Search for low mass dark photon radiated from a muon in Proton-Proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector

By

YAO YAO

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Physics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

S. Mani Tripathi, Chair

Michael Mulhearn

Maxwell Chertok

Committee in Charge

2023

Abstract

A search for a low mass dark photon below 1 MeV which is radiated from a muon using proton-proton collision data at a center-of-mass energy $\sqrt{s} = 13$ TeV is presented. The data was collected with the CERN CMS detector in 2018 corresponding to an integrated luminosity of 59.8 fb^{-1}. Such a low mass dark photon has no available decay channel to standard model particles, and therefore is stable. The dark photon is assumed to directly interact with detector materials through pair-production, and its small kinetic mixing leads to it depositing energy outside the CMS Electromagnetic Calorimeter. The search is performed by looking for events whose final state is composed of two muons and a photon-like shower in the CMS Hadronic Calorimeter, with the 3-body mass consistent with the Z boson. The accessible range of kinetic mixing in this search is between 0.032 to 0.32. No significant deviation from estimated backgrounds is observed with the 2018 data. The dark photon with kinetic mixing between 0.063 to 0.32 and mass lower than 1 MeV is excluded at 95% confidence level.

To my parents,

who support my dream of doing scientific research, and who encourage me to take on challenges and make my own decisions in my career path.

And to Reyer, my dear boyfriend,

who has always supported me since we met in 2019 and celebrates with me for every single achievement in graduate school.

Contents

List of Figures

List of Tables

Acknowledgments

I started graduate school back in 2016, and it has been a long journey. I went through the first year's general graduate physics classes, the second year's specific physics classes related to particles and detectors. Then I moved to CERN for two and half years and got involved in the CMS detector's operations and upgrades. I moved back to Davis in December 2020 after the first wave of COVID-19. Since September 2019, I focused on the dark photon search and it has taken me more than three years to accomplish. Finally, in May 2023 my thesis has been endorsed by the CMS Exotic physics analysis group.

Over the past 7 years I met a lot of kind people in Davis, at CERN, and at Fermilab. They've helped me, both mentally and physically, on-work and off-work. It is with their support that I am able to reach the end of the graduate school. There are a million thanks that I want to say.

I want to start with people that support me to achieve this degree. I want to first thank Prof. Mani Tripathi, who brings up this interesting research idea, that a dark photon may shower in the calorimeters. Mani is very supportive in my research, replying to my questions on the research promptly and meet with me whenever I feel frustrated. He also encourages me to present my results at conferences. Mani always attends the meetings when I report the analysis progress to the CMS exotic analysis group and helps pushing the reviewing progress of my analysis. Those meetings are usually 5 am in the morning on Mondays for California, so they are never easy for me or Mani.

Together with Mani, I want to thank all the Davis CMS group professors for being supportive in research opportunities. Prof. Robin Erbacher helped me to connect to the CSC group experts in the CMS experiment got me involved in the CMS operations and CSC Phase II upgrades. She visits students and postdocs located at CERN frequently and asks us if we need any help locally. She also contacted Fermilab cmslpc physicists and I got to spend a summer at Fermilab last year. During my second year when I am not sure what to work on, I worked in Prof. Maxwell Chertok's lab for measuring thermal conductivity of aluminum carbon fibers for the CMS tracker, and it was a fun experience with Max as well as his other students.

I also want to thank the postdocs and students in the Davis CMS group. Dr. Devin Taylor, though has left this field in late 2020, gave me great help in starting this research project, understanding the backgrounds of this search and basic analysis strategies. He also supported me with his computing knowledge. Dr. Frank Jensen works on an excited tau physics search, and he is always willing to discuss with me whenever I had trouble understanding the CMS physics object related tutorials and also about the datacards and fittings results from the Higgs combine tool. Dr. Fengwangdong Zhang and Dr. Johan Bonilla are based at CERN and they were very hospitable during my stay there. I mostly overlapped with Davis CMS students Dr. Christine McLean, Christopher Brainerd, Dr. Zhangqier Wang, Dr. Grace Haza, Dr. Brendan Regnery, Dr. Troy Welton at CERN, and Wei Wei at Fermilab. We had productive time when we shared the office, and fun time off-work. All the current postdocs and students gave me crucial advice when I was looking

for a postdoc position.

I want to thank all the experts, postdocs and students in the CSC group in the CMS experiment. Dr. Armando Lanaro, the CSC project manager, supervised me during the CSC electronics replacement for the Phase II upgrade. I had a great time working with CSC technicians and other students. He also supervised me on a CSC trigger primitive study project and help me get connected to the CSC database experts. After the CSC trigger primitive study, I collaborated with Dr. Katerina Kuznetsova and student Osvaldo Colin in developing an on-line trigger primitive quality monitoring system, but unfortunately due to my research schedule, I left the development in late 2020. When I was in the CSC team, I gained great support from senior graduate students there, as well as the postdocs who are very good at communicating with students and managing the CSC lab.

I want to thank Dr. David Yu, who was the previous jet+X analysis group convener and who gave me many good suggestions and brilliant ideas on this analysis. He pointed out that the hadronic calorimeters' absolute energy scale is calibrated by pions and how it explains the 1.4 ratio of dark photon jet's energy to generated dark photon's energy in the hadronic calorimeters. He also pointed out the HCAL electronic noise, which was a large background in this search.

Aside from research, I want to thank all the people that I met locally in Davis and at CERN. My roommates were always willing to listen to me when I feel frustrated and drove me to shopping, hiking, etc. We celebrated every moment that worth celebrating, and I wish all the best for their future careers. I want to thank all the physics students in my year,

especially Dr. Rose Baunach and Dr. Yuxi Zhang, for discussing the class work together and getting through all the tough homework. I also want to thank the local Davis people who hosted or attended Intercultural Mosaics, which helped me improve my English over the weekends in my first two years and gave me a chance to get in touch with local community.

At the end, I want to thank UC Davis during COVID, for it requested frequent COVID tests and reacted fast in remote teaching.

Chapter 1

The Dark Matter Problem

The existence of unobserved mass in the universe has been revealed by several independent cosmological studies including the rotation curves of galaxies, gravitational lensing, hot gas in clusters, and cosmic microwave background. Dark matter is used as a label to represent whatever particles that may account for the missing mass in the universe. It is "dark" because it is not luminous in cosmological observations that measure light spectra of celestial objects. Dark matter is currently observed through its gravitational effects, and no other interactions has been so far detected between dark matter and standard model particles. Therefore, a great variety of dark matter candidates particles have been proposed in theory, including weakly interacting massive particles (WIMPs) such as heavy neutrinos and stable super-symmetric particles, axions, and dark photons. A massive dark photon, which is the topic of the study in this dissertation, is a dark matter candidate within a certain mass range.

Some evidences of dark matter from cosmological observations are briefly introduced in

Chapter 1.1. Some theoretical ideas of dark matter candidates and corresponding experiments are briefly introduced in Chapter 1.3. Dark photon and its related experiments are introduced in Chapter 1.4.

1.1 Evidence for existence of dark matter

The earliest evidence of dark matter in the universe comes from the observation of the speed of the luminous matter of the galaxies through Doppler shifts, that luminous matter in a galaxy move faster than the prediction based on Newtonian gravity given the galaxy's visible mass. The discrepancy of a galaxy's observed and predicted rotation curve implies a potential answer that the actual mass of a galaxy is much higher than its observed mass. An example is the rotation curve measurement of the nearby galaxy M33 [1] in Figure 1.1. The observed rotation curve is fitted over three parameters: the amplitudes of halo, stellar disc and gas's contribution to the curve.

Another compelling evidence comes from gravitational lensing. Gravitational lensing is a phenomenon in which light bends around a massive celestial object due to the distortion in local space-time caused by the celestial object's strong gravitational field. It was predicted by Albert Einstein's theory of general relativity, and was first observed in 1979, that two quasars with similar distances and spectra were found to be very close to each other [2]. The two quasars in the image are proven to be a single quasar, and its light distortion is caused by a cluster of galaxies located between the telescope and the quasar.

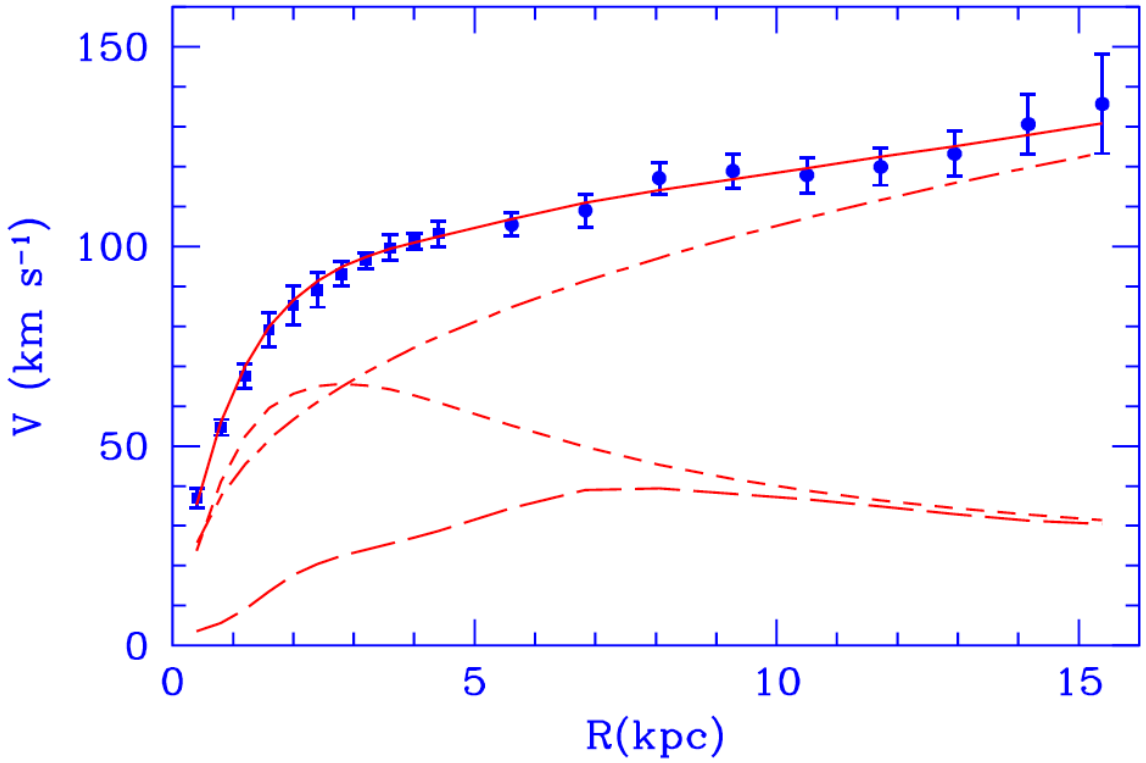


Figure 1.1: The galaxy M33's rotation curve (points) with its best-fitted model (continuous line). Halo contribution is illustrated as dot-dashed lines. The rotation curve of stellar disc is illustrated as short-dashed lines. Gas contribution is illustrated as long-dashed lines. [1].

Since the mass of the celestial object that the light passes by is correlated to the light's bending angle, gravitational lensing becomes an effective method to measure the actual mass of a massive celestial object. Recently, the Hubble telescope has observed more gravitational lensing phenomena in the universe. An example is the gravitational lensing phenomenon of the massive galaxy cluster Cl 0024+17 (ZwCl 0024+1652) in Figure 1.2, and its distribution of dark matter is calculated by comparing the mass predicted by gravitational lensing with the measured luminous mass [3].

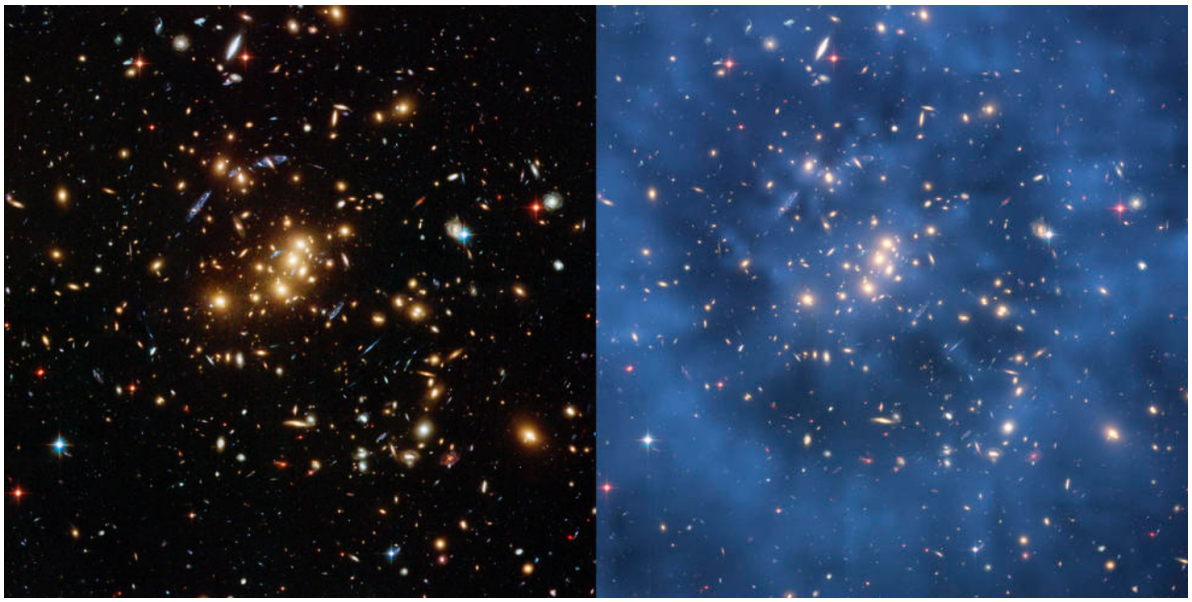


Figure 1.2: Gravitational lensing of the massive galaxy cluster Cl 0024+17 (ZwCl 0024+1652). (Left) The blue arcs are from the effect of gravitational lensing. They are actually from the galaxies located behind the cluster. (Right) The blue shading is a calculation result and it indicates the location and density of dark matter[3].Credits: NASA, ESA, M.J. Jee and H. Ford (Johns Hopkins University)

Studies of the cosmic microwave background (CMB) not only prove the existence of dark matter, but also calculate the amount of dark matter compared to luminous matter in the universe. CMB consists of radiation from the early days of the universe, which is red shifted to the microwave range due to the expansion of the universe. Its amplitude is even in all directions (isotropic) with extremely small density fluctuations. The amplitude of density perturbations of the CMB is determined by the matter-dominated era when all particles in the universe formed a hot soup and moved at non-relativistic speed. It was thus demonstrated that the hot soup had a total mass that was larger than the mass of luminous matter in the universe. Without the exceed mass, the first galaxy's formation would be much

slower than what we observed. The calculation of dark matter density with the measurement of on cosmological scales of anisotropy in the CMB can be found in [4] by WMAP and and [5] by Planck collaborations.

1.2 The Standard Model

To explain the dark matter, many theoretical particle models are proposed. Before going though those dark matter candidates, I want to briefly introduce the so-far best-adapted particle model – the Standard Model (SM), and the elementary particles that have been discovered. The SM is a theory that describes strong, weak and electromagnetic interactions between elementary particles. The theory was finalized in mid-1970s and it successfully predicted the existence of the Higgs boson and also accurately defined the ratio between the masses of the W and Z bosons, which along with the photon, are the force carriers f the electroweak theory. The table 1.3 lists all the elementary particles that have been discovered in the experiments thus far. However, the SM provides no description of gravitational interactions, and it also can not accommodate massive neutrinos, while neutrinos have been measured to have mass in light of their flavor oscillations.

The SM describes the interactions between elementary particles via exchanged mediators. For example, in an electron elastic scattering where an electron A hits an electron B, the interaction is described as a virtual photon being exchanged between A and B and modifying their momentum. In the SM, strong interactions happen by exchanging gluons,

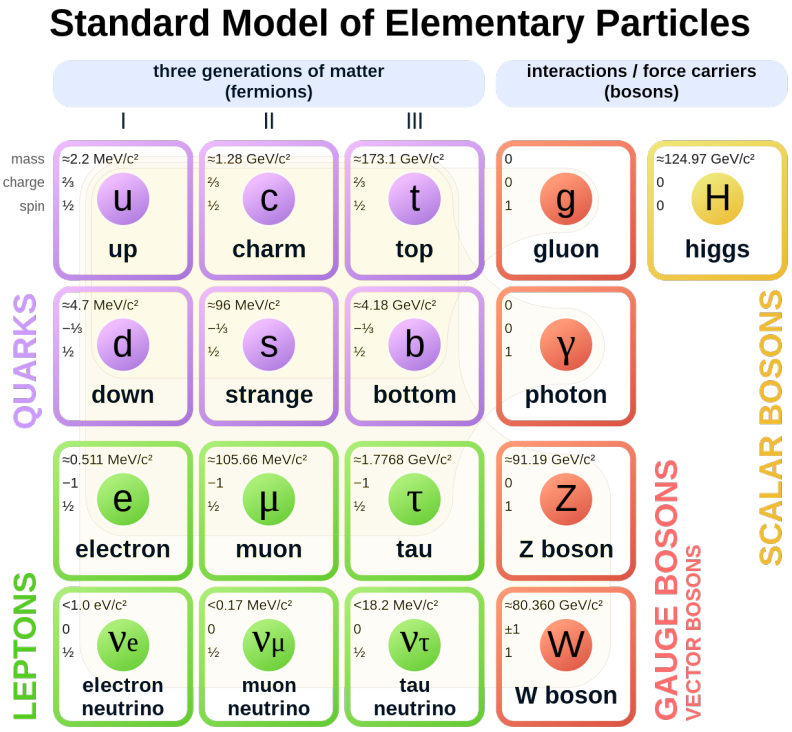


Figure 1.3: Elementary particles that have been discovered up to now. The particles are categorized as quarks, leptons, gauge bosons and the Higgs boson which is the only scalar boson.

weak interactions happen by exchanging W or Z bosons, and electromagnetic interactions happen by exchanging photons.

The elementary particles in the SM can be categorized as quarks, leptons, gauge bosons, and the Higgs boson. The quarks and leptons are fermions which have non-integer spin. The gauge bosons have non-zero integer spin, while the Higgs boson has spin 0. There are three generations of quarks and leptons. The higher generation particles have the same properties as their lower generation, except that they are more massive. The spin, mass and charge of the particles are listed in the Table 1.3.

Quarks have color charge, weak charge and electromagnetic charge, which means that they can experience strong, week and electromagnetic interactions. In each generation, there is a quark that carries +2/3 electromagnetic charge, and a quark that carries -1/3 electromagnetic charge. A single quark in principle cannot exist alone, and they tend to combine with other quarks to form hadrons. If a quark and anti-quark combine, they form a meson. If three quarks combine, they form a baryon. Among the quarks, the top quark is the heaviest quark and it decays through weak interactions to lighter quarks. Leptons all have weak charge. Electron, muon and tau have -1 electromagnetic charge, while neutrinos are electromagnetically neutral.

Gauge bosons can be categorized by the type of interactions that they are involved. The photon is the mediator of electromagnetic interactions. The W and Z bosons are the mediators of weak interactions. The W boson is responsible for charged current flavor change, while the Z boson participates in neutral current interactions. There are 8 type of gluons in the SM because the gauge group of quantum chromodynamics is SU(3). The 8 types of gluons are not distinguishable in experiments because they lose their color properties during the process of hadronization. Both gluons and photons are mass-less, while the W and Z bosons' masses have been measured to be about 80.4 GeV and 91.2 GeV respectively.

The Higgs boson is the only scalar boson in the SM. It couples to massive particles, and the strength of the coupling is proportional to the square of the mass of the particle. While the SM has been highly successful in explaining nearly all the measurements made in particle physics, it does not have a dark matter candidate in its particle multiplet.

1.3 Dark matter candidates

There are many theoretical models that include dark matter as a component. One of the most popular proposals is that dark matter is likely made of non-baryonic weakly interacting massive particles (WIMPs). The CMB perturbations suggest that the mass of WIMPs as dark matter candidates should not be too light (less than 10 GeV) [6]. There is a further restriction on the dark matter's self-annihilation cross section given the observed dark matter density, which is also discussed in [6]. WIMP is a generic name for theoretical particles that were thermally produced in the early Universe and constitute cold dark matter.

Heavy neutrinos can contribute to dark matter and are a type of WIMPs. It is proposed that additional heavy neutrinos (above 1 keV) kinetically mix with the standard model neutrinos but have no electroweak interactions [7]. Detection approaches include neutrino-electron scattering[8], low temperature bolometers[9], and atomic excitation measurements.[10].

Supersymmetry purposes that each fermion(boson) has a super-partner boson(fermion) with the same quantum numbers, but different masses. A supersymmetric extension of the standard model predicts a few neutral super-partners named neutralinos. Heavy neutralinos can decay to lighter neutralinos along with some standard model particles while the lightest neutralino does not have decay channels in theory and becomes a potential dark matter candidate. Direct production searches, such as chargino-neutralino production with the chargino decaying to a W boson and the lightest SUSY particle (LSP), while the neutralino

decaying to a Higgs boson and the LSP have been searched for at the LHC [11]. Indirect searches, via observing excess gamma-ray production in astrophysical sources have been performed by the Fermi Large Area Telescope (FermilLAT) [12]. Other indirect searches include measurements of the muon g-2 experiment [13].

