal modeling linearity	$\pm 2.1\%$
δ_{BG}	Background modeling (absolute uncertainty)	Eq. 6.4
	Signal efficiency uncertainty	
$\delta_{\varepsilon\text{EL}}$	EL signal efficiency	Eq. 5.3
$\delta_{\varepsilon\text{SD}}$	SD signal efficiency	Eq. 5.4
$\delta_{\varepsilon\text{DD}}$	DD signal efficiency	Eq. 5.5
	Signal cross section uncertainty	
$\delta_{S^2\text{EL}}$	EL signal soft survival factor	$\pm 2\%$
$\delta_{S^2\text{SD}}$	SD signal soft survival factor	$\pm 10\%$
$\delta_{S^2\text{DD}}$	DD signal soft survival factor	$\pm 50\%$
	Signal model parameter uncertainty	
$\delta_{\Delta m0,1}$	Signal model parameter $p_{\Delta m0,1}$ (absolute uncertainty)	Table 5.1
$\delta_{\sigma_{\text{CB}}0,1}$	Signal model parameter $p_{\sigma_{\text{CB}}0,1}$ (absolute uncertainty)	Table 5.1
$\delta_{\alpha_{\text{L,R}}}$	Signal model parameter $p_{\alpha_{\text{L,R}}}$ (absolute uncertainty)	Table 5.1
$\delta_{n_{\text{L,R}}}$	Signal model parameter $p_{n_{\text{L,R}}}$ (absolute uncertainty)	Table 5.1

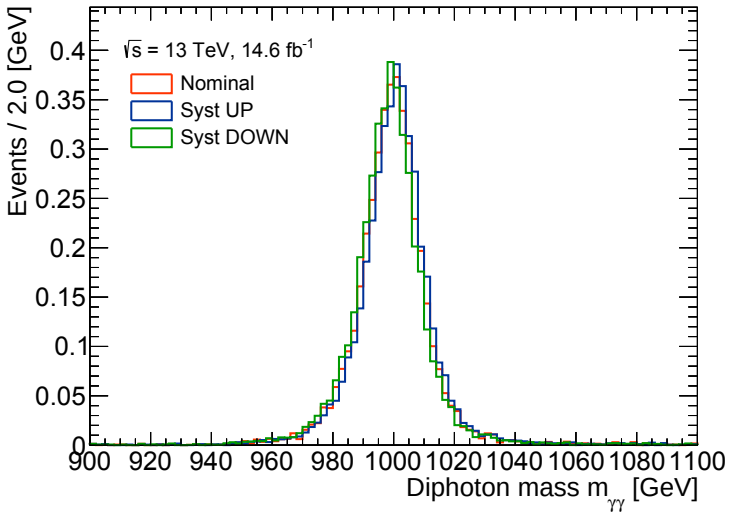
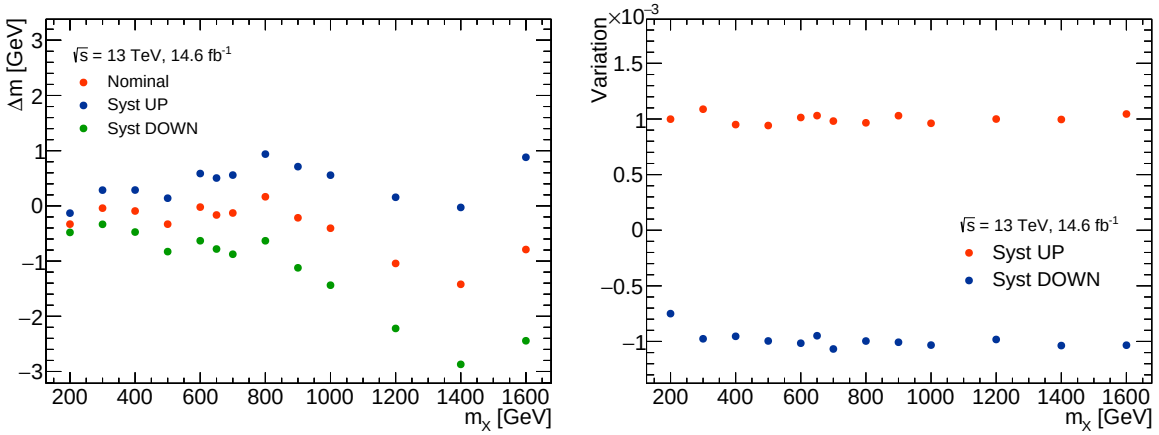


Figure 7.1: Invariant mass distributions of the $m_X = 1000$ GeV simulated signal after the selection obtained with the nominal and systematically shifted photon energy scales.



(a) $\langle m_{\gamma\gamma} \rangle - m_X$ (b) $\left(\left\langle m_{\gamma\gamma}^{\text{up/down}} \right\rangle - \langle m_{\gamma\gamma}^{\text{nominal}} \rangle \right) \Big/ \langle m_{\gamma\gamma}^{\text{nominal}} \rangle$

Figure 7.2: (a) The mean of the $m_{\gamma\gamma}$ distribution subtracted by m_X as a function of m_X for the signal samples after the whole selection for the nominal and systematically shifted photon energy scales. (b) The relative difference of the mean between the shifted sample and the nominal sample after the whole selection.

  - Photon identification efficiency uncertainty δ_{ID} (see Sec. 3.4).
  - Photon isolation efficiency uncertainty δ_{iso} (see Sec. 3.4).
  - Photon trigger efficiency uncertainty [111] (see Sec. 3.1).
- Pileup reweighting described in Sec. 3.2. A shifting in the pileup reweighting of the simulation is performed to cover the difference between the predicted and measured inelastic pp cross section [146]. This primarily affects the signal yield. The μ distribution derived from the data is shifted by ±3% before reweighting the MC sample according to the distribution.

7.2 Proton object-related uncertainty

The proton object-related systematic uncertainty is described in this section. The position of the truth proton is smeared to create an AFP track object in the nominal AFP simulation as described in Sec. 3.2.

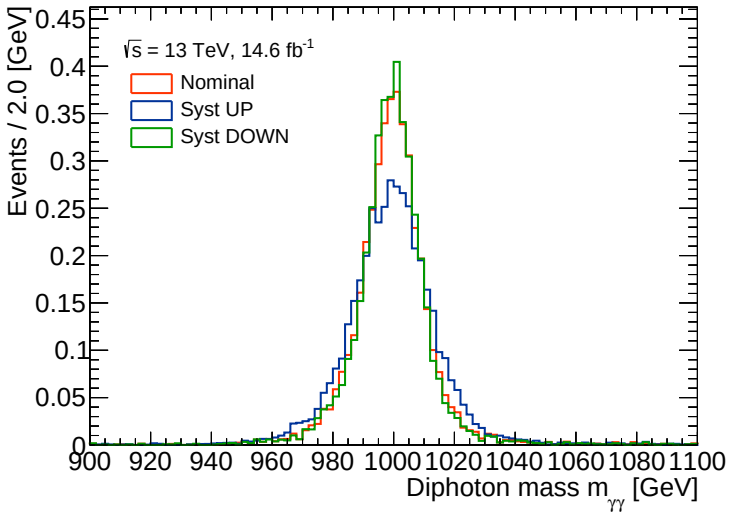
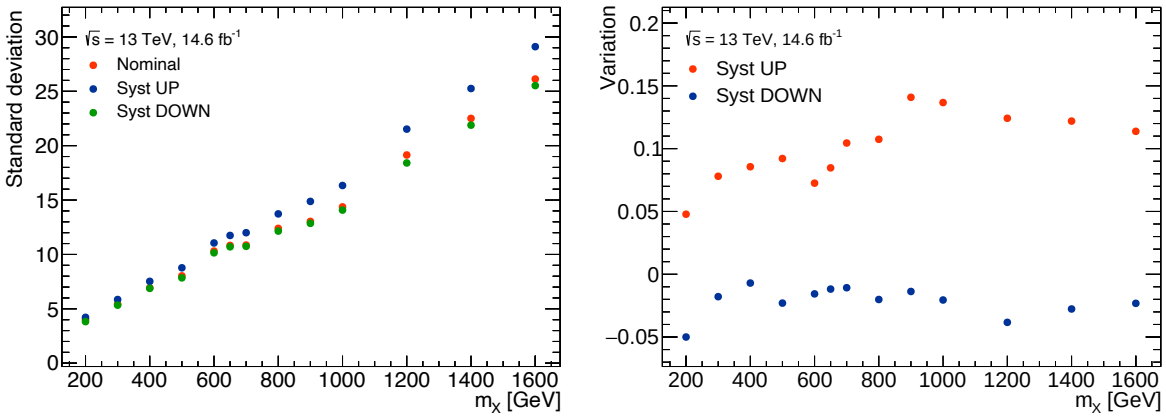


Figure 7.3: Invariant mass distributions of the $m_X = 1000$ GeV simulated signal after the selection obtained with nominal and systematically shifted photon energy resolutions.



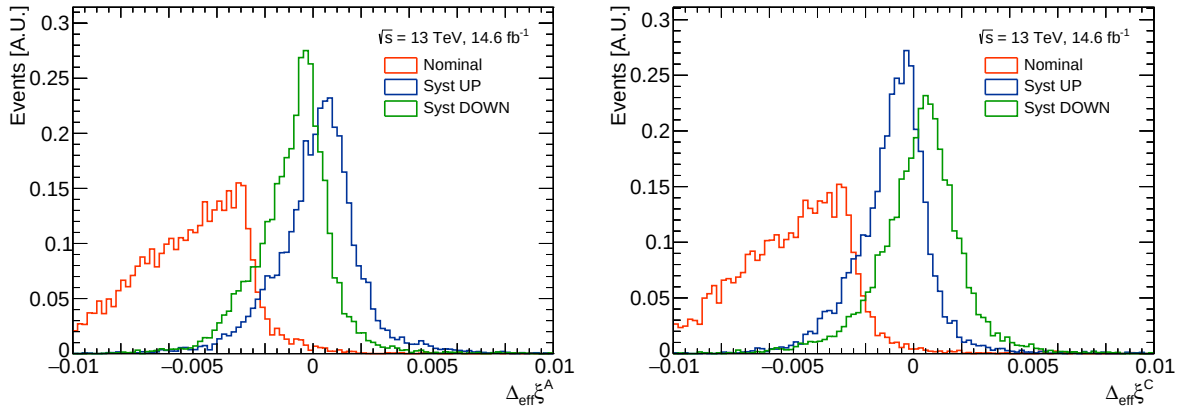
(a) Standard deviation (b) Relative difference

Figure 7.4: (a) The standard deviation of the $m_{\gamma\gamma}$ distribution as a function of m_X for the signal samples after the whole selection for the nominal and systematically shifted photon energy resolutions. (b) The relative difference of the standard deviation between the shifted sample and the nominal sample after the whole selection.

However, in tests for some systematic uncertainty sources, the simulation starts by creating cluster objects (see Sec. 3.5) with similar smearing. In this section, such samples are referred to as the *clustered MC* samples. The following systematic uncertainty sources are considered.

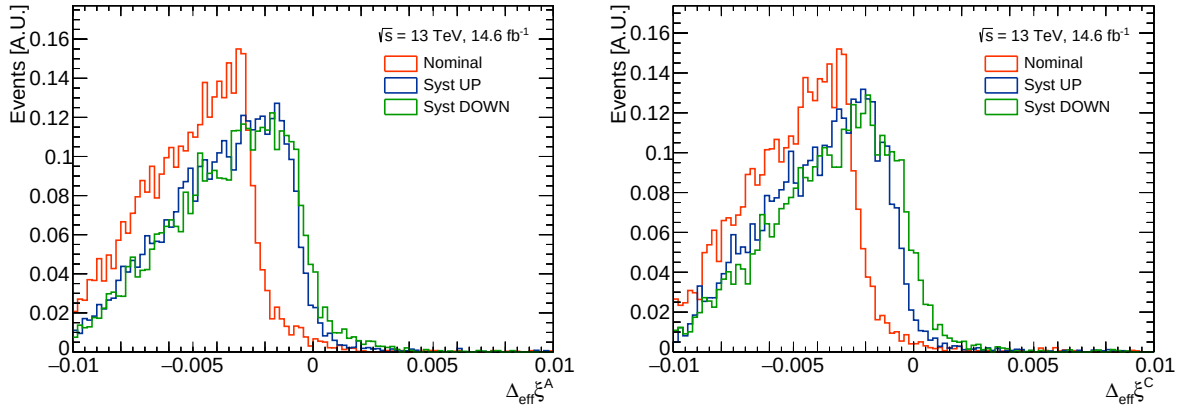
- Beam optics and beam angle uncertainty in the proton transport simulation (see Sec. 3.5). Fig. 7.5 shows the $\Delta_{\text{eff}}\xi$ distributions of the nominal and systematically shifted signal MC samples for the 1000 GeV ALP. Despite the significant shift, most signal events are selected because of the correction term described in Sec. 4.4. Fig. 7.7 (a) shows that the beam optics systematic uncertainties are below 5% for all the mass points.
- AFP alignment uncertainty.
  - AFP global alignment uncertainty is 300 μm as described in Sec. 2.3.1. Fig. 7.6 shows the $\Delta_{\text{eff}}\xi$ distributions of the nominal and systematically shifted signal MC samples for the 1000 GeV ALP. Although the dominant source of the ξ_{AFP} uncertainty described in Sec. 3.5 is the beam optics uncertainty, the global alignment uncertainty becomes the dominant source of the signal yield uncertainty, after applying the correction term in the event selection (see Sec. 4.4) as shown in Fig. 7.7. Such a difference in behaviour from the beam optics uncertainty is due to the correlation between the A-side and C-side, which can be found by comparing Figs. 4.11 and C.1. The global alignment variation shifts $\xi^{\text{A}}_{\text{AFP}}$ and $\xi^{\text{C}}_{\text{AFP}}$ towards the same direction, changing the selection efficiency drastically at some particular ALP mass points.
  - AFP local alignment uncertainty (see Sec. 2.3.1).
  - Roman pot rotation uncertainty. The pots are rotated about the z-axis by 4 mrad to evaluate the uncertainty.
- Proton reconstruction efficiency uncertainty.
  - The nominal efficiency is $92 \pm 2\%$ as described in Sec. 3.5.
  - Potential efficiency deterioration by showering from signal protons. The tag-and-probe measurement of the nominal proton reconstruction efficiency might not be fully sensitive to such effects. Therefore, the following three configurations in the proton reconstruction process are conservatively tightened, and the decrease in the number of events is regarded as the showering effect.
    - The allowed transverse distance between AFP SiT clusters to reconstruct an AFP track is decreased from the default of 500 μm to 400 μm. The clustered MC samples are used.
    - The allowed transverse distance between tracks in the NEAR and FAR stations to reconstruct a proton is tightened from the default of 2 mm to 1 mm.
    - The minimum required number of clusters to reconstruct an AFP track is increased from the default of 2 to 3 for the data sample to quantify the decrease of proton reconstruction efficiency as detailed in Appendix F. The derived value is used as the proton reconstruction efficiency and applied for the shift as per the nominal efficiency uncertainty case.
- AFP track smearing uncertainty. A Gaussian width $\sigma_x = 10$ μm was used for smearing the truth proton position to reconstruct a track. This reflects the spacial resolution of the AFP 3D-Si sensors. However, the dilepton analysis [66] found that $\sigma_x = 50$ μm width provided reasonable modeling of the signal for the selection criteria $\xi_{\text{AFP}} \in [0.035, 0.08]$ and $|\Delta\xi| < 0.005$. Thus, the configuration is changed to $\sigma_x = 50$ μm to consider this as a source of the systematic uncertainty.[1] Fig. 7.8 shows that most signal events pass the matching selection $\Delta_{\text{eff}}\xi < 0.004$ even after the shifting.
- Pileup proton overlay uncertainty. The nominal signal MC samples do not have pileup proton objects even at the truth level because generating pileup events with reasonable uncertainty is difficult.

[1] This is because the origin of the protons is the same as the dilepton analysis in that they are photon-induced. Although the true ξ distribution is different from the dilepton analysis, the $\sigma_x = 50$ μm is used because this systematic uncertainty is not dominant even if the variation doubles.



(a) A-side (b) C-side

Figure 7.5: $\Delta_{\text{eff}}\xi$ distributions of the $m_X = 1000$ GeV simulated signal before the selection obtained with nominal and systematically varied beam optics configuration.



(a) A-side (b) C-side

Figure 7.6: $\Delta_{\text{eff}}\xi$ distributions of the $m_X = 1000$ GeV simulated signal before the selection obtained with nominal and systematically varied global alignment.

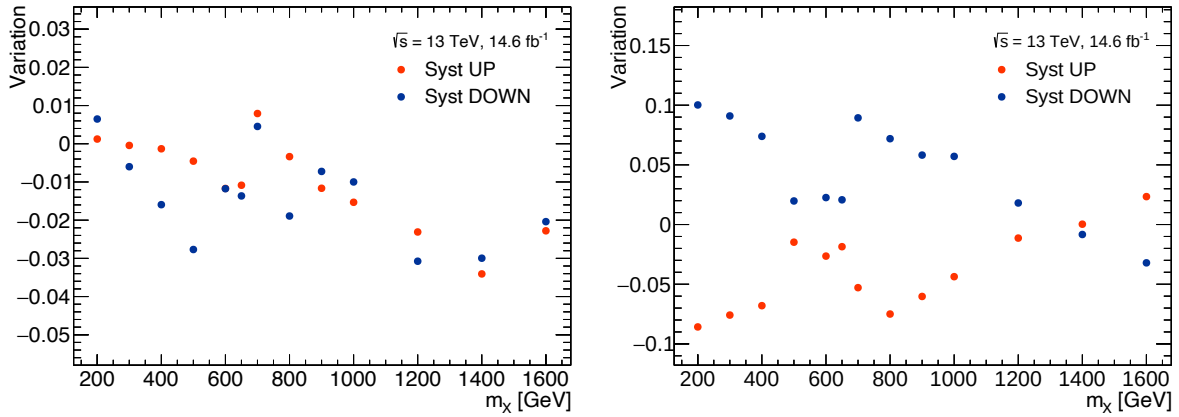
Therefore, the reconstruction probability of protons that should have failed to be reconstructed due to the existence of pileup might not be fully considered in the nominal proton reconstruction efficiency estimation. To consider this possibility, pileup effects on clustering and tracking results are taken from the data, and the cluster and track objects are added to the nominal signal sample and subsequently used in proton reconstruction to see the effect.

- Clustered MC uncertainty. The results of the nominal and clustered MC samples are compared.
- Selection correction uncertainty $\delta_{0.1\xi_{\gamma\gamma}}$. The potential failure of uncertainty evaluation due to the difference of $\xi_{\gamma\gamma}$ and ξ_{AFP} selection thresholds is mentioned in Sec. 4.4 and evaluated in Sec. D.

7.3 Fitting-related uncertainty

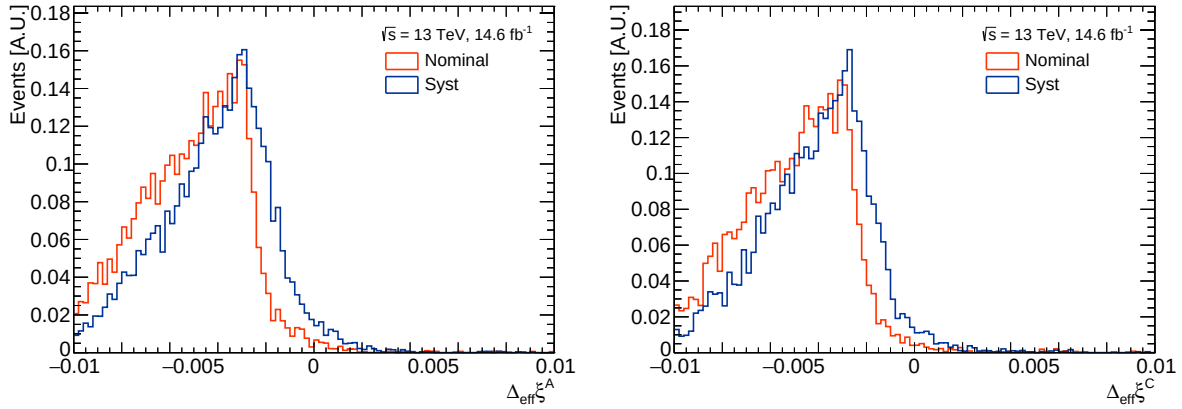
This section describes the systematic uncertainty originating from the function fittings.

- Signal model parameter uncertainty (see Table 5.1).
- Signal modeling linearity uncertainty δ_{smodel}. This is described in Sec. 5.1.1.