An axion is a theoretical Nambu-Goldstone boson that comes from introducing a new global U(1) symmetry that is spontaneously broken to the standard model. The new U(1) symmetry potentially solves the strong-CP problem in Quantum chromodynamics(QCD). The quantity closely related to the phase of the QCD vacuum (the angle θ) needs to be extremely small[14]. Experimentally, an axion's mass range is restricted to be lower than 10^{-3} eV due to the exclusion of its induced rare meson decays as well as the limits from Supernova observations [15]. This implies that an axion is not as massive as a WIMP, and therefore, the way that it accounts for dark matter is different from WIMPs. A theory that θ was at a larger value in the early days of the Universe could lead to the production of a non-thermal axion from Peccei-Quinn field and it makes an axion still a viable dark matter candidate[14]. The axion dark matter experiment (ADMX G2) is the first and currently the only running experimental effort for detecting axions[16]. It tries to detect axions by converting them into microwave photons under a strong magnetic field with extraordinarily sensitive microwave receivers.

1.4 The Proposed Dark Photon

A dark photon is a hypothetical U(1) gauge boson, which has been proposed as a mediator between Standard Model (SM) particles and the dark sector. It is allowed to have a small mass, however, there will be kinetic mixing between a dark broken Abelian gauge symmetry and the SM hypercharge [17]. In a U(1) theory, the kinetic term in the Lagrangian is given by $L = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu}$, where $F_{\mu\nu}$ is the electromagnetic field tensor. After kinetically mixing with a dark photon, the relevant kinetic terms [17] are

$$L \subset -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} - \frac{1}{4}Z_D^{\mu\nu}Z_{D\mu\nu} + \frac{1}{2}\frac{\epsilon}{cos\theta_W}Z_{D\mu\nu}F^{\mu\nu} + \frac{1}{2}m_D^2 Z_D^{\mu}Z_{D\mu} \tag{1.1}$$

where $Z_D^{\mu\nu}$ is the dark photon field tensor, θ_W is the Weinberg mixing angle, and ϵ is the kinetic mixing coefficient. The third term in Equation 1.1 represents the kinetic mixing, and the fourth is a mass term.

A dark photon has similar properties to a SM photon. A dark photon's mass m_D and kinetic mixing ϵ are unknown parameters. Existing constraints on dark photons contributed by experimental, cosmological, and astrophysical experiments are presented in Figure 1.4 for the dark photon lighter than 0.1 MeV. High energy experiments usually search for a dark photon decays or escape the detector. Therefore, the constraints and prospects for a dark photon are placed where a dark photon has mass higher than several GeVs, and kinetic mixing is higher than 10^{-6}. The constraints and prospects are presented in Figure1.5. The dashed lines in these two plots represent the expected limits from future experiments if a

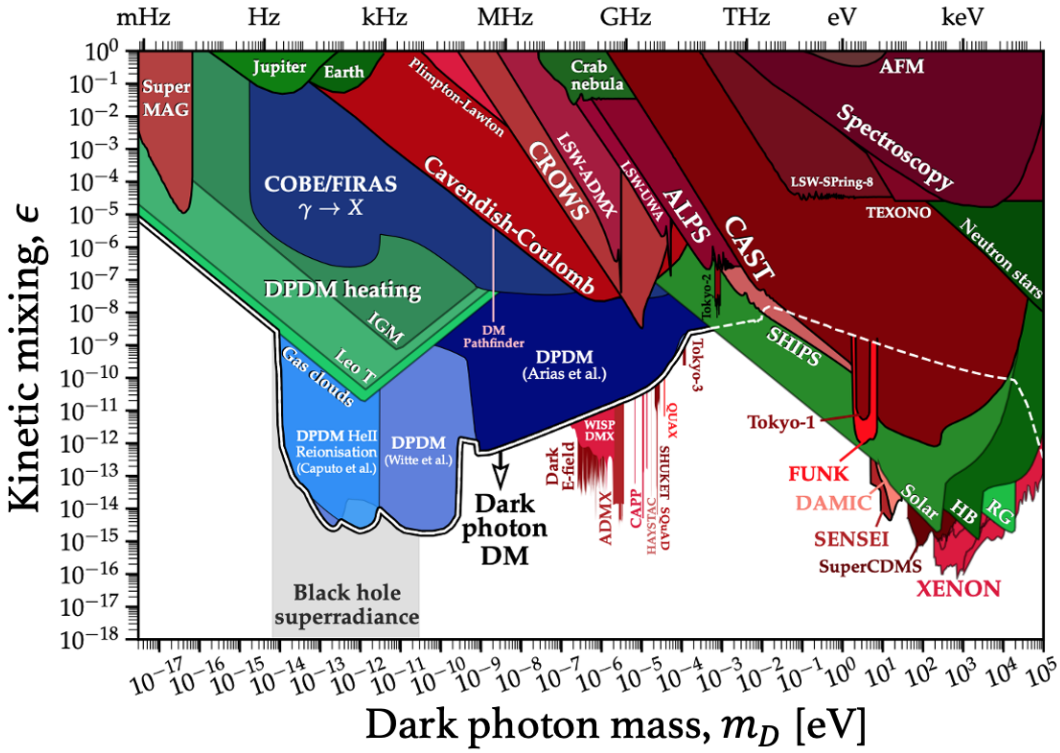


Figure 1.4: Current constraints on the dark photon's mass, m_D and kinetic mixing parameter with the SM photon, ϵ, in the region m_D less than 0.1 MeV[18]. The blue bounds are contributed by cosmological observations. The red bounds are contributed by experiments. The green bounds are contributed by astrophysical detection. For the details of each experiment, please refer to [18].

dark photon is not found in that experiment. Probing to low kinetic mixing in high energy physics requires a big amount of data. For the CMS experiment, the lowest mass limit on long-lived particles, which can be a dark photon potentially, is so far at 0.6 GeV[19]. This search looks for a Higgs boson that decays to a pair of long-lived particles, and the each long-lived particle decays into a pair of muons. Therefore, the limits on the dark photon's kinetic mixing and mass are correlated with the branching ratio of the Higgs boson to a pair of dark photon.

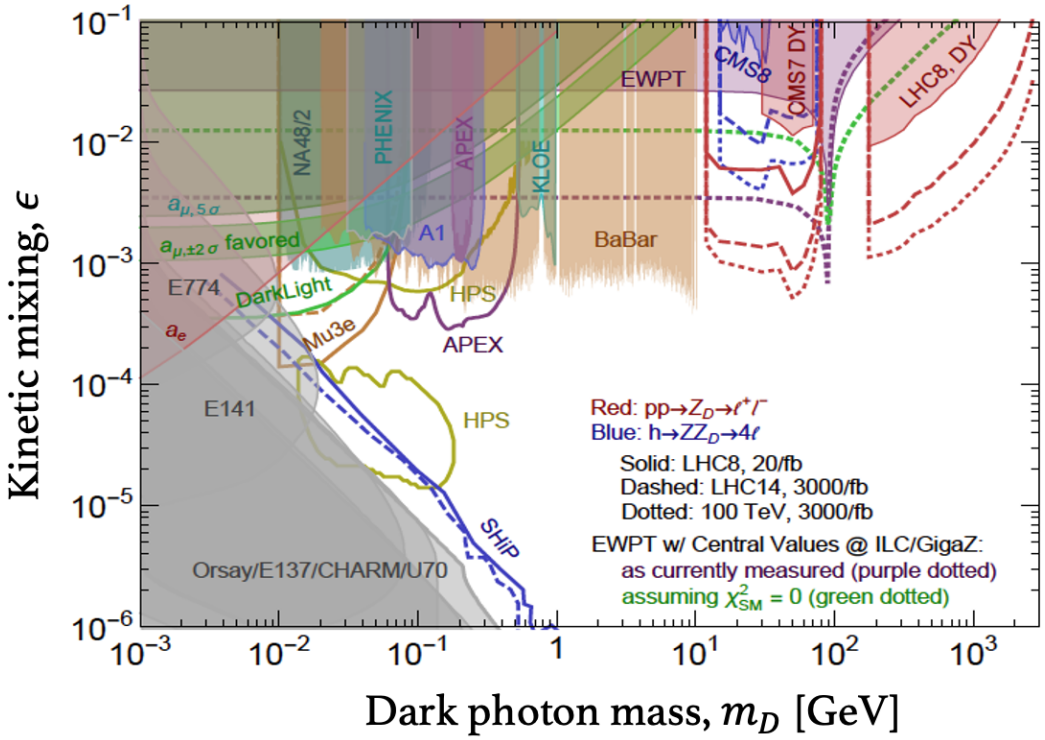


Figure 1.5: Summary of dark photon constraints and prospects with the dark photon mass greater than 1 MeV [17]. The region where m_D is higher than 10 GeV is uniquely contributed by high energy collisions experiments, while the region with lower mass is contributed by precision QED observables and searches at $B-$ and Φ factories, beam dump experiments, and fixed target-experiments. The dashed lines are limits from perspective experiments.

This work searches for dark photons that emerge from the high energy proton-proton (pp) collisions generated by the Large Hadron Collider (LHC) at CERN. The collisions are recorded by the CMS detector. The details of the LHC and the CMS detector are introduced in Section 2. We search for a dark photon that is generated together with two muons from the pp collisions. The dark photon, within a certain range of kinetic mixing, initiates a shower in the calorimeters (see Section 2.2.2 2.2.3) of the CMS detector through pair production

of an electron and a positron. The dark photon's interaction with detecting materials and generation of an electron-positron pair is shown in Figure 1.6.

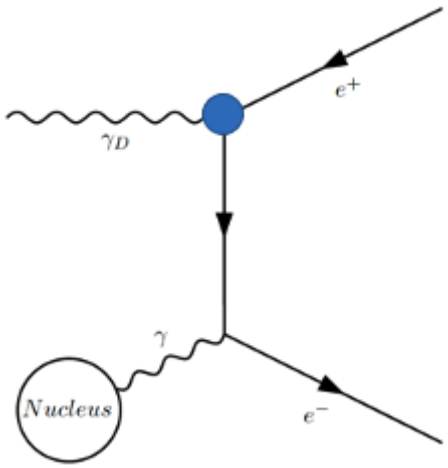


Figure 1.6: A high energy dark photon interacts with detector materials' nucleus and generates a pair of electron and positron. The electron and positron radiate SM photons when they pass through the detector materials. This process is called bremsstrahlung. A emitted photon again generates a pair of electrons. The bremsstrahlung and pair production take place alternatively until photons fall below the pair production threshold, and electrons loss energy mainly through other processes like ionization and elastic scattering.

The dark photon pair production cross section is defined in Equation 1.2. We define a dark factor f_D $(= \epsilon^2)$ to describe the ratio between the cross section of a dark photon pair production to the cross section of a SM photon pair production. α is the fine-structure constant that quantifies the strength of the electromagnetic interaction between elementary charged particles, r_e is the classic electron radius, and Z is atomic number. $P(E, Z)$ is some complex-valued function that depends on the energy and atomic number.

$$\sigma_{\gamma_D \to ee} = f_D \sigma_{\gamma_{SM} \to ee} = \epsilon^2 \alpha r_e^2 Z^2 P(E, Z) \tag{1.2}$$

For this search, the range of the kinetic mixing parameter to be probed, as well as the reasons behind looking for dark photon showers in hadronic calorimeters (see Section 2.2.3), are directly related to the CMS detector's structure. Therefore, they are discussed after Section 2.

Chapter 2

The CMS Detector at the Large Hadron Collider

The Large Hadron Collider(LHC) is a 27 km circumference circular accelerator that is designed to provide proton-proton (pp) collisions up to 14 TeV center-of-mass energy [20]. The LHC is located underground at the Swiss-French boarded near Geneva, and is currently the most powerful particle accelerator under operation. During its operation from 2016 to 2018 (Run 2), it produced pp collisions with a center-of-mass energy at 13 TeV, and delivered in total about 160 fb^{-1} integrated luminosity[21]. The LHC has started Run 3 data taking since 2022 with a center-of-mass energy of 13.6 TeV.

The Compact Muon Solenoid (CMS) detector consists of a complex set of sub-detectors that measure and reconstruct different types of particles that emerge from the collisions. The LHC has four pp beam crossing points. The CMS detector surrounds one of the high

luminosity collision points, which is located 100 meters underground near Cessy in France. During Run 2, the CMS detector successfully collected about 147 fb^{-1} of collision data. Within that data set, about 138 fb^{-1} is validated as having good quality for general physics analyses[21].

In this chapter, I will introduce some basic concepts of collider physics that are relevant to the high luminosity pp collisions. I will also introduce the CMS detector components that are related to the dark photon search, including the muon chambers, the Hadron Calorimeters (HCAL), the Electromagnetic Calorimeters (ECAL), the trackers, event triggers, as well as the particle reconstruction method known as the Particle-Flow algorithm.

2.1 The CERN accelerator complex

The European Organization for Nuclear Research (CERN) was established in 1954. The high energy proton beams that are used for collisions in the LHC are produced, accelerated and stored by the CERN accelerator complex, as shown in Figure 2.1. The process of creating collisions starts at the proton source. Before 2020, the proton source was a bottle of hydrogen gas located at one end of the Linear Accelerator 2 (Linac2). The electron in a hydrogen atom is stripped off by electric field, and the remaining proton is delivered to the Linac2[22]. The Linac2 accelerates the protons to 50 MeV and delivers them to the Proton Synchrotron Booster(PSB). The PSB in turn accelerates the protons to 1.4 GeV and injects them into the Proton Synchrotron (PS)[23]. The PS further accelerates the protons

to 25 GeV and delivers them to the Super Proton Synchrotron (SPS)[24]. The SPS is the second largest accelerator in the CERN accelerator complex. It has been in operation since 1976 with a center-of-mass energy at 450 GeV. In 1983, the UA1 and UA2 collaborations discovered the W and Z bosons [25] [26] [27] [28] with the $SP\bar{P}S$ when it was configured to be a proton-antiproton collider. Now the SPS [29] serves as the last accelerator before the proton beams are injected into the LHC.

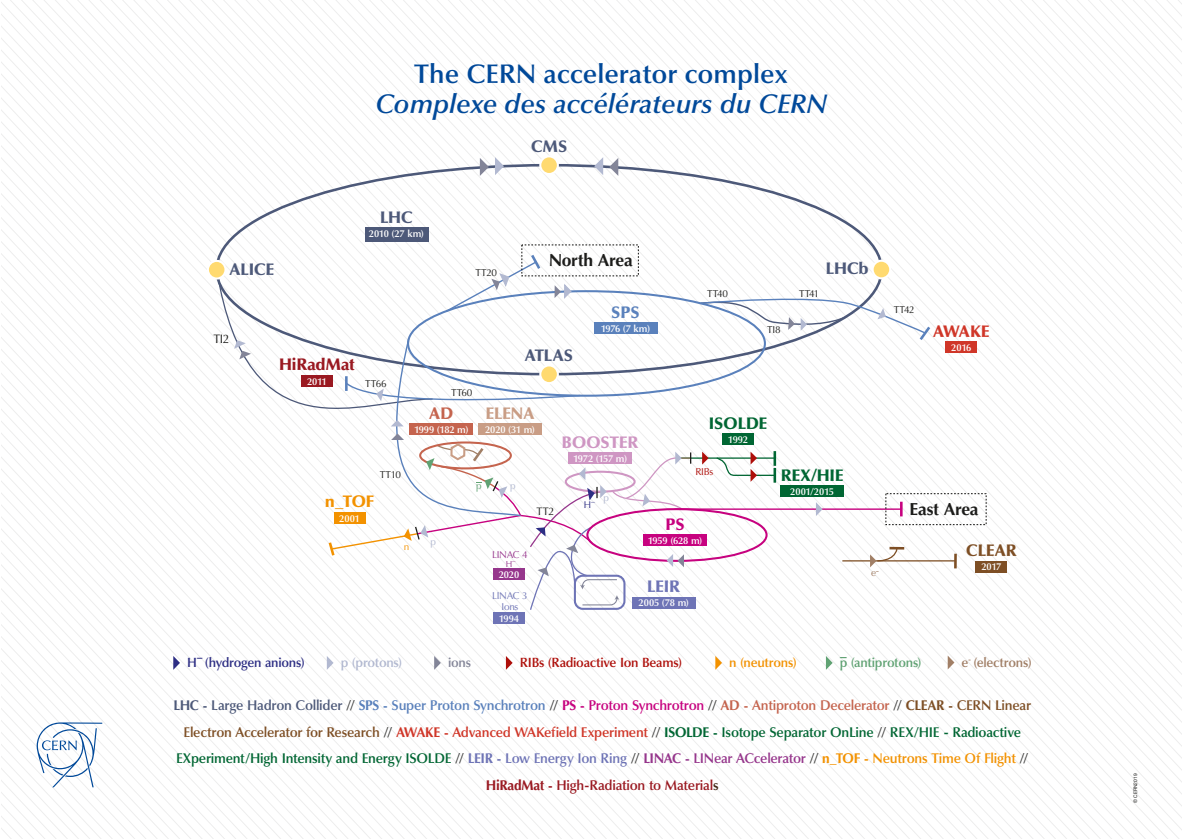


Figure 2.1: The CERN accelerator complex layout [30].

The center-of-mass energy refers to the total energy of the proton-proton system in the center-of-mass frame. In the pp collisions in Run 2, each proton beam is accelerated to

6.5 TeV, and they generate collisions at a center-of-mass energy 13 TeV. The Mandelstam variables s, t, u represent the four-momentum squared q^2 of the exchanged boson between two incoming particles in three different categories. The s-channel is indicated in the Feynman diagram in Figure 2.2 [31]. $\sqrt{s}$ in the center-of-mass frame equals to the sum of the energy of two incoming particles.

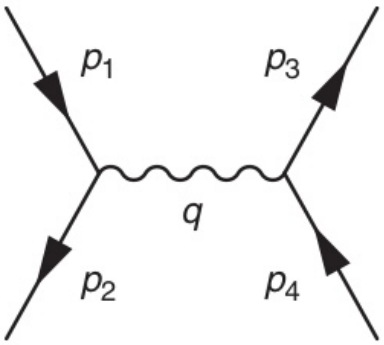


Figure 2.2: The Mandelstam variable s-channel [31].

The luminosity of LHC directly indicates the number of pp collisions that took place in the LHC with a linear relationship as shown in Equation 2.1.

$$N_{event} = L_{total}\sigma \tag{2.1}$$

where σ is the total cross section of the pp collisions, including elastic, diffractive, and inelastic processes. The instantaneous luminosity (L) of the machine depends on the beam parameters [20]:

$$L = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi\epsilon_n \beta^*} F \tag{2.2}$$

where N_b is the number of particles per bunch, n_b is the the number of bunches per beam. The revolution frequency is denoted by f_{rev}, the relativistic gamma factor by γ_r. The instantaneous luminosity is inversely proportional to the normalized transverse beam emittance ϵ_n and the beta function at the collision point β^*. Finally, F is a geometric reduction factor which occurs due to the fact that the beams do not collide head-on, but rather at a small crossing angle at the interaction point(IP).

The LHC accelerates the protons to 6.5 TeV with an acceleration system composed of 400 MHz Radiofrequency (RF) cavities. It accommodates 1232 superconducting dipoles that bend the trajectory of the proton beams with a 8.3 Tesla magnetic field. It also accommodates 474 superconducting quadrupoles to focus the beams[32].

The LHC stores the proton beams and maintains them at 6.5 TeV during a physics run. The bunches in the proton beams are spaced 25 ns apart considering the timing resolution of the detectors for recording each collision as well as allowing for maximum number of bunches in a beam to reach high luminosity. The peak beam luminosity can reach about $10^{34}\text{cm}^{-1}\text{s}^{-1}$ where the CMS detector is located. The luminosity decays during a physics run mainly due to the collisions reducing the bunch intensity. The LHC can store the proton beams for 24 to 48 hours for a stable physics run. When the luminosity drops below a certain threshold, the protons are transported to an external absorber located far away, referred to as a beam dump. The beam injection and beam dump locations in the LHC are shown in Figure 2.3.

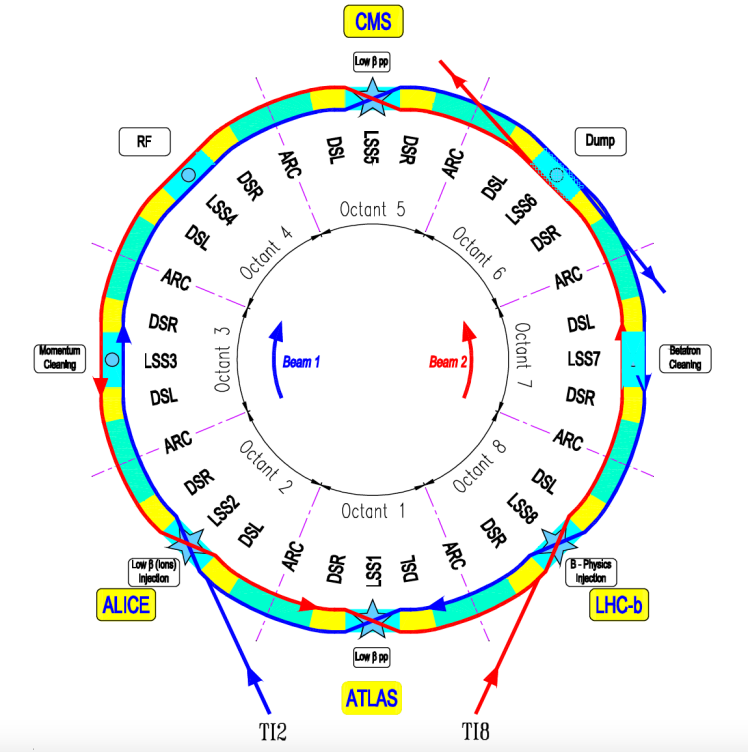


Figure 2.3: Schematic layout of the LHC ring[20].