(a) Beam optics uncertainty (b) Global alignment uncertainty

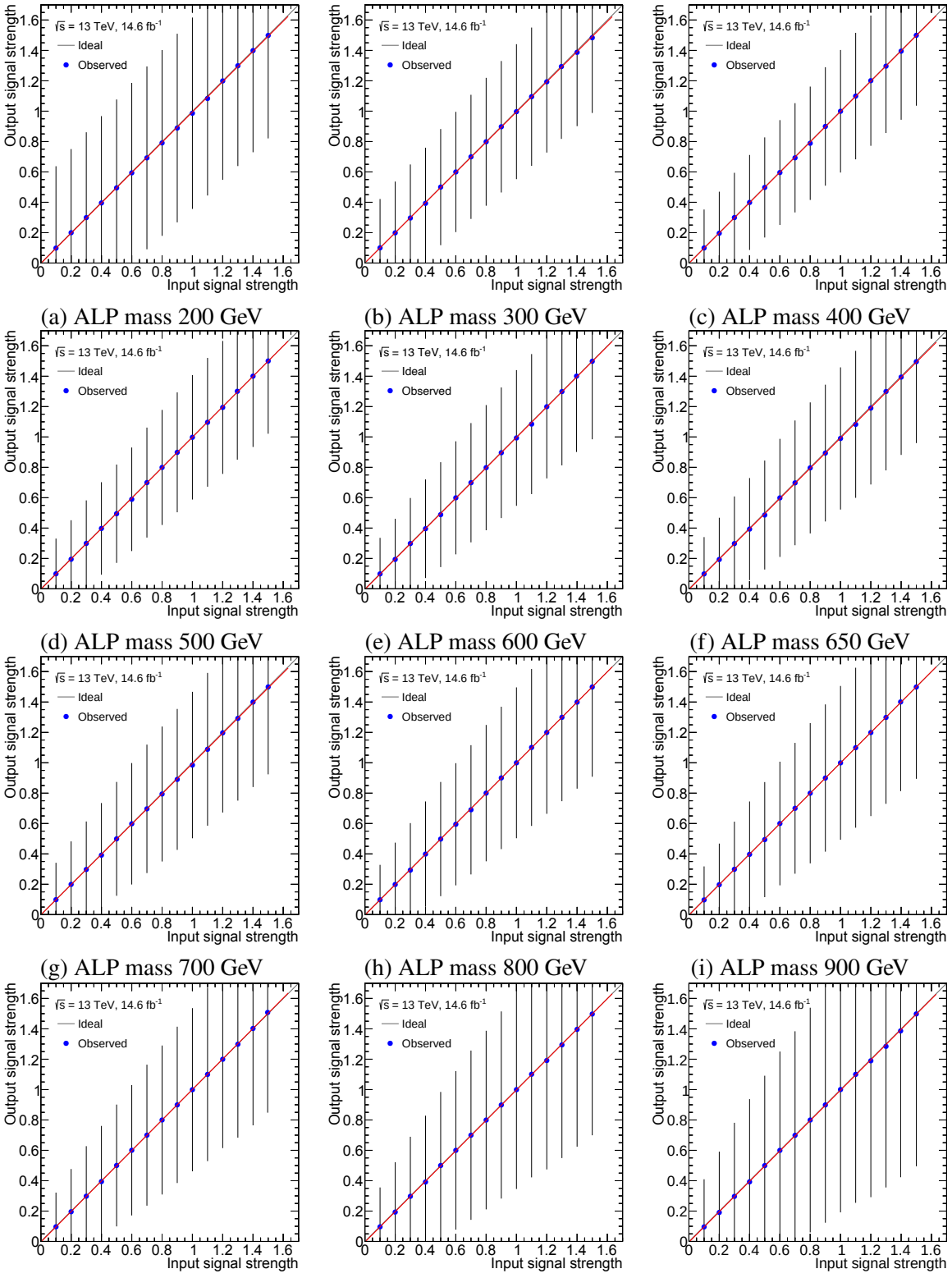
Figure 7.7: The relative difference in the number of signal events between the shifted and nominal signal samples after the whole selection. The upper and lower (a) beam optics and (b) global alignment systematic variations are applied to the signal samples.



(a) A-side (b) C-side

Figure 7.8: $\Delta_{\text{eff}}\xi$ distributions of the $m_X = 1000$ GeV simulated signal before the selection obtained with nominal and systematically varied track smearing width in the AFP simulation.

- Background modeling uncertainty δ_{BG}. Eq. 6.4 is used as the absolute signal yield uncertainty originating from background modeling.
- The potential bias of the s+b fitting is evaluated with a signal injection test, as shown for the case without backgrounds in Sec. 5. Asimov samples are generated following the s+b function in Eq. 6.2. In contrast to Sec. 6.4.2, the Asimov sample has only one event for each fine mass point, every 2 GeV, but is weighted to follow the distribution function. The signal and background parameters for the generation are set as determined in Sec. 5 and 6.1, except for N_X, which is varied to test the linearity of the fit result. The signal function parameters μ_{CB}, σ_{CB}, $n_{\text{L,R}}$, and $\alpha_{\text{L,R}}$ are fixed to distinguish the sources of systematic uncertainties, while the other parameters are set to be variable when fitting. The signal strength μ defined in Eq. 5.6 is compared between the input and out values in Fig. 7.9 for each hypothetical ALP mass. The maximum relative difference from 1 is only $\delta_{\text{sbfit}} = 0.006$. Thus, this kind of fitting uncertainty is neglected.



(a) ALP mass 200 GeV (b) ALP mass 300 GeV (c) ALP mass 400 GeV

(d) ALP mass 500 GeV (e) ALP mass 600 GeV (f) ALP mass 650 GeV

(g) ALP mass 700 GeV (h) ALP mass 800 GeV (i) ALP mass 900 GeV

(j) ALP mass 1000 GeV (k) ALP mass 1200 GeV (i) ALP mass 1400 GeV

Figure 7.9: The s+b signal injection test results. The input and output signal strengths are compared for each hypothetical ALP mass. The error bars represent the statistical error for 14.6 fb^{-1}.

7.4 Theoretical uncertainty

The dominant source of theoretical uncertainty of the ALP signal events is the uncertainty of the soft survival factor S^2 described in Sec. 3.2, estimated to be 2%, 10%, and 50% for the EL, SD, and DD signal cross sections, respectively, and the other theoretical uncertainties, such as experimental uncertainty due to structure function and higher order corrections, are found to be negligible compared to the soft survival factor, concluded referring to Ref. [147]. The small uncertainty of EL S^2 is due to its large pp impact parameter, resulting in small model variation of the soft QCD interaction [148]. The uncertainty of SD and DD S^2 are estimated in the evaluation of the short-distance component of S^2 [149].

8 Statistical Procedure and Results

In this section, a statistical model is built, and the search results and exclusion regions for the signal events are presented through statistical tests. First, the statistical model, including the systematic uncertainty effect, is constructed in Sec. 8.1 based on the maximum likelihood method. The effects of the systematic uncertainty on the s+b fitting results are evaluated in Sec. 8.2. Then, statistical tests are performed using dedicated test statistics. The search results and the 95% exclusion limits in the m_X-f^{-1} plane are described in Sec. 8.3 with the test methods. The 441 data events after the whole selection in $m_{\gamma\gamma} \in [150, 1600]$ GeV, which is compatible with the prediction from the background sample, 429 ± 8 events, are used for the tests.

8.1 Statistical model

The bare s+b function in Eq. 6.2 does not consider the effect of systematic uncertainty. Another statistical model is constructed in this section to incorporate systematic uncertainty. The s+b function is rewritten as

$$\mathcal{F}\left(m_{\gamma\gamma}; \boldsymbol{\sigma}_X, m_X, N_\mathrm{b}, \boldsymbol{a}, \boldsymbol{\theta}\right) = f_X\left(m_{\gamma\gamma}; \boldsymbol{x}_X\left(m_X, \boldsymbol{\theta}_\mathrm{CB}\right)\right) \cdot N_X\left(\boldsymbol{\sigma}_X; m_X, \boldsymbol{\theta}_{N_X}\right) + f_\mathrm{b}\left(m_{\gamma\gamma}; \boldsymbol{a}\right) \cdot N_\mathrm{b}, \quad (8.1)$$

where f_X, f_b, N_X, and N_b stand for the probability density function (PDF) and yield as concisely used in Eq. 6.2, and $\boldsymbol{a} = \{a, \alpha_0\}$ denotes the background shape parameters. $\boldsymbol{\theta} = \left\{\boldsymbol{\theta}_\mathrm{CB}, \boldsymbol{\theta}_{N_X}\right\}$ is a set of *nuisance parameters* for systematic uncertainties, where

$$\begin{aligned}
\boldsymbol{\theta}_\mathrm{CB} &= \left\{\theta_{\mu_\mathrm{CB}0}, \theta_{\mu_\mathrm{CB}1}, \theta_{\sigma_\mathrm{CB}0}, \theta_{\sigma_\mathrm{CB}1}, \theta_{\alpha_\mathrm{L}}, \theta_{\alpha_\mathrm{R}}, \theta_{n_\mathrm{L}}, \theta_{n_\mathrm{R}}, \theta_\mathrm{scale}, \boldsymbol{\theta}_{\sigma_\mathrm{CB}}\right\}, \\
\boldsymbol{\theta}_{\sigma_\mathrm{CB}} &= \left\{\theta_k \mid k \in S_{\sigma_\mathrm{CB}}\right\}, \\
S_{\sigma_\mathrm{CB}} &= \{\sigma_\mathrm{scale}, \mathrm{res}, \sigma_\mathrm{PRW}, \}
\end{aligned}$$

affect the signal model parameters, and

$$\begin{aligned}
\boldsymbol{\theta}_{N_X} &= \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \theta_\mathrm{BG}\}, \\
\boldsymbol{\theta}_i &= \{\theta_k \mid k \in S_i\}, \\
S_1 &= \left\{\mathrm{lumi}, \mathrm{ID}, \mathrm{iso}, \mathrm{PRW}, \mathrm{optics}, \mathrm{glob}, \varepsilon_\mathrm{AFP}, \mathrm{Nclust}, \mathrm{smear}, 0.1\xi_{\gamma\gamma}, \mathrm{smodel}\right\}, \\
S_2 &= \left\{\varepsilon\mathrm{EL}, \varepsilon\mathrm{SD}, \varepsilon\mathrm{DD}, S^2\mathrm{EL}, S^2\mathrm{SD}, S^2\mathrm{DD}\right\}
\end{aligned}$$

affect signal yield.

The signal model f_X is the DSCB function (Eq. 5.1) as discussed in Sec. 5. The systematic uncertainty

$\boldsymbol{\theta}_{\mathrm{CB}}$ of its parameters $\boldsymbol{x}_X$ are considered in the fitting,[1]

$$
\begin{aligned}
\boldsymbol{x}_X\left(m_X, \boldsymbol{\theta}_{\mathrm{CB}}\right) &= \left\{\mu_{\mathrm{CB}}(m_X, \theta_{\mu_{\mathrm{CB}}0,1}), \sigma_{\mathrm{CB}}(m_X, \theta_{\sigma_{\mathrm{CB}}0,1}), \alpha_{\mathrm{L,R}}(\theta_{\alpha_{\mathrm{L,R}}}), n_{\mathrm{L,R}}(\theta_{n_{\mathrm{L,R}}})\right\}, \\
\mu_{\mathrm{CB}}(m_X, \theta_{\mu_{\mathrm{CB}}0,1}) &= \left[m_X + \left(p_{\Delta m0} + \delta_{\Delta m0}\theta_{\mu_{\mathrm{CB}}0}\right) + \left(p_{\Delta m1} + \delta_{\Delta m1}\theta_{\mu_{\mathrm{CB}}1}\right) m_X\right] K_{\mathrm{scale}}\left(\theta_{\mathrm{scale}}\right), \\
\sigma_{\mathrm{CB}}(m_X, \theta_{\sigma_{\mathrm{CB}}0,1}) &= \left[\left(p_{\sigma_{\mathrm{CB}}0} + \delta_{\sigma_{\mathrm{CB}}0}\theta_{\sigma_{\mathrm{CB}}0}\right) + \left(p_{\sigma_{\mathrm{CB}}1} + \delta_{\sigma_{\mathrm{CB}}1}\theta_{\sigma_{\mathrm{CB}}1}\right) m_X\right] \prod_{k \in S_{\sigma_{\mathrm{CB}}}} K_k(\theta_k), \\
\alpha_{\mathrm{L,R}}(\theta_{\alpha_{\mathrm{L,R}}}) &= p_{\alpha_{\mathrm{L,R}}} + \delta_{\alpha_{\mathrm{L,R}}}\theta_{\alpha_{\mathrm{L,R}}}, \\
n_{\mathrm{L,R}}(\theta_{n_{\mathrm{L,R}}}) &= p_{n_{\mathrm{L,R}}} + \delta_{n_{\mathrm{L,R}}}\theta_{n_{\mathrm{L,R}}},
\end{aligned}
$$

where

$$
\begin{aligned}
K_k(\theta_k) &= K\left(\theta_k; \delta_k^{\mathrm{up}}, \delta_k^{\mathrm{down}}\right), \\
K\left(\theta; \delta^{\mathrm{up}}, \delta^{\mathrm{down}}\right) &= \left[r\left(\theta; \delta^{\mathrm{up}}, \delta^{\mathrm{down}}\right)\right]^{\theta}, \\
r\left(\theta; \delta^{\mathrm{up}}, \delta^{\mathrm{down}}\right) &= \begin{cases} 1 + \delta^{\mathrm{up}}, & \text{if } \theta > 0 \\ 1/(1 - \delta^{\mathrm{down}}), & \text{if } \theta < 0 \end{cases}.
\end{aligned} \tag{8.2}
$$

The subscripts up and down stand for the upper and lower uncertainty. K denotes a function characterizing the effect of the systematic uncertainties, which are generally asymmetric. The expression of r is an approximation ensuring that the modifications to the signal event yield for $\theta = \pm 1$ correspond to the $\pm 1\sigma$ variations used to define the uncertainties. The actual expression for r is interpolated smoothly between the cases $\theta > 0$ and $\theta < 0$ using HistFactory [150] to avoid numerical problems at $\theta = 0$.

The signal yield

$$
N_X\left(\boldsymbol{\sigma}_X; m_X, \boldsymbol{\theta}_{N_X}\right) = L_{\mathrm{int}}\hat{\sigma}_X\left(\boldsymbol{\sigma}_X; m_X, \boldsymbol{\theta}_2\right) \prod_{k \in S_1} K_k(\theta_k) + \delta_{\mathrm{BG}}(m_X)\theta_{\mathrm{BG}} \tag{8.3}
$$

is characterized by signal cross sections

$$
\begin{aligned}
\boldsymbol{\sigma}_X &= \left\{\sigma_{\mathrm{EL}}, \sigma_{\mathrm{SD}}, \sigma_{\mathrm{DD}}\right\}, \\
\hat{\sigma}_X\left(\boldsymbol{\sigma}_X; m_X, \boldsymbol{\theta}_2\right) &= \sum_{i \in \{\mathrm{EL,SD,DD}\}} \left(\sigma_i(m_X)\varepsilon_i(m_X)K_{\varepsilon i}(m_X, \theta_{\varepsilon i})K_{S^2 i}(\theta_{S^2 i})\right), \\
\sigma_i(m_X) &= \mu\sigma_{\mathrm{std}}^i(m_X),
\end{aligned}
$$

where

$$
K_k(m_X, \theta_k) = K\left(\theta_k; \delta_k^{\mathrm{up}}(m_X), \delta_k^{\mathrm{down}}(m_X)\right),
$$

μ is signal strength, and σ_{std}^i is described in Sec. 5.2.2.[2]

The background model f_{b} is the function of Eq. 6.1 as discussed in Sec. 6. Its parameters $\boldsymbol{a} = \{a, \alpha_0\}$ and background yield N_{b} are all treated as free parameters in the fit.

The overall likelihood, including extended and Gaussian constraint terms, is given by

$$
L\left(\mu, \boldsymbol{\nu}; m_X, \left\{m_{\gamma\gamma,i}\right\}\right) = e^{-(N_X + N_{\mathrm{b}})}\left[\prod_{i=1}^{M} \mathcal{F}\left(m_{\gamma\gamma,i}; \boldsymbol{\sigma}_X(\mu), m_X, N_{\mathrm{b}}, \boldsymbol{a}, \boldsymbol{\theta}\right)\right] \prod_{\vartheta \in \boldsymbol{\theta}} e^{-\vartheta^2/2}, \tag{8.4}
$$

[1] The correlation among the signal parameters is ignored because their systematic uncertainty has a negligible effect on the results as described in Sec. 8.2.

[2] The correlation among the soft survival factors is found to be negligible. The test was performed using an additional constraint, $\theta_{S^2\mathrm{EL}} = \theta_{S^2\mathrm{SD}} = \theta_{S^2\mathrm{DD}}$, in the evaluation of the systematic uncertainty effects in Sec. 8.2. The difference in the result from the nominal evaluation was negligible, so it was concluded that the presence or absence of the correlation does not affect the results.

where $\boldsymbol{\nu} \equiv \{N_\mathrm{b}, \boldsymbol{a}, \boldsymbol{\theta}\}$, M is the number of events in the dataset, and $m_{\gamma\gamma,i}$ is the diphoton mass for a given event i. Unbinned fitting is performed for the statistical tests.

8.2 Effects of systematic uncertainty

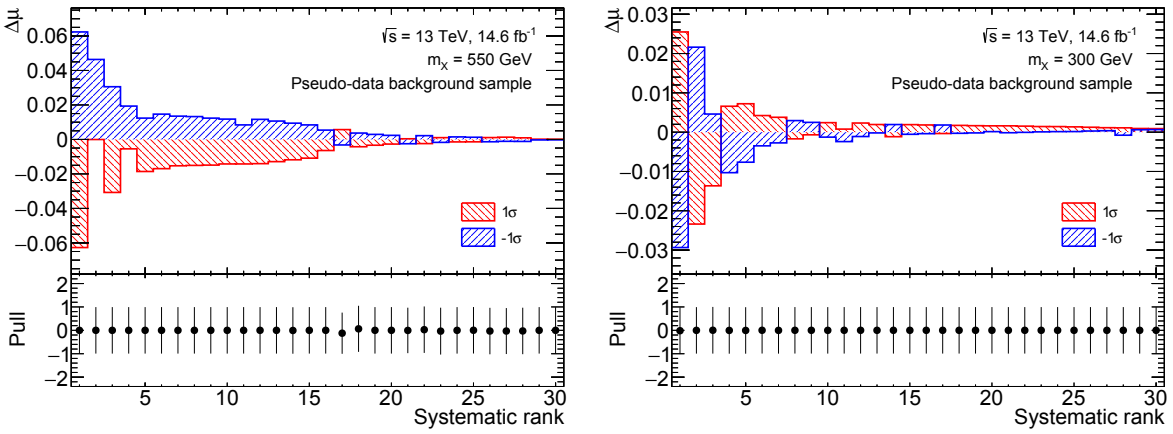
The effect of each systematic uncertainty on the signal strength is evaluated using the background sample with a similar size as the data represented in Fig. 6.4 (a). It can validate the method's implementation by comparing the results with the relative qualitative impact of each source of systematic uncertainty expected from the statistical model in Sec 8.1 and the systematic uncertainty in Table 7.1. First, the s+b fit using the likelihood in Eq. 8.4 is performed with all the parameters not fixed, called *unconditional fit*. The set of the resulting value and error is referred to as *pull* for each nuisance parameter $\vartheta \in \boldsymbol{\theta}$. Then, the following procedure is repeated for each ϑ.

> Starting from the resulting parameters of the unconditional fit, the s+b fit is performed again, fixing ϑ to the upper edge of the error bar of the pull and floating other parameters. The same is performed for the lower edge of the error bar. The differences in the signal strength μ of the two fits from the unconditional fit $\Delta\mu$ are regarded as the effect of the systematic uncertainty corresponding to ϑ.

Fig. 8.1 shows the pulls (lower panel) and effects (upper panel) of each nuisance parameter distinguished by the horizontal axis. The results are obtained by the fits setting the hypothetical ALP mass as $m_X = 550$ GeV and 300 GeV. The systematic uncertainties are ranked according to the effects, and the source of the systematic uncertainty corresponding to each rank is shown in Table 8.1 and Table 8.2, respectively. The two mass points are in the mass ranges with locally concentrated events and no excess of events, respectively. Fig. 8.2 compares the $m_{\gamma\gamma}$ distribution of the sample and the background-only fit (fixing $\mu = 0$) result. The dominant systematic uncertainty for $m_X = 550$ GeV is the AFP alignment systematics. At the 300 GeV part without event excess, the rank of background modeling uncertainty, which is absolute uncertainty (see Eq. 8.3), is increased, and the rank of the other signal yield uncertainties is decreased, as expected. Asymmetric behaviours for some nuisance parameters such as the second bin in Fig. 8.1 (a) are due to the asymmetric input systematic values listed in Table 7.1. Some systematic uncertainties show the same sign of effects for both the upper and lower shifting tests. Two cases can cause such behaviour. The first case is that the input systematic uncertainty is entirely one-sided, in which the interpolation algorithm at $\theta = 0$ in Eq. 8.2 the factor causes such a specific behaviour. The second case is that the systematic uncertainty is a signal modeling parameter uncertainty corresponding to $\boldsymbol{\theta}_\mathrm{CB}$. In this case, there is no need for the effects to have different signs.

In many other analyses, the background distribution in the blinded kinematic region (see Sec. 4.4), referred to as signal region (SR), is estimated using background MC samples, so the difference from the data is measured in another region close to the SR, referred to as control region (CR), and reflected by adding a constant term to ϑ in Eq. 8.4. This is often expressed as "the auxiliary measurements constrain the nuisance parameters". In such cases, the pull information is used to quantify the level of the constraints. However, since no CR is used in this analysis because the background distribution is determined in a fully data-driven way, there is no such CR-originated constraint. The pulls in Fig. 8.1 are almost 0 ± 1 for all the nuisance parameters, so it is confirmed before unblinding the SR that there is no significant constraint.

For $m_X = 550$ GeV, the conditional fit result is $\mu = 0.683 \pm 0.348$. On the other hand, if all the nuisance parameters are fixed, i.e. the bare function Eq. 6.2 is used, the result becomes 0.685 ± 0.337. This error corresponds to the effect of statistical uncertainty, and the systematic uncertainty effect is $\sqrt{0.348^2 - 0.337^2} \approx 0.087$. Thus, the statistical uncertainty dominates.



(a) $m_X = 550$ GeV (b) $m_X = 300$ GeV

Figure 8.1: The ranking plots created with hypothetical ALP mass (a) 550 GeV corresponding to the mass range with events locally concentrated, and (b) 300 GeV corresponding to the range with almost no event, created using the background sample represented in Fig. 6.4 (a). The nuisance parameters' pulls and effects are displayed in descending order rank corresponding to the horizontal axis starting from 1. The source of the systematic uncertainty corresponding to each rank is listed in Table 8.1 for (a) and Table 8.2 for (b). A red (blue) bar shows the effect evaluated by fixing the nuisance parameter to the upper (lower) edge of the error bar of the pull.

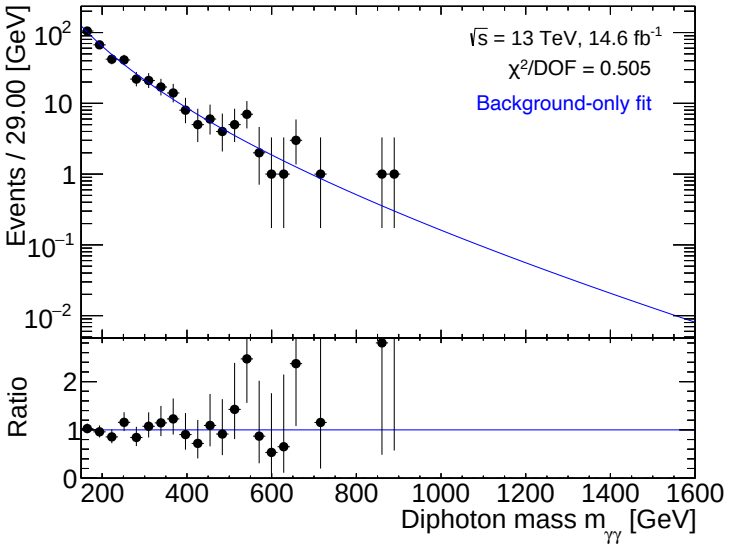


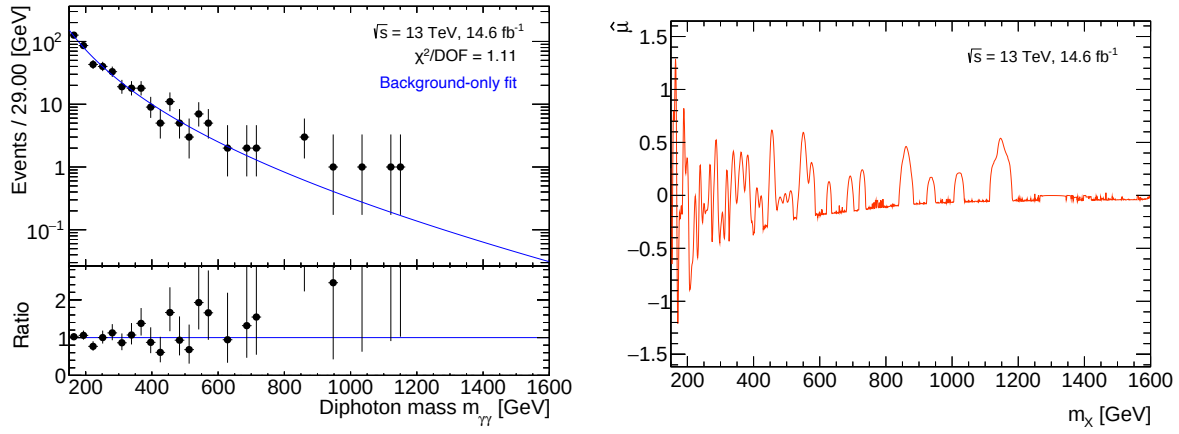
Figure 8.2: Diphoton mass distribution of the pseudo-data background sample represented in Fig. 6.4 (a) and its background-only fit result.

Table 8.1: Ranking of the effects of systematic uncertainty on the signal strength for hypothetical ALP mass 550 GeV. The pseudo-data background sample represented in Fig. 6.4 (a) is used.

Rank	Nuisance parameter	Systematic uncertainty source
1	θ_{glob}	AFP global alignment
2	θ_{Nclust}	Required number of AFP clusters
3	$\theta_{S^2\text{SD}}$	SD signal soft survival factor
4	θ_{optics}	Beam optics
5	$\theta_{\varepsilon_{\text{AFP}}}$	Proton reconstruction efficiency
6	θ_{PRW}	Pileup reweighting
7	θ_{lumi}	Integrated luminosity
8	$\theta_{\varepsilon\text{EL}}$	EL signal efficiency
9	$\theta_{0.1\xi_{\gamma\gamma}}$	Selection correction
10	θ_{smodel}	Signal modeling linearity
11	θ_{smear}	AFP track smearing
12	$\theta_{\varepsilon\text{SD}}$	SD signal efficiency
13	θ_{iso}	Photon isolation efficiency
14	θ_{BG}	Background modeling (absolute uncertainty)
15	θ_{ID}	Photon identification efficiency
16	$\theta_{S^2\text{EL}}$	EL signal soft survival factor
17	θ_{res}	Photon energy resolution
18	$\theta_{n_{\text{R}}}$	Signal model parameter $p_{n_{\text{R}}}$ (absolute uncertainty)
19	$\theta_{S^2\text{DD}}$	DD signal soft survival factor
20	$\theta_{\varepsilon\text{DD}}$	DD signal efficiency
21	θ_{scale}	Photon energy scale
22	$\theta_{\alpha_{\text{R}}}$	Signal model parameter $p_{\alpha_{\text{R}}}$ (absolute uncertainty)
23	$\theta_{\sigma_{\text{scale}}}$	Photon energy scale
24	$\theta_{n_{\text{L}}}$	Signal model parameter $p_{n_{\text{L}}}$ (absolute uncertainty)
25	$\theta_{\alpha_{\text{L}}}$	Signal model parameter $p_{\alpha_{\text{L}}}$ (absolute uncertainty)
26	$\theta_{\sigma_{\text{CB}}0}$	Signal model parameter $p_{\sigma_{\text{CB}}0}$ (absolute uncertainty)
27	$\theta_{\sigma_{\text{PRW}}}$	Pileup reweighting
28	$\theta_{\sigma_{\text{CB}}1}$	Signal model parameter $p_{\sigma_{\text{CB}}1}$ (absolute uncertainty)
29	$\theta_{\Delta m0}$	Signal model parameter $p_{\Delta m0}$ (absolute uncertainty)
30	$\theta_{\Delta m1}$	Signal model parameter $p_{\Delta m1}$ (absolute uncertainty)

Table 8.2: Ranking of the effects of systematic uncertainty on the signal strength for hypothetical ALP mass 300 GeV. The pseudo-data background sample represented in Fig. 6.4 (a) is used.

Rank	Nuisance parameter	Systematic uncertainty source
1	θ_{scale}	Photon energy scale
2	θ_{BG}	Background modeling (absolute uncertainty)
3	θ_{res}	Photon energy resolution
4	$\theta_{n_{\text{R}}}$	Signal model parameter $p_{n_{\text{R}}}$ (absolute uncertainty)
5	$\theta_{\Delta m0}$	Signal model parameter $p_{\Delta m0}$ (absolute uncertainty)
6	θ_{glob}	AFP global alignment
7	$\theta_{\Delta m1}$	Signal model parameter $p_{\Delta m1}$ (absolute uncertainty)
8	$\theta_{\sigma_{\text{CB}}0}$	Signal model parameter $p_{\sigma_{\text{CB}}0}$ (absolute uncertainty)
9	$\theta_{\sigma_{\text{scale}}}$	Photon energy scale
10	$\theta_{\alpha_{\text{R}}}$	Signal model parameter $p_{\alpha_{\text{R}}}$ (absolute uncertainty)
11	θ_{Nclust}	Required number of AFP clusters
12	$\theta_{S^2\text{SD}}$	SD signal soft survival factor
13	$\theta_{\varepsilon_{\text{AFP}}}$	Proton reconstruction efficiency
14	$\theta_{\sigma_{\text{PRW}}}$	Pileup reweighting
15	$\theta_{\varepsilon\text{EL}}$	EL signal efficiency
16	θ_{PRW}	Pileup reweighting
17	$\theta_{\sigma_{\text{CB}}1}$	Signal model parameter $p_{\sigma_{\text{CB}}1}$ (absolute uncertainty)
18	θ_{lumi}	Integrated luminosity
19	$\theta_{0.1\,\xi_{\gamma\gamma}}$	Selection correction
20	θ_{smear}	AFP track smearing
21	θ_{smodel}	Signal modeling linearity
22	θ_{iso}	Photon isolation efficiency
23	$\theta_{\varepsilon\text{SD}}$	SD signal efficiency
24	θ_{ID}	Photon identification efficiency
25	$\theta_{\alpha_{\text{L}}}$	Signal model parameter $p_{\alpha_{\text{L}}}$ (absolute uncertainty)
26	$\theta_{S^2\text{EL}}$	EL signal soft survival factor
27	$\theta_{n_{\text{L}}}$	Signal model parameter $p_{n_{\text{L}}}$ (absolute uncertainty)
28	θ_{optics}	Beam optics
29	$\theta_{S^2\text{DD}}$	DD signal soft survival factor
30	$\theta_{\varepsilon\text{DD}}$	DD signal efficiency



(a) Background-only fit

(b) Signal strength

Figure 8.3: Fit results to the unblinded data. (a) Background-only fit result with the data distribution. (b) Unconditional fit results $\hat{\mu}$ as a function of m_X.

8.3 Statistical tests and results

In this section, statistical tests are performed using the unblinded data to search for a BSM resonance and set the exclusion limits using the likelihood in Eq. 8.4, abbreviated as

$$L(\mu, \boldsymbol{\nu}) \equiv L\left(\mu, \boldsymbol{\nu}; m_X, \{m_{\gamma\gamma,i}\}\right),$$

for each hypothetical ALP mass m_X. The signal strength μ is called the parameter of interest (POI) to distinguish it from the other parameters $\boldsymbol{\nu}$ and is used to define the hypothesis. To define the test statistics, the profile likelihood ratio [151] for the hypothetical signal strength μ_{hypo} is defined as

$$\lambda(\mu_{\text{hypo}}) = \frac{L\left(\mu_{\text{hypo}}, \hat{\hat{\boldsymbol{\nu}}}(\mu_{\text{hypo}})\right)}{L(\hat{\mu}, \hat{\boldsymbol{\nu}})},$$

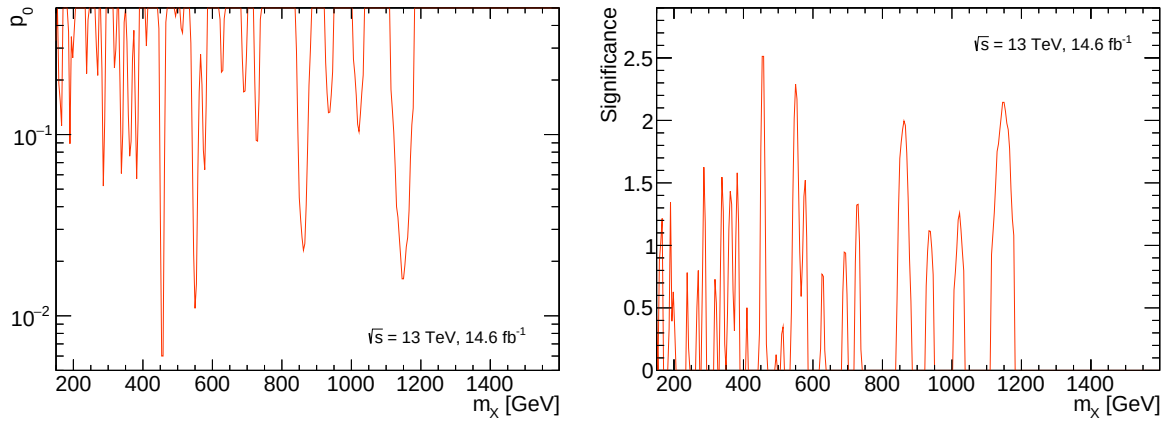
where $\hat{\mu}$ and $\hat{\boldsymbol{\nu}}$ are the maximum-likelihood estimators (MLEs) of the unconditional fit, and $\hat{\hat{\boldsymbol{\nu}}}(\mu_{\text{hypo}})$ is the MLEs for the *conditional fit* fixing the signal strength as μ_{hypo}. Fig. 8.3 shows the $m_{\gamma\gamma}$ distribution of the data, in comparison to its background-only fit result (fixing $\mu = 0$), and the resulting signal strength of the unconditional fit for each m_X.[3]

The test statistics used in Sec. 8.3.1 and 8.3.2 are based on a statistic

$$t_{\mu_{\text{hypo}}} = -2 \ln \lambda(\mu_{\text{hypo}}).$$

It follows the χ^2 distribution if the true signal strength $\mu_{\text{true}} = \mu_{\text{hypo}}$ or noncentral χ^2 distribution if $\mu_{\text{true}} \neq \mu_{\text{hypo}}$. This leads to the use of so-called asymptotic formulae [151] instead of toy MC samples with a large number of events to obtain the distributions of the test statistics. Thus, the test statistics based on $t_{\mu_{\text{hypo}}}$ have been conventionally used in the ATLAS collaboration. However, in this research, the asymptotic formulae are not used for the nominal results because it is not valid for the number of data

[3]The fit result becomes unstable in a higher mass range, especially above 1000 GeV. This is because there are local mass ranges with no event. The fits are judged to be converged if events with extremely small weights are added every 1 GeV to the sample. However, the fit results are still unstable because the fit result converges not on a differentiable minimum but on a non-smooth point due to the artificial modification of the likelihood to prevent negative PDF. Still, this means the unstable behaviour is seen when $\hat{\mu} < 0$, and there is no problem because the test statistic does not use the unconditional fit result when $\hat{\mu} < 0$.



(a) p-value (b) Significance

Figure 8.4: Test results for null hypothesis $\mu = 0$ using the unblinded data as functions of m_X. (a) p-value p_0. (b) Significance Z_0.

events corresponding to 14.6 fb^{-1}, although the results using the asymptotic formulae are in Appendix G for comparison. Instead, a primitive method is used in which the toy MC samples normalized to 14.6 fb^{-1} are generated based on the statistical model described in Sec. 8.1 to create the distributions of the test statistics assuming the null hypotheses.

8.3.1 Significance

For the search for a BSM resonance, the null hypothesis $\mu_{\text{hypo}} = 0$ is used. The test statistic is defined as

$$q_0 = \begin{cases} t_0 & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases} .$$

q_0 is set as 0 when $\hat{\mu} < 0$ because the negative signal strength is not physical when searching for a BSM resonant contribution. 1000 toy MC samples were made to create the distribution following the PDF $f\left(q_0 \mid 0\right)$, where $f\left(q_0 \mid \mu_{\text{true}}\right)$ is the conditional PDF of q_0 under the condition $\mu = \mu_{\text{true}}$.

Then, the p-value p_0 and significance Z_0 are calculated as

$$p_0 = \int_{q_0}^{\infty} f\left(q_0' \mid 0\right) \mathrm{d}q_0',$$
$$Z_0 = \Phi^{-1}\left(1 - p_0\right),$$

for each m_X corresponding to the toy MC samples, where

$$\Phi(x) \equiv \frac{1}{2}\left(1 + \operatorname{erf} \frac{x}{\sqrt{2}}\right)$$

is the cumulative distribution function of the normalized Gaussian distribution. Fig. 8.4 shows the results as functions of hypothetical ALP mass m_X. No excess of significance from 3 is observed, meaning no evidence of a BSM resonance. The maximum observed significance is 2.51 at $m_X = 454$ GeV. Fig. G.1 shows the results using the asymptotic formulae for comparison. Fig. 8.5 shows the ranking plot created by the same method as in Sec. 8.2, tested at $m_X = 454$ GeV.

The local significance of 2.51 is relatively large. Still, the look-elsewhere effect (LEE) [152, 153], considering that many trials increase the probability of a low-probability event occurring, should also be

Table 8.3: Ranking of the effects of systematic uncertainty on the signal strength for hypothetical ALP mass 454 GeV. The unblinded data is used.

Rank	Nuisance parameter	Systematic uncertainty source
1	θ_{glob}	AFP global alignment
2	θ_{Nclust}	Required number of AFP clusters
3	$\theta_{S^2\text{SD}}$	SD signal soft survival factor
4	θ_{optics}	Beam optics
5	$\theta_{n_{\text{R}}}$	Signal model parameter $p_{n_{\text{R}}}$ (absolute uncertainty)
6	$\theta_{\varepsilon\text{EL}}$	EL signal efficiency
7	θ_{PRW}	Pileup reweighting
8	θ_{lumi}	Integrated luminosity
9	$\theta_{\varepsilon_{\text{AFP}}}$	Proton reconstruction efficiency
10	θ_{BG}	Background modeling (absolute uncertainty)
11	$\theta_{0.1\xi_{\gamma\gamma}}$	Selection correction
12	θ_{smodel}	Signal modeling linearity
13	θ_{iso}	Photon isolation efficiency
14	$\theta_{\varepsilon\text{SD}}$	SD signal efficiency
15	θ_{smear}	AFP track smearing
16	θ_{ID}	Photon identification efficiency
17	θ_{scale}	Photon energy scale
18	$\theta_{\alpha_{\text{R}}}$	Signal model parameter $p_{\alpha_{\text{R}}}$ (absolute uncertainty)
19	$\theta_{S^2\text{EL}}$	EL signal soft survival factor
20	θ_{res}	Photon energy resolution
21	$\theta_{S^2\text{DD}}$	DD signal soft survival factor
22	$\theta_{\varepsilon\text{DD}}$	DD signal efficiency
23	$\theta_{\Delta m0}$	Signal model parameter $p_{\Delta m0}$ (absolute uncertainty)
24	$\theta_{\Delta m1}$	Signal model parameter $p_{\Delta m1}$ (absolute uncertainty)
25	$\theta_{\sigma_{\text{PRW}}}$	Pileup reweighting
26	$\theta_{\sigma_{\text{CB}}0}$	Signal model parameter $p_{\sigma_{\text{CB}}0}$ (absolute uncertainty)
27	$\theta_{\sigma_{\text{scale}}}$	Photon energy scale
28	$\theta_{\sigma_{\text{CB}}1}$	Signal model parameter $p_{\sigma_{\text{CB}}1}$ (absolute uncertainty)
29	$\theta_{\alpha_{\text{L}}}$	Signal model parameter $p_{\alpha_{\text{L}}}$ (absolute uncertainty)
30	$\theta_{n_{\text{L}}}$	Signal model parameter $p_{n_{\text{L}}}$ (absolute uncertainty)

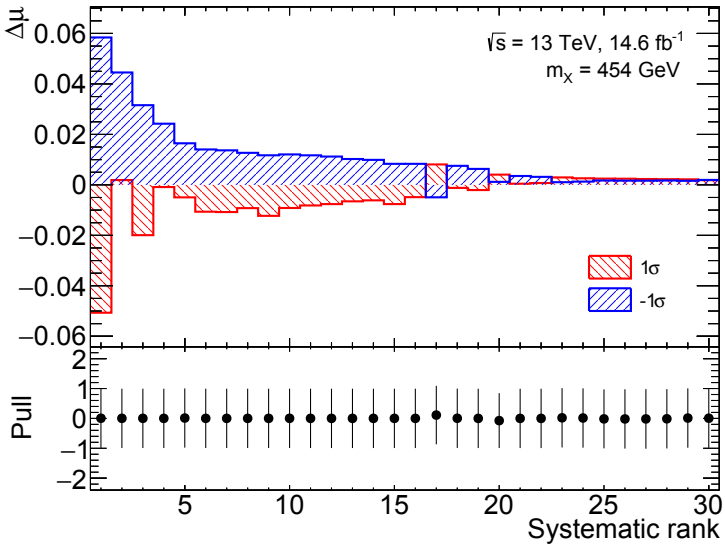


Figure 8.5: The ranking plot created with a hypothetical ALP mass of 454 GeV. The nuisance parameters' pulls and effects are displayed in descending order rank corresponding to the horizontal axis starting from 1. The source of the systematic uncertainty corresponding to each rank is listed in Table 8.3. A red (blue) bar shows the effect evaluated by fixing the nuisance parameter to the upper (lower) edge of the error bar of the pull.

large in this case. This is because the $m_{\gamma\gamma}$ resonance peak of the expected signal events is very narrow compared to the mass range where the BSM resonance is searched, and many hypothetical m_X are tested. The statistical significance of the maximum local deviation is evaluated as follows.[4] The maximum q_0 with respect to m_X, $q_{0,\max}$, is calculated for each toy MC sample and filled into a histogram to create the distribution following the PDF of $q_{0,\max}$ under the condition that $\mu = 0$ for any m_X, $f_{\max}\left(q_{0,\max}\right)$. Then, using the observed $q_{0,\max}$ in the data, $q_{0,\max} = 6.03$, the global p-value p_{glob} and global significance Z_{glob} are calculated as

$$p_{\text{glob}} = \int_{q_{0,\max}}^{\infty} f\left(q'_{0,\max}\right) \mathrm{d}q'_{0,\max},$$
$$Z_{\text{glob}} = \Phi^{-1}\left(1 - p_{\text{glob}}\right).$$

Fig. 8.6 shows the $f_{\max}\left(q_{0,\max}\right)$ and the observed $q_{0,\max}$. The observed $q_{0,\max}$ is found to be smaller than the median of the distribution, so it is concluded that the data $m_{\gamma\gamma}$ distribution is consistent with the background-only hypothesis.

8.3.2 Exclusion limits

To make the exclusion limits in the m_X-f^{-1} plane, null hypotheses,

$$\mu_{\text{hypo}} \in \{0.10, 0.20, 0.40, 0.70, 1.00, 1.50, 2.00, 3.00, 4.00, 5.00, 6.20, 7.40, 8.70, 9.99\}, \tag{8.5}$$

[4] The procedure is the same as the one used in Ref. [154], although the process is documented only internally.