2.2 The CMS detector

The Compact Muon Solenoid (CMS) detector is a general-purpose detector that is designed to measure various types of physics processes occurring in high luminosity pp collisions. It is cylindrical in shape, with a diameter of about 15 meters, length of about 21 meters, and weights about 14,000 tonnes[33]. The CMS detector envelopes the IP with detecting materials as compact as possible, and leaves almost no gap for a particle to go through without depositing energy in some sub-detector. It also has a powerful solenoid that provides a 3.8 Tesla magnetic field inside of the solenoid to bend the trajectory of charged particles

that are products of the collisions.

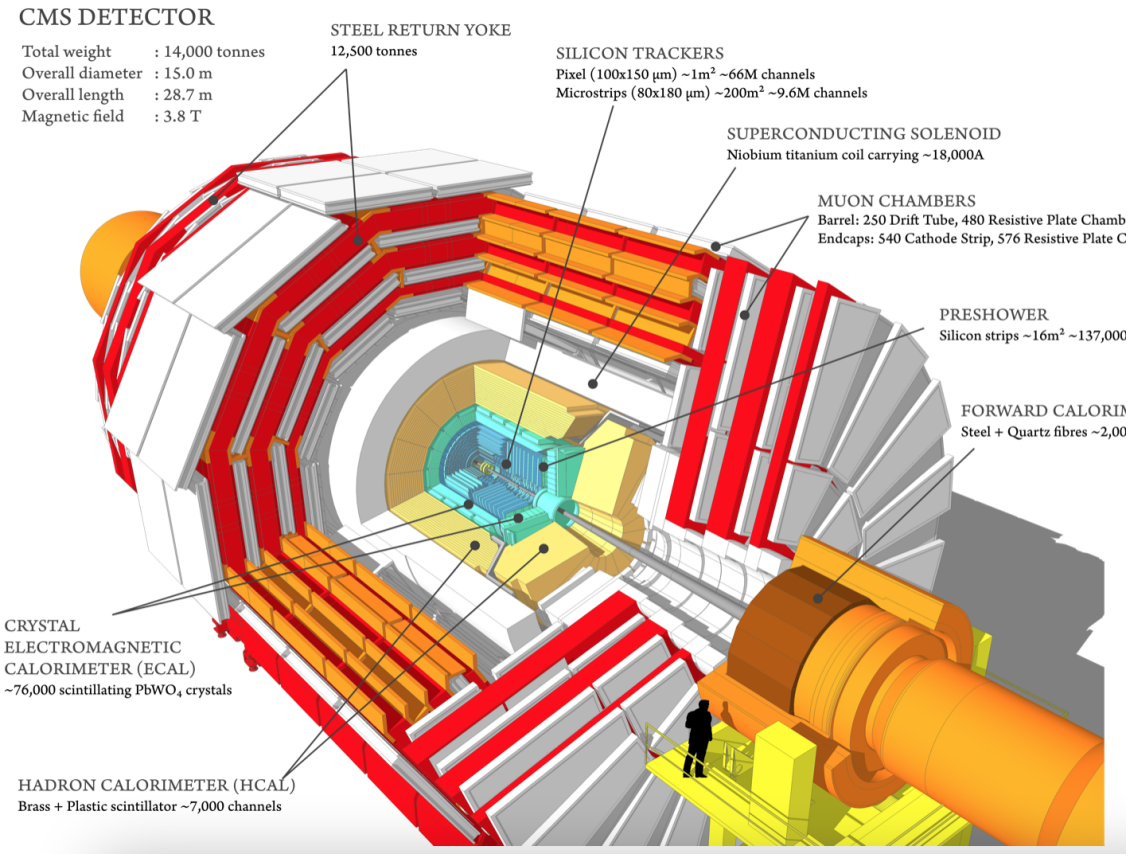


Figure 2.4: Schematic layout of the CMS detector[33].

The detector is best described by moving radially outward from the IP. The first detectors encountered by particles are made of silicon, deployed in the form of pixellated layers followed by stripline ones. These detectors measure charged tracks with high precision. The silicon tracker is composed of about 124 millions pixels, each measuring 100 *150 μm [34] and 10 millions micro-strips[35], which inner-strip distance varies from 80 μm to 205 μm and strip length is 4 times the inner-strip distance for all types from the inner-most layer to the outer-

most layer[36]. In the barrel region, the electromagnetic calorimeter (ECAL) is located out of the silicon trackers to initiate and measure particles' electromagnetic showers. In the endcap region, a fine-grained photon-sensitive (pre-shower) detector is installed in front of the endcap ECAL. It initiates photon showers and provides a higher position resolution for forward photon reconstructions. The hadronic calorimeter (HCAL) is located out of the ECAL, aiming to initiate and measure hadronic showers from the hadrons that penetrate through the ECAL. The superconducting solenoid is located around the barrel HCAL. The muon detectors are located outside of the solenoid. These are instrumented with several different types of sensors: 1) Drift Tubes (DT) for the barrel region, 2) Cathode Strip Chambers (CSC) for the endcap region, and 3) Resistive Plate Chambers (RPC) for the overlapping region between the endcap and barrel and the gaps between muon chambers. In Run 3, Gas Electron Multipliers (GEM) have been installed in the endcap in front of the CSCs to increase the muon resolution in the forward region. In the CMS detector, the solenoid produces a 3.8 T magnetic field inside of the solenoid and 1.9 T opposite direction magnetic field out of the solenoid. Therefore, steel return yokes, interspersed with the muon chambers, guide the magnetic field back to the solenoid. A brief demonstration of particle tracks and energy deposit is shown in Figure 2.5. A muon from the collisions usually penetrates through all the sub-detectors, and leaves a track in the trackers and the muon detectors due to its charge. On its trajectory, a muon may radiate some low energy photons and maybe detected in the calorimeters. A hadron from the collisions usually penetrates through the trackers and the ECAL and deposits most of its energy in the HCAL. It leaves a track in the trackers

if a hadron is charged. An electron or photon from the collisions usually deposits its energy in the ECAL, while the electron leaves a track in the trackers. The details of the particle reconstruction using the sub-detector information will be mentioned in Section 2.2.6.

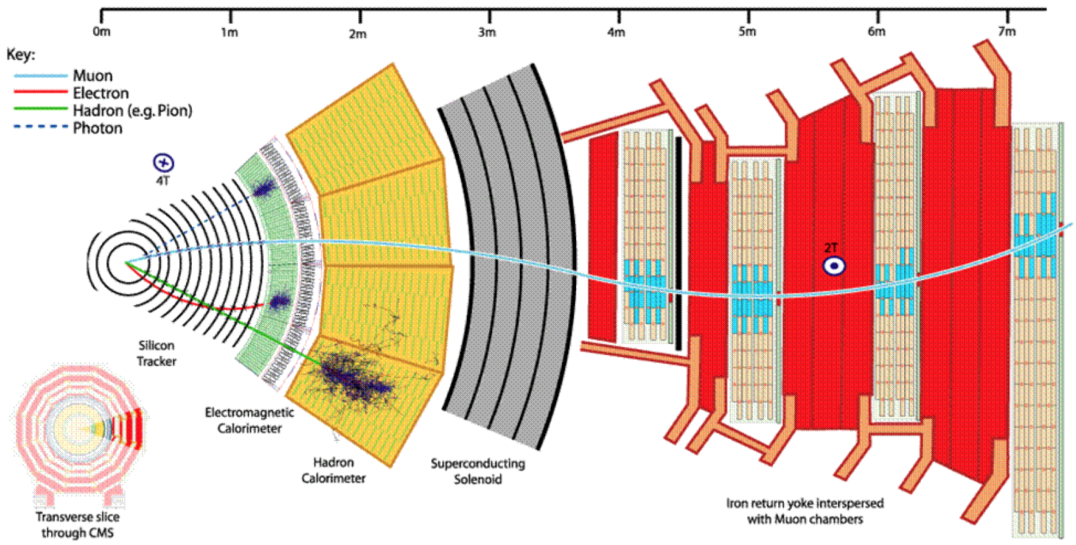


Figure 2.5: Particle reconstruction in the CMS [37]. Particles are identified as muons, charged hadrons, neutral hadrons, electrons and photons based on their different sub-detector signatures.

2.2.1 Trackers

The CMS tracking detector uses silicon sensors. A reverse bias voltage is applied to the silicon sensors during operation. When a charged particle passes through a silicon sensor, it creates electron-hole pairs from the silicon lattice along its way. The ejected electrons are then collected by the sensor and form a small electric signal.

Trackers are composed of pixel detectors as the innermost tracker and silicon strip detec-

tors as the outer tracker. The pixel detector has a finer segmentation than the silicon strip detector. Therefore it has a better spatial resolution while having a higher read-out load due to a massive channel count. The layouts of the pixel detector and the silicon strip detector are shown in Figure 2.6[38] and Figure 2.7[39].

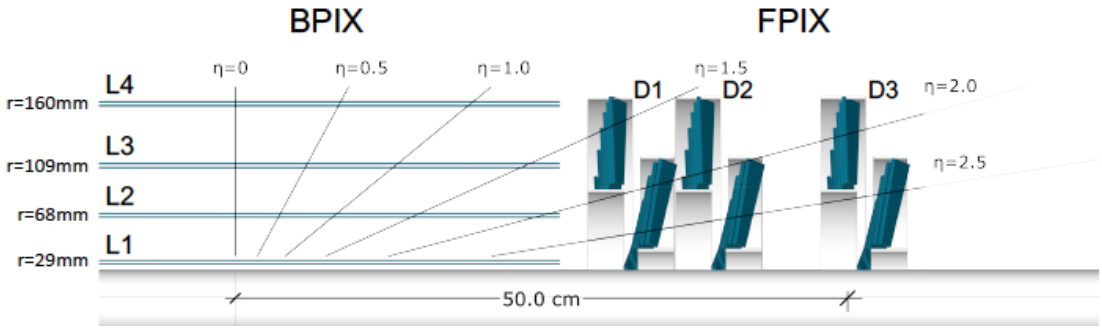


Figure 2.6: The layout of the phase-1 pixel detector[38].

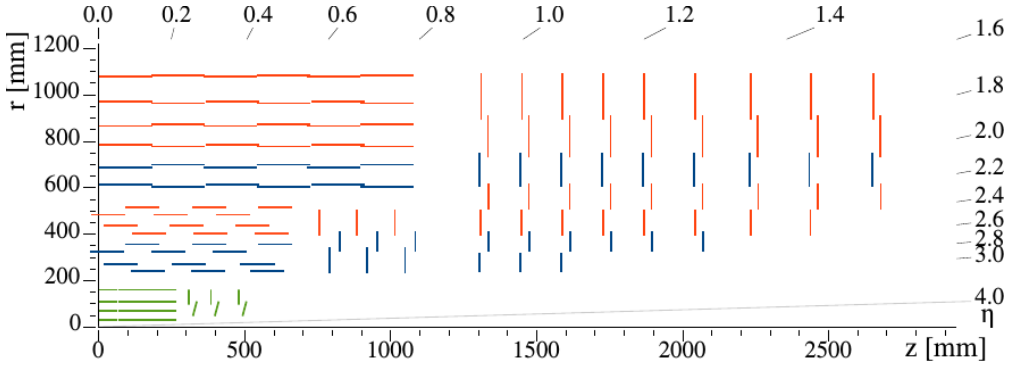


Figure 2.7: The layout of the phase-1 tracker. The green lines are the pixel detector. The blue lines are the double-sided strip modules, and the red lines are the single-sided strip modules[39].

The phase-1 pixel detector is located closest to the beam pipe. The barrel pixel (BPIX) has 4 barrel layers located at a radii of 29, 68, 109 and 160 mm from beam pipe. The forward

pixel (FPIX) has 3 disks in each endcap, located at 291, 396, 516 mm from the center of the detector and with radial coverage from 45 to 161 mm. In the FPIX, the outer rings are rotated by 20 degree and the inner rings are tilted by 12 degree respect to the beam line. The design optimizes the spatial resolution of a track under the condition that charged particles are bent by the magnetic field. The pixel detector has 4-hit coverage up to $|\eta|$ at 2.5. The pixel detector has in total 124 million readout channels (pixels), with each pixel size 100 * 150 μm[38].

The silicon sensors of the pixel detector use N-in-N approach, as shown in Figure 2.8[40]. Due to the high radiation environment, the N-type in a standard PN diode becomes progressively less n-type under continuous hadronic radiation, and it eventually inverts to effective p-type during operation. With the N-in-N approach, after type inversion, the bulk silicon depletes from the strip and therefore, signals can still be collected.

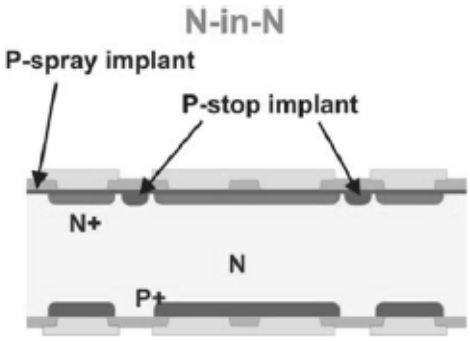


Figure 2.8: The N-in-N approach silicon sensor layout[40].

The silicon strip detector reaches out to a radius of 110 centimetres. It is composed of double-sided modules (blue in Figure 2.7) that have two layers of strips mounted back to

back, and single-sided modules (red in Figure 2.7). It has a total of about 10 million detector strips and provides a 6-hit coverage up to $|\eta|$ at 2.4[41].

2.2.2 Electromagnetic calorimeter

The Electromagnetic Calorimeter (ECAL) is located outside of the trackers and it uses over 75,000 scintillating lead tungstate crystals ($PbWO_4$) to measure the energy deposit of electrons and photons passing through the detector[42]. The ECAL barrel reads out the scintillation through silicon avalanche photo-diodes, while the endcap uses vacuum photo-triodes. There is an ECAL pre-shower silicon strip detector located in front of the endcap crystals. It can help better distinguish neutral pions from photons in forward region.

The ECAL crystals cover up to $|\eta|$ at 3.0, while energy of photons and electrons can be precisely measured when $|\eta| < 2.6$. The ECAL barrel and endcap split at $|\eta| = 1.48$. The size of each crystal front face is about 22 * 22 mm, and it matches the radius of a electromagnetic shower cone (Molière radius) in $PbWO_4$. It corresponds to $\Delta\eta$ * $\Delta\phi$ = 0.0175 * 0.0175 in barrel and $\Delta\eta$ * $\Delta\phi$ = 0.05 * 0.05 in endcap[42]. The ECAL contains 26 radiation lengths at $\eta = 0$. The expected energy resolution (in %) of the $PbWO_4$ calorimeter with respect to the energy is shown in Figure 2.9[42]. The energy resolution percentage is expected to reduce and become stable after about 200 GeV.

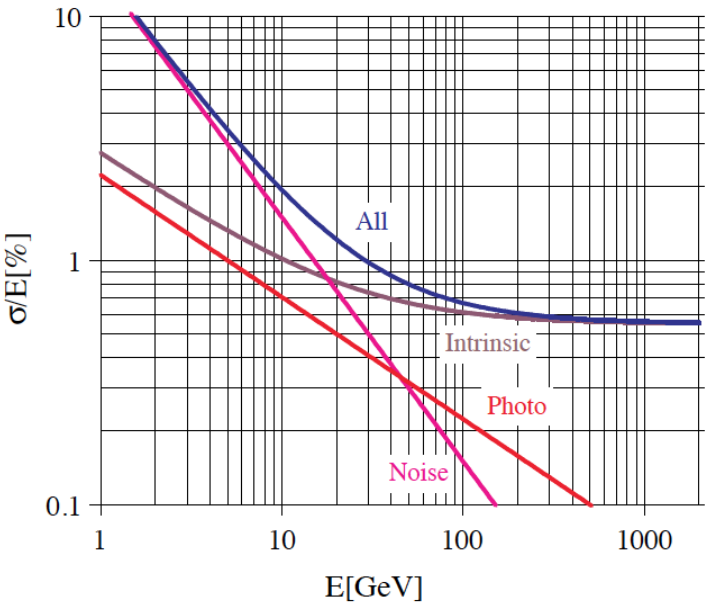


Figure 2.9: The expected energy resolution of PbWO4 crystal calorimeter with respect to photon/electron energy. The overall $PbWO_4$ resolution is contributed by 1)the noise part that contains the contributions from electronic noise and pileup energy, 2)the intrinsic part that includes the shower containment and a constant term of 0.55%[42].

2.2.3 Hadronic calorimeter

The geometry of the hadronic calorimeter in the CMS detector is restricted by the size of the ECAL inside of the HCAL as well as the diameter of the solenoid. With a limited diameter, the HCAL needs to maximize its number of interaction length to contain the hadronic showers. Therefore, the HCAL is designed as a sampling calorimeter with absorber plates (about 5 cm thick in barrel and 8 cm thick in endcaps) and 4 mm thick plastic scintillator tiles[43]. Since the HCAL is housed within the magnet, the absorber plates are made of brass and stainless steel that are non-magnetic materials. The plastic scintillator tiles are connected to green double-clad wavelength-shifting fibers that collect the scintillation light,

re-emit at a longer wavelength, and guide this secondary light to hybrid photo-diodes (HPD). An HPD amplifies the gain of photons and convert them into electric signals. An HPD tolerates high magnetic field and has a gain of $2 * 10^3$[43].

The HCAL has 4 sections, the HCAL barrel (HB), the HCAL endcap (HE), the HCAL outer (HO) and the HCAL forward (HF). The HB covers up to $|\eta <|$ 1.3 and the HE covers up to $|\eta| < 3.0$. They are the two sections that are relevant to the dark photon search, and I will introduce more about their hardware design. The HO is located out of the solenoid due to the HB depth in the solenoid is not enough to contain the hadronic showers completely. The HF is located 6 meters downstream of the HE endcaps and is designed for forward physics studies.

The HB has a layer of 7 cm thick stainless steel absorber in the innermost and outermost to provide structural support, and it has fourteen brass absorbers. The brass absorbers' density is 8.83g/cm^3. It has radiation length $X_0 = 1.49$ cm and nuclear interactive length 16.42 cm. The absorbers provide a total of 5.82 λ_I at $\eta = 0$ and 10.6 λ_I at $|\eta| = 1.3$[43].

The scintillators are divided into segments that each covers $(\Delta\eta, \Delta\phi) = (0.087, 0.087)$ in the HB[43]. Each η division corresponds to an HB tower, as shown in Fig 2.10. The HCAL's absolute energy scale is calibrated with 50 GeV pions. The energy resolution from simulation of the HB is about $100\%/\sqrt{E} + 5\%$ for hadrons. Figure 2.11 shows the expected energy resolution of high energy pions, where the inverse square-root behavior and the constant 5% contributions are visible [44].

The HPDs are segmented and each device can provide optical-to-electrical signal con-

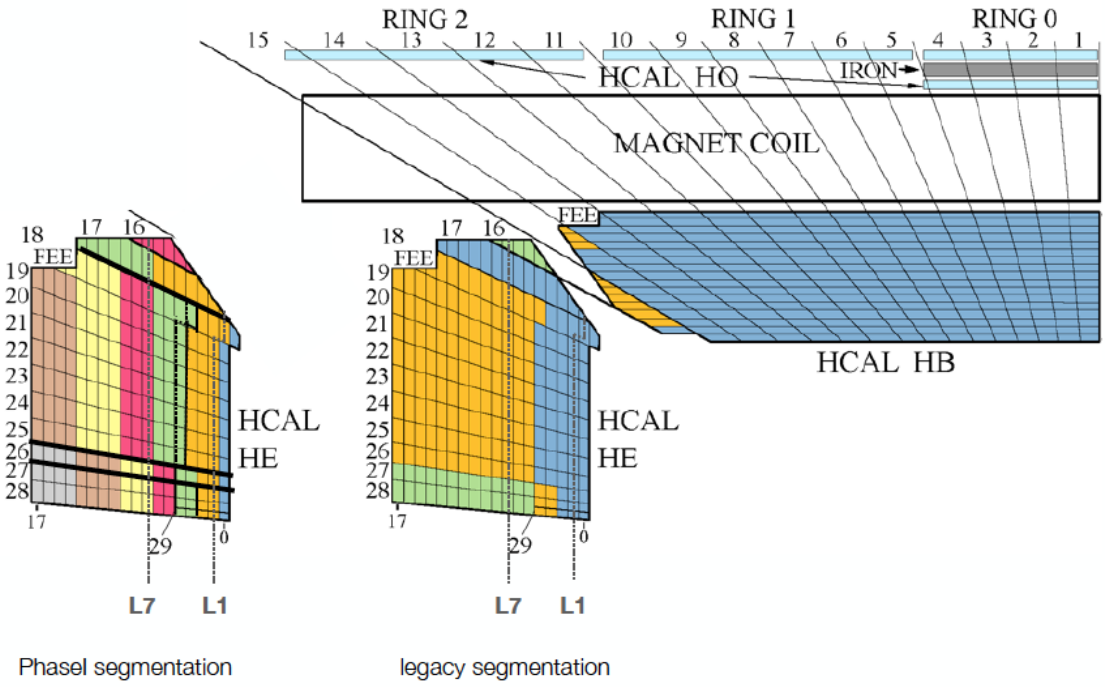


Figure 2.10: Segmentation of the HB, HE and HO detectors before(left top) and after(right bottom) the SiPM installation . With HPDs, the HE had 2 to 3 depth layers. With SiPMs, the HE has 6 to 7 depth layers[45].

versions for 18 scintillator segments. Each read-out box (RBX) collects information from four HPDs. A typical hadronic shower deposits most of its energy in two or three layers of the HCAL. During the *pp* collision operation, HPDs showed occasional electrical discharges, which were enhanced by the strong magnetic field from the solenoid at certain orientations. This problem in the HPDs is a source of significant high-amplitude noise, which affected this search for dark photons. This will be discussed later.

HPDs were the only viable solution when the HCAL was designed. In late 2000s, Silicon Photo Multipliers (SiPM) technique emerged and has quickly become a better choice for the

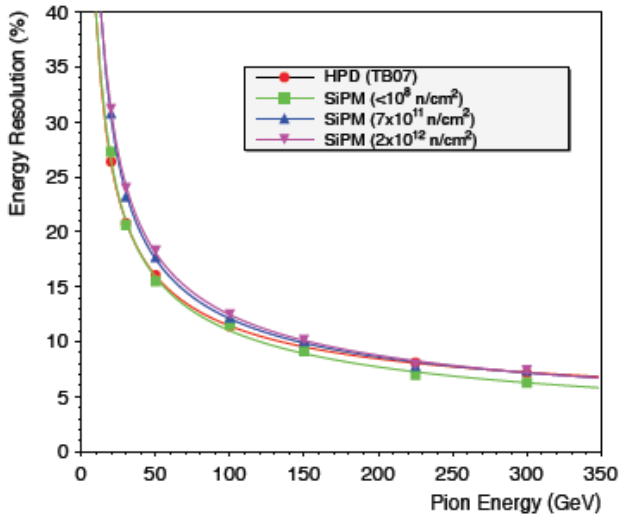


Figure 2.11: The HCAL energy resolution of a pion from simulation with respect to the pion's energy.

optical-to-electrical conversion. SiPMs are a multi-pixel Geiger-mode Avalanche Photodiode device, and are an ideal replacement for HPDs. They provide a 10^4 to 10^6 gain and do not need high operating voltage[44]. Their high signal-to-noise performance makes it feasible to increase longitudinal segmentation (along R) in the HCAL towers, which improves the Particle flow (PF) reconstruction performance of the hadrons. The longitudinal segmentation helps the PF reconstruction in multiple aspects, including identifying the energy of pileup particles that mostly deposit their energy in the first layer of the HCAL, and distinguishing prompt muon from the muons from hadron decay in the HCAL. During the first long stop (LS1) between 2013 and 2015, HO replaced all its HPDs with SiPMs, and in early 2018, HE did the same. The HB SiPMs have been installed in 2019 during LS2[44]. The segmentation in the HE and HB is compared before and after the SiPMs replaced the HPDs in Figure

2.10.

2.2.4 Muon chambers

DT, CSC and RPC are the three types of muon chambers that were actively involved in the muon measurement in Run 2. The muon chambers are multi-layer gas-ionization detectors. When a muon passes through a chamber, it ionizes the gas and electrons and gas ions are attracted respectively to anodes and cathodes under the electric field provided by applied high voltage. The CMS detector uses three types of muon chambers due to differences in average muon fluxes and distributions of muon momenta in different locations. The location of muon chambers is shown in Figure 2.12. DTs mainly cover barrel region up to $|\eta| < 1.2$, while CSCs cover endcap region where $0.9< |\eta| < 2.4$.

The DTs are composed of long aluminum drift cells, as indicated in Figure 2.13. When a muon pass through a DT cell and ionize a gas molecule, the drift time is measured to the anode wire in the center of the cell, and the location of a muon can be calculated. The spatial resolution for a single DT cell is 250 μm[46]. A DT chamber is made of 3 super-layers (SL) that each comprises 4 staggered layers of parallel cells. Two of the SLs have their anode wires parallel to beam pipe (R-ϕ) and precisely measure a muon's ϕ. One SL has anode wires perpendicular to R and z, and precisely measures a muon's z. A muon's R is determined by the location of the layer. A muon's spatial resolution in DTs can reach 100 μm per 8-layer chamber[46]. There are four stations of DT chambers, with the inner 3

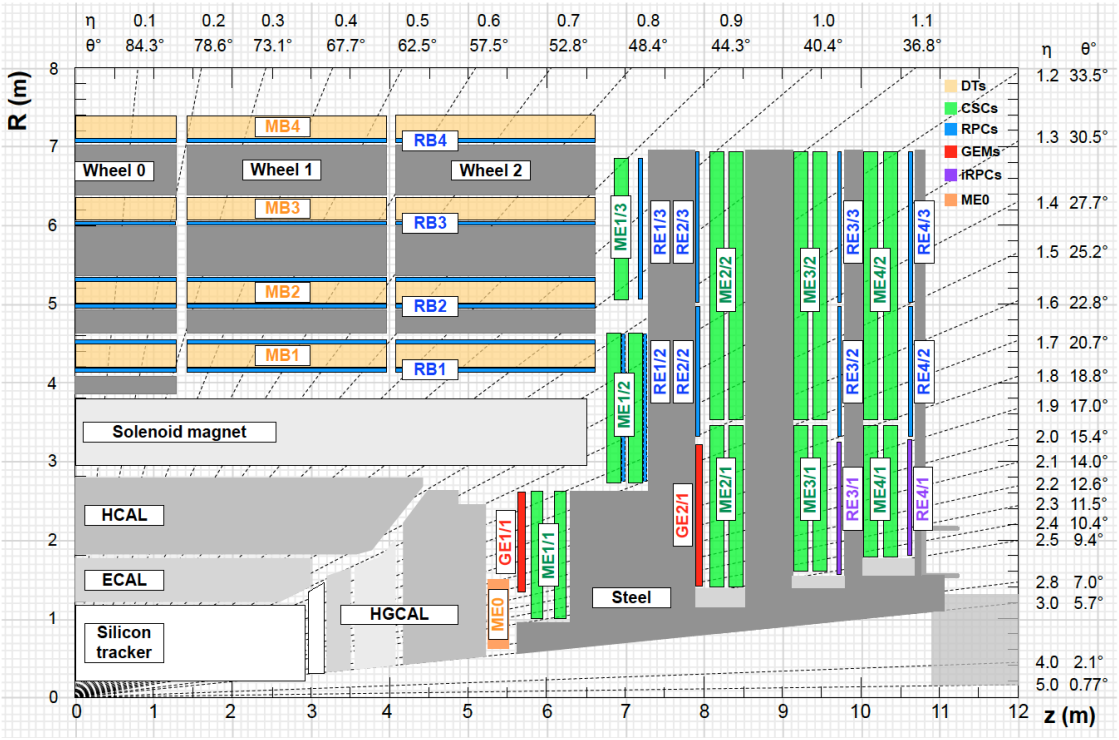


Figure 2.12: A demonstration of location of muon chambers in an R-z plane, including the Phase-2 upgrades (RE3/1, RE4/1, GE1/1, GE2/1, ME0) that were not involved in Run 2[46].

stations consist of 60 DT chambers each, and the outer station consists of 70 DT chambers. Along the z direction, there are 5 wheels of DT chambers covering η from -1.2 to 1.2. In total, DT has about 195,000 sensitive wires[47].

Each CSC is composed of 6 layers of anodes and cathodes crossing each other. Cathodes are finely segmented and placed radially and provide a measurement of muon ϕ. Anodes are evenly spaced with 3.12 mm in each layer and provide a measurement of muon position in the bending plane. The spatial resolution of a muon measured by a CSC chamber ranges between 50 and 140 μm. The time resolution is 3 ns per chamber[46]. The finely

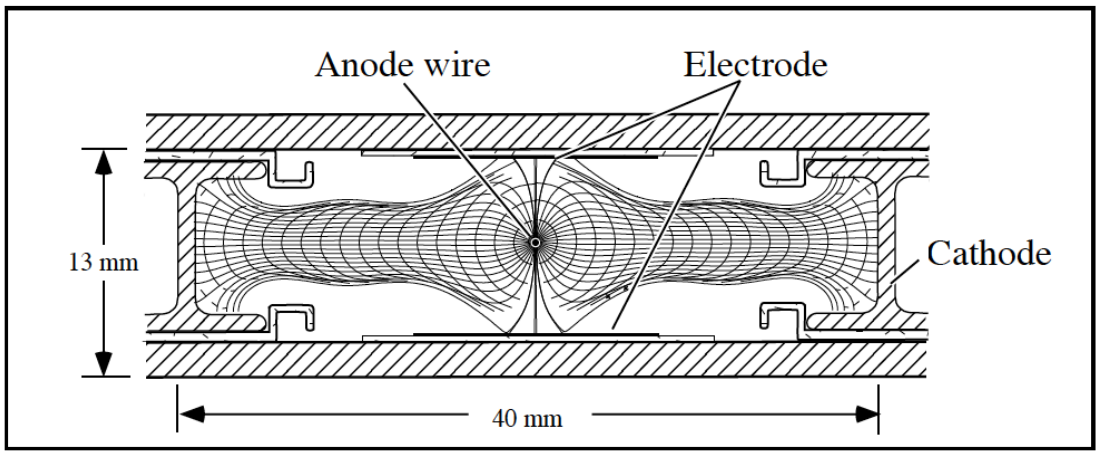


Figure 2.13: Transverse view of a DT cell[47].

segmented cathodes and anodes suppress the performance deterioration due to the large and non-uniform magnetic field in the endcap affecting the drift time of electrons and gas ions.

There are 4 disks of CSCs on each endcap. Among the 4 disks, the disk that is closest to the IP has 3 rings (ME1/1, ME1/2, ME1/3), while the other 3 disks each has 2 rings. The chambers in the rings that are closer to the beam pipe receive much more hits than the others. This is the reason why the CMS experiment plans to install the GEM detector in front of the CSCs (ME1/1) that are closest to the IP, to increase the muon track spatial resolution in the forward region in the high-luminosity LHC (HL-LHC). For Run 2, CSCs cover $|\eta|$ from 0.9 to 2.4 for both end-caps. CSCs in total have 210,816 anode channels and 273,024 cathode channels[46].

RPCs are double-gap gas chambers, which have a fast response and provide a timing resolution of 1.5 ns for a passing muon[46]. RPCs provide fast tracking information to the muon trigger system, including identifying candidate muon tracks, the correct bunch crossing

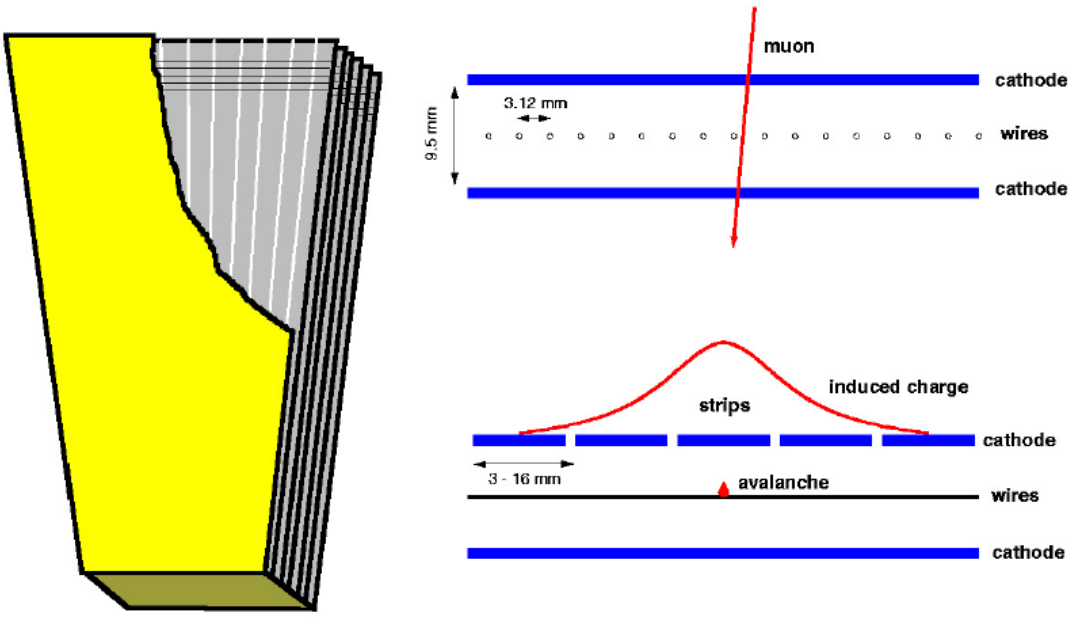


Figure 2.14: Left: A CSC is composed of 6 layers of anodes (horizontal lines) and cathodes (radially outward)[46]. Right: A demonstration of avalanche on the wire and induced charge on the strips when a high energy muon passes through a layer of CSC[46].

that a candidate muon track belongs to, as well as an estimation of muon momentum. RPC chambers are installed in both barrel and endcaps, inserted in front of and behind DTs and CSCs. During Run 2, the RPCs cover up to $|\eta| < 1.9$. More RPCs are planned to be installed in the forward region for the HL-LHC.

2.2.5 Trigger and Data Acquisition System

The data acquisition system in the CMS experiment is responsible for converting the detected signals collected by the sub-detectors into data used for physics studies. It begins by digitizing signals from the sub-detectors and establishing a data pathway. It reduces the data

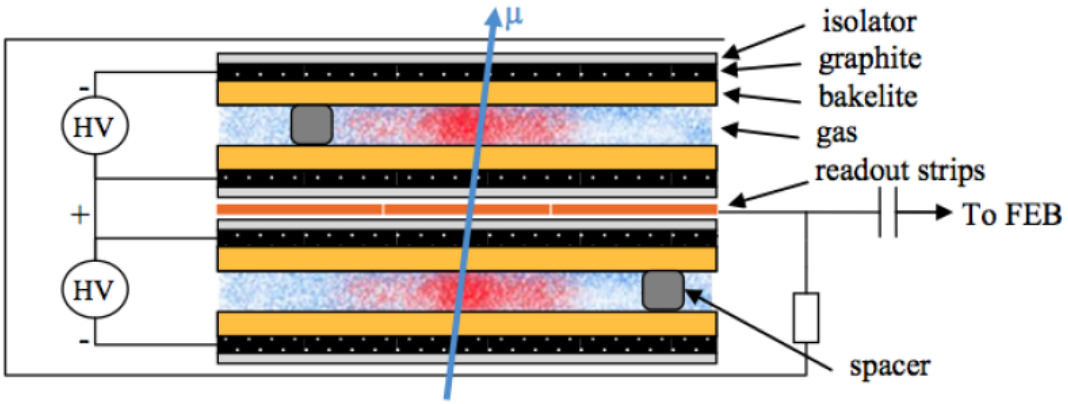


Figure 2.15: A schematic representation of the layers of an RPC detector[46].

to a manageable amount for recording and storage through event selections with the trigger system. It delivers the data to local storage at the experiment site (P5) and subsequently transfers it to Tier-0 (located at the CERN site) for offline storage.

The CMS experiment uses a two-level trigger to record proton-proton collision events that are of interest to particular searches. The Level-1 trigger(L1T) is hardware based. It pre-selects events with an output rate up to 100 kHz using fast and partial information from DT, CSC, RPC, HB, HE, HF and ECAL. The High level trigger(HLT) is software based, and it processes the events that pass L1T with a CPU farm and further selects events with an output rate up to 1 kHz[48]. In Run 2, more than 400 different HLT paths were employed to filter and categorize events based on various combinations of reconstructed objects included in the definition of a trigger type[49]. Since the object reconstruction in the HLT is done online, it implies that the property of a physics object, for example, the momentum of an electron, is calculated with the fast information provided by local sub-detectors. An HLT

path can decide whether to store an event or not. Due to limited processing time and computing resources, not all the information contained in an event can be used in the online reconstruction. Also, the online reconstruction algorithm cannot be too complicated, hence it sometimes adopts a simplified version of the offline reconstruction algorithm.

The data used in this thesis search is first selected by a L1 muon trigger. It pre-selects events based on the muon tracks delivered by barrel, overlap and endcap muon track finders, which combine information from DT, CSC and RPC[48]. The Muon HLT triggers have 2 steps. In Level-2, muons are reconstructed using the information only from muon detectors, and events are filtered based on the HLT path's requirement on muons. In Level-3, information from the full detector is used to judge if the event satisfies the requirements of the HLT path. For example, an isolated muon trigger requires a muon that is spatially apart from any electromagnetic or hadronic showers. In this case, information from the calorimeters is used.

2.2.6 Particle Flow Algorithm

Particles measured by the detector are identified as electrons (positrons), photons, muons, charged hadrons and neutral hadrons, based on the particle's track in the tracker and the muon detectors, and its energy deposition in the ECAL and the HCAL. At each bunch crossing, many particles are detected at the same time in sub-detectors, and a particle flow(PF) algorithm has been developed to identify and reconstruct each particle. I will

briefly introduce the basic elements and the idea of the PF algorithm in this section. The PF algorithm has many details, for example, the special strategies for reducing mis-identification of each type of particles, which will not be mentioned here.

Charged-particle tracks are reconstructed using an iterative Kalman Filtering (KF) algorithm. The KF algorithm reconstruct the tracks in three stages: 1) generating track seed: find a few hits that obviously link and look like a charged-particle trajectory, 2) building trajectory: find more hits in other tracker layers along this potential trajectory, 3) making a fit with all those hits to determine the direction and momentum of the charged particle[37]. The KF algorithm is efficient when measuring energetic charged particles, while the iterative KF algorithm can fit the tracks with lower energy while keeping the mis-reconstructed track rate low. The iterative KF algorithm relax the track quality criteria required for a track in each iteration, like the number of hits, quality of the track seed, etc. After each iteration, the fitted hits are masked from the rest of fitting.

An electron track is reconstructed with both ECAL-based seeding approach and tracker seeding approach. The track reconstruction for electrons is complicated due to the fact that an electron radiates photons in the tracker. Therefore, an electron trajectory in the tracker can have a kink or multiple kinks. An electron and its bremsstrahlung photons' energy deposits (clusters) in the ECAL are grouped as a supercluster. The ECAL-based seeding approach uses the supercluster's energy and direction as seed, and then extrapolate it to the tracker. It works efficiently if the electron has large energy and is isolated, which means that only low energy bremsstrahlung photons were radiated. The track-based electron seeding

uses all the seeds found by the iterative KF algorithm as mentioned before, and extrapolates the track to the superclusters in the ECAL. If the track momentum and the ECAL supercluster energy is not consistent, we consider that energetic photons are radiated and are not contained in the supercluster. The electron track is reconstructed again with a Gaussian-sum filter (GSF) algorithm which is better adapted to electrons that have substantial energy losses[37].

The PF algorithm reconstructs muon in two approaches. A muon is seeded from the hits in the DT and CSC detector, and later on fits all the hits including those in the RPC along the muon trajectory. We call it a standalone muon. A standalone muon is matched to a track in the tracker to improve the muon momentum resolution, and we call it a global muon. A tracker muon is a muon trajectory that extrapolates from a tracker seed to the muon detector and finds at least one muon segment matches. Global muons, combined with tracker muons, have a total reconstruction efficiency of about 99%, and very often these two methods pick the same segments in the muon detectors and the same track in the tracker[37].

Calorimeter clusters are reconstructed from cluster seeds, which are the cells with an energy higher than their neighbouring cells and higher than a seed threshold. From the seed, its neighbouring cells are grouped into the cluster if the neighbouring cell have an energy that is at least twice the noise level. Next, from the clustered cells, the algorithm continues looking for neighbouring cells and the cluster keeps growing. A topological cluster is formed and, since it may arise from the energy deposits from N particles, a Gaussian-mixture model is used to fit the 2-D energy distribution of the cluster[37].

Isolated neutral hadrons and photons are identified easily since they show up as calorimeter clusters that do not extrapolate to any charged particle track. However, if they overlap with a charged particle, their energy can only be calculated as the excess of deposit energy to the track momentum. Calorimeter cluster energy needs to be calibrated to improve the neutral energy resolution. For the photon energy, due to the threshold in ECAL clustering algorithm, the energy excess cannot represent the overlapping photon energy accurately. The calibration of the absolute photon energy is performed, both in simulation and data, using an abundant π^0 sample with a di-photon resonance fit. For the neutral hadron energy, the calibration is a bit more complicated. The initial calibration of HCAL energy uses 50 GeV charged pions that do not interact in the ECAL. The calibration of neutral hadron energy involves both ECAL and HCAL energy re-calibration. A simulated K_L^0 sample is used to fit the calibration coefficients for ECAL energy and HCAL energy. The neutral hadron calibration in general affects less than 10% of the measured energy[37].