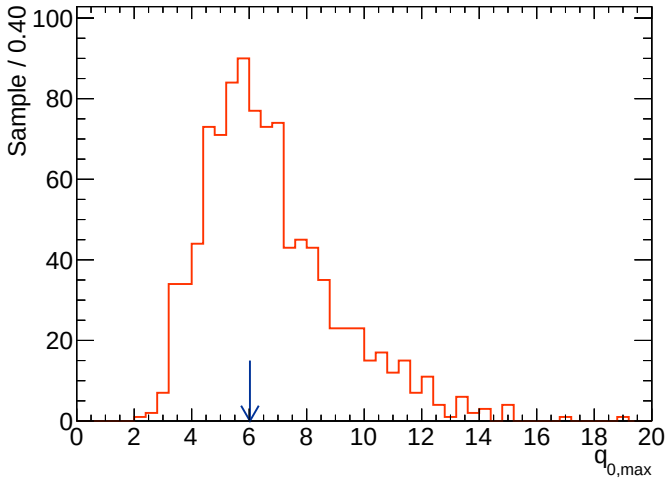


Figure 8.6: The PDF $f_{\max}\left(q_{0,\max}\right)$ (red) and the observed $q_{0,\max} = 6.03$ (blue). The p-value is $p_{\text{glob}} = 0.56$.

are tested using the test statistic

$$\tilde{q}_{\mu_{\text{hypo}}} = \begin{cases} -2\ln\dfrac{L\left(\mu_{\text{hypo}}, \hat{\hat{\nu}}(\mu_{\text{hypo}})\right)}{L\left(0, \hat{\hat{\nu}}(0)\right)} & \hat{\mu} < 0 \\ t_{\mu_{\text{hypo}}} & 0 \le \hat{\mu} \le \mu_{\text{hypo}} \\ 0 & \hat{\mu} > \mu_{\text{hypo}} \end{cases}.$$

If $\hat{\mu} < 0$, the unconditional fit is replaced by the conditional fit assuming $\mu = 0$, while if $\hat{\mu} > \mu_{\text{hypo}}$, the test statistic is set as 0 because this case is not worth considering in the context of rejecting the hypothesis that there is a BSM contribution ($\mu_{\text{hypo}} > 0$). Each null hypothesis is tested for each hypothetical m_X.

To derive the observed exclusion limits, 1000 toy MC samples were made every 4 GeV of the hypothetical ALP mass m_X to create a distribution following the PDF $g(\tilde{q}_{\mu_{\text{hypo}}} \mid \mu_{\text{hypo}})$, where $g(\tilde{q}_{\mu_{\text{hypo}}} \mid \mu_{\text{true}})$ is the conditional PDF of $\tilde{q}_{\mu_{\text{hypo}}}$ under the condition $\mu = \mu_{\text{true}}$. In addition, the same number of samples were made with the same interval to create a distribution following the PDF $g(\tilde{q}_{\mu_{\text{hypo}}} \mid 0)$, which corresponds to the alternative hypothesis. Then,

$$\text{CL}_\text{s} = \frac{\displaystyle\int_{\tilde{q}_{\mu_{\text{hypo}}}}^{\infty} g\left(\tilde{q}'_{\mu_{\text{hypo}}} \mid \mu_{\text{hypo}}\right) \mathrm{d}\tilde{q}'_{\mu_{\text{hypo}}}}{\displaystyle\int_{\tilde{q}_{\mu_{\text{hypo}}}}^{\infty} g\left(\tilde{q}'_{\mu_{\text{hypo}}} \mid 0\right) \mathrm{d}\tilde{q}'_{\mu_{\text{hypo}}}}$$

is used for the tests rather than the p-value to prevent non-physical results [155, 156]. The root of an equation of μ_{hypo}, $\text{CL}_\text{s} = 0.05$, is regarded as the observed 95% upper limits $\mu_{95(\text{obs.})}$ of signal strength μ. The solid black line in Fig. 8.7 represents the results.

The observed limits are roughly within the 2σ confidence intervals centered on the expected limits, which are calculated only from the nominal statistical model. To derive the expected limits, as in the case of the observed limits, 10000 toy samples were made every 40 GeV to create a distribution following the PDF $g(\tilde{q}_{\mu_{\text{hypo}}} \mid \mu_{\text{hypo}})$, and the same number of samples with the same interval were made to a distribution following the PDF $g(\tilde{q}_{\mu_{\text{hypo}}} \mid 0)$. Then, the CL_s value is calculated for each pair of the samples to create the distribution of CL_s with 10000 entries for each μ_{hypo} and m_X. The median, $\Phi(\pm 1)$-quantiles, and

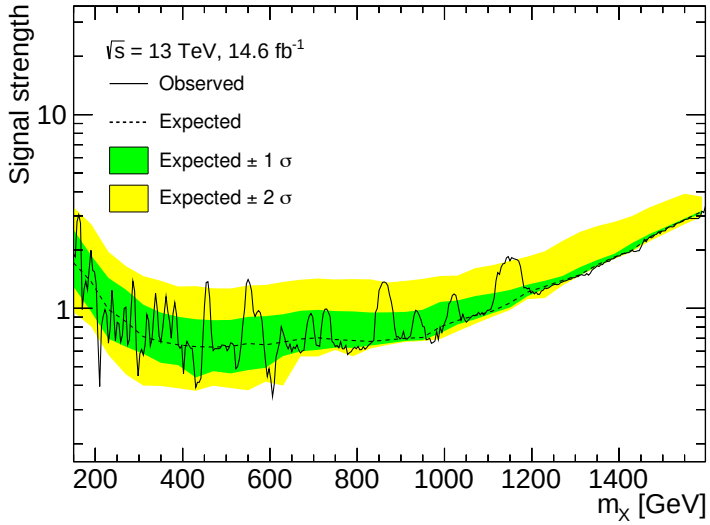


Figure 8.7: The expected and observed upper limits of signal strength μ at 95% confidence level as functions of hypothetical ALP mass m_X. The 1σ and 2σ confidence intervals are also illustrated by the colorized bands.

$\Phi(\pm 2)$-quantiles of the distribution are recorded, and they are denoted $\mathrm{CL_s}$, $\mathrm{CL_s}^{\pm 1\sigma}$, and $\mathrm{CL_s}^{\pm 2\sigma}$. The equations of μ_{hypo},

$$\begin{aligned} \mathrm{CL_s} &= 0.05, \\ \mathrm{CL_s}^{\mp 1\sigma} &= 0.05, \\ \mathrm{CL_s}^{\mp 2\sigma} &= 0.05, \end{aligned}$$

are solved to obtain $\mu_{95(\mathrm{exp.})}$, $\mu^{\pm 1\sigma}_{95(\mathrm{exp.})}$, and $\mu^{\pm 2\sigma}_{95(\mathrm{exp.})}$, corresponding to the expected limit, its 1σ confidence interval, and 2σ confidence interval, respectively. The dashed line and the colorized bands in Fig. 8.7 represent them.

A sharp downward swing around 200 GeV extends far beyond the 2σ band, but this is due to the difference in the number of samples created for the observed and expected limits. If the same number of samples is used, the relation between the band and observed limits looks more reasonable, as found comparing Figs. 8.9 and A.5. The step of the lower edge of the 2σ band around 650 GeV is caused by the small number of μ_{hypo} points (see Eq. 8.5) and the imprecise interpolation of μ_{hypo} dependencies in $\mathrm{CL_s}$.

The upper limits μ_{95} are converted to the limits of signal yield N_X and cross section σ_X as

$$\begin{aligned} N_{95} &= \mu_{95} N_{\mathrm{std}}, \\ \sigma_{95} &= \mu_{95} \sigma_{\mathrm{std}}, \end{aligned}$$

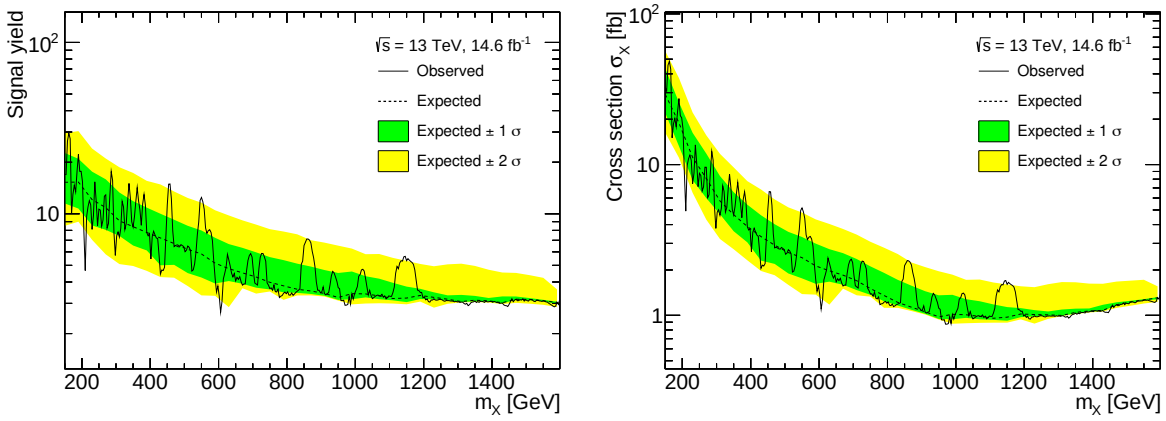
where N_{std} and σ_{std} are defined in Sec. 5.2.2, and

$$\sigma_X = \sum_{i \in \{\mathrm{EL,SD,DD}\}} \sigma_i.$$

Fig. 8.8 shows these results. Furthermore, according to Eq. 1.4, the limits of the ALP coupling constant f^{-1} are calculated as

$$f^{-1}_{95} = \sqrt{\mu_{95}} f^{-1}_{\mathrm{std}},$$

assuming the ALP's branching ratio to two photons is 100%. $f^{-1}_{\mathrm{std}} = 0.05\ \mathrm{TeV}^{-1}$ is the specified coupling constant when generating the signal MC samples. Fig. 8.9 shows the results. The ALP natural width



(a) Signal yield limits (b) Cross section limits

Figure 8.8: The expected and observed upper limits of (a) signal yield N_X and (b) signal cross section σ_X at 95% confidence level as functions of hypothetical ALP mass m_X. The 1σ and 2σ confidence intervals are also illustrated by the colorized bands.

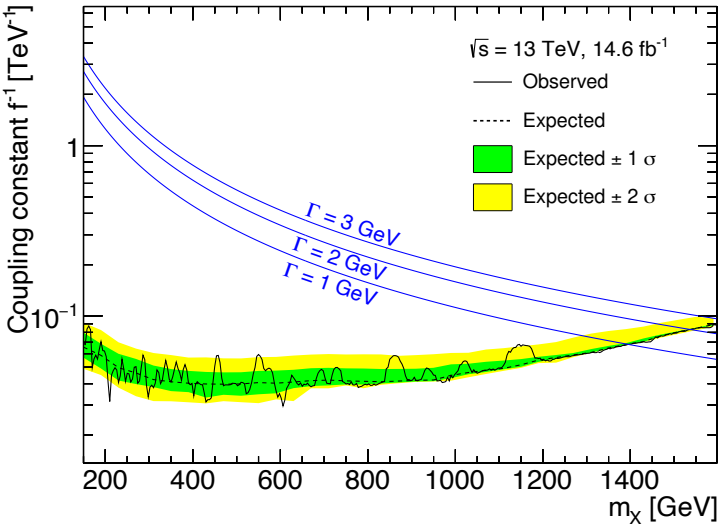


Figure 8.9: The expected and observed upper limits of the ALP coupling constant f^{-1} at 95% confidence level as functions of hypothetical ALP mass m_X. The 1σ and 2σ confidence intervals are also illustrated by the colorized bands. Contours of the ALP natural width Γ expressed in Eq. 1.2 are illustrated by the smooth blue solid lines.

is $\Gamma = 1$ GeV for $m_X \approx 1400$ GeV and $\Gamma = 3$ GeV for $m_X \approx 1600$ GeV. Such widths are small enough to be ignored because the detector resolution dominates them. The exclusion of a coupling constant f^{-1} much larger than the illustrated Γ contours is not justified in this research because the narrow-width approximation in the signal modeling does not hold. Fig. G.2 and G.3 show the results using the asymptotic formulae for comparison, although these are only for the expected limits.[5]

[5]This is because $\sigma_{\hat{\mu}}$ cannot be calculated when the signal strength $\hat{\mu}$ is negative due to the instability of the unconditional fit (see Appendix G).

9 Discussion and Conclusion

This work is one of the first analyses to correlate information from the forward proton detectors in terms of the kinematics with the central detector in collider experiments at the energy frontier, serving as a demonstration of the proton tagging for background rejection in a search for a BSM particle. In this section, the results are compared with the previous research. Then, the prospect is described while showing the significance of this work. Finally, this research is concluded.

9.1 Comparison with previous research

The limits of the coupling constant in Fig. 8.9 can be compared with the CMS-TOTEM results [74] in Fig. 1.9. Fig. 9.1 shows the comparison of the expected limits. The upper limits exceed the CMS results for $m_X < 700$ GeV and $m_X > 1100$ GeV. One reason for this difference is the set operation in the event selection. This research uses the OR selection, while the CMS-TOTEM uses the AND selection. The signal efficiency of the analyses is qualitatively the same as Fig. 4.19, so the CMS-TOTEM analysis has sensitivity in a limited mass range, while this analysis is sensitive in a wider range. In addition, this research generally has more stringent limits considering the difference of the integrated luminosity, 14.6 fb^{-1} for this research and 102.7 fb^{-1} for the CMS-TOTEM analysis. One reason for such higher sensitivity is that this research performs the resonance search with unbinned fits making use of the narrowness of the $m_{\gamma\gamma}$ distribution, while the CMS-TOTEM just performs the counting experiment with a single bin.

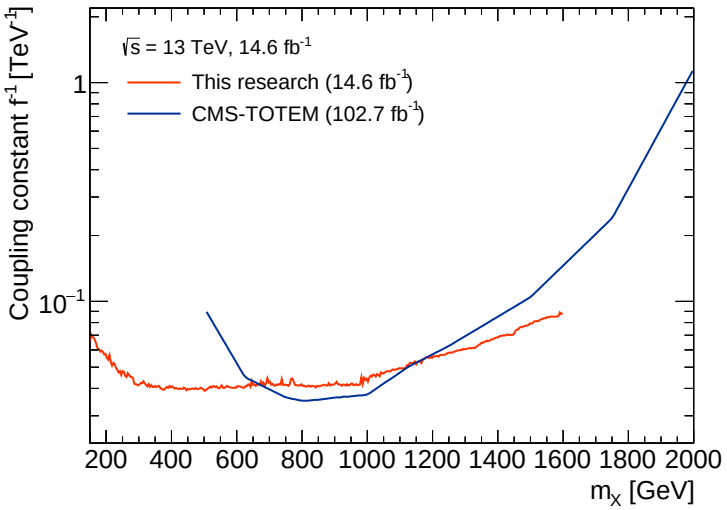


Figure 9.1: The comparison of the expected exclusion limits of ALP coupling constant f^{-1} from the CMS-TOTEM analysis results [74].

9.2 Prospect

9.2.1 Experimental setup

The physics data-taking of the LHC has been restarted with $\sqrt{s} = 13.6$ TeV in 2022 (Run 3). The Run 3 is planned to operate until 2025 to collect the data with up to 300 fb^{-1}, which is about 20 times higher than the integrated luminosity used in this research. This reduces the statistical uncertainty that was dominant in this research to about $1/\sqrt{20}$ times. There is no significant change in the instantaneous luminosity, and the luminosity leveling [157], an operation that dynamically adjusts the β-function during the run, has been introduced. Thus, the combinatorial background rate is comparable to this research. Furthermore, high-luminosity LHC (HL-LHC) experiment is planned to take place using the LHC from 2029. In HL-LHC, the data-taking is expected to perform at $\sqrt{s} = 14$ TeV to collect up to 4000 fb^{-1} of integrated luminosity with the instantaneous luminosity of 5×10^{34} cm^{2}s^{-1}.

Higher instantaneous luminosity means more pileup events and a concomitant increase in trigger rate. However, the trigger rate is limited due to finite storage capacity. Therefore, the ATLAS collaboration renovates the detector and trigger system in the Phase-I upgrade before Run 3 and Phase-II upgrade before HL-LHC to cope with the high luminosity of HL-LHC. For example, the EM calorimeter has a new trigger readout system in the Phase-I upgrade. The higher granularity of the cell division for the trigger readout was realized to improve the object identification in the triggers. The author has been engaged in the performance evaluation and calibration of this EM trigger readout system [158–160].

The AFP ToF detectors have been reinstalled to take the correct data this time. Furthermore, a plan to install new AFP SiT detectors in different positions from the SiTs used in Run 2 is being discussed for HL-LHC. The CMS-TOTEM is also planning to increase the PPS stations.

9.2.2 Event selection

This research has a lot of room for optimization of the event selection conditions. Since this research is the first search for new particles using AFP, the event selection was not aggressively optimized while the establishment of the background estimation etc., has been focused. Especially, the correction $0.1\xi_{\gamma\gamma}$ deteriorates the sensitivity. This was introduced considering the large beam optics uncertainty, but it is expected to be overestimated. For example, the $\Delta\xi$ distribution of the dilepton analysis for $\xi_{\mathrm{AFP}} \in [0.02, 0.12]$ in Fig. 2.17 is as narrow as 0.002. But it does not realize unless $\xi_{\mathrm{AFP}} = 0.02$ for all protons, as long as the beam optics uncertainty of ξ_{AFP} is 10%. Thus, this uncertainty is just initial and preliminarily and is being improved.

The acoplanarity selection can also be improved. Since the SD and DD signal samples were generated very late, as described in Sec. 3.2, the selection optimization carries over to the subsequent analysis.

The OR selection has the advantage of covering a wide mass range than the AND selection, but the AND selection contributes very strongly to background reduction and becomes valuable as the integrated luminosity increases. Then, it is conceivable to combine the AND and XOR selection results, which are mutually exclusive, to increase the sensitivity compared to the OR selection. Moreover, it is also possible to combine such results and the usual high-mass diphoton resonance search with results imposing the NOR condition on $\xi_{\gamma\gamma}$ to the latter. An initial study was performed in Ref. [161], and this approach is promising.

The primary vertex information can also be used for the event selection. For example, one can use the information about the number of high-energy tracks around the primary vertex or the tracks' η for the event selection. Furthermore, as the primary vertex of an EL signal event where both two photons are

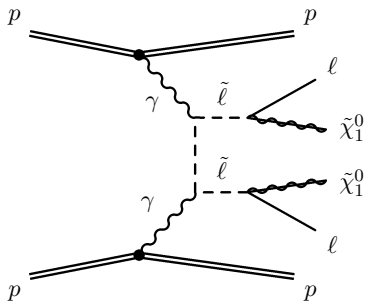


Figure 9.2: Feynman diagram of the event $\gamma\gamma \to \tilde{\chi}_1^0 \ell \tilde{\chi}_1^0 \ell$ in the pp collisions [162].

unconverted cannot be selected correctly as described in Sec. 3.3, it could be efficient to use the primary vertex reconstructed by the non-pointing or calo-vertexing mentioned in the same section can be used for the dedicated selection in combination with the photon pointing. In addition, the vertex position information from the AFP ToF detectors can also be used. To select events, one can compare the primary vertex position with the ToF vertex information. But this is feasible basically only when using the AND selection condition because the ToF vertexing needs protons on both A and C-sides.

9.2.3 Other analyses using the AFP detectors

AFP detectors are suited for analyses whose sensitivity has been limited by pileup and underlying events. Possible dark matter search using the AFP detectors is described in Refs. [162, 163]. It targets mildly compressed slepton $\tilde{l}$ and neutralino dark matter $\tilde{\chi}_1^0$. Fig. 9.2 shows the Feynman diagram of the signal event. The primary strategy is to constrain the kinematics using the relation,

$$\begin{aligned} W_{\text{miss}}^2 &= p_{\text{miss}}^2, \\ p_{\text{miss}} &= p_{\gamma 1} + p_{\gamma 2} - p_{\ell 1} - p_{\ell 2}, \\ \left(p_{\gamma 1} + p_{\gamma 2}\right)^2 &= m_{\gamma\gamma}^2 = \sqrt{s\xi_1\xi_2}, \end{aligned}$$

where W_{miss} and p_{miss} are the mass and 4-momentum of the invisible $2\tilde{\chi}_1^0$ system, p_{xi} is the 4-momentum of the i-th particle x, and the last relation uses Eq. 4.2. Ref. [162] expects good sensitivity with 100 fb^{-1} of luminosity, but Ref. [163] expects it is barely sensitive even with 300 fb^{-1}. The former focuses on the photon-initiated diboson production ($\gamma\gamma \to WW$) [164] as the background, while the latter determines the dominant backgrounds to be the SD dilepton events (see Fig. 1.8) and pileup events. However, the AFP ToF detectors are not considered in these evaluations. If the ToF is used, the ToF vertex is matched to the primary vertex reconstructed by the ATLAS detector responses, and the pileup background is reduced roughly by a factor of 10. This is the same as the matching selection in this research in that it requires consistency between the ATLAS detector information and the AFPs information. Thus, the multiple assignment procedure for the background sample creation described in Sec. 6.1 could be applied to model background. The difference from this research is that the matching is not performed for each side.

As exemplified above, there is a lot of possible application of the methodology developed in this research to BSM searches using AFP. On the other hand, the AFP can also be used to identify the event type when an excess from SM is observed in another BSM search not using the AFP. If the signal event is photon initiated, the event can have associated forward protons and possibly fewer tracks around the primary vertex. Therefore, information such as ξ_{AFP}, ToF vertex position, and track multiplicity can be used to see if the excess originates from photon-induced events.

9.3 Conclusion

A diphoton resonance search has been performed using the ATLAS and AFP detectors with the 14.6 fb^{-1} data of the LHC pp collisions in 2017 at $\sqrt{s}$ = 13 TeV. A diphoton trigger was used, and the photons were required to have high transverse momentum and to be back-to-back in the x-y plane. In the signal event, photons originating from the beam protons scatter with each other, mediated by an ALP. Thus, this research is also a light-by-light scattering search. The protons remain intact or dissociate. Intact protons that lose roughly 2% to 10% of their energy in the event are detected by AFP, and their fractional energy loss ξ are measured. The ξ is measured from the diphoton system as well, and the difference between the two ξ values is used for the event selection to reduce the background events.