As discussed, most of the particle reconstruction requires linking between different PF elements. Tracks are linked to calorimeter clusters to identify electrons or charged hadrons, and to muon detectors to identify muons. Clusters are linked to form superclusters. ECAL and HCAL clusters are linked to identify and calibrate neutral hadrons. In addition, the link algorithm also links PF elements that might be correlated based on their locations, for example, linking two close tracks to reconstruct a secondary vertex, or linking multiple HCAL showers to an ECAL cluster. PF blocks are formed with the link algorithm, and each block contain PF elements resulting from one or a few particles. In each block, muons

are first identified, and then the electrons and its associated bremsstrahlung photons. After that, isolated high energy photons and neutral hadrons are identified. The associated PF elements are removed in each step. The remaining PF elements are used to cross-identify charged hadrons, non-isolated photons and neutral hadrons, with the calibration of cluster energies[37].

After all PF candidates are identified, missing transverse momentum of an event can be calculated as the negative sum of all PF candidates' transverse momentum. Its magnitude is usually called missing transverse energy and is abbreviated as "MET". Typical searches use these PF objects as the starting point for event selection and further analysis.

Chapter 3

The Signal, Backgrounds, Data and Reconstructed Physics Objects

In this chapter, we continue to discuss the signal of the search, which is two muons and a dark photon generated from pp collisions. The details of the signal and its simulation are mentioned in Section 3.1 and Section 3.2. We will also discuss the backgrounds, which are the processes that happen in pp collisions that look similar to the signal's signature in the detectors. The simulation of backgrounds and the data that we use in this analysis will be described in Section 3.3 and Section 3.4.

In the search, we use some reconstructed physics objects that are defined and recommended by the CMS collaboration for physics analysis purposes. In Section 3.5 we will discuss the selection of muon triggers, the concept of pileup and primary vertices, and muon IDs. We will also examine jets, which are clusters of reconstructed particle-flow candidates.

Clustering particles helps to reconstruct a high energy quark or gluon that generates multiple electromagnetic or hadronic showers along the direction of its propagation before hadronization. This strategy is broadly used to identify the origin of the showers. In this search, a dark photon shower in the calorimeters is also clustered into a jet. The jet that is dominated by a dark photon should have fewer constituents than a hadronic jet. The properties of a dark photon jet are discussed in Section 3.5.5.

3.1 The signal of the search

We search for events in which two muons and a dark photon are generated from pp collisions. Two s-channel Feynman diagrams, shown in Figure 3.1, are the only ones at the lowest order. A dark photon is assumed to interact with the detector materials through pair production, as discussed in Section 1.4. Lowering the dark photon's kinetic mixing coefficient ϵ suppresses the cross section of pair-production, which implies that a dark photon may escape the ECAL. Within a certain range of the ϵ parameter, some of the dark photons will initiate a shower in the HCAL that is located outside the ECAL. If ϵ is even smaller than this range, most dark photons will penetrate through without interacting with the calorimeters and escape into the muon detectors, or beyond.

For the search, we look for a dark photon that initiates an electromagnetic shower in the HCAL. An electromagnetic shower is narrower than a hadronic shower, so it is possible to be distinguished. The background of the search is dominated by the SM Drell-Yan process that

two muons are generated from a pair of quarks through a mediator. If a hadronic shower from any gluon or quark is detected at the same time, such a signature can look like the signal. Considering the background, we focus on final state radiation (FSR) in which case one of the two muons emits a dark photon. The 3-body mass of the muons and the dark photon should be resonant near the Z boson mass and is distinguishable from the background.

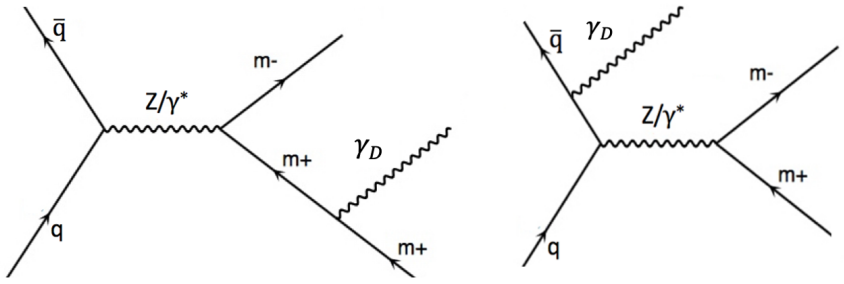


Figure 3.1: Feynman diagrams of s-channel of $q\bar{q} \to \mu^+\mu^-\gamma_D$. The left diagram is Final State Radiation (FSR), where the dark photon (γ_D) recoils from a muon. The right diagram is Initial State Radiation (ISR), where the dark photon recoils from a quark.

The range of f_D that we can probe is constrained by the radius of the HCAL, as well as the cross section of the $pp \to \mu^+\mu^-\gamma_D$ process. By studying simulated events, we observe that the expected signal is well below backgrounds if f_D is lower than 0.001. Also, a dark photon cannot be distinguished from SM photons if it showers in the ECAL. When f_D is higher than 0.1, most dark photons shower in the ECAL and cannot be distinguished. This analysis therefore considers $f_D \subset (0.001, 0.1)$.

We restrict the dark photon mass to $m_D < 1$ MeV. This upper limit is imposed to ensure that it cannot decay into an e^+e^- pair, which are the lightest charged massive particles in

the SM. This criterion makes a dark photon stable and thus, has an opportunity to deposit energy in the detectors.

Below 1 MeV, the analysis strategy is independent of the mass of the dark photon, because we only analyze events with a high energy jet with $p_T > 10$ GeV. At these energies, the kinematic distributions of dark photons will not change due to its small mass compared to the mass of a muon. The cross section of the process is also independent of the dark photon's mass, due to its small mass compared to the collision's center-of-mass energy.

3.2 Signal simulation

Signal events at leading order (LO) from the process $pp \to \mu\bar{\mu}\gamma_D$ are simulated through a couple of steps. The computations of cross sections and the generation of hard scattering events are accomplished using MADGRAPH5_aMCNLO [50] generator with a dark photon (γ_D) that is defined in the Hidden Abelian Higgs Model [17]. Since we focus on the physics that is related to the dark photon, the influence from a dark Higgs in the model is turned off by setting an extremely large dark Higgs mass and an extremely small singlet-higgs coupling constant as instructed by the model's documentation. The details of parameters are listed in Table 3.1. We fix the dark photon mass at 0.5MeV for our analysis, and 11 different ϵ's are sampled between $\sqrt{0.001} \sim \sqrt{0.1}$, as listed in Table 3.3.

The events are generated with kinematic thresholds to avoid events in the sample that would not have passed the triggering criteria, or would be below the reconstruction limit.

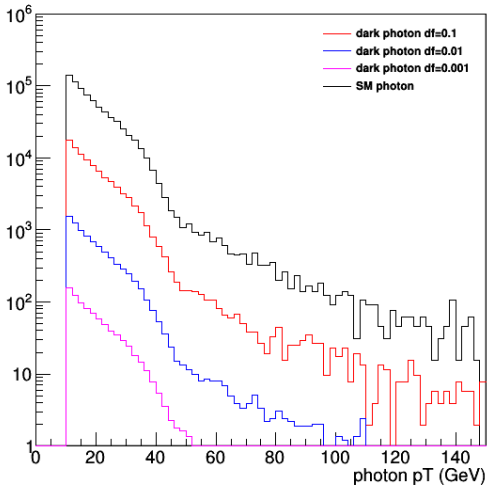
Parameter	Value
mass of a dark photon (M_{γ_D})	5×10^{-4} GeV
mass of a dark Higgs boson (M_{h_s})	1×10^{4} GeV
kinetic mixing between γ_D and γ_{SM} (ϵ)	$\sqrt{0.001} \sim \sqrt{0.1}$
singlet-higgs coupling (κ)	1×10^{-10}
dark Higgs decay width (auto)	5.237950 GeV
dark photon decay width	0 GeV

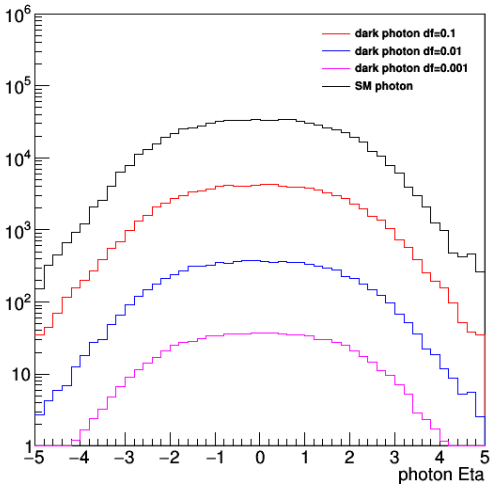
Table 3.1: HAHM_MG5model_v3 parameters for simulating the signal process.

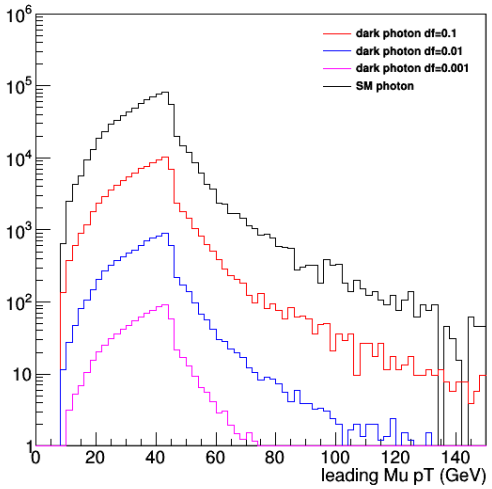
The details of the kinematic thresholds of generated events are listed in Table 3.2 and the basic kinematic are plotted in Figure 3.2

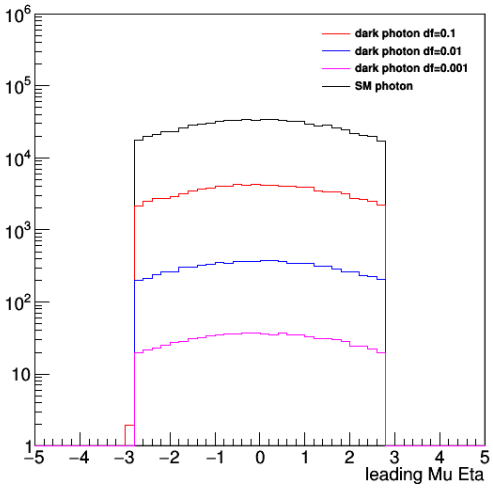
Parameter	Value
minimum p_T for a dark photon	10 GeV
minimum p_T for a charged lepton	8 GeV
minimum invariant mass of l+l- (same flavour) lepton pair	10 GeV
maximum $\|\eta\|$ for a charged lepton	2.8
minimum ΔR between a muon and a dark photon	0.4
minimum ΔR between leptons	0.4

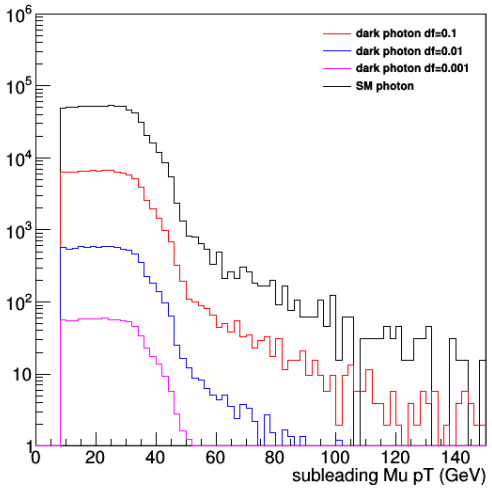
Table 3.2: Important kinematic thresholds of signal simulation.

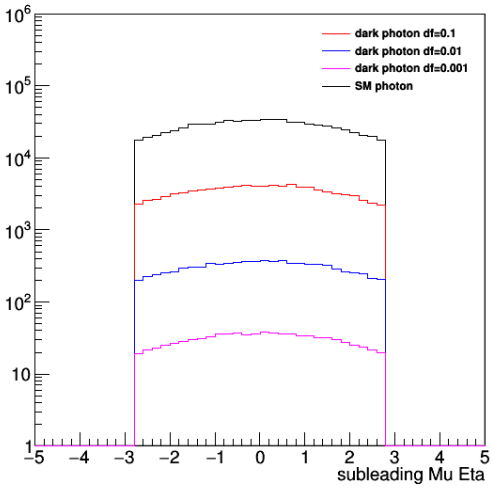

dark photon df=0.1
dark photon df=0.01
dark photon df=0.001
SM photon
photon pT (GeV)


dark photon df=0.1
dark photon df=0.01
dark photon df=0.001
SM photon
photon Eta


dark photon df=0.1
dark photon df=0.01
dark photon df=0.001
SM photon
leading Mu pT (GeV)


dark photon df=0.1
dark photon df=0.01
dark photon df=0.001
SM photon
leading Mu Eta


dark photon df=0.1
dark photon df=0.01
dark photon df=0.001
SM photon
subleading Mu pT (GeV)


dark photon df=0.1
dark photon df=0.01
dark photon df=0.001
SM photon
subleading Mu Eta

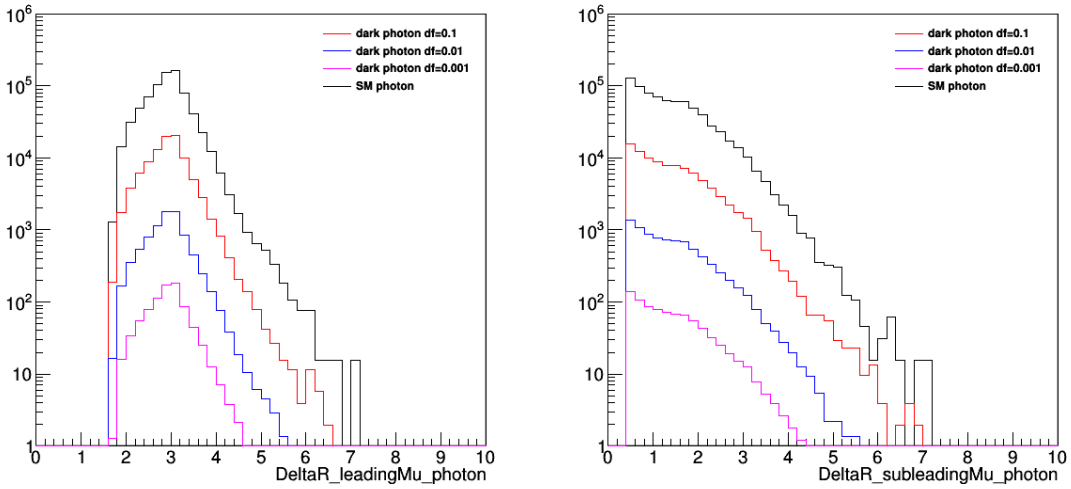


Figure 3.2: Generator level kinematics extracted from LHE files of signals with dark factor value 0.001, 0.01, 0.1 and a standard model photon are compared. The first three rows are p_T, η of dark photon/SM photon, leading muon and subleading muon. Δ R between dark photon/SM photon with each muon are plotted at the last row. Number of events are scaled to 2018 integrated luminosity.

MADGRAPH5_aMC@NLO generates an LHE (Les Houches Event) file containing the kinematic information of incoming and outgoing particles for each hard scattering. PYTHIA8 [51] [52] takes the LHE file as input and simulates the parton showers and hadronization for the events based on the proton's parton distributions. A group of settings of hadronization simulation named TuneCP5 is applied. The Next-to-Next-to leading order parton distribution function (NNPDF)3.1 [53] [54] is used for all simulation steps that need information of parton distribution.

GEANT4 [55] simulates particles' interaction with detector materials. A dark photon interacts with detector materials through pair production. The cross section of the pair production is calculated as the cross section of the pair production from a SM photon multiplied by f_D. The process is implemented in `SimG4Core/CustomPhysics/` package in the CMS software (CMSSW) from version `10_6_X`.

The signal samples we use for the analysis are listed in Table 3.3. We scan through 11 dark factors from 0.001 to 0.1. The cross section varies from 0.0174 pb to 1.98 pb.

3.3 Background simulation

SM processes with two muons and a jet generated from pp collisions are backgrounds to this analysis. The dominant background is Drell-Yan production of a $\mu^+\mu^-$ pair, as discussed earlier. This is included in the `DYJetsToLL` simulation. The sample listed in Table 3.4, `DYJetsToLL_M-50`, simulates events with di-lepton mass heavier than 50 GeV and includes

Dark Factor	Data set Name	Xsec (pb)
0.1000	/ZpMuMu_MZp-0p0005_Df-0p1000_TuneCP5_13TeV-madgraph-pythia8/ RunIISummer20UL18MiniAOD-darkfactor_0p1_106X_upgrade2018 _realistic_v11_L1v1-v2/MINIAODSIM	1.979
0.0631	/ZpMuMu_MZp-0p0005_Df-0p0631_TuneCP5_13TeV-madgraph-pythia8/ RunIISummer20UL18MiniAOD-darkfactor_0p0631_106X_upgrade2018 _realistic_v11_L1v1-v1/MINIAODSIM	1.188
0.0398	/ZpMuMu_MZp-0p0005_Df-0p0398_TuneCP5_13TeV-madgraph-pythia8/ RunIISummer20UL18MiniAOD-darkfactor_0p0398_106X_upgrade2018 _realistic_v11_L1v1-v1/MINIAODSIM	0.7304
0.0251	/ZpMuMu_MZp-0p0005_Df-0p0251_TuneCP5_13TeV-madgraph-pythia8/ RunIISummer20UL18MiniAOD-darkfactor_0p0251_106X_upgrade2018 _realistic_v11_L1v1-v1/MINIAODSIM	0.4498
0.0158	/ZpMuMu_MZp-0p0005_Df-0p0158_TuneCP5_13TeV-madgraph-pythia8/ RunIISummer20UL18MiniAOD-darkfactor_0p0158_106X_upgrade2018 _realistic_v11_L1v1-v1/MINIAODSIM	0.2812
0.0100	/ZpMuMu_MZp-0p0005_Df-0p0100_TuneCP5_13TeV-madgraph-pythia8/ RunIISummer20UL18MiniAOD-darkfactor_0p0100_106X_upgrade2018 _realistic_v11_L1v1-v2/MINIAODSIM	0.1763
0.0063	/ZpMuMu_MZp-0p0005_Df-0p0063_TuneCP5_13TeV-madgraph-pythia8/ RunIISummer20UL18MiniAOD-darkfactor_0p0063_106X_upgrade2018 _realistic_v11_L1v1-v1/MINIAODSIM	0.1108
0.0040	/ZpMuMu_MZp-0p0005_Df-0p0040_TuneCP5_13TeV-madgraph-pythia8/ RunIISummer20UL18MiniAOD-darkfactor_0p0040_106X_upgrade2018 _realistic_v11_L1v1-v1/MINIAODSIM	0.06970
0.0025	/ZpMuMu_MZp-0p0005_Df-0p0025_TuneCP5_13TeV-madgraph-pythia8/ RunIISummer20UL18MiniAOD-darkfactor_0p0025_106X_upgrade2018 _realistic_v11_L1v1-v1/MINIAODSIM	0.04385
0.0016	/ZpMuMu_MZp-0p0005_Df-0p0016_TuneCP5_13TeV-madgraph-pythia8/ RunIISummer20UL18MiniAOD-darkfactor_0p0016_106X_upgrade2018 _realistic_v11_L1v1-v2/MINIAODSIM	0.02782
0.0010	/ZpMuMu_MZp-0p0005_Df-0p0010_TuneCP5_13TeV-madgraph-pythia8/ RunIISummer20UL18MiniAOD-darkfactor_0p0010_106X_upgrade2018 _realistic_v11_L1v1-v1/MINIAODSIM	0.01742

Table 3.3: A list of signals with various dark factor are simulated under 2018 Ultra Legacy conditions. MiniAOD is a custom data format that is used in CMS to save event information.

0, 1 or 2 jets in the final state. It is simulated up to next-to-leading order (NLO).

The sub-dominant background is a $t\bar{t}$ pair with extra jets. The branching ratio of a top quark decaying to a muon, an anti-neutrino and a bottom quark is about 11.4% [56], therefore

about 1% of $t\bar{t}$ events will produce a $\mu^+\mu^-$ pair, thus becoming a potential background. The listed sample TTJets simulates events with up to 2 extra jets and is simulated up to NLO.

Process	Data set Name	Xsec (pb)
Z($l\bar{l}$) + jets	/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIISummer 20UL18MiniAOD-106X_upgrade2018_realistic_v11_L1v1-v1/MINIAODSIM	6077.22
$t\bar{t}$ + jets	/TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8/ RunIISummer20 UL18MiniAOD-106X_upgrade2018_realistic_v11_L1v1-v1/MINIAODSIM	831.76

Table 3.4: Background simulation under 2018 Ultra Legacy conditions.

3.4 Data

The analysis is performed with `SingleMuon` Primary Datasets (PDs). The `SingleMuon` PDs record events that pass at least one muon-related high level trigger (HLT). We further select the events that are produced in the good luminosity sections that are approved for CMS physics analysis, corresponding to an integrated luminosity of 59.83 fb^{-1}. Table 3.5 lists the datasets that record the 2018 `SingleMuon` data. The triggers that correspond to the dataset, as wells to the search, will be discussed in Section 3.5.1.

Process	Data set Name
SingleMuon	/SingleMuon/Run2018A-UL2018_MiniAODv2-v3/MINIAOD /SingleMuon/Run2018B-UL2018_MiniAODv2-v2/MINIAOD /SingleMuon/Run2018C-UL2018_MiniAODv2-v2/MINIAOD /SingleMuon/Run2018D-UL2018_MiniAODv2-v3/MINIAOD

Table 3.5: Data under 2018 Ultra Legacy conditions.

3.5 Defining Reconstructed Physics Objects

3.5.1 Muon triggers

As we've discussed in Section 2.2.5, the trigger system is composed of L1 hardware triggers and HLT software triggers. In this search, we use the single muon L1 trigger and isolated single muon HLT triggers. As recommended by the CMS Muon Physics Object Group, in searching for muons with intermediate p_T we use `HLT_IsoMu24_v*` for 2018 data. This trigger is not pre-scaled, which means that all the collision events meeting the trigger requirements were recorded. The requirement of muon isolation in the trigger distinguishes prompt muons that are generated from primary vertices from those that originate from in-flight weak decays of hadrons.

3.5.2 Pileup and primary vertices

Within one bunch crossing of the LHC, there are more than 20 pp interactions that take place on average. The phenomenon is known as "pileup". Primary vertices are the pp interaction points along the beam. The locations of the primary vertices are reconstructed by first selecting the tracks from charged particles, and then clustering the tracks that appear to originate from the same spot and finally fitting for each spot [57]. In an event, the vertex with the largest Σp_T of its charged particles is selected as the leading vertex (LV) while other vertices become pileup contributors (PUs). Prior or later bunch crossings can also contribute to pileup due to the limitations in the detector's time resolution.

3.5.3 Muons

As we've discussed in Section 2.2.4, muons are detected and reconstructed with an $|\eta|$ coverage up to 2.4 in the CMS experiment. Also, there are three types of muon tracks that are fitted with the information from the muon detectors and the trackers. The offline PF algorithm [37] provides a better identification of muons with information from all CMS subdetectors. For the analysis, we use the muons that pass PF tight muon ID. This ID focuses on selecting prompt muons that originate from primary vertices. It suppresses muons that decay from other particles and muon tracks that are caused by punch-through charged hadrons. In addition, a tight muon PF Isolation ID is applied to the leading muon in an event to make sure that the muon that fires the trigger is properly reconstructed. More details of muon reconstruction and selection criteria can be found in [58].

3.5.4 Jets

We use `AK4PFCHS` jets, that have $p_T > 15$ GeV and $|\eta| < 2.5$, for the analysis. Particles are reconstructed through PF algorithm as leptons, photons, neutral hadrons and charged hadrons, and they are called PF candidates. The PF candidates are clustered into `AK4` jets with anti-k$_T$ algorithm (radius parameter $R = 0.4$) [59] [60]. An algorithm for the subtraction of charged hadrons (CHS) removes charged hadrons that are associated with pileup vertices. With CHS, about 2/3 of the pileup jet energy is eliminated since the charged to neutral hadron ratio is approximately 2:1 in pp collisions. The energy of a reconstructed

jet is further calibrated to suppress the effects from pileup, uneven detector responses and residual jet energy scale differences between data and MC [61].

3.5.5 Dark photon jets

A dark photon can be reconstructed within an `AK4PFCHS` jet if it deposits energy in either ECAL or HCAL. To study the features of a jet from a dark photon, we identify a reconstructed jet as a dark photon jet if $\Delta R(\gamma_{Dgen}, jet_{reco}) < 0.2$.

A dark photon may start its electromagnetic shower at any depth in the calorimeters. With PF algorithm, a dark photon that showers in the ECAL is identified as a photon, and a dark photon that showers in the HCAL is identified as a neutral hadron. Therefore, a reconstructed jet from a dark photon is expected to have a single neutral constituent that contributes significantly to its momentum. Figure 3.3 shows a distribution of a dark photon's contribution to the its jet's p_T if the jet has a neutral hadron and has small neutral EM energy contribution. A rare case occurs when a dark photon deposits part of its energy at the edge of ECAL and its remaining energy in HCAL. In this scenario, two particles may be reconstructed by PF algorithm and the dark photon's energy is split between them.

As is the usual case, the HCAL material sandwich (brass and scintillator) has a higher response to electron-induced showers than hadron-induced showers. The intrinsic e/h ratio is about 1.4 for CMS HCAL [62]. Thus we expect a reconstructed dark photon jet to have energy about 1.4 times higher than the dark photon's generated energy. The e/h value is

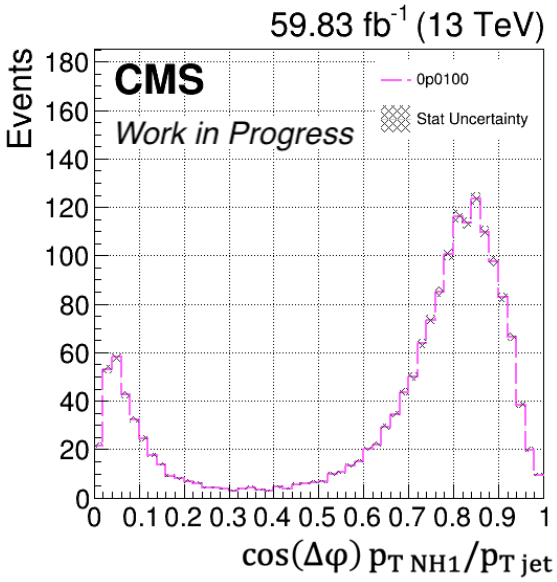


Figure 3.3: The distribution of the leading neutral hadron's contribution to the dark photon jet's p_T with a signal sample that dark factor equals to 0.01. The distribution is calculated by projecting the p_T of the leading neutral hadron onto the jet axis and dividing it by the p_T of the jet itself. The $\Delta\varphi$ in the X-axis represents the φ-angle difference between the jet axis and the leading neutral hadron.The peak around 0.85 indicates that a majority of dark photons are identified as neutral hadrons and contribute significantly to the jet's transverse momentum. The peak at lower end is contributed by the dark photons that deposit energy in ECAL. Those jets have high neutral EM energy fraction.

extracted from test beam calibrations that have measured the HCAL responses with pion beams and electron beams (e/π) [62] when the fraction of electromagnetic shower in a pion decay is already known.

We identify dark photon showers in HCAL when its reconstructed dark photon jet has a high ratio between its energy deposited in HCAL and the dark photon's generated energy (E_{HCAL}/E_{γ_D}) and a low ratio between its energy deposited in ECAL and the dark photon's generated energy (E_{ECAL}/E_{γ_D}). We identify this jet as an HCAL dark photon jet, and it is

the signal that we are interested in. As we see in Figure 3.4, the hot spot along the X axis around 1.4 is contributed by dark photons that shower in HCAL, while the one along the Y axis around 1.0 is contributed by dark photons that shower in ECAL. Details of various properties of such a jet are listed in Appendix A.

3.5.6 The HBHE noise filters

The HBHE (HCAL Barrel and HCAL End-cap) also generate spurious noise due to a variety of reasons, including electronics failures, spikes, or readout issues in the HCAL detecting system. During 2016 and 2017, these were mostly associated with the Hybrid Photo Diodes (HPDs) and Readout Boxes (RBXs). In 2018, the endcap HCAL replaced the HPDs with Silicon Photo-Multipliers (SiPMs) to ameliorate this problem. An `HBHENoiseFilter` was developed to distinguish such noise induced fake jets based on their hit multiplicity, pulse shape, and timing. An `HBHENoiseIsoFilter` further improves the filter by rejecting events that have an isolated signal with a high hit multiplicity in a single RBX or HPD, or a single hit without any adjacent hits [63].

Almost all selected signal events (as described in Section 4) pass the `HBHENoiseFilter`, but about 40% of them fail the `HBHENoiseIsoFilter` due to the dark photon's narrow shower topology. However, this filter is still applied in the analysis to remove HCAL noise during event selection.

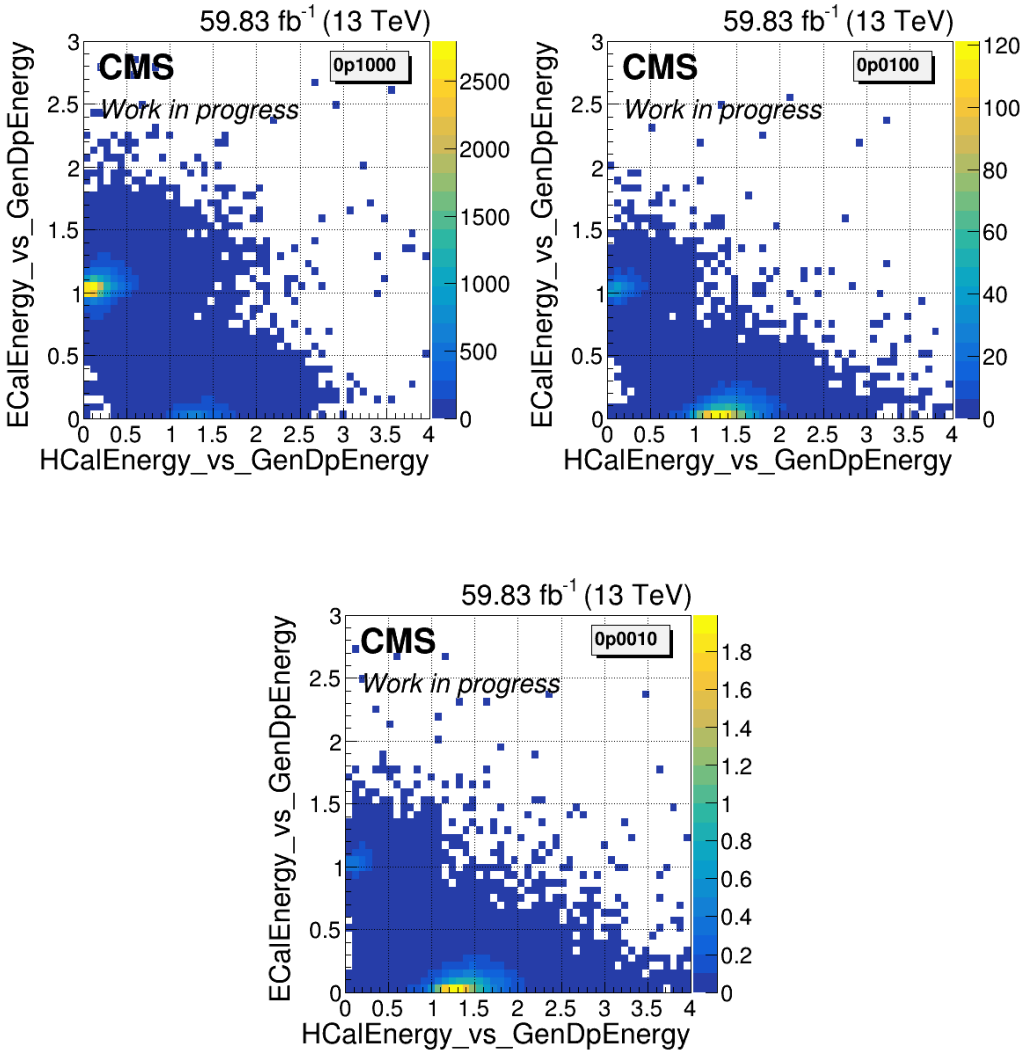


Figure 3.4: The X axis is the dark photon jet's HCAL energy (neutral hadron energy + charged hadron energy) divided by its generated dark photon energy. The Y axis is the dark photon jet's ECAL energy (neutral em energy + charged em energy) divided by its generated dark photon energy. The distribution has two peaks which indicates that most dark photons either deposit in ECAL or HCAL. The e/h ratio for CMS HCAL causes the peak at X axis around 1.4 instead of 1. The events in the top left/top right/bottom plot are from the signal with dark factor equals to 0.1/0.01/0.001. With a lower dark factor, more events drift to the peak on the X axis.

Chapter 4

Signal event selection

This chapter describes the selection of events in the signal search region, based on muons, jets, and other reconstructed objects. Note that in the simulated datasets the search for HCAL dark photon jets, as described in in Section 3.5.5, identifies jets that are co-linear to a generated dark photon. However, in the accumulated collision data events in the signal search region are selected based on information only from reconstructed objects. The jets that are selected without generated dark photon information are called dark photon jet candidates. We will discuss the efficiency of selecting true dark photon jets, and the percentage of wrongly selected jets, later in Section 4.4. The signal events selected in the analysis are required to pass the isolated muon triggers that are described in Section 3.5.1. They are also required to pass the `HBHENoiseIsoFilter` as described in Section 3.5.6.

4.1 Muon pairs

We first select a pair of oppositely charged muons that pass the tight Muon Id and are within $|\eta| < 2.4$. The highest p_T positively charged muon and the highest p_T negatively charged muon are identified as a muon pair candidate. Once the muon pair is determined, we select events in which the two muons are not co-linear and the leading muon (the muon that has higher p_T) is PF isolated. We also require that $\Delta R_{\mu_1,\mu_2} > 0.4$ and that the leading muon passes tight PF Isolated working point.

We further require that the leading muon has $p_T > 26$ GeV. In this region trigger efficiency is estimated by the muon POG to be higher than 75% with respect to the offline muons that pass muon tight ID and PF based isolation requirements [64]. A cut on sub-leading muon ($p_T > 10$ GeV) is placed to match the settings of signal generation as described in Table 3.2.

4.2 FSR HCAL dark photon jet candidates

In the signal samples, the process $pp \to \mu^+\mu^-\gamma_D$ is simulated, but only the events where a dark photon radiates from a muon and showers in HCAL are of interest for the analysis.

An `AK4PFCHS` jet is selected as a candidate for the FSR dark photon jet candidate if the three body mass ($M_{\mu\mu j}$) falls between 85 GeV and 120 GeV, and the di-muon mass ($M_{\mu\mu}$) falls between 55 GeV and 80 GeV. These selection criteria are determined based on the distribution of ISR and FSR events, as shown in Figure 4.1. The event distributions with the masses for ISR and FSR are independent of dark factors. Among the selected jets, the

one with the highest p_T is chosen as the candidate dark photon jet if it satisfies the following conditions: 1) $p_T > 30$ GeV, and 2) $\Delta R_{\mu,j} > 0.4$ for both muons.

The selection of $p_T > 30$ GeV for the jets is necessary due to the significant contribution from pileup jets below this threshold. Many analyses establish a prerequisite of $p_T > 30$ GeV even after applying PUJetId[65]. Without this p_T threshold, in a control region (CR), we observe a larger number of events passing the selection in data compared to the simulated DYJetsToLL process. This discrepancy arises because more observed events, as opposed to simulated events, select a jet with low p_T and high $|\eta|$, as illustrated in Figure 4.2.

The CR encompasses events with $M_{\mu\mu}$ around the Z boson resonance (80 to 100 GeV) and $M_{\mu\mu j}$ between 120 and 145 GeV. These criteria ensure that the selected jets possess similar kinematics to those in the signal region. Further details about the control region, labeled as CR_j, can be found in Table 4.5.

The spatial separation requirement of $\Delta R_{\mu,j} > 0.4$ is implemented to prevent the selection of SM photons which are softly radiated from the muons in the calorimeters, as dark photon jets.

A strong co-relation between a jet's energy fractions and its $E_{HCAL,ECAL}$ to E_{γ_D} ratio provides an efficient method for selecting HCAL dark photon jet candidates without generator-level information. The majority of HCAL dark photon jets have neutral hadron energy fraction (NHEF) greater than 0.6 and neutral electromagnetic energy fraction (NEEF) less than 0.2. To identify a typical FSR HCAL dark photon jet, we select jets with the following criteria: $85 < M_{\mu\mu j} < 120$ GeV, $55 < M_{\mu\mu} < 80$ GeV, $1.3 < E_{HCAL}/E_{\gamma_D} < 1.6$ and

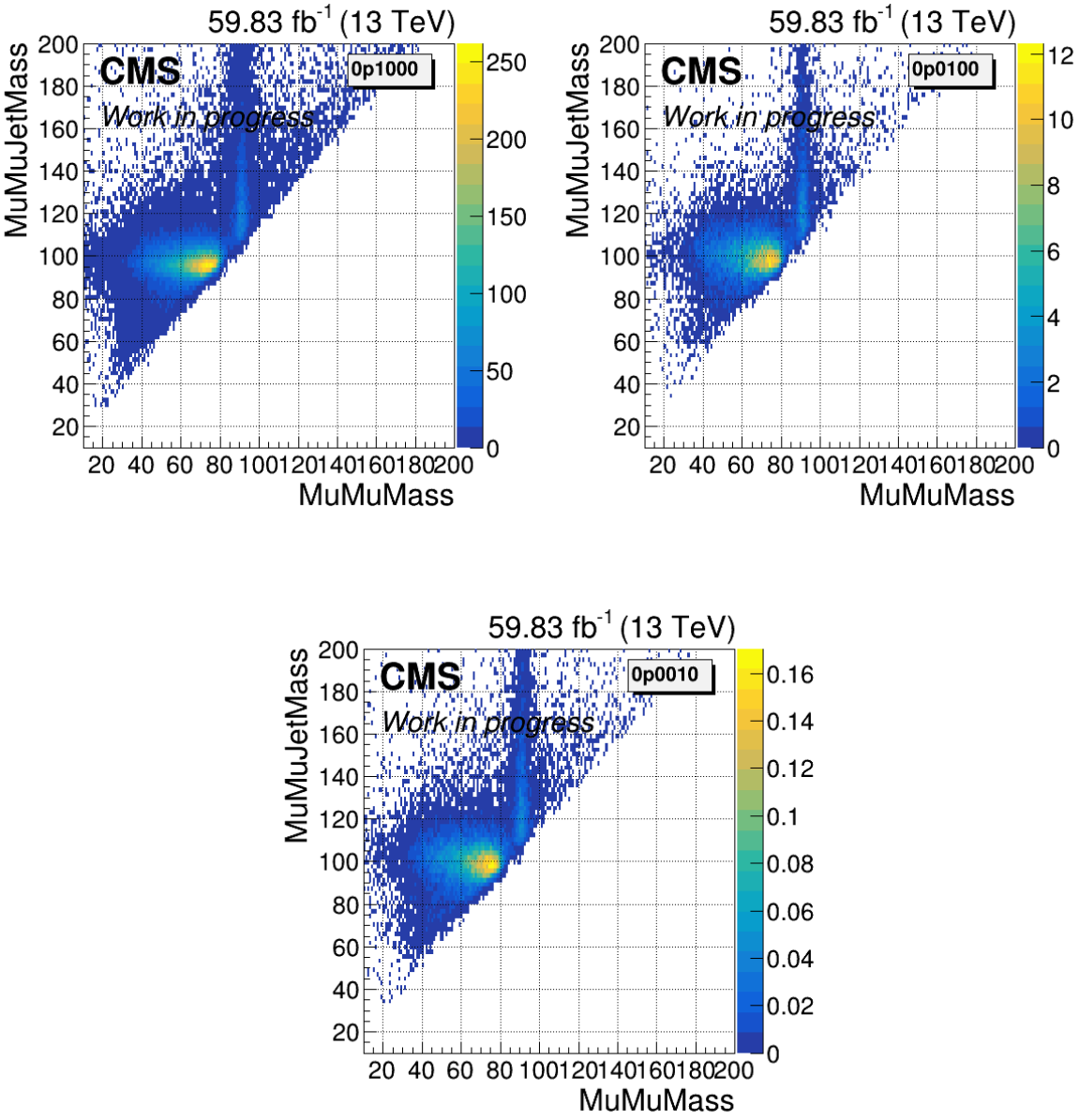


Figure 4.1: The plots are generated using a signal sample with dark factors equal to (top left) 0.1, (top right) 0.01, (bottom) 0.001. Events are primarily concentrated around two peaks. Events with a three-body mass, $M_{\mu\mu j}$, between 85 GeV to 120 GeV are attributed to the FSR process, while events with a di-muon mass, $M_{\mu\mu}$, around the mass of the Z boson (approximately 91 GeV) are attributed to the ISR process. It is important to note that the distribution shown in the plots is independent of the dark factors.