The search was performed with statistical tests on the diphoton mass distributions of the samples. The target mass range is $[150, 1600]$ GeV. The signal and background PDFs are modelled to fit them into the samples. The signal modeling was performed using the signal MC samples generated by SuperChic. The DSCB function was used to model the resonant peak, and each parameter was determined with systematic uncertainty by interpolating the fit results to the MC samples. The background modeling was performed in a fully data-driven manner. The dominant source of background is the coincident matching between the diphoton object and protons from diffractive pileup events. The background sample was made from the data reassigning the proton sets to diphoton from the other events. It is used to model the background and evaluate the systematic uncertainty of the modeling. Background events other than the combinatorial background are found to be negligible.

Various systematic uncertainties were studied. The AFP global alignment is the source of the dominant systematic uncertainty of signal yield of a resonant peak. Most of the uncertainty in ξ measured by AFP is due to a lack of knowledge about the magnetic fields between the ATLAS detector and the AFPs, and a correction was made to the event selection conditions to suppress this effect. Statistical uncertainty is finally found to be the dominant source of total uncertainty.

The results were expressed as the significance and exclusion limits. The statistical tests were performed for each hypothetical ALP mass point m_X. The maximum observed local significance was 2.51 at m_X = 454 GeV, and there was no excess from the standard model. The exclusion limits were set for the signal yield, signal cross section, and ALP coupling constant with two photons assuming the branching ratio is 100%. The observed upper limits of the coupling constant were about 0.04 TeV^{-1} in the mass range m_X = 200 to 1000 GeV. The limits are compared with the previous research where both outgoing protons are required to match with the diphoton kinematics. This research has more sensitivity in a wider range than the previous research because requiring at least one proton matching extends the mass range with high acceptance. Moreover, the statistical modeling specialized in diphoton resonance search makes use of the narrow-width of the resonant $m_{\gamma\gamma}$ peak and enhances the sensitivity. This research is the first analysis specialized for BSM search using the AFP. There is a lot of possibility of optimizing event selection criteria for future analysis with higher integrated luminosity and applying the methodology developed in this research to other analyses.

A Auxiliary plots

Auxiliary plots are collected in this appendix. Detailed plots not covered in each section are shown below. Each plot is referenced in each section.

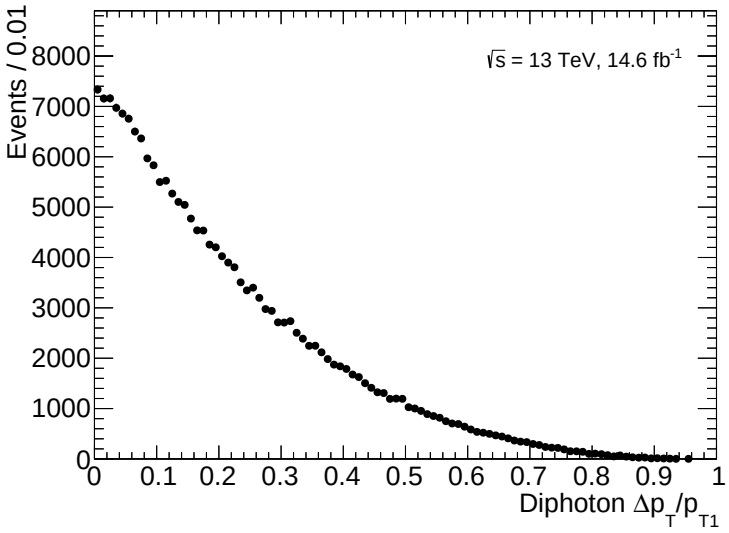
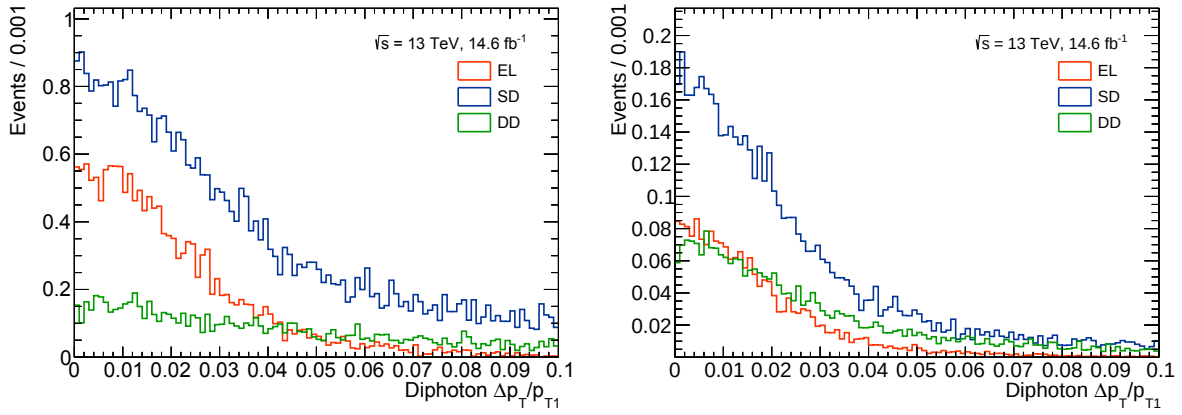


Figure A.1: The distribution of p_{T} difference between the detector level two photons relative to the leading photon's p_{T} in the data before the acoplanarity selection.



(a) m_X = 300 GeV (b) m_X = 1200 GeV

Figure A.2: The distribution of p_{T} difference between the detector level two photons relative to the leading photon's p_{T} in the signal MC samples before the acoplanarity selection.

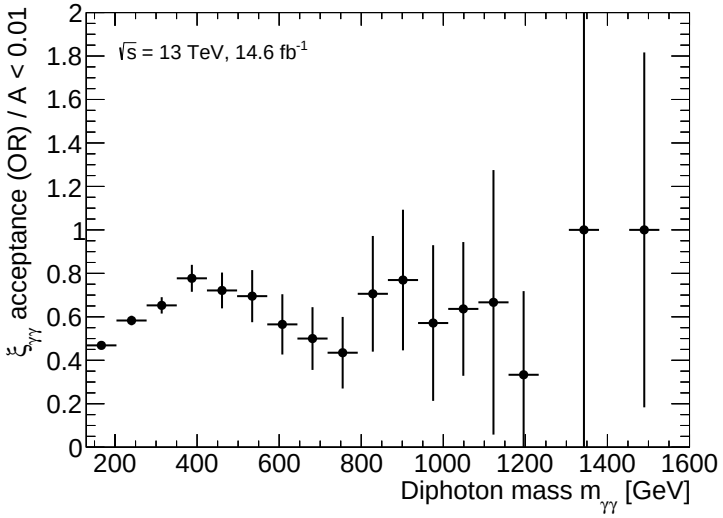
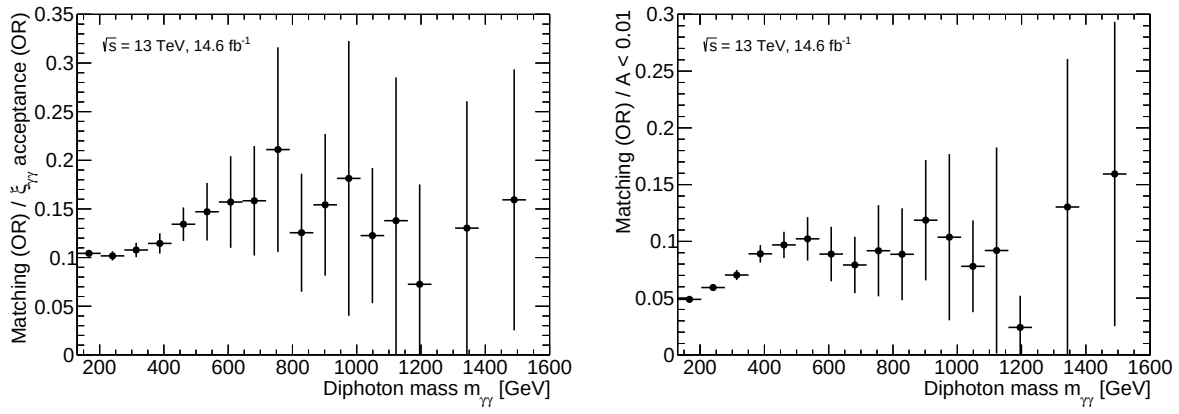


Figure A.3: $m_{\gamma\gamma}$ dependency of the $\xi_{\gamma\gamma}$ cut efficiency derived from the data for the OR condition dividing the $m_{\gamma\gamma}$ distribution after the $\xi_{\gamma\gamma}$ cut by the distribution just after the acoplanarity cut.



(a) Matching efficiency (b) Overall efficiency

Figure A.4: (a) $m_{\gamma\gamma}$ dependency of the matching selection efficiency derived from the data for the OR condition dividing the $m_{\gamma\gamma}$ distribution after the matching by the distribution just after the $\xi_{\gamma\gamma}$ cut. (b) $m_{\gamma\gamma}$ dependency of the overall selection efficiency derived from the data and the multi-shifted data for the OR condition dividing the $m_{\gamma\gamma}$ distribution after the matching by the distribution just after the acoplanarity cut.

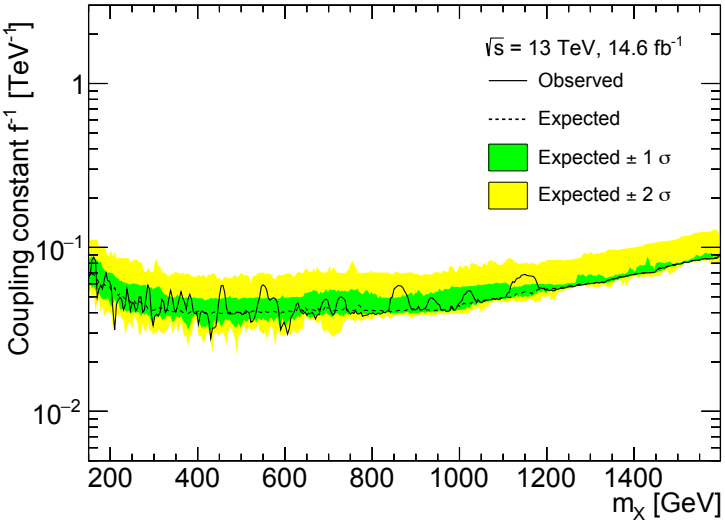


Figure A.5: The expected and observed upper limits of the ALP coupling constant f^{-1} at 95% confidence level as functions of hypothetical ALP mass m_X. The 1σ and 2σ confidence intervals are also illustrated by the colorized bands. The bands are made from 1000 toy MC samples generated for each μ_{hypo} and m_X.

B Auxiliary tables

Auxiliary tables are collected in this appendix. Detailed tables not covered in each section are shown below. Each table is referenced in each section.

Table B.1: PYTHIA configuration for single-vertex background estimation.

Parameter	**Process**		
	SD $\to \gamma\gamma$	SD $\to jj$	ND $\to jj$
`Beams:eCM`	`13000.`	`13000.`	`13000.`
`Diffraction:doHard`	`on`	`on`	`on`
`Diffraction:sampleType`	`4`	`4`	`4`
`PhaseSpace:pTHatMin`	`40.`	`40.`	`40.`
`PhaseSpace:mHatMin`	`130.`	`130.`	`130.`
`PromptPhoton:ffbar2gammagamma`	`on`		
`PromptPhoton:gg2gammagamma`	`on`		
`SigmaDiffractive:PomFlux`		`1`	`1`
`PDF:PomSet`		`4`	`4`
`HardQCD:all`		`on`	`on`

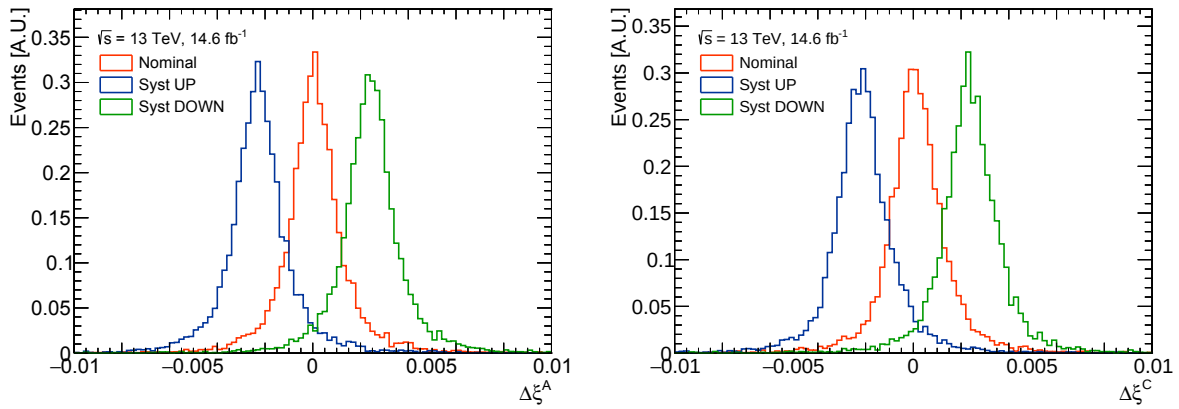
Table B.2: Background efficiency model parameters.

	p	q
$i = 0$	295	16.9
$i = 1$	−41.7	14.4
$i = 2$	−8.76	0.0347
$i = 3$	4.54	2.92
$i = 4$	−0.593	0.00404

C Systematic variations of $\Delta\xi$ distribution

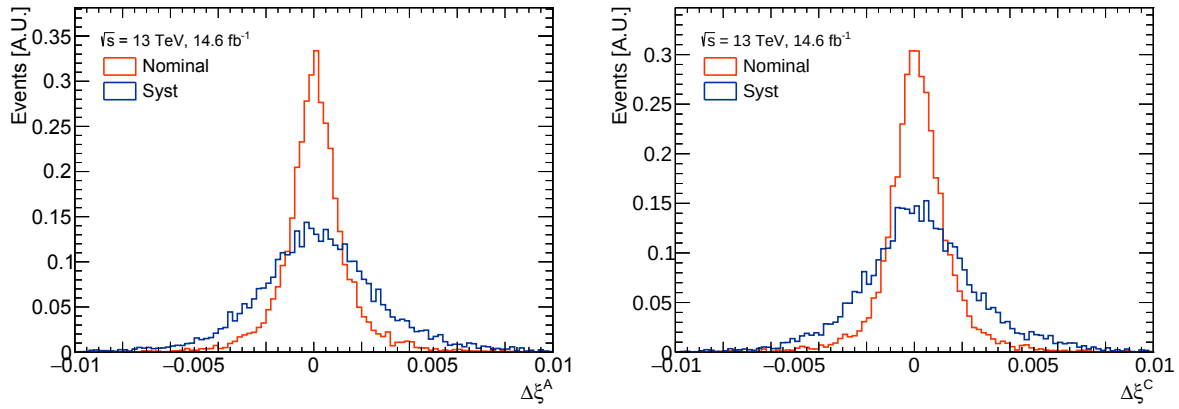
The global alignment uncertainty affects the signal yield by shifting the $\Delta\xi$ distributions. Fig. C.1 shows the $\Delta\xi^{\pm}$ distributions of the signal events where the global alignment configuration in the simulation is redefined as the nominal value added or subtracted by its uncertainty. The distributions represent the nominal case, upper-shifted case, and lower-shifted case.

The AFP track smearing uncertainty described in Sec 7.2 also affects the signal yield by changing the width of the $\Delta\xi$ distributions. Fig. C.2 shows the $\Delta\xi^{\pm}$ distributions of the signal events where the AFP track smearing configuration in the simulation is redefined as the nominal value added or subtracted by its uncertainty. The nominal case and the systematically variated case are compared.



(a) A-side (b) C-side

Figure C.1: $\Delta\xi$ distributions of the m_X = 1000 GeV simulated signal obtained with nominal and systematically varied global alignment parameters. The distributions are normalized by the integrals.



(a) A-side (b) C-side

Figure C.2: $\Delta\xi$ distributions of the m_X = 1000 GeV simulated signal obtained with nominal and systematically varied AFP track smearing configurations in the AFP simulation. The distributions are normalized by the integrals.

D Loose $\xi_{\gamma\gamma}$ cut effect evaluation

For the signal event simulation, the ξ_{AFP} resolution for a given true ξ originates from some modeling of the AFP resolution and uncertainty. However, the tail of the distribution may not actually follow the model primarily constructed using Gaussian distributions. The $\xi_{\gamma\gamma}$ selection is set looser than the ξ_{AFP} selection as described in Sec. 4. This can cause the situation that the non-negligible number of events are selected at the tail of ξ_{AFP} distribution, considering the events where $\xi_{\gamma\gamma}$ is outside of the ξ_{AFP} selection range but inside the $\xi_{\gamma\gamma}$ selection range. Fig. D.1 shows such a situation. Such events at the tail of ξ_{AFP} distribution are possible not to be modelled well. That is, the evaluated systematic uncertainty could be incorrect for the events. Therefore, additional systematic uncertainty is needed to take this situation into account.

D.1 Proton energy loss fraction resolution

Suppose the $\xi_{\gamma\gamma}$ resolution is much worse than the ξ_{AFP} resolution. In that case, the additional systematic uncertainty does not need to be set because the well-modelled photon kinematics uncertainty dominates the $\Delta\xi$ uncertainty. The $\xi_{\gamma\gamma}$ and ξ_{AFP} resolutions are investigated for each ALP mass. Fig. D.2 and D.3 shows the distributions of relative difference of reconstructed $\xi_{\gamma\gamma}$ and ξ_{AFP} from the true values. The ξ_{AFP} resolution is better than the $\xi_{\gamma\gamma}$ resolution for the nominal case. However, since the difference is not tremendous, the situation of interest cannot be naively ignored.

D.2 Threshold changing test

The signal MC samples for the nominal and systematically varied cases are used to compare the number of events remaining when the entire selection is applied for each of the nominal $\xi_{\gamma\gamma}$ selection and the $\xi_{\gamma\gamma}$

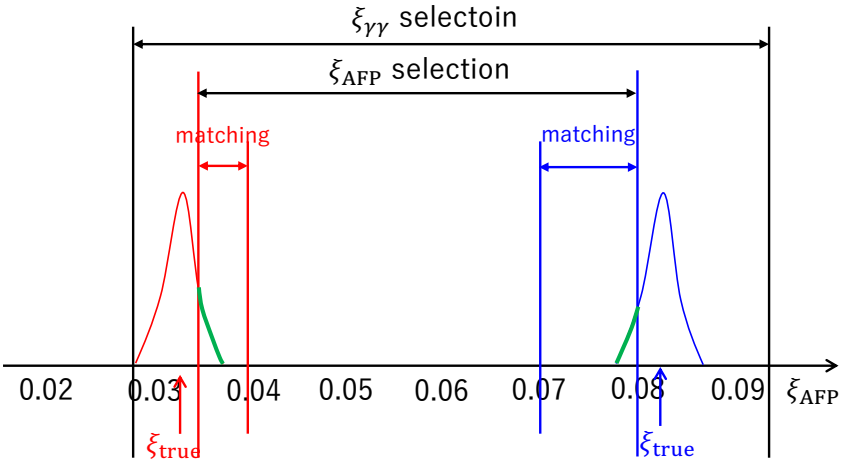
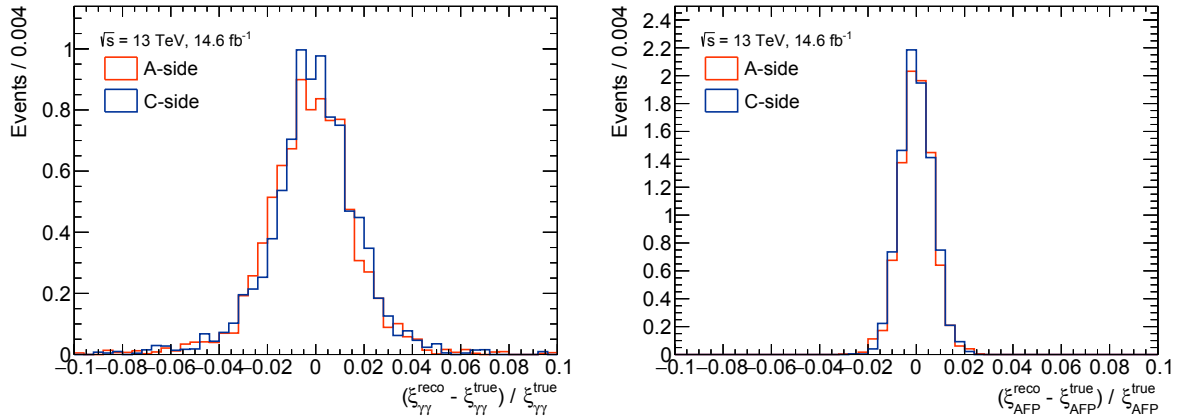
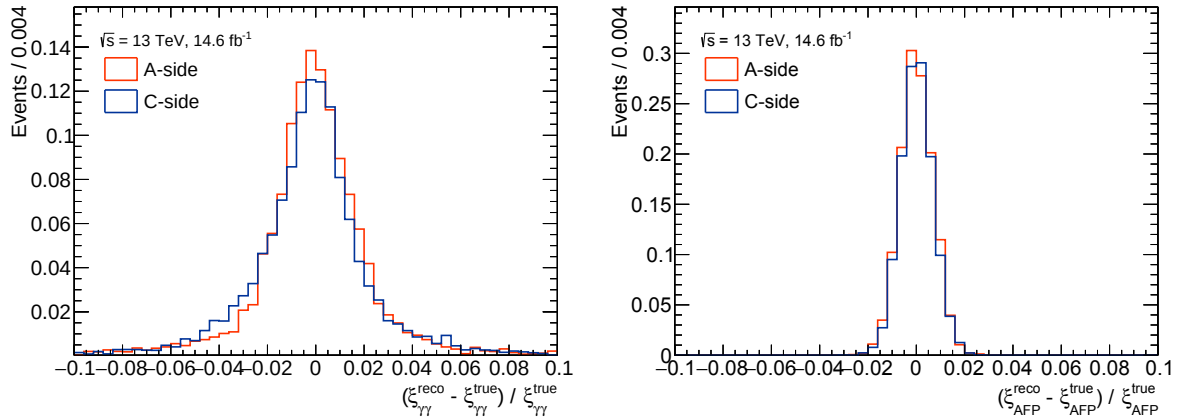


Figure D.1: ξ_{AFP} distributions for events where $\xi_{\gamma\gamma}$ is outside of the ξ_{AFP} selection range but inside the $\xi_{\gamma\gamma}$ selection range. The green parts show the events which are selected by the matching criteria.



(a) $\left(\xi_{\gamma\gamma}^{\text{reco}} - \xi_{\gamma\gamma}^{\text{true}}\right)/\xi_{\gamma\gamma}^{\text{true}}$ (b) $\left(\xi_{\text{AFP}}^{\text{reco}} - \xi_{\text{AFP}}^{\text{true}}\right)/\xi_{\text{AFP}}^{\text{true}}$

Figure D.2: Distributions of the relative difference of reconstructed (a) $\xi_{\gamma\gamma}$ and (b) ξ_{AFP} from the truth values evaluated using the 200 GeV ALP signal MC sample.



(a) $\left(\xi_{\gamma\gamma}^{\text{reco}} - \xi_{\gamma\gamma}^{\text{true}}\right)/\xi_{\gamma\gamma}^{\text{true}}$ (b) $\left(\xi_{\text{AFP}}^{\text{reco}} - \xi_{\text{AFP}}^{\text{true}}\right)/\xi_{\text{AFP}}^{\text{true}}$

Figure D.3: Distributions of the relative difference of reconstructed (a) $\xi_{\gamma\gamma}$ and (b) ξ_{AFP} from the truth values evaluated using the 1400 GeV ALP signal MC sample.

selection with the same condition as the xi_{AFP} selection. The latter has fewer events left, and its selection range $[0.035, 0.08]$ is regarded as the safe range in which the uncertainty modeling is correct, as assumed in the dilepton analysis [66]. The difference in the number of events between the two cases is regarded as the event loss in case the $\xi_{\gamma\gamma}$ selection is tightened and interpreted as the number of events where the ξ_{AFP} may be mismodelled. For each ALP mass m_X, the event loss fractions are shown in Fig. D.4.

Only 2.2% of all events are possible to be mismodelled in terms of the ξ_{AFP} distribution when considering only the nominal case. The 2.2% events may not have a reliable systematic uncertainty. It can be naturally assumed that it is conservative enough if the signal yield uncertainty of such events is evaluated as 100%. That is, the nominal $\xi_{\gamma\gamma}$ selection described in Sec. 4.2 can be used, and 2.2% of the signal yield uncertainty should be considered.

Although the event loss fractions of the samples with some systematic variations are larger than the nominal case, they should not be considered as the uncertainty originating from the difference of the selection range between the nominal $\xi_{\gamma\gamma}$ selection and ξ_{AFP} selection. For example, when the systematic variation for the track smearing configuration is applied, the ξ_{AFP} distribution becomes more expansive, and the matching range tends to select a part of the distribution closer to its center. Therefore, even if the number of failed

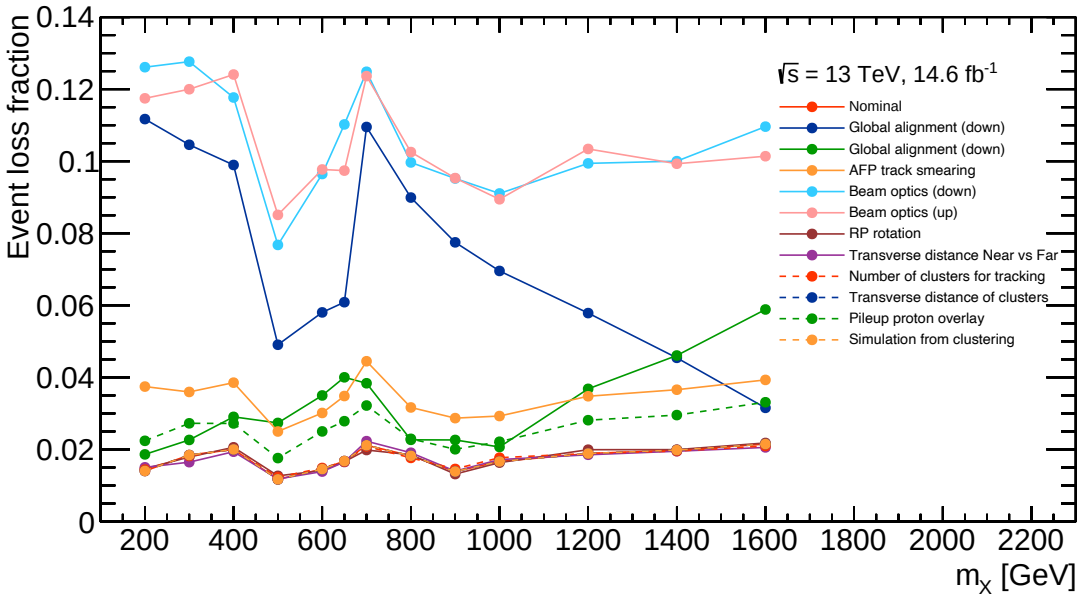


Figure D.4: Event loss fractions when changing the $\xi_{\gamma\gamma}$ selection from the nominal one to the condition identical to the ξ_{AFP} selection. Samples with systematic variations (see Sec. 7) are also tested.

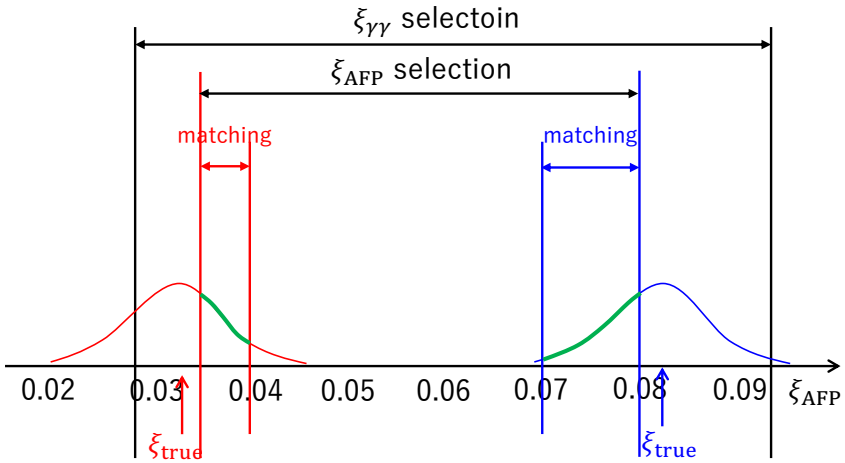


Figure D.5: ξ_{AFP} distributions for events where $\xi_{\gamma\gamma}$ is outside of the ξ_{AFP} selection range but inside the $\xi_{\gamma\gamma}$ selection range. The green parts show the events which are selected by the matching criteria.

events increases by the selection threshold change, it only means that the number of events with better ξ_{AFP} modeling increases. Fig. D.5 shows such a situation. In the other case, when the systematic variation for AFP global alignment or the beam optics configuration is applied, the ξ_{AFP} distribution is shifted. The number of failed events increases only when the peak is systematically shifted towards the ξ_{AFP} selection range. Fig. D.6 shows such a situation. The core of the distribution is more likely to be selected than the nominal problematic case, so even if the number of failed events increases by the selection threshold change, it only means that the number of events with better ξ_{AFP} modeling increases.

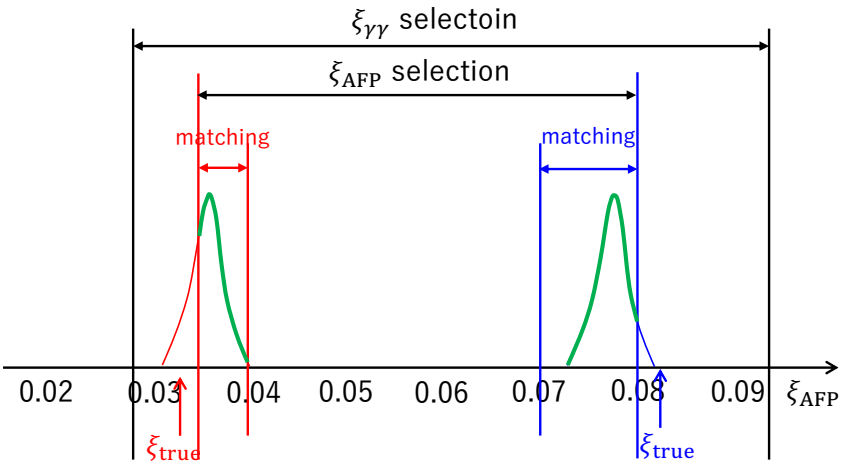


Figure D.6: ξ_{AFP} distributions for events where $\xi_{\gamma\gamma}$ is outside of the ξ_{AFP} selection range but inside the $\xi_{\gamma\gamma}$ selection range. The green parts show the events which are selected by the matching criteria.

E Luminosity dependency of combinatorial background

The event selection efficiency after the diphoton selection depends on the number of interactions per BC μ as described in Sec. 6.1. In addition, the selection efficiency differs depending on diphoton kinematics because the matching uses the diphoton information. The efficiency of the selection requiring only A(C)-side matching is modelled as

$$\varepsilon_{\mathrm{A,C}}(\mu, \xi^{\pm}_{\gamma\gamma}) = 1 - \left(1 - \varepsilon_{\mathrm{A,C}}(1, \xi^{\pm}_{\gamma\gamma})\right)^{\mu} \approx \mu\varepsilon_{\mathrm{A,C}}(1, \xi^{\pm}_{\gamma\gamma}),$$

where $\varepsilon_{\mathrm{A,C}}(1, \xi^{\pm}_{\gamma\gamma})$ is the efficiency per pp collision depending only on the diphoton kinematics. The matching efficiency is low enough for the approximation to hold. Using them, the OR selection efficiency can be written as

$$\varepsilon_{\mathrm{OR}}(\mu, \xi^{\pm}_{\gamma\gamma}) = \varepsilon_{\mathrm{A}}(\mu, \xi^{\pm}_{\gamma\gamma}) + \varepsilon_{\mathrm{C}}(\mu, \xi^{\pm}_{\gamma\gamma}) - \varepsilon_{\mathrm{AND}}(\mu, \xi^{\pm}_{\gamma\gamma}),$$

where $\varepsilon_{\mathrm{AND}}(\mu, \xi^{\pm}_{\gamma\gamma})$ is the AND selection efficiency. If $\varepsilon_{\mathrm{AND}}(\mu, \xi^{\pm}_{\gamma\gamma})$ is negligible, $\varepsilon_{\mathrm{OR}}(\mu, \xi^{\pm}_{\gamma\gamma})$ can be regarded to be proportional to μ. The event weight $R \equiv \mu_{\gamma\gamma}/\mu_p$ discussed in Sec. 6.1 uses this approximation.

In general, as μ increases, the proportionality breaks down. $\varepsilon_{\mathrm{OR}}(\mu, \xi^{\pm}_{\gamma\gamma})$ is convex upward and monotonically increases with respect to μ. Using the fact that it converges to 1 when $\mu \to \infty$, the uncertainty of the event weight is calculated. Fig. E.1 shows $\varepsilon_{\mathrm{OR}}(\mu)$ for a diphoton in case the efficiency is not proportional to μ. μ_{low} and μ_{high} are $\min\{\mu_{\gamma\gamma}, \mu_p\}$ and $\max\{\mu_{\gamma\gamma}, \mu_p\}$, respectively. The left (a) case is that the efficiency saturates for both $\mu_{\gamma\gamma}$ and μ_p. The lengths of blue and red lines, in this case, are $\delta/\mu_{\mathrm{high}}$ and $\delta/\mu_{\mathrm{low}}$, respectively, where $\delta = \mu_{\mathrm{high}} - \mu_{\mathrm{low}}$. The more general case (b) is that the efficiency is between the proportionality and saturation for $\mu_{\gamma\gamma}$ and μ_p. The lengths of blue and red lines, in this case, are $\varepsilon_{\mathrm{low}}\delta/\mu_{\mathrm{high}}$ and $\varepsilon_{\mathrm{low}}\delta/\mu_{\mathrm{low}}$, respectively, where $\varepsilon_{\mathrm{low}} = \varepsilon_{\mathrm{OR}}(\mu_{\mathrm{low}})$. Since $\varepsilon_{\mathrm{low}} < 1$, the systematic deviation is maximum in case of (a). Therefore, if $\mu_p \geq \mu_{\gamma\gamma}$, the event weight has

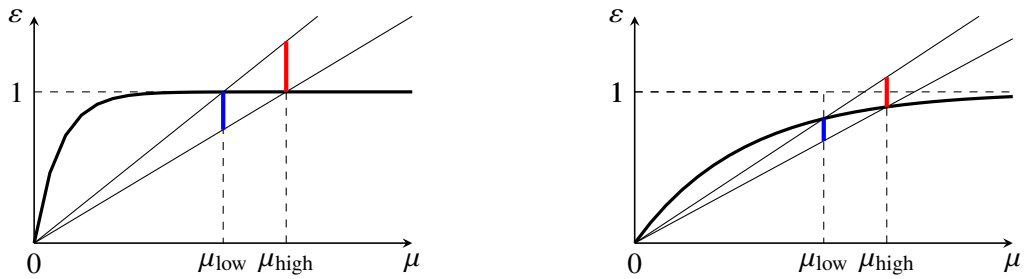
$$\frac{\delta}{\mu_{\mathrm{high}}} = 1 - R$$

of upper uncertainty. On the other hand, if $\mu_p < \mu_{\gamma\gamma}$, the event weight has

$$\frac{\delta}{\mu_{\mathrm{low}}} = R - 1$$

of lower uncertainty. Thus, the event weight with the systematic uncertainty is $R^{+(1-R)}_{-0}$ for $R \leq 1$ and $R^{+0}_{-(R-1)}$ for $R > 1$. This means that the systematic uncertainty of R can be conservatively evaluated in any case by comparing the cases with and without the event weight R applied.

Fig. E.2 compares the background templates with only the upward and the downward shifts according to these uncertainties with the nominal template. The uncertainty is about 0.5%. This uncertainty is neglected because the function fitting in this research is insensitive to the number of background events and non-local changes in the distribution shape.



(a) Fast convergence (b) Slow convergence

Figure E.1: $\varepsilon_{\mathrm{OR}}(\mu)$ for a diphoton. Two different ramps represent the badness of the assumption that the efficiency is proportional to μ.

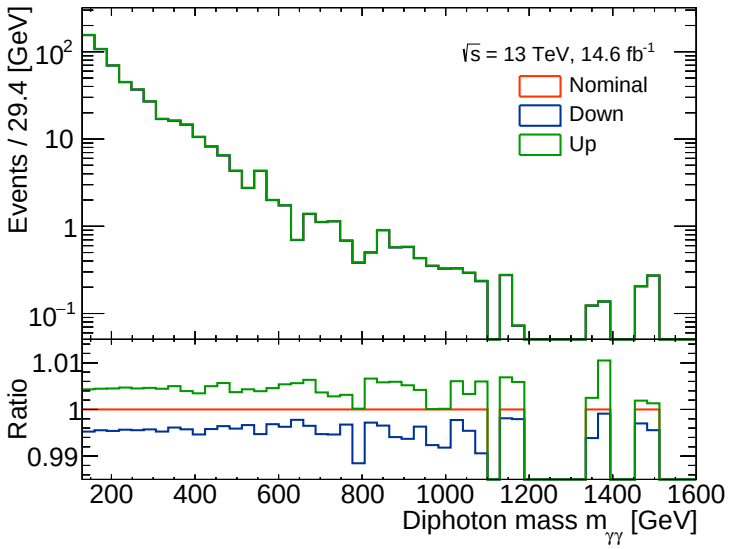
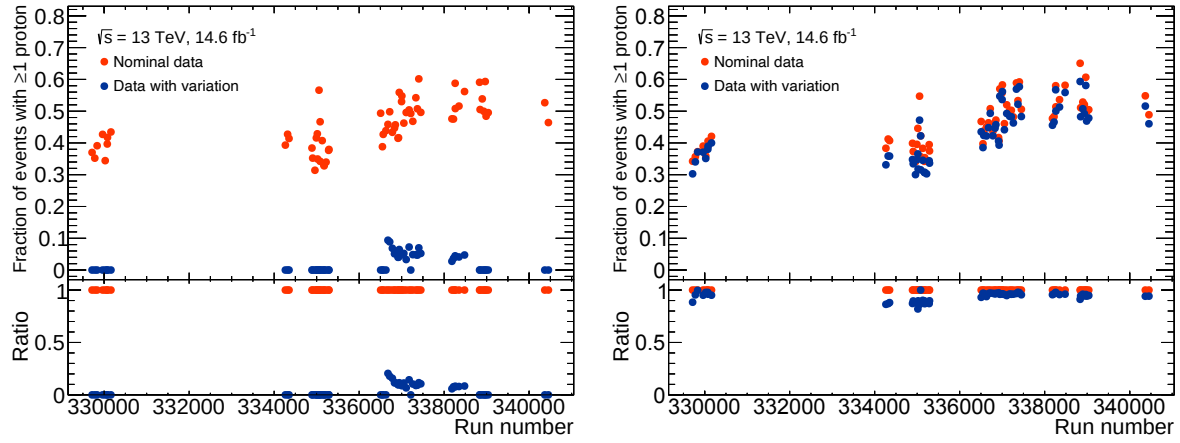


Figure E.2: The background templates with only the upward and the downward shifts considering the $\mu_{\gamma\gamma}/\mu_p$ uncertainties compared with the nominal template.

F Variation of requirement on the number of AFP clusters

In the proton object-related systematic uncertainty evaluation, changing the minimum required number of clusters to reconstruct a track is found to have relatively significant effects on the signal yield. Thus, the effect is evaluated from the data with the real clustering rather than the clusters obtained just by smearing the truth proton positions in the clustered MC samples. However, the resulting difference between the nominal and changed cases is huge because one of the A-side AFP FAR station planes was not working permanently. Fig. F.1 shows the fraction of events with at least one reconstructed proton for each run number[1] in data for each side. The efficiency of the A-side is low when the change is applied. The test requires at least three clusters to reconstruct a track, while the nominal case requires at least two. In addition, proton reconstruction for each side requires tracks from both NEAR and FAR stations for both cases. Therefore, protons cannot be reconstructed on A-side unless clusters are in all three working planes of FAR station. This is why the A-side efficiency is very low in this case.

Thus, the relative change of the number of events between the nominal and changed cases after the whole selection is used as the proton reconstruction efficiency uncertainty, where the final set operation in the event selection requires only the C-side selection result. Fig. F.2 shows the cut flow of the changed data. The decrease of events on the C-side is 6.6% of the original number of events in this research illustrated in Fig. 4.16. This value is used as the systematic uncertainty of proton reconstruction efficiency.



(a) A-side (b) C-side

Figure F.1: Fraction of events with at least one reconstructed proton for each run number in data for (a) A-side and (b) C-side. The red points show the nominal data, and the blue points show the changed data.

[1] Run number is the identifier of the LHC pp collision run.

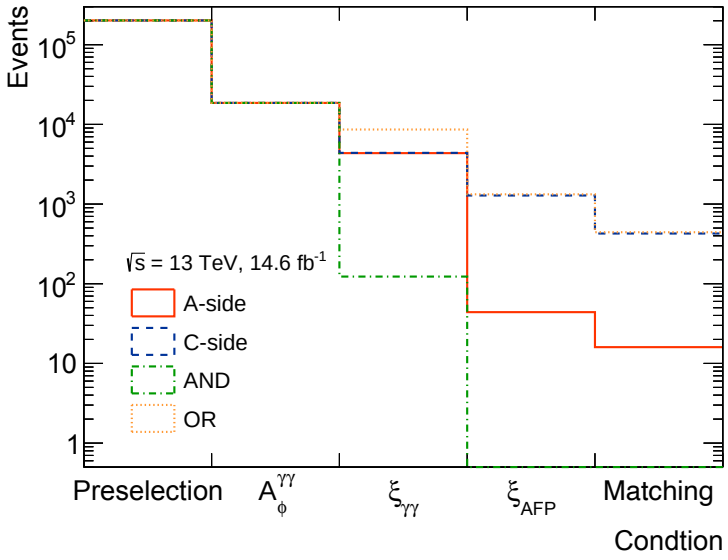


Figure F.2: Cut flow of the changed data after the preselection. The values for $\xi_{\gamma\gamma}$ and ξ_{AFP} are the numbers of events when the set operation is performed without the latter selection.

G Results from asymptotic formulae

In this section, the results in Sec. 8.3 are replaced by the asymptotic formulae for comparison. The formulae are derived and detailed in Ref. [151].

The significance and p-value are calculated as

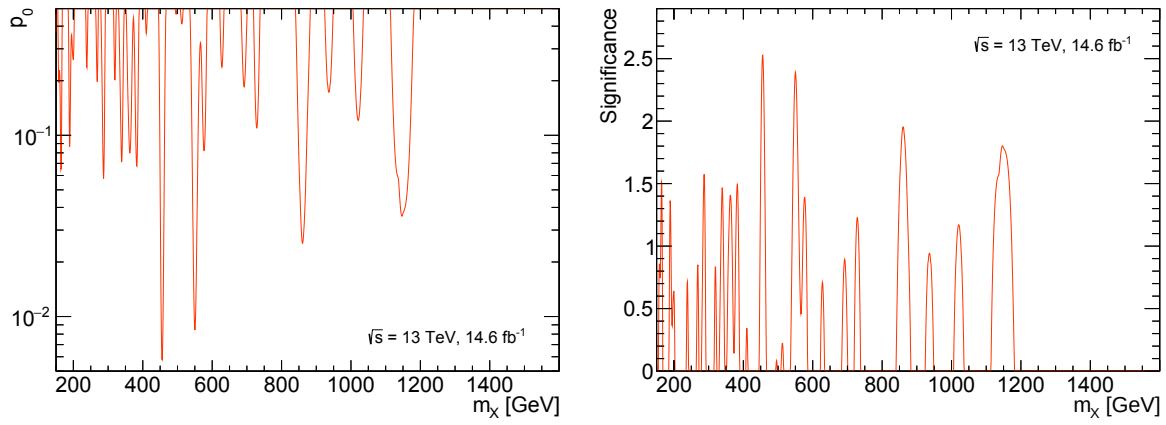
$$Z_0 = \sqrt{q_0},$$
$$p_0 = 1 - \Phi\left(Z_0\right),$$

by the asymptotic formulae. Fig. G.1 shows the results. The maximum observed significance is 2.53 at $m_X = 456$ GeV.