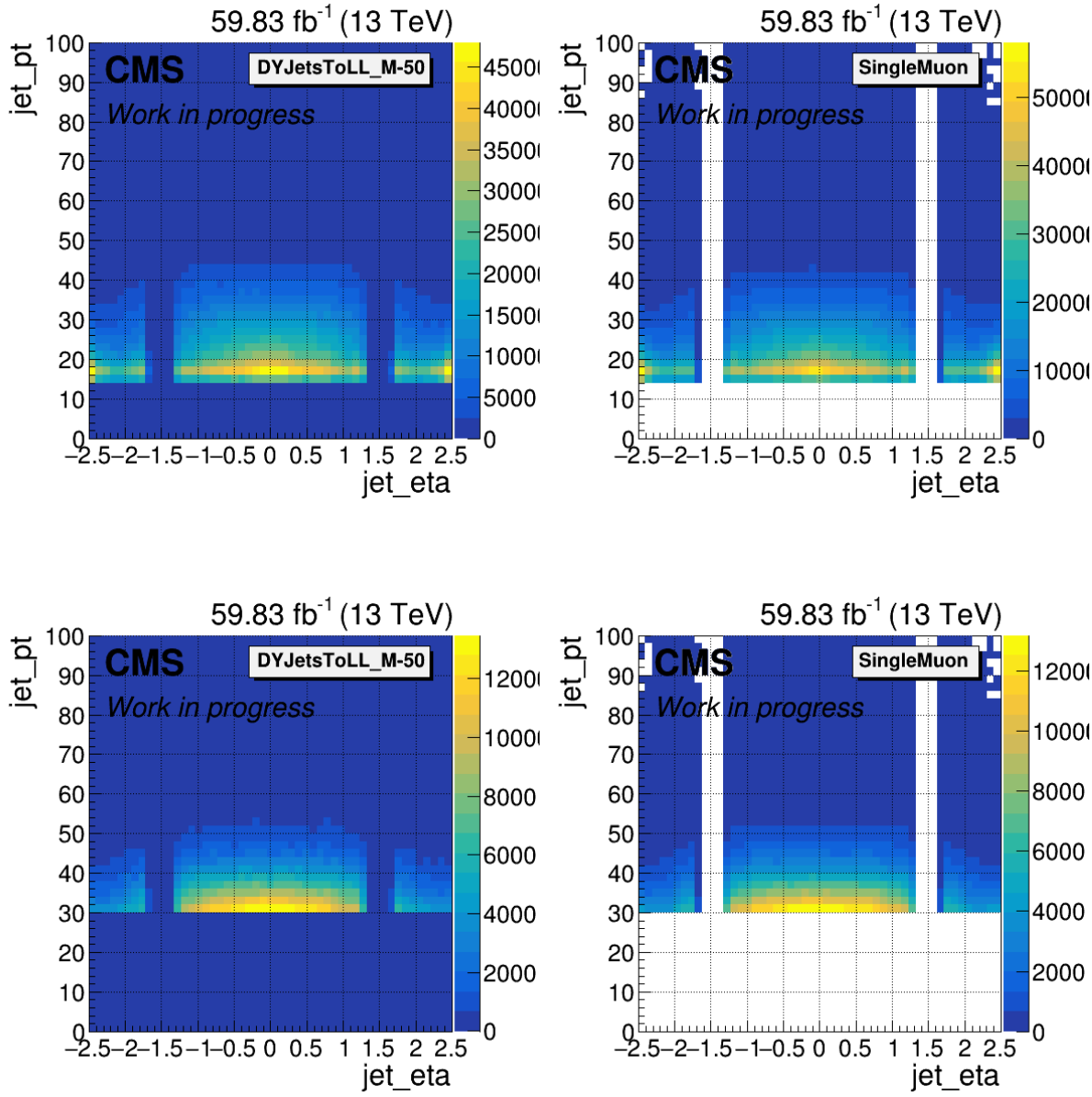


Figure 4.2: The top figures show the p_T and η distributions of events in the CR for DYJetsToLL_M-50 MC (left) and Single Muon (right) before applying the p_T cut at 30 GeV. The bottom plots display the corresponding distributions after p_T cut. In the top plots, both the Single Muon and Drell-Yan simulation exhibit peaks at high jet η. However, the peaks reach over 55,000 events per bin in the Single Muon dataset, while in the Drell-Yan simulation, they reach 45,000 events per bin. After implementing the p_T cut at 30 GeV, as seen in the bottom plots, the number of events per bin over the jet p_T and η are very similar in both the Single Muon and Drell-Yan simulation datasets.

$E_{ECAL}/E_{\gamma_D} < 0.2$. The NHEF and NEEF distributions of these selected jets are plotted in Figure 4.3. This property enables the direct differentiation of HCAL dark photon jets from those that shower in the ECAL. By examining the E_{HCAL}/E_{γ_D} distributions of jets with NEEF < 0.2 and NHEF > 0.6, we observe that only those with high E_{HCAL}/E_{γ_D} and low E_{ECAL}/E_{γ_D} values pass the selection, as shown in Figure 4.4. Thus, we effectively utilize NEEF < 0.2 and NHEF > 0.6 to select HCAL dark photon jets.

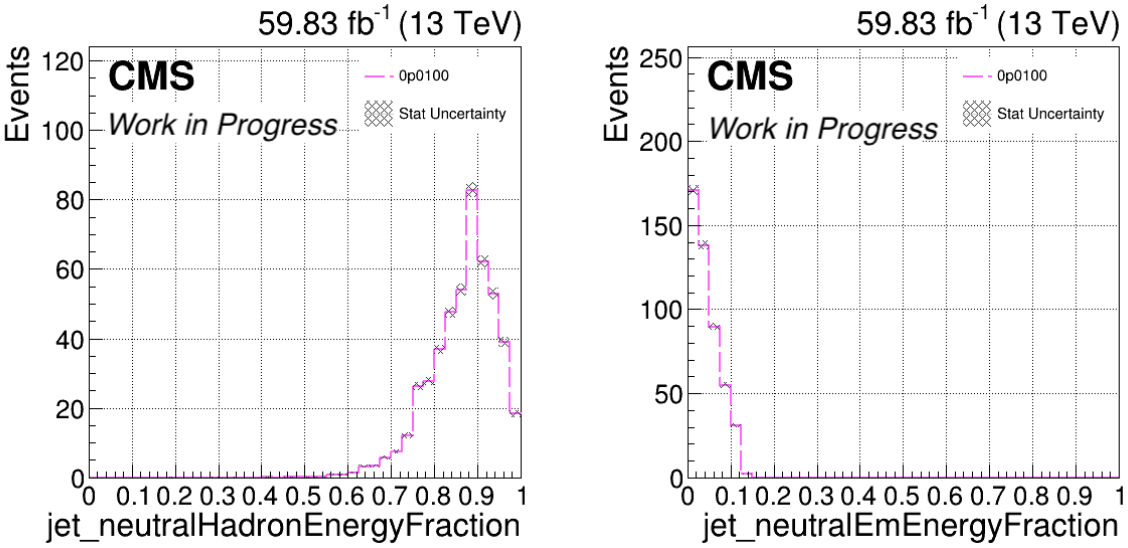


Figure 4.3: A distribution of generated dark photon's (a) NHEF (b) NEEF if the dark photon jet is from the FSR process and deposits most of its energy in HCAL.

The region where the barrel and end-cap HCALs overlap, i.e. $1.3 < |\eta| < 1.7$, is excluded from the signal region due to its suspiciously high probability of having a dark photon jet cluster compared to barrel and end-cap regions. This is indicated in Figure 4.5, where it is likely that mis-matching of tracks with energy showers in the HCAL causes a high neutral

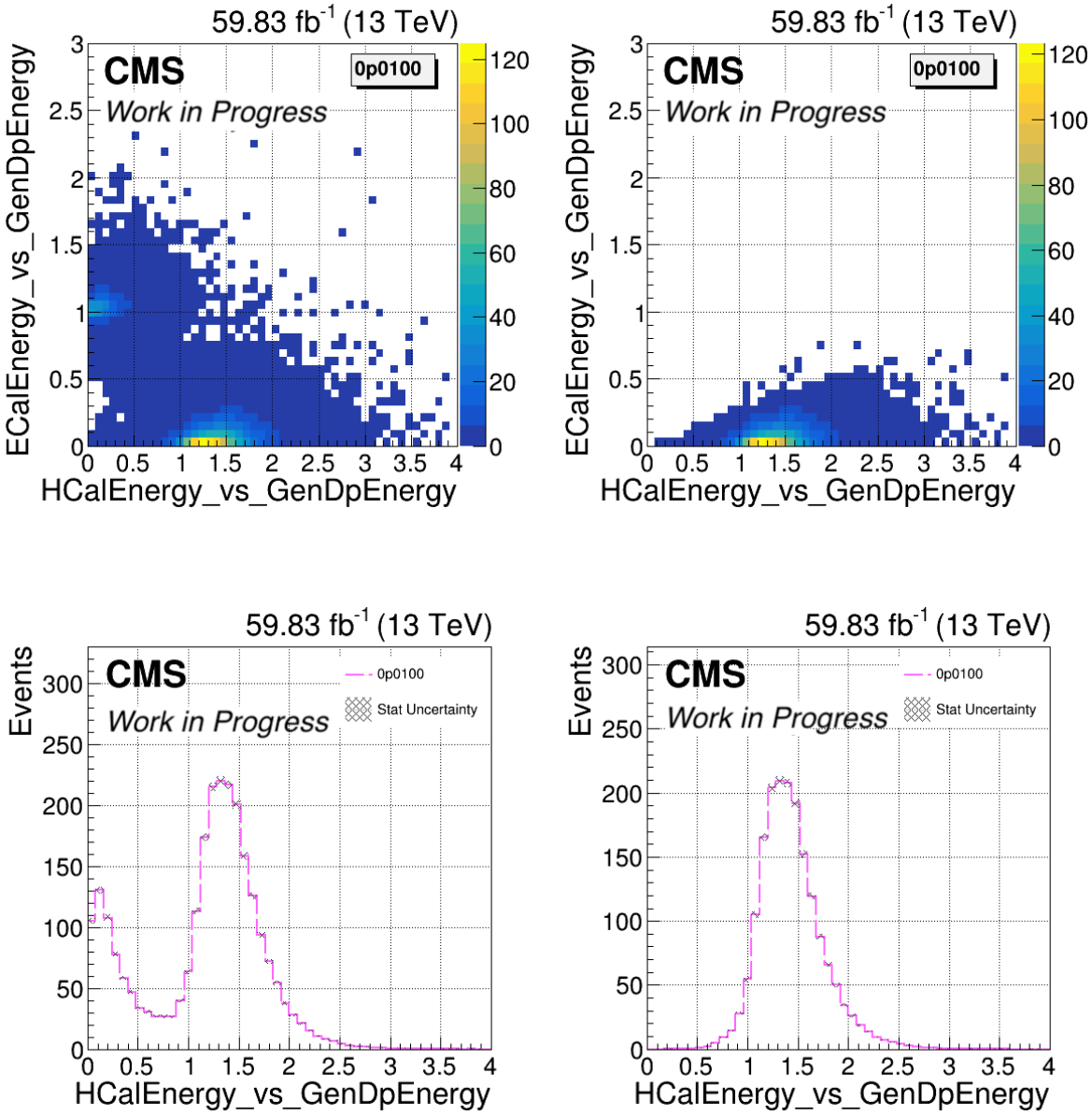


Figure 4.4: Top left: The distribution of events before applying the NHEF and NEEF selections. Bottom left: A projection of the X-axis from the top left plot. Top right: The distribution of events after applying the NHEF and NEEF selections. Bottom right: A projection of the X-axis from the top right plot. By comparing the two bottom plots, it becomes evident that the NHEF and NEEF selections retain the majority of events where the dark photon deposits its energy in HCAL, while effectively rejecting events where the dark photon deposits its energy in ECAL.

hadron energy left in the jet. We lose about 15% of the dark photon acceptance after removing the corresponding η region according to generated dark photon's η distribution.

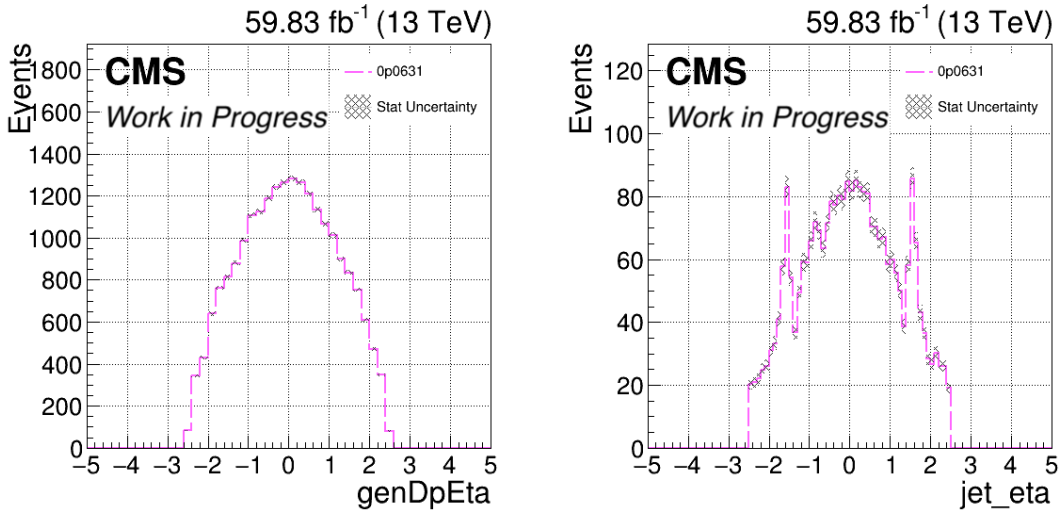


Figure 4.5: (Left) The distribution of the generated dark photon's eta if a dark photon is from the FSR process and its dark photon jet is found. (Right) The distribution of the dark photon jet's eta if the dark photon jet is from the FSR process and the jet deposits most of its energy in HCAL.

4.3 Other selections on jets

An event that contains a B-tagged jet is eliminated to reduce the second dominating background $t\bar{t}$ jets. A jet is a B-tagged if 1) its $p_T \geq 20$ GeV; 2) its $|\eta| < 2.5$; 3) it passes tightLepVetoID; 4) it passes DeepCSV tight working point [66].

We choose not to apply JetId nor PUJetId for reasons explained below. JetId is developed to retain most real jets in the simulation while rejecting most fake jets arising from noise in

the calorimeter electronics [67]. JetId requires neutral hadron energy fraction (NHEF) to be less than 90%, while a non-negligible fraction of dark photon jets have NHEF higher than 90%. PUJetId is developed to remove the jets from pileup. It uses a multivariate method that relies on the vertex and jet shape information. However, a big fraction of simulated dark photon jets either fail loose PUJetId or only pass loose PUJetId. The distributions of the JetId and PUJetId for FSR HCAL dark photon jets with $p_T > 30$ GeV are shown in Figure 4.6. Dark photon jets failing both IDs can be caused by the high neutral energy fraction in the jet. Since we develop a multivariate method that also uses jet shape information, and it is trained with the backgrounds that contain pileup jets simulation, we decide not to use either IDs.

4.4 Signal purity and jet mis-selection

The events that pass the selections mentioned in Section 4 do not necessarily contain all the FSR HCAL dark photon jets. Among the selected events, some of them might not have an FSR HCAL dark photon jet, or the event has an FSR HCAL dark photon jet but a wrong jet is selected as candidate. Table 4.1 lists the purity for each signal sample to highlight these concerns. Purity is defined as the ratio of the number of simulated signal events in which their FSR HCAL dark photon jets are selected using the aforementioned strategies to the total number of selected simulated signal events in the signal region. Overall the purity decreases when dark factor decreases, but they are all above 90%. The purity metric

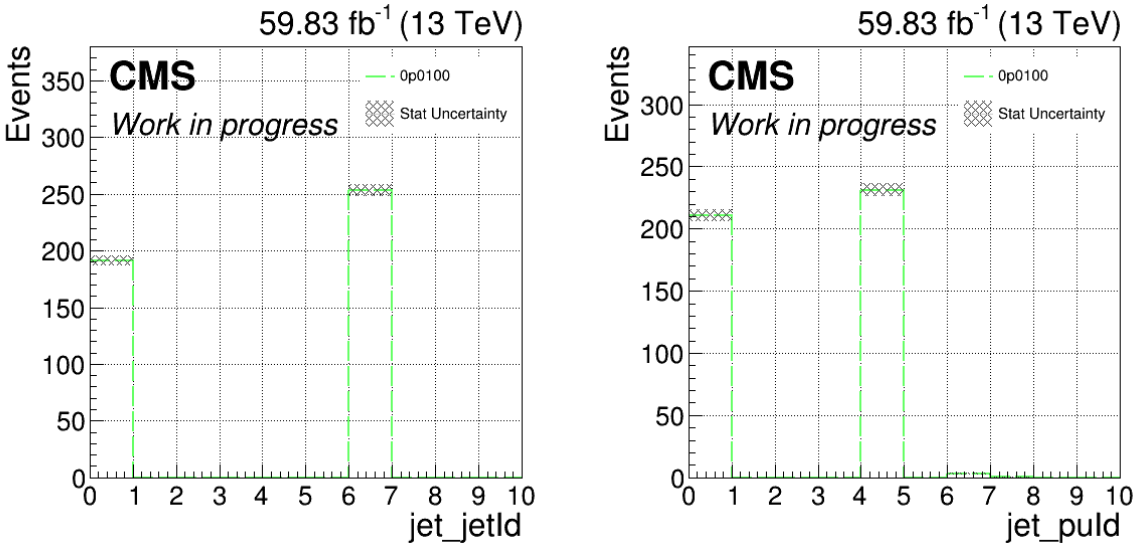


Figure 4.6: The FSR HCAL dark photon jets' Jet ID and Pileup ID distribution. The Jet ID and Pileup ID is defined in the NanoAOD as the following: jetId = 6 when a jet passes tight JetID and tightLepVeto ID, jetId = 2 when a jet passes tight JetID but fails tightLepVeto ID, jetId = 0 when a jet fails tight JetId and tightLepVeto ID, puId = 7/6/4 when a jet passes tight/medium/loose puId, puId = 0 when a jet fails all puIds.

provides a brief measure of how well the selection strategies identify FSR HCAL dark photon jets from other jets in the event. It indicates the effectiveness and reliability of the selection criteria.

4.5 Multi-variate Analysis

We develop a forest of Gradient Boosted Decision Trees (BDTs) with the TMVA [68] package to classify signal events and background events that are simulated in the SR. We also require that the BDTs can be applied to the control regions where the $M_{\mu\mu}$ is around the Z boson

Dark Factor	purity
0.1000	0.97
0.0631	0.98
0.0398	0.98
0.0251	0.97
0.0158	0.97
0.0100	0.96
0.0063	0.95
0.0040	0.94
0.0025	0.94
0.0016	0.92
0.0010	0.91

Table 4.1: Signal purity for each signal sample.

mass resonance. The control region is defined for the purpose of measuring the pileup and HBHENoiseIsoFilter leakage in Section 5.2 and Section 5.3. Therefore, the variables that are related to the muons' kinematics are kept away from the BDTs. The signal events are those that have a generated dark photon matching with a jet as described in Section 3.5.5 and passes the selections as described in Section 4. The background events include not only the FSR events but also the control region events whose jet p_T, η and energy distributions are similar to the jets in the SR. The control region is defined as $M_{\mu\mu}$ between 80 and 100 GeV, $M_{\mu\mu j}$ between 120 and 145 GeV and pass the other selections as described in Section 4, and they are also summarized in Table 4.5 as BDT$_{BackgroundCR}$. We are able to include the control region into the BDTs because the input variables are muon irrelevant. The distributions of the input variables of the control region events look the same as the SR events. It dramatically increases the number of background events being involved in the training.

We only model one discriminator for all the dark factors because the pair-production

generated by a dark photon in principle only depends on the energy of the dark photon. To validate this argument, distributions of input variables from signal events with different dark factors are compared and discussed in Section 4.6.

In the training, signal and background events are weighted separately and will eventually have the same amount of total weight. The signal events are from signal samples that dark factor ranges from 0.001 to 0.1. Each signal event is assigned the same weight event-wisely when we form a signal dataset. The background events are from `DYJetsToLL_M-50` and `TTJets` simulation, as mentioned in Section 3.3. Each background event is weighted by the cross section of the process times the event's generated weight divided by the total generated weight of the data set of the process.

The input variables are listed in Table 4.2. Hyper-parameters settings and optimization of hyper-parameters are discussed in Section 4.7. A 10-fold cross validation is used to check the stability of the training. A preliminary training result is presented in Section 4.8.

4.6 Input variables

We train the BDTs with a list of variables that are shown in Table 4.2. The training focuses on jet properties and minimize the information from muons. The first column lists the muon related variables, which are their spatial separations to the dark photon jet. $\mu 1$ represents the leading muon. $\mu 2$ represents the sub-leading muon. The second column lists the variables that are defined in MiniAOD about a dark photon jet in signal simulations

and a dark photon jet candidate in background simulations. N90(60) is defined as number of constituents carrying a 90(60)% of the total jet energy. The third column lists the self-defined variables that focus on the properties of neutral and charged constituents in a jet. k_T of a jet constituent is defined as the projection of the constituent's transverse momentum onto the jet's transverse momentum. $k_T = p_{T\,constituent} * cos(\phi)$, where ϕ is the polar angle between the jet p_T and the jet constituent p_T. NH1 and CH1 represents the leading neutral hadron and the leading charged hadron in the jet constituents. $\Delta R_{NH1,j}$ represents the spatial separation between the jet constituent and the jet axis.