The limits of the signal strengths $\mu^{N\sigma}_{95(\text{exp.})}$ are calculated as the root of the equation of μ_{hypo},

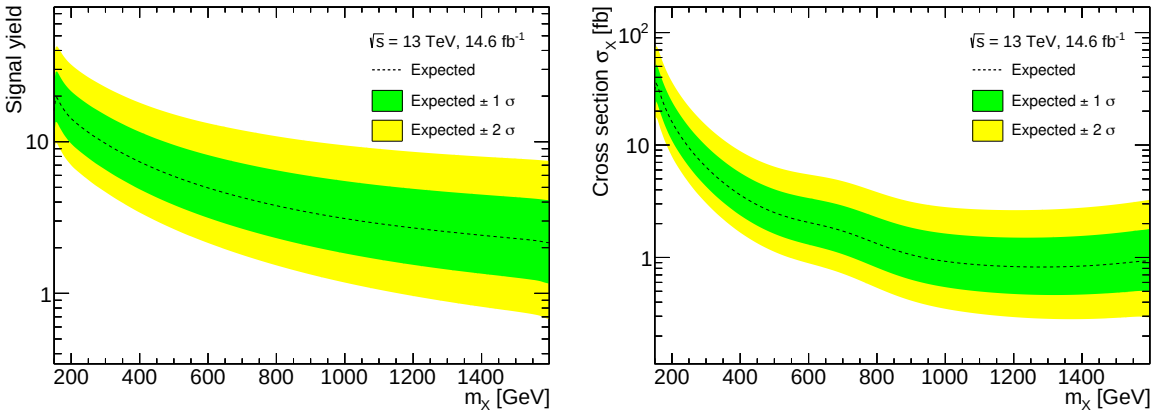
$$\mu_{\text{hypo}} = \sigma_{\hat{\mu}}\left[\Phi^{-1}\left(1 - 0.05\Phi(N)\right) + N\right],$$

where $\sigma_{\hat{\mu}} = \mu_{\text{hypo}}/\sqrt{\tilde{q}_{\mu_{\text{hypo}},\text{A}}}$, and $\tilde{q}_{\mu_{\text{hypo}},\text{A}}$ is the calculated $\tilde{q}_{\mu_{\text{hypo}}}$ from an Asimov sample generated assuming $\mu_{\text{true}} = 0$ [165]. This is converted to the limits of the signal strength, cross section, and coupling constant, shown in Figs. G.2 and G.3.



(a) p-value (b) Significance

Figure G.1: Test results for null hypothesis $\mu = 0$ using the unblinded data as functions m_X. (a) p-value p_0. (b) Significance Z_0. The asymptotic formulae are used.



(a) Signal yield limits N_{95} (b) Cross section limits σ_{95}

Figure G.2: Expected exclusion limits of signal yield and cross section.

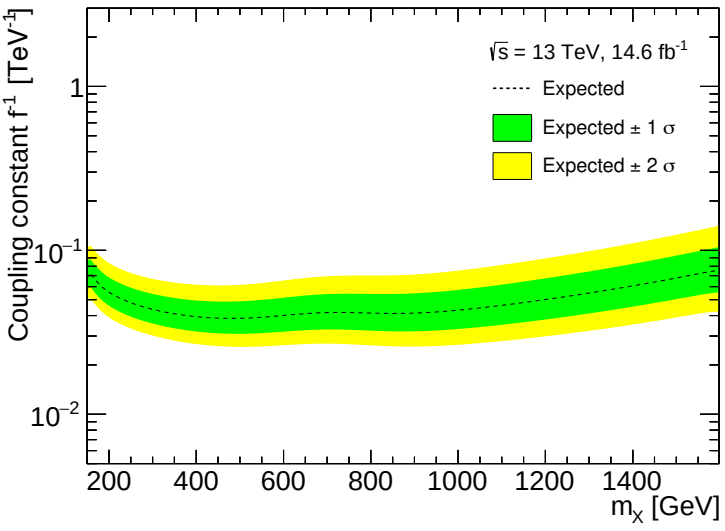


Figure G.3: Expected exclusion limits in the m_X-f^{-1} plane.

Acknowledgements

I would like to express my gratitude to the referees Prof. Toshinori Mori, Prof. Takeo Moroi, Prof. Yutaka Ushiroda, Associate Prof. Wataru Ootani, and Associate Prof. Taku Gunji, who review this thesis. I would also like to thank Prof. Junichi Tanaka (my supervisor) and Assistant Prof. Yuji Enari, for their careful advice on this thesis. I also thank the SPRING GX program of the University of Tokyo for supporting this research.

Everyone at ICEPP gave me a lot of advice for my research. Prof. Tanaka, Assistant Prof. Enari, Associate Prof. Yasuyuki Okumura, and Assistant Prof. Tatsuya Masubuchi have always consulted me about this research. The Ref. [22] shared by Prof. Shoji Asai was one of the triggers for me to start this research. Prof. Tanaka and Associate Prof. Koji Terashi helped me get used to physics analysis for about a year after I joined the analysis team. Dr. Takuya Nobe told me about the ATLAS EFT framework, and Dr. Shion Chen gave me ideas for future analysis using AFP. Assistant Prof. Masahiko Saito consulted me about statistical test methods. Associate Prof. Ryu Sawada and Associate Prof. Tetsuro Mashimo provided advice on computing. All the secretaries also helped me a lot.

I would also like to thank the members of the analysis team to which I belong. As an analysis contact, Dr. André Sopczak (Prague CTU) allowed me to proceed with the analysis smoothly. Dr. Peter J Bussey (University of Glasgow), as one of the earliest members of the team, gave me various comments on the analysis. Dr. Rafał Staszewski (Krakow IFJ PAN), based on his experience with the dilepton analysis [66], gave me a lot of helpful advice for almost all phases of the analysis. Research Prof. Marek Taševský (Prague AS) kindly answered many questions about the AFP and SuperChic. Mr. Denis-Patrick Odagiu, a CERN summer student, made the prototype of my analysis software. Several other undergraduate and master's students also participated in this analysis and produced exciting results.

The editorial board members Dr. Simone Pagan Griso (Berkeley LBNL), Associate Prof. Deepak Kar (University of Witwatersrand), and Dr. Mustafa Schmidt (the University of Wuppertal) provided insightful comments and made the analysis robust during the ATLAS approval stage of this research. The analysis sub-group conveners, Dr. Danijela Bogavac (CERN) and Dr. Christian Sander (DESY), and the analysis group conveners, Dr. Flavia De Almeida Dias (Nikhef, Amsterdam) and Dr. Tamara Vazquez Schroeder (CERN) provided many comments on this analysis and approved it smoothly. As members of the dilepton analysis [66], Dr. Lydia Beresford (DESY) and Dr. Jesse Liu (University of Cambridge) gave me many comments and answered my questions.

Dr. Lucian Harland-Lang (University College London), the author of SuperChic, updated SuperChic for this research and answered many questions about it. Research Prof. Taševský often intermediated us. Dr. Ewelina Maria Lobodzińska (DESY), Dr. Aleksandra Dimitrievska (Berkeley LBNL), William Patrick Mccormack (Berkeley LBNL), and Dr. Pagan Griso proceeded the implementation of SuperChic in the ATLAS software.

Dr. Chris Hayes (the University of Michigan), Dr. Mungo Davide (University of Toronto) and Dr. Chris Meyer (Indiana University) supported the introduction of analysis tools. Dr. Ahmed Tarek (Michigan SU) and Dr. Anthony Morley (CERN) answered many questions about the tool and sample productions. Nicolas Berger (LAPP Annecy) has generously supported me as the author of the function fitting tool. Dr. Kiley Kennedy (Princeton University) introduced some materials as an analyser of calo-vertexing.

Dr. Krzysztof Ciesla (Krakow AGH-UST) and Dr. Staszewski answered software questions related to AFP and helped with updates. Mr. Adrien Auriol (University of Birmingham) and Miss Maura Barros (LIP) helped me to include the pileup effect in the AFP simulation for the systematic uncertainty evaluation.

KEK-IINAS Iwate Collider School 2022 taught me much about MC sample generation methodology. Especially, Associate Prof. Kentarou Mawatari (Iwate University) showed me how to generate signal MC samples using MadGraph. Also, Mr. Atsuya Niki (University of Tokyo), whom I met at the school, invited me to a seminar and gave me a chance to study.

Finally, my family has contributed greatly to my research life.

Many other people also helped me. This research was completed thanks to all the above people. I appreciate all the contributions. Thank you very much.

Bibliography

[1] W. Heisenberg and H. Euler, *Folgerungen aus der Diracschen Theorie des Positrons*, Z. Physik **98** (1936) 714, arXiv: physics/0605038 [physics.hist-ph] (cit. on p. 7).

[2] R. R. Wilson, *Scattering of 1.33 Mev Gamma-Rays by an Electric Field*, Phys. Rev. **90** (1953) 720 (cit. on p. 7).

[3] G. Jarlskog et al., *Measurement of Delbrück Scattering and Observation of Photon Splitting at High Energies*, Phys. Rev. D **8** (1973) 3813 (cit. on p. 7).

[4] M. Schumacher, F. Smend, and I. Borchert, *Delbrück scattering of 2.75 MeV photons by lead*, Phys. Lett. B **59** (1975) 134 (cit. on p. 7).

[5] S. Z. Akhmadaliev et al., *Delbrück scattering at energies of 140–450 MeV*, Phys. Rev. C **58** (1998) 2844 (cit. on p. 7).

[6] S. Z. Akhmadaliev et al., *Experimental Investigation of High-Energy Photon Splitting in Atomic Fields*, Phys. Rev. Lett. **89** (2002) 061802, arXiv: hep-ex/0111084 (cit. on p. 7).

[7] T. Inada, *Search for axion-like particles using strong pulsed magnets at SPring-8*, PhD thesis: The University of Tokyo, 2015, URL: https://doi.org/10.15083/00073282 (cit. on p. 7).

[8] S. Kamioka, *Search for vacuum magnetic birefringence with a high repetitive pulsed magnet*, PhD thesis: The University of Tokyo, 2020, URL: https://www.icepp.s.u-tokyo.ac.jp/research/thesis.html (cit. on p. 7).

[9] J. S. Schwinger, *On Gauge Invariance and Vacuum Polarization*, Phys. Rev. **82** (1951) 664 (cit. on p. 7).

[10] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001 (cit. on pp. 7, 19).

[11] ATLAS Collaboration, *Evidence for light-by-light scattering in heavy-ion collisions with the ATLAS detector at the LHC*, Nature Phys. **13** (2017) 852, arXiv: 1702.01625 [hep-ex] (cit. on pp. 7, 15).

[12] E. Fermi, *Sulla teoria dell' urto tra atomi e corpuscoli elettrici*, Nuovo Cim **2** (1925) 323, arXiv: hep-th/0205086 (cit. on p. 7).

[13] C. Weizsäcker, *Ausstrahlung bei Stößen sehr schneller Elektronen*, Z. Physik **88** (1934) 612 (cit. on pp. 7, 8).

[14] E. J. Williams, *Nature of the High Energy Particles of Penetrating Radiation and Status of Ionization and Radiation Formulae*, Phys. Rev. **45** (1934) 729 (cit. on p. 7).

[15] A. Baltz et al., *The physics of ultraperipheral collisions at the LHC*, Phys. Rept. **458** (2008) 1 (cit. on p. 7).

[16] S. R. Klein and P. Steinberg, *Photonuclear and Two-Photon Interactions at High-Energy Nuclear Colliders*, Annual Review of Nuclear and Particle Science **70** (2020) 323 (cit. on p. 7).

[17] E. Fermi, *Über die Theorie des Stoßes zwischen Atomen und elektrisch geladenen Teilchen*, Z. Physik **29** (1924) 315 (cit. on p. 7).

[18] E. J. Williams, *Correlation of certain collision problems with radiation theory*, Kong. Dan. Vid. Sel. Mat. Fys. Med. **13** (1935) 1 (cit. on p. 8).

[19] Z.-L. Ma, Z. Lu, and L. Zhang, *Validity of equivalent photon spectra and the photoproduction processes in p-p collisions*, Nucl. Phys. B **974** (2022) 115645 (cit. on p. 8).

[20] V. Budnev, I. Ginzburg, G. Meledin, and V. Serbo, *The two-photon particle production mechanism. Physical problems. Applications. Equivalent photon approximation*, Phys. Rept. **15** (1975) 181 (cit. on p. 8).

[21] K. Piotrzkowski, *Tagging two-photon production at the CERN Large Hadron Collider*, Phys. Rev. D **63** (2001) 071502 (cit. on p. 8).

[22] S. Fichet, G. von Gersdorff, B. Lenzi, C. Royon, and M. Saimpert, *Light-by-light scattering with intact protons at the LHC: from standard model to new physics*, JHEP **02** (2015) 165, arXiv: 1411.6629 [hep-ph] (cit. on pp. 9, 13, 145).

[23] C. Baldenegro, S. Fichet, G. von Gersdorff, and C. Royon, *Searching for axion-like particles with proton tagging at the LHC*, JHEP **06** (2018) 131, arXiv: 1803.10835 [hep-ph] (cit. on pp. 10, 12, 13, 86).

[24] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on pp. 10, 21, 25).

[25] ATLAS Collaboration, *ATLAS Forward Proton Phase-I Upgrade: Technical Design Report*, ATLAS-TDR-024; CERN-LHCC-2015-009, 2015, URL: https://cds.cern.ch/record/2017378 (cit. on pp. 10, 27, 28, 31).

[26] R. Bruce et al., *New physics searches with heavy-ion collisions at the CERN Large Hadron Collider*, J Phys G Nucl Part Phys **47** (2020) 060501, arXiv: 1812.07688 [hep-ph] (cit. on p. 11).

[27] CMS and TOTEM Collaborations, *CMS-TOTEM Precision Proton Spectrometer*, CERN-LHCC-2014-021; TOTEM-TDR-003; CMS-TDR-13, 2014, URL: https://cds.cern.ch/record/1753795 (cit. on p. 10).

[28] CMS Collaboration, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004 (cit. on p. 10).

[29] TOTEM collaboration, *The TOTEM experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08007 (cit. on p. 10).

[30] FP420 R&D Collaboration, *The FP420 R&D Project: Higgs and New Physics with forward protons at the LHC*, JINST **4** (2009) T10001, arXiv: 0806.0302 [hep-ex] (cit. on p. 10).

[31] R. L. Workman et al., *Review of Particle Physics*, PTEP **2022** (2022) 083C01 (cit. on pp. 11, 38).

[32] C. A. Baker et al., *Improved Experimental Limit on the Electric Dipole Moment of the Neutron*, Phys. Rev. Lett. **97** (2006) 131801, arXiv: hep-ex/0602020 (cit. on p. 11).

[33] C. Abel et al., *Measurement of the Permanent Electric Dipole Moment of the Neutron*, Phys. Rev. Lett. **124** (2020) 081803, arXiv: 2001.11966 [hep-ex] (cit. on p. 11).

[34] R. D. Peccei and H. R. Quinn, CP *Conservation in the Presence of Pseudoparticles*, Phys. Rev. Lett. **38** (1977) 1440 (cit. on p. 11).

[35] R. D. Peccei and H. R. Quinn,
Constraints imposed by CP *conservation in the presence of pseudoparticles*,
Phys. Rev. D **16** (1977) 1791 (cit. on p. 11).

[36] S. Weinberg, *A New Light Boson?* Phys. Rev. Lett. **40** (1978) 223 (cit. on p. 11).

[37] F. Wilczek, *Problem of Strong P and T Invariance in the Presence of Instantons*,
Phys. Rev. Lett. **40** (1978) 279 (cit. on p. 11).

[38] W. A. Bardeen, S.-H. Tye, and J. Vermaseren,
Phenomenology of the new light Higgs boson search, Phys. Lett. B **76** (1978) 580 (cit. on p. 12).

[39] W. A. Bardeen, R. Peccei, and T. Yanagida, *Constraints on variant axion models*,
Nucl. Phys. B **279** (1987) 401 (cit. on p. 12).

[40] H. Leutwyler, *The ratios of the light quark masses*, Phys. Lett. B **378** (1996) 313,
arXiv: hep-ph/9602366 (cit. on p. 12).

[41] K. Agashe, A. Delgado, M. J. May, and R. Sundrum, *RS1, custodial isospin and precision tests*,
JHEP **2003** (2003) 050, arXiv: hep-ph/0308036 (cit. on p. 13).

[42] L. Randall and R. Sundrum, *Large Mass Hierarchy from a Small Extra Dimension*,
Phys. Rev. Lett. **83** (17 1999) 3370, arXiv: hep-ph/9905221 (cit. on p. 13).

[43] K. Agashe, R. Contino, and A. Pomarol, *The minimal composite Higgs model*,
Nucl. Phys. B **719** (2005) 165, arXiv: hep-ph/0412089 (cit. on p. 13).

[44] M. Bauer, M. Neubert, and A. Thamm, *Collider probes of axion-like particles*,
JHEP **12** (2017) 044, arXiv: 1708.00443 [hep-ph] (cit. on pp. 13, 15).

[45] K. Schmieden, *Searches for Axion Like Particles at the LHC*, PoS **LHCP2021** (2021) 005
(cit. on p. 14).

[46] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1,
arXiv: 1207.7214 [hep-ex] (cit. on p. 14).

[47] CMS Collaboration,
Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,
Phys. Lett. B **716** (2012) 30, arXiv: 1207.7235 [hep-ex] (cit. on p. 14).

[48] ATLAS Collaboration, *Search for extra dimensions in diphoton events from proton–proton collisions recorded at* $\sqrt{s} = 7$ *TeV in the ATLAS detector at the LHC*,
New J. Phys. **15** (2013) 043007, arXiv: 1210.8389 [hep-ex] (cit. on p. 14).

[49] ATLAS Collaboration, *Search for high-mass diphoton resonances in* pp *collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, Phys. Rev. D **92** (2015) 032004, arXiv: 1504.05511 [hep-ex]
(cit. on p. 14).

[50] ATLAS Collaboration,
Search for resonances in diphoton events at $\sqrt{s} = 13$ *TeV with the ATLAS detector*,
JHEP **09** (2016) 001, arXiv: 1606.03833 [hep-ex] (cit. on pp. 14, 82).

[51] ATLAS Collaboration, *Search for new phenomena in high-mass diphoton final states using* $37\,fb^{-1}$ *of proton–proton collisions collected at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*,
Phys. Lett. B **775** (2017) 105, arXiv: 1707.04147 [hep-ex] (cit. on p. 14).

[52] ATLAS Collaboration, *Search for resonances decaying into photon pairs in* $139\,fb^{-1}$ *of* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Lett. B **822** (2021) 136651,
arXiv: 2102.13405 [hep-ex] (cit. on pp. 14, 16, 17, 37, 42, 67, 82, 88, 91).

[53] CMS Collaboration, *Search for signatures of extra dimensions in the diphoton mass spectrum at the Large Hadron Collider*, Phys. Rev. Lett. **108** (2012) 111801, arXiv: 1112.0688 [hep-ex] (cit. on p. 14).

[54] CMS Collaboration, *Search for diphoton resonances in the mass range from* 150 *to* 850 *GeV in pp collisions at* $\sqrt{s}$ = 8 *TeV*, Phys. Lett. B **750** (2015) 494, arXiv: 1506.02301 [hep-ex] (cit. on p. 14).

[55] CMS Collaboration, *Search for Resonant Production of High-Mass Photon Pairs in Proton–Proton Collisions at* $\sqrt{s}$ = 8 *and* 13 *TeV*, Phys. Rev. Lett. **117** (2016) 051802, arXiv: 1606.04093 [hep-ex] (cit. on p. 14).

[56] CMS Collaboration, *Search for high-mass diphoton resonances in proton–proton collisions at* 13 *TeV and combination with* 8 *TeV search*, Phys. Lett. B **767** (2017) 147, arXiv: 1609.02507 [hep-ex] (cit. on p. 14).

[57] CMS Collaboration, *Search for physics beyond the standard model in high-mass diphoton events from proton–proton collisions at* $\sqrt{s}$ = 13 *TeV*, Phys. Rev. D **98** (2018) 092001, arXiv: 1809.00327 [hep-ex] (cit. on p. 14).

[58] CMS Collaboration, *Evidence for light-by-light scattering and searches for axion-like particles in ultraperipheral PbPb collisions at* $\sqrt{s_{NN}}$ = 5.02 *TeV*, Phys. Lett. B **797** (2019) 134826, arXiv: 1810.04602 [hep-ex] (cit. on p. 15).

[59] ATLAS Collaboration, *Observation of Light-by-Light Scattering in Ultraperipheral Pb+Pb Collisions with the ATLAS Detector*, Phys. Rev. Lett. **123** (2019) 052001, arXiv: 1904.03536 [hep-ex] (cit. on pp. 15, 16).

[60] ATLAS Collaboration, *Measurement of light-by-light scattering and search for axion-like particles with* 2.2 nb^{-1} *of Pb+Pb data with the ATLAS detector*, JHEP **03** (2021) 243, arXiv: 2008.05355 [hep-ex] (cit. on pp. 15–17), Erratum: JHEP **11** (2021) 050.

[61] F. Neuhaus, *Search for Axions at the LHC*, PhD thesis: Johannes Gutenberg-Universität Mainz, 2022, URL: https://doi.org/10.25358/openscience-7448 (cit. on p. 15).

[62] S. Knapen, T. Lin, H. K. Lou, and T. Melia, *Searching for Axionlike Particles with Ultraperipheral Heavy-Ion Collisions*, Phys. Rev. Lett. **118** (2017) 171801, arXiv: 1607.06083 [hep-ph] (cit. on p. 15).

[63] D. Aloni, C. Fanelli, Y. Soreq, and M. Williams, *Photoproduction of Axionlike Particles*, Phys. Rev. Lett. **123** (2019) 071801, arXiv: 1903.03586 [hep-ph] (cit. on p. 15).

[64] NA64 Collaboration, *Search for Axionlike and Scalar Particles with the NA64 Experiment*, Phys. Rev. Lett. **125** (2020) 081801, arXiv: 2005.02710 [hep-ex] (cit. on p. 15).

[65] Belle II Collaboration, *Search for Axionlike Particles Produced in* e^+e^- *Collisions at Belle II*, Phys. Rev. Lett. **125** (16 2020) 161806, arXiv: 2007.13071 [hep-ex] (cit. on p. 15).

[66] ATLAS Collaboration, *Observation and Measurement of Forward Proton Scattering in Association with Lepton Pairs Produced via the Photon Fusion Mechanism at ATLAS*, Phys. Rev. Lett. **125** (2020) 261801, arXiv: 2009.14537 [hep-ex] (cit. on pp. 15–17, 33, 35, 37, 79, 80, 85, 103, 136, 145).

[67] ATLAS Collaboration, *Measurement of exclusive* $\gamma\gamma \to \ell^+\ell^-$ *production in proton–proton collisions at* $\sqrt{s}$ = 7 *TeV with the ATLAS detector*, Phys. Lett. B **749** (2015) 242, arXiv: 1506.07098 [hep-ex] (cit. on p. 15).

[68] ATLAS Collaboration, *Measurement of the exclusive $\gamma\gamma \to \mu^+\mu^-$ process in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **777** (2018) 303, arXiv: 1708.04053 [hep-ex] (cit. on p. 15).

[69] ATLAS Collaboration, *Observation of Centrality-Dependent Acoplanarity for Muon Pairs Produced via Two-Photon Scattering in Pb+Pb Collisions at $\sqrt{s_{NN}} = 5.02$ TeV with the ATLAS Detector*, Phys. Rev. Lett. **121** (2018) 212301, arXiv: 1806.08708 [hep-ex] (cit. on p. 15).

[70] CMS Collaboration, *Exclusive $\gamma\gamma \to \mu^+\mu^-$ production in proton–proton collisions at $\sqrt{s} = 7$ TeV*, JHEP **01** (2012) 052, arXiv: 1111.5536 [hep-ex] (cit. on p. 15).

[71] CMS Collaboration, *Search for exclusive or semi-exclusive $\gamma\gamma$ production and observation of exclusive and semi-exclusive e^+e^- production in pp collisions at $\sqrt{s} = 7$ TeV*, JHEP **11** (2012) 080, arXiv: 1209.1666 [hep-ex] (cit. on p. 15).

[72] CMS and TOTEM Collaborations, *Observation of proton-tagged, central (semi)exclusive production of high-mass lepton pairs in pp collisions at 13 TeV with the CMS-TOTEM precision proton spectrometer*, JHEP **07** (2018) 153, arXiv: 1803.04496 [hep-ex] (cit. on p. 15).

[73] CMS and TOTEM Collaborations, *First Search for Exclusive Diphoton Production at High Mass with Tagged Protons in Proton–Proton Collisions at $\sqrt{s} = 13$ TeV*, Phys. Rev. Lett. **129** (2021) 011801, arXiv: 2110.05916 [hep-ex] (cit. on p. 16).

[74] CMS Collaboration, *Search for high-mass exclusive diphoton production with tagged protons in PPS*, CMS-PAS-EXO-21-007; TOTEM-NOTE-2022-005, 2022, URL: https://cds.cern.ch/record/2810862 (cit. on pp. 16, 17, 57, 123).

[75] J. Pequenao, *Computer generated image of the whole ATLAS detector*, CERN-GE-0803012, 2008, URL: https://cds.cern.ch/record/1095924 (cit. on pp. 18, 22).

[76] R. Garoby, *Multiple splitting in the PS: results and alternative filling schemes*, 2001, URL: https://cds.cern.ch/record/567169 (cit. on p. 20).

[77] E. Mobs, *The CERN accelerator complex - August 2018*, OPEN-PHO-ACCEL-2018-005, 2018, URL: https://cds.cern.ch/record/2636343 (cit. on p. 20).

[78] M. Trzebiński, *Machine optics studies for the LHC measurements*, SPIE **9290** (2014) 929026, ed. by R. S. Romaniuk, arXiv: 1408.1836 [physics.acc-ph] (cit. on pp. 21, 27).

[79] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, ATLAS-TDR-19; CERN-LHCC-2010-013, 2010, URL: https://cds.cern.ch/record/1291633 (cit. on p. 22), Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009, 2012, URL: https://cds.cern.ch/record/1451888.

[80] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, JINST **13** (2018) T05008, arXiv: 1803.00844 [physics.ins-det] (cit. on p. 22).

[81] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017 (cit. on p. 23).

[82] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021, 2019, URL: https://cds.cern.ch/record/2677054 (cit. on pp. 23, 99).

[83] S. van der Meer, *Calibration of the effective beam height in the ISR*, CERN-ISR-PO-68-31; ISR-PO-68-31, 1968, URL: https://cds.cern.ch/record/296752 (cit. on p. 23).

[84] P. Grafström and W. Kozanecki, *Luminosity determination at proton colliders*, Prog. Part. Nucl. Phys. **81** (2015) 97 (cit. on p. 23).

[85] J. Pequenao, *Computer generated image of the ATLAS Liquid Argon*, CERN-GE-0803016, 2008, URL: https://cds.cern.ch/record/1095928 (cit. on p. 24).

[86] ATLAS Collaboration, *ATLAS LAr Calorimeter Phase-II Upgrade: Technical Design Report*, ATLAS-TDR-027; CERN-LHCC-2017-018, 2017, URL: https://cds.cern.ch/record/2285582 (cit. on p. 24).

[87] M. Aharrouche et al., *Response uniformity of the ATLAS liquid argon electromagnetic calorimeter*, Nucl. Instrum. Meth. A **582** (2007) 429, arXiv: 0709.1094 [physics.ins-det] (cit. on p. 25).

[88] M. Aleksa et al., *ATLAS Combined Testbeam: Computation and Validation of the Electronic Calibration Constants for the Electromagnetic Calorimeter*, ATL-LARG-PUB-2006-003, ATL-COM-LARG-2006-003, 2006, URL: https://cds.cern.ch/record/942528 (cit. on p. 25).

[89] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys. J. C **77** (2017) 317, arXiv: 1611.09661 [hep-ex] (cit. on p. 26).

[90] ATLAS Collaboration, *ATLAS Forward Detectors for Measurement of Elastic Scattering and Luminosity: Technical Design Report*, ATLAS-TDR-18; CERN-LHCC-2008-004, 2008, URL: https://cds.cern.ch/record/1095847 (cit. on pp. 27, 30).

[91] S. A. Khalek et al., *The ALFA Roman Pot detectors of ATLAS*, JINST **11** (2016) P11013, arXiv: 1609.00249 [physics.ins-det] (cit. on p. 27).

[92] M. Trzebiński, *Prospects for Proton-Proton Measurements with Tagged Protons in ATLAS*, 2019, arXiv: 1909.10827 [physics.ins-det] (cit. on p. 27).

[93] M. Trzebiński, *The ATLAS ALFA and AFP detectors - the experimental challenge of measuring forward protons at the LHC*, ATL-LUM-SLIDE-2013-252, 2013, URL: https://cds.cern.ch/record/1548122 (cit. on p. 28).

[94] A. Auriol, *Performances of AFP detector during LHC Run 2*, ATL-FWD-SLIDE-2021-725, 2021, URL: https://cds.cern.ch/record/2798061 (cit. on pp. 28, 32, 46).

[95] I. López Paz, *The one-armed ATLAS forward proton detector: Characterization, Installation, Commissioning and Performance*, PhD thesis: Barcelona, Autonoma U., 2018, URL: https://ddd.uab.cat/record/195362 (cit. on p. 27).

[96] The ATLAS IBL Collaboration, *Prototype ATLAS IBL modules using the FE-I4A front-end readout chip*, JINST **7** (2012) P11010, arXiv: 1209.1906 [physics.ins-det] (cit. on p. 29).

[97] G.-F. Dalla Betta et al., *Small pitch 3D devices*, PoS **Vertex 2016** (2017) 028 (cit. on p. 29).

[98] P. A. Erland, *ATLAS Forward Proton DetectorsStatus and Plans*, PoS **VERTEX2018** (2019) 007 (cit. on p. 29).

[99] ATLAS Collaboration, *Performance of the ATLAS Forward Proton Time-of-Flight Detector in 2017*, ATL-FWD-PUB-2021-002, 2021, URL: https://cds.cern.ch/record/2749821 (cit. on pp. 29, 30).

[100] M. Trzebiński, *ATLAS Forward Proton Detector*, 2017, URL: https://indico.cern.ch/event/682159/ (cit. on p. 31).

[101] P. R. Newman, *The ATLAS AFP Spectrometer*, ATL-FWD-SLIDE-2021-546, 2021, URL: https://cds.cern.ch/record/2781707 (cit. on p. 31).

[102] M. Kocian, *Readout and trigger for the AFP detector at ATLAS experiment*, JINST **12** (2017) C01077 (cit. on pp. 30, 37).

[103] ATLAS Collaboration, *Dijet production in $\sqrt{s} = 7$ TeV pp collisions with large rapidity gaps at the ATLAS experiment*, Phys. Lett. B **754** (2016) 214, arXiv: 1511.00502 [hep-ex] (cit. on p. 31).

[104] M. Rijssenbeek et al., *Public Forward Detector Plots for Collision Data*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ForwardDetPublicResults (cit. on pp. 32, 33, 38, 45).

[105] G. Valentino et al., *Semiautomatic beam-based LHC collimator alignment*, Phys. Rev. ST Accel. Beams **15** (2012) 051002 (cit. on p. 33).

[106] C. Zamantzas et al., "The LHC beam loss monitoring system's data contribution to other systems," *2007 IEEE Nuclear Science Symposium Conference Record*, vol. 3, 2007 2331, URL: https://ieeexplore.ieee.org/document/4436612 (cit. on p. 33).

[107] G. Valentino et al., *Final implementation, commissioning, and performance of embedded collimator beam position monitors in the Large Hadron Collider*, Phys. Rev. Accel. Beams **20** (2017) 081002 (cit. on p. 33).

[108] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **76** (2016) 292, arXiv: 1603.05598 [hep-ex] (cit. on p. 33).

[109] J. Liu, *The Alignment of the ATLAS Forward Proton Detector*, 2020, URL: https://indico.cern.ch/event/868940/contributions/3813694/ (cit. on p. 34).

[110] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, JINST **15** (2020) P04003, arXiv: 1911.04632 [physics.ins-det] (cit. on p. 37).

[111] ATLAS Collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, Eur. Phys. J. C **80** (2020) 47, arXiv: 1909.00761 [hep-ex] (cit. on pp. 37, 101).

[112] L. A. Harland-Lang, V. A. Khoze, and M. G. Ryskin, *Exclusive physics at the LHC with SuperChic 2*, Eur. Phys. J. C **76** (2016), arXiv: 1508.02718 [hep-ph] (cit. on pp. 38, 86).

[113] L. A. Harland-Lang, V. A. Khoze, and M. G. Ryskin, *Exclusive LHC physics with heavy ions: SuperChic 3*, Eur. Phys. J. C **79** (2019), arXiv: 1810.06567 [hep-ph] (cit. on p. 38).

[114] L. A. Harland-Lang, M. Tasevsky, V. A. Khoze, and M. G. Ryskin, *A new approach to modelling elastic and inelastic photon-initiated production at the LHC: SuperChic 4*, Eur. Phys. J. C **80** (2020), arXiv: 2007.12704 [hep-ph] (cit. on pp. 38, 39).

[115] L. A. Harland-Lang and M. Tasevsky, *Searching for axionlike particles with proton tagging and rapidity gaps at the LHC*, Phys. Rev. D **107** (2023) 033001, arXiv: 2208.10526 [hep-ph] (cit. on pp. 38, 77).

[116] L. Harland-Lang,
The proton in high definition: revisiting photon-initiated production in high energy collisions,
JHEP **2020** (2020), arXiv: 1910.10178 [hep-ph] (cit. on p. 38).

[117] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*,
SciPost Phys. Codebases (2022) 8, arXiv: 2203.11601 [hep-ph] (cit. on p. 39).

[118] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159,
arXiv: 1410.3012 [hep-ph] (cit. on p. 40).

[119] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244,
arXiv: 1207.1303 [hep-ph] (cit. on p. 40).

[120] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*,
ATL-PHYS-PUB-2016-017, 2016, URL: https://cds.cern.ch/record/2206965
(cit. on p. 40).

[121] D. J. Lange, *The EvtGen particle decay simulation package*,
Nucl. Instrum. Meth. A **462** (2001) 152 (cit. on p. 40).

[122] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823,
arXiv: 1005.4568 [physics.ins-det] (cit. on p. 40).

[123] GEANT4 Collaboration, S. Agostinelli, et al., *GEANT4 – a simulation toolkit*,
Nucl. Instrum. Meth. A **506** (2003) 250 (cit. on p. 40).

[124] ATLAS Collaboration,
The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim,
ATL-PHYS-PUB-2010-013, 2010, URL: https://cds.cern.ch/record/1300517
(cit. on p. 40).

[125] A. Rosenfeld and J. L. Pfaltz, *Sequential Operations in Digital Picture Processing*,
J. ACM **13** (1966) 471 (cit. on p. 41).

[126] ATLAS Collaboration,
Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2,
Eur. Phys. J. C **77** (2017) 673, arXiv: 1704.07983 [hep-ex] (cit. on p. 41).

[127] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*,
Nucl. Instrum. Meth. A **262** (1987) 444 (cit. on p. 41).

[128] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, Eur. Phys. J. C **77** (2017) 332,
arXiv: 1611.10235 [hep-ex] (cit. on p. 41).

[129] ATLAS Collaboration, *Measurement of Higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector*,
Phys. Rev. D **90** (2014) 112015, arXiv: 1408.7084 [hep-ex] (cit. on pp. 41, 91).

[130] ATLAS Collaboration, *Search for nonpointing and delayed photons in the diphoton and missing transverse momentum final state in 8 TeV pp collisions at the LHC using the ATLAS detector*,
Phys. Rev. D **90** (2014) 112005, arXiv: 1409.5542 [hep-ex] (cit. on p. 41).

[131] ATLAS Collaboration, *Search for displaced photons produced in exotic decays of the Higgs boson using 13 TeV pp collisions with the ATLAS detector*, CERN-EP-2022-096, 2022,
arXiv: 2209.01029 [hep-ex] (cit. on p. 41).

[132] ATLAS Collaboration, *Search in diphoton and dielectron final states for displaced production of Higgs or Z bosons with the ATLAS detector in $\sqrt{s}$ = 13 TeV pp collisions*, ATLAS-CONF-2022-051, 2022, URL: https://cds.cern.ch/record/2816334 (cit. on p. 41).

[133] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, JINST **14** (2019) P12006, arXiv: 1908.00005 [hep-ex] (cit. on pp. 42–44).

[134] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002, 2008, URL: https://cds.cern.ch/record/1099735 (cit. on pp. 42, 43).

[135] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: 1603.02934 [hep-ex] (cit. on p. 42).

[136] ATLAS Collaboration, *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016*, Eur. Phys. J. C **79** (2019) 205, arXiv: 1810.05087 [hep-ex] (cit. on pp. 43, 86).

[137] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **79** (2019) 639, arXiv: 1902.04655 [hep-ex] (cit. on p. 43).

[138] A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, 2007, arXiv: physics/0703039 [physics.data-an] (cit. on p. 44).

[139] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Phys. Lett. B **659** (2008) 119, arXiv: 0707.1378 [hep-ph] (cit. on p. 44).

[140] R. Staszewski and J. Chwastowski, *Transport simulation and diffractive event reconstruction at the LHC*, Nucl. Instrum. Meth. A **609** (2009) 136, arXiv: 0906.2868 [physics.ins-det] (cit. on p. 45).

[141] ATLAS Collaboration, *Search for Scalar Diphoton Resonances in the Mass Range 65–600 GeV with the ATLAS Detector in pp Collision Data at $\sqrt{s}$ = 8 TeV*, Phys. Rev. Lett. **113** (2014) 171801, arXiv: 1407.6583 [hep-ex] (cit. on p. 67).

[142] M. Oreglia, *A Study of the Reactions $\psi' \to \gamma\gamma\psi$*, SLAC-R-236, 1980 (cit. on p. 67).

[143] R. Edgar, D. Amidei, C. Grud, and K. Sekhon, *Functional Decomposition: A new method for search and limit setting*, 2018, arXiv: 1805.04536 [physics.data-an] (cit. on p. 90).

[144] ATLAS Collaboration, *Search for boosted diphoton resonances in the 10 to 70 GeV mass range using 138 fb^{-1} of 13 TeV pp collisions with the ATLAS detector*, CERN-EP-2022-183, 2022, arXiv: 2211.04172 [hep-ex] (cit. on pp. 91, 95).

[145] E. Todesco and J. Wenninger, *Large Hadron Collider momentum calibration and accuracy*, Phys. Rev. Accel. Beams **20** (2017) 081003 (cit. on p. 99).

[146] ATLAS Collaboration, *Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s}$ = 13 TeV with the ATLAS Detector at the LHC*, Phys. Rev. Lett. **117** (2016) 182002, arXiv: 1606.02625 [hep-ex] (cit. on p. 101).

[147] S. Bailey and L. A. Harland-Lang, *Modeling W^+W^- production with rapidity gaps at the LHC*, Phys. Rev. D **105** (2022) 093010, arXiv: 2201.08403 [hep-ph] (cit. on p. 107).

[148] L. A. Harland-Lang, V. A. Khoze, and M. G. Ryskin,
Elastic photon-initiated production at the LHC: the role of hadron-hadron interactions,
SciPost Phys. **11** (2021) 064, arXiv: 2104.13392 [hep-ph] (cit. on p. 107).

[149] L. A. Harland-Lang, V. A. Khoze, and M. G. Ryskin,
The photon PDF in events with rapidity gaps, Eur. Phys. J. C **76** (2016) 255,
arXiv: 1601.03772 [hep-ph] (cit. on p. 107).

[150] K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke,
HistFactory: A tool for creating statistical models for use with RooFit and RooStats,
CERN-OPEN-2012-016, 2012, URL: https://cds.cern.ch/record/1456844
(cit. on p. 110).

[151] G. Cowan, K. Cranmer, E. Gross, and O. Vitells,
Asymptotic formulae for likelihood-based tests of new physics, Eur. Phys. J. C **71** (2011) 1554,
arXiv: 1007.1727 [physics.data-an] (cit. on pp. 115, 143).

[152] E. Gross and O. Vitells, *Trial factors for the look elsewhere effect in high energy physics*,
Eur. Phys. J. C **70** (2010) 525, arXiv: 1005.1891 [physics.data-an] (cit. on p. 116).

[153] O. Vitells and E. Gross, *Estimating the significance of a signal in a multi-dimensional search*,
Astropart. Phys. **35** (2011) 230, arXiv: 1105.4355 [astro-ph.IM] (cit. on p. 116).

[154] ATLAS Collaboration, *Search for pairs of highly collimated photon-jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **99** (2019) 012008,
arXiv: 1808.10515 [hep-ex] (cit. on p. 118).

[155] T. Junk, *Confidence level computation for combining searches with small statistics*,
Nucl. Instrum. Meth. A **434** (1999) 435, arXiv: hep-ex/9902006 (cit. on p. 119).

[156] A. L. Read, *Presentation of search results: the CL_S technique*, J. Phys. G **28** (2002) 2693
(cit. on p. 119).

[157] S. Fartoukh et al., *LHC Configuration and Operational Scenario for Run 3*,
CERN-ACC-2021-0007, 2021, URL: https://cds.cern.ch/record/2790409
(cit. on p. 124).

[158] G. Tateno, *LHC-ATLAS Phase-I upgrade: calibration and simulation of a new trigger readout system for the Liquid Argon calorimeter*, JINST **15** (2020) C05060 (cit. on p. 124).

[159] G. Tateno, *Calibration and performance evaluation of the liquid argon calorimeter trigger readout in the Phase-I upgrade of the LHC-ATLAS experiment*,
MA thesis: The University of Tokyo, 2020,
URL: https://www.icepp.s.u-tokyo.ac.jp/research/thesis.html (cit. on p. 124).

[160] G. Aad et al.,
The Phase-I trigger readout electronics upgrade of the ATLAS Liquid Argon calorimeters,
JINST **17** (2022) P05024 (cit. on p. 124).

[161] T. Chobola,
Study of light-by-light scattering with the ATLAS Forward Proton (AFP) Detector at CERN,
CERN-THESIS-2020-058, 2020, URL: https://hdl.handle.net/10467/88719
(cit. on p. 124).

[162] L. Beresford and J. Liu,
Search Strategy for Sleptons and Dark Matter Using the LHC as a Photon Collider,
Phys. Rev. Lett. **123** (2019) 141801, arXiv: 1811.06465 [hep-ph] (cit. on p. 125).

[163] L. A. Harland-Lang, V. A. Khoze, M. G. Ryskin, and M. Tasevsky, *LHC searches for Dark Matter in compressed mass scenarios: challenges in the forward proton mode*, JHEP **04** (2019) 010, arXiv: 1812.04886 [hep-ph] (cit. on p. 125).

[164] ATLAS Collaboration, *Observation of photon-induced W^+W^- production in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector*, Phys. Lett. B **816** (2021) 136190, arXiv: 2010.04019 [hep-ex] (cit. on p. 125).

[165] A. Armbruster, *Discovery of a Higgs Boson with the ATLAS Detector*, PhD thesis: Michigan U., 2013, URL: https://cds.cern.ch/record/1553771 (cit. on p. 143).