Muons	Jet	Self-defined
$\Delta R_{j,\mu 2}$	p_{Tj}	$k_{T\,NH1}/p_{T\,jet}$
$\Delta R_{j,\mu 1}$	E_T	$k_{T\,NH1}$
	NConstituents	$\sum k_{T\,NH}/p_{T\,jet}$
	N90	$\sum k_{T\,NH}$
	N60	$Average(\Delta R_{NH,j})$
	NeutralMultiplicity	$\Delta R_{NH1,j}$
	NeutralHadronMultiplicity	$k_{T\,CH1}/p_{T\,jet}$
	NeutralEmEnergyFraction	$k_{T\,CH1}$
	NeutralHadronEnergyFraction	$\sum k_{T\,CH}/p_{T\,jet}$
	ChargedMultiplicity	$\sum k_{T\,CH}$
	ChargedHadronMultiplicity	$Average(\Delta R_{CH,j})$
	ChargedEmEnergyFraction	$\Delta R_{CH1,j}$
	ChargedHadronEnergyFraction	

Table 4.2: Variables being used as inputs to a preliminary BDT training.

The distributions of input variables as well as basic kinematics of signal events with different dark factors are compared in Appendix A. We checked the distributions for dark factors equals to 0.001, 0.006, 0.01, 0.06 and 0.1. Most of them are similar except the dark photon jet's η. As we can see in Figure 4.7, most dark photon jets with a higher dark factor

peak around $|\eta|$ at 0, while for those with a lower dark factor there is a dip around $|\eta|$ at 0. We propose that it is caused by the variation in the depth of detecting material with $|\eta|$. The CMS's cylindrical geometry leads to a larger depth at higher $|\eta|$, which gives a dark photon a higher interaction probability. For high dark factor events, most dark photons are contained in the calorimeters, therefore the jet's η looks normal, while for low dark factor events, a fraction of dark photons escape HCAL at low $|\eta|$ but would have been contained at high $|\eta|$.

Since jet's η is not an input variable and none of the input variables are correlated with it, we believe that our BDT training provides an unbiased model among the samples with different dark factors.

4.7 Hyper parameters optimization

The following hyper-parameters: number of trees, minimum node size, shrinkage and maximum depth of the trees are optimized using the TMVA package[68]. Other hyper-parameters including bagged sample fraction, number of cuts and the way of treating negative weights need to be manually determined by the user. Final values of those hyper-parameters are shown in Table 4.3. Other BDTs parameters, which are not mentioned, have their default values as indicated in [68].

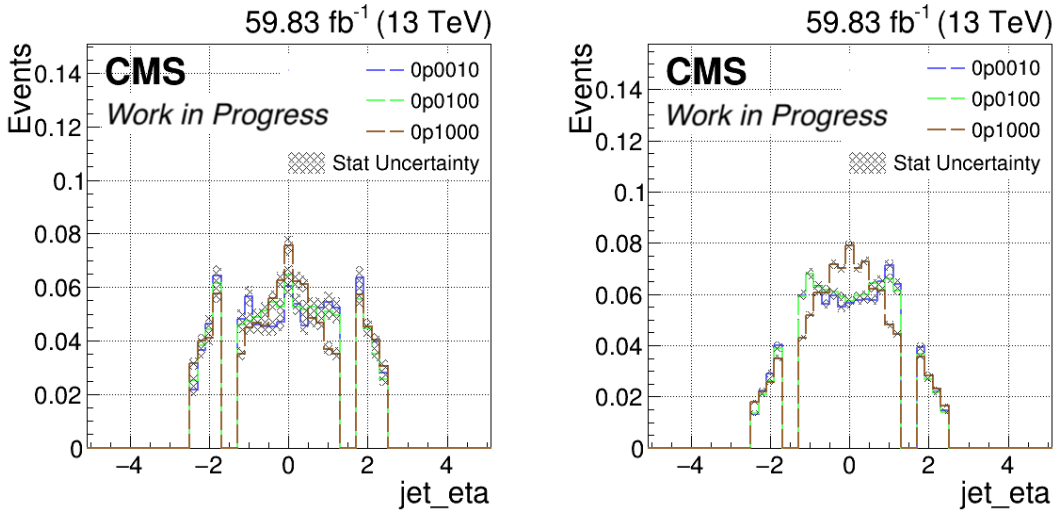


Figure 4.7: (Left) A normalized distribution of an FSR HCAL dark photon jet's η after all selections for a signal region event are applied. (Right) A normalized distribution of an FSR HCAL dark photon jet's η without HBHENoiseIsoFilter and jet $p_T > 30$ GeV selections, but a basic $p_T > 15$ GeV selection is applied. It is for the purpose of gaining more statistics to clearly see the dip at low $|\eta|$ with small dark factor.

Hyper-parameter	Value
NTrees	505
MaxDepth	3
MinNodeSize	15.5
Shrinkage	0.05
BaggedSampleFraction	0.6
NCuts	20
NegWeightTreatment	Pray

Table 4.3: Hyperparameters for BDTs after optimization.

4.8 Preliminary training results

We randomly pick 50% of the signal events and 50% of the background events to train the BDTs, and the others are used to test the BDTs. A 10-fold cross validation procedure is

applied to reduce over-training. It divides the signal and background events in the training set into 10 folds with same weight of signal and background for each fold. Each result is gained by training 9 folds, and is evaluated with the remaining fold. Therefore, 10 sets of selections are generated and we gain a model whose performance is equivalent to the average performance of the 10 sets of selections. We achieve an average ROC (receiver operating characteristic curve) integral 0.913 with standard deviation of 0.011. The overall training and testing distributions for both signal and background is plotted in Figure 4.8.

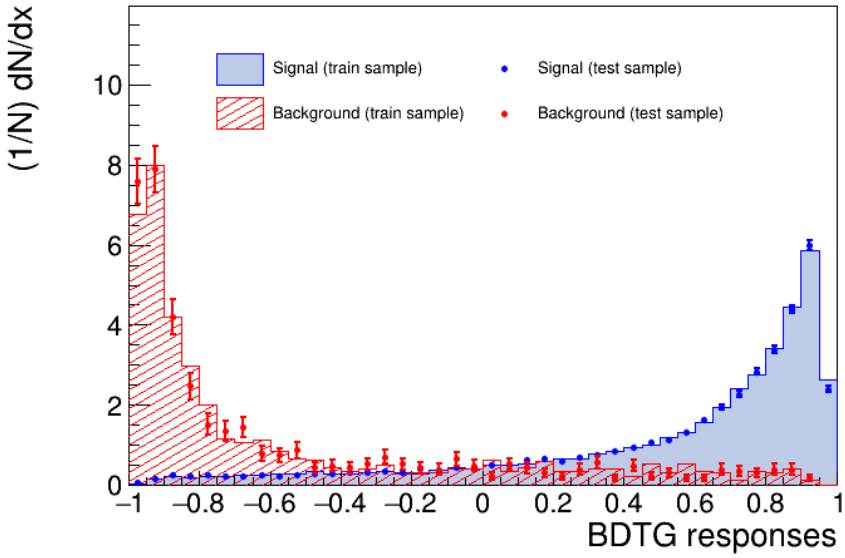


Figure 4.8: The testing and training distributions of both signal and background samples that are involved in the BDTs. It indicates that the BDT model can distinguish signal from background and it does not have over-training issue.

The BDT threshold for identifying an event with a dark photon jet is determined by maximizing the signal significance $s/\sqrt{s+b}$, as well as maintaining sufficient statistics. The threshold depends on the number of expected signal events and background events before the

BDT selections. For different dark factors we expect different number of signal events, but since we will open the box only once, we set this threshold by picking the expected number of signal events when dark factor is at 0.004. The threshold at BDT score is 0.71. Figure 4.9 shows that before the BDT, we expect about 2485 background events and 39 signal events when dark factor is at 0.004. The threshold at 0.71 maximizes the signal significance in this scenario. With the threshold, we gain the amount of expected signal events and background events in Table 4.4 in the SR.

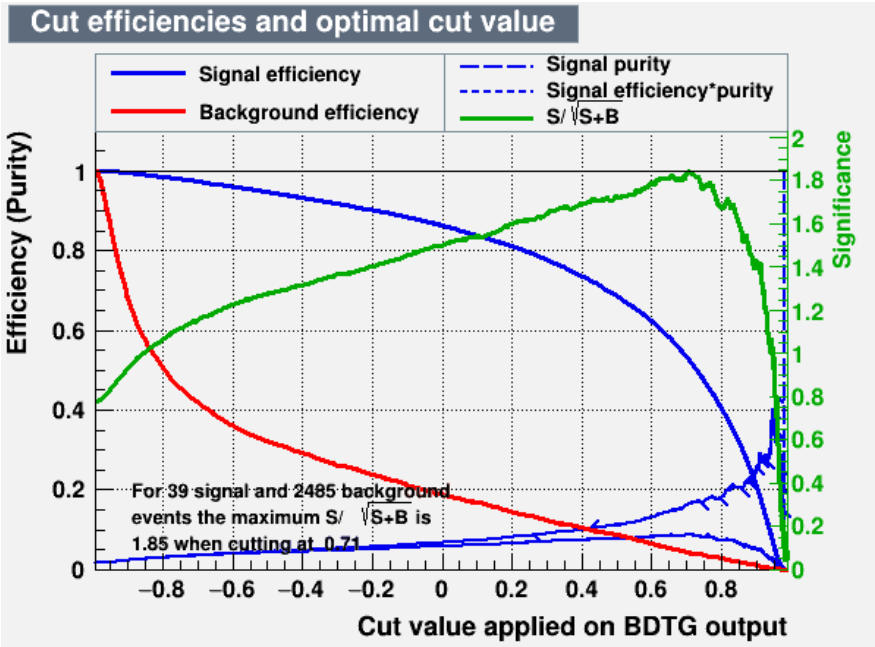


Figure 4.9: Cut efficiencies and optimal cut value if we use the expected number of signal events with dark factor 0.004. The number of background events keeps the same. The optimal cut is at 0.7115. In the context of the TMVA package, the signal purity on the plot is defined as the number of signal events divided by the sum of signal and background events. The signal purity is the blue line that raise up when Cut value approaches 1, and the signal efficiency*purity is the lowest blue line in the figure.

Note that the model is trained and tested on the dark photon jets that have generated

dark photons match with them. When we search for dark photon jet candidates in the SR without generator level information, we may have some events that select a jet that is not from a generated dark photon. The ratio of the dark photon jets out of the selected jets in SR is above 95% for all signal samples after applying the BDT model. The number of expected signal events and signal significance listed in Table 4.4 are calculated without this correction.

dark factor	Expected Signal	Expected DYJets	Expected TTJets	Signal Significance
0.0010	1.4	96	2.3	0.14
0.0016	3.4	96	2.3	0.34
0.0025	7.8	96	2.3	0.75
0.0040	20	96	2.3	1.84
0.0063	42	96	2.3	3.5
0.0100	87	96	2.3	6.4
0.0158	172	96	2.3	10.4
0.0251	291	96	2.3	14.8
0.0398	402	96	2.3	18.0
0.0631	468	96	2.3	19.7
0.1000	398	96	2.3	17.9

Table 4.4: Number of expected signal and background events, and signal significance after applying BDT result to the signal and background events in the SR. They are scaled to the 2018 integrated luminosity 58.93 fb^{-1}.

Selections	CR_j	BDT_{Sig}	BDT_{BkgSR}	BDT_{BkgCR}	SR_{Final}
HLT_IsoMu24	✓	✓	✓	✓	✓
$\Delta R_{gen\ dp,j} < 0.2$		✓			
$55GeV < M_{\mu\mu} < 80GeV$		✓	✓		✓
$85GeV < M_{\mu\mu j} < 120GeV$		✓	✓		✓
$80GeV < M_{\mu\mu} < 100GeV$	✓			✓	
$120GeV < M_{\mu\mu j} < 145GeV$	✓			✓	
$\mu^+\mu^-$ both pass tight muon ID	✓	✓	✓	✓	✓
$p_{T\mu 1} > 26GeV$	✓	✓	✓	✓	✓
$p_{T\mu 2} > 10GeV$	✓	✓	✓	✓	✓
$\|\eta\| < 2.4$ for both muons	✓	✓	✓	✓	✓
$\mu 1$ pass tight pfIsoId	✓	✓	✓	✓	✓
$\Delta R_{\mu 1\mu 2} > 0.4$	✓	✓	✓	✓	✓
$\Delta R_{\mu j} > 0.4$ for both muons	✓	✓	✓	✓	✓
jet $p_T > 30GeV$		✓	✓	✓	✓
b jet veto	✓	✓	✓	✓	✓
HBHENoiseFilter	✓	✓	✓	✓	✓
HBHENoiseIsoFilter	✓	✓	✓	✓	✓
jet $\|\eta\|$ (0 - 1.3, 1.7 - 2.5)	✓	✓	✓	✓	✓
jet $NHEF > 0.6$		✓	✓	✓	✓
jet $NEEF < 0.2$		✓	✓	✓	✓
BDT Threshold at 0.7115					✓

Table 4.5: Summary table of the selections that are used in Section 4. CR_j is the control region for checking data and MC discrepancy in jet's high η and low p_T, as discussed in Section 4.2. BDT_{Signal}, $BDT_{BackgroundSR}$ and $BDT_{BackgroundCR}$ are the signal and background datasets for the BDT training. SR_{Final} is the definition of the signal region after all selections applied.

Chapter 5

Background Estimation, Scale Factors, and Systematic Uncertainties

The contribution of background events in the analysis are from SM processes, pileup and HCAL noise. For the SM processes, the background events are mostly from the Drell-Yan process, in which a pair of muons is produced along with extra jets, and from $t\bar{t}$ events with extra jets. The estimation of these backgrounds relies on simulations, as described in Section 3.3. A control region that barely contains any signal events is selected to validate that we are not missing any significant processes. The details are discussed in Section 5.1.

Pileup is simulated and included in the `DYJetsToLL` and `TTJets` MC datasets. Therefore, pileup effect is trained and tested together with the SM backgrounds in the BDT. However, we find discrepancy between data and MC in the control region that has a similar jet p_T, η and energy distribution as the signal region. The details are discussed in Section 5.2. The

discrepancy is at jet $|\eta|$ between 0.9 and 1.3, and it is plotted in Figure 5.4. It can be caused by the fact that the pileup is not modeled very well in the signal region where we select a jet that is dominated by neutral hadron energy, and it needs to be further investigated. For now, we remove the events in which the selected jet falls into this $|\eta|$ region.

The HCAL noise usually can be filtered away at almost 100% efficiency by applying the HCAL noise filters, which have been developed by the CMS jet noise study group. However, since we look into the region that prefers an isolated high energy deposit in the HCAL, we are highly likely to pick jets that are actually noise but leak through the noise filters. The noise happens occasionally and randomly in some HCAL channels during data taking, hence the HCAL noise cannot be simulated in a the same manner as pileup. An ABCD method has been developed to estimate the HCAL noise filter leakage. It should also cover the potentially remaining pileup after BDT selection. The details are discussed in Section 5.3.

In order to keep track of all the control regions being used in this section, we provide a summary of the selections for each control region in Table 5.1.

Scale factors from multiple sources are calculated to cover the difference between MC and data. Systematic uncertainties from multiple sources are also calculated. The details are discussed in Section 5.4.

Selections	CR_0	SR_{final}	CR_B	CR_C	CR_D
HLT_IsoMu24	✓	✓	✓	✓	✓
$55GeV < M_{\mu\mu} < 80GeV$	✓	✓		✓	
$85GeV < M_{\mu\mu j} < 120GeV$		✓		✓	
$80GeV < M_{\mu\mu} < 100GeV$			✓		✓
$120GeV < M_{\mu\mu j} < 145GeV$			✓		✓
$140GeV < M_{\mu\mu j}$	✓				
$\mu^+\mu^-$ both pass tight muon ID	✓	✓	✓	✓	✓
$p_{T\mu 1} > 26GeV$	✓	✓	✓	✓	✓
$p_{T\mu 2} > 10GeV$	✓	✓	✓	✓	✓
$\|\eta\| < 2.4$ for both muons	✓	✓	✓	✓	✓
$\mu 1$ pass tight pfIsoId	✓	✓	✓	✓	✓
$\Delta R_{\mu 1 \mu 2} > 0.4$	✓	✓	✓	✓	✓
$\Delta R_{\mu j} > 0.4$ for both muons	✓	✓	✓	✓	✓
jet $p_T > 30GeV$	✓	✓	✓	✓	✓
b jet veto	✓	✓	✓	✓	✓
HBHENoiseFilter	✓	✓	✓	✓	✓
HBHENoiseIsoFilter	✓	✓	✓		
Fail HBHENoiseIsoFilter				✓	✓
jet $\|\eta\|$ (0 - 1.3, 1.7 - 2.5)	✓	✓	✓	✓	✓
jet $\|\eta\|$ (0 - 0.9, 1.7 - 2.5)		✓	✓	✓	✓
jet $NHEF > 0.6$	✓	✓	✓	✓	✓
jet $NEEF < 0.2$	✓	✓	✓	✓	✓
BDT Threshold at 0.7115		✓	✓	✓	✓

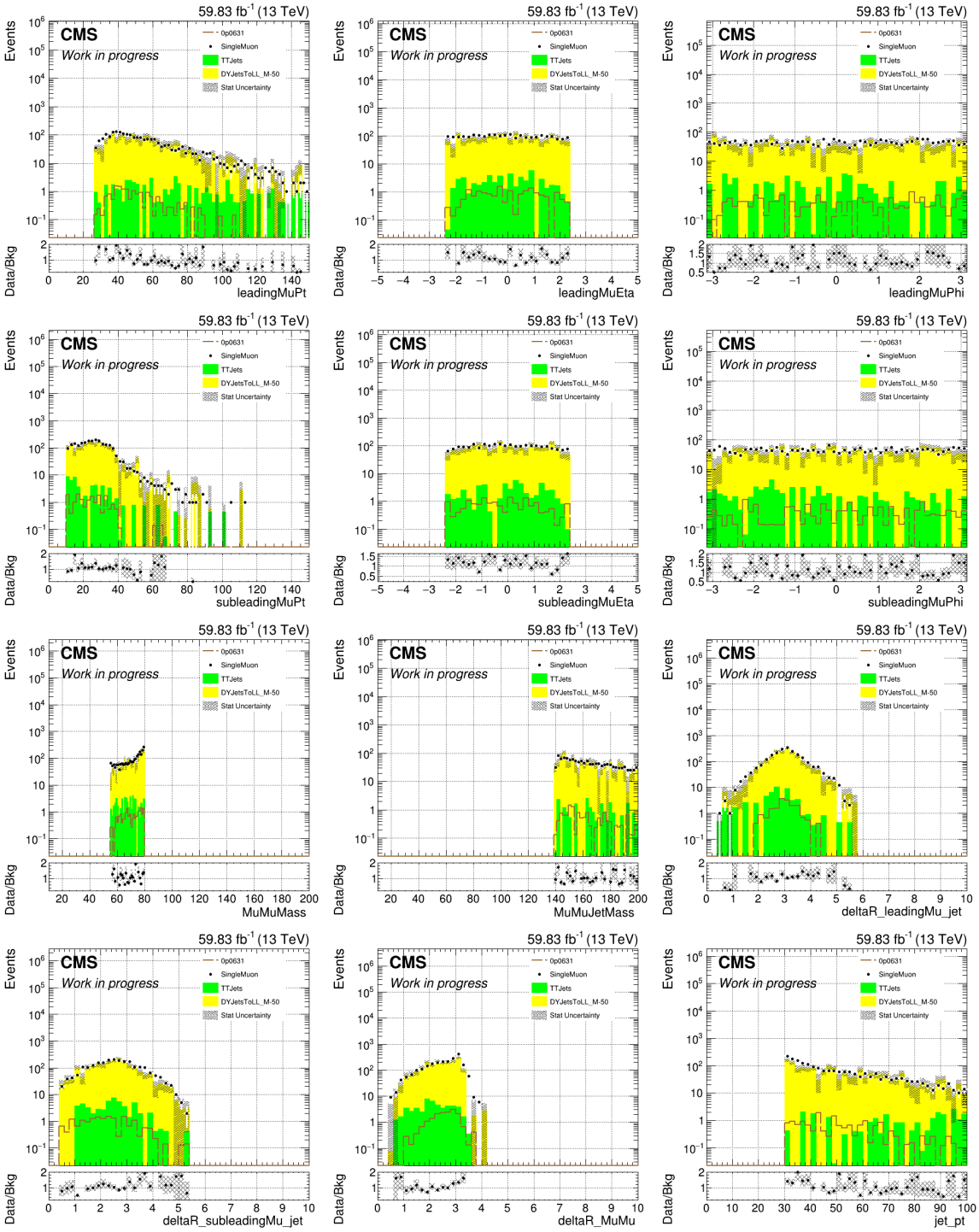
Table 5.1: Summary table of the selections that are used in Section 5. CR_0 is the control region for checking data and the standard model MC discrepancy, as discussed in Section 5.1. SR_{Final} is the definition of the signal region after all selections applied. CR_B, CR_C and CR_D control regions that used in the ABCD method for measuring HCAL isolated noise leakage and it is discussed in Section 5.3.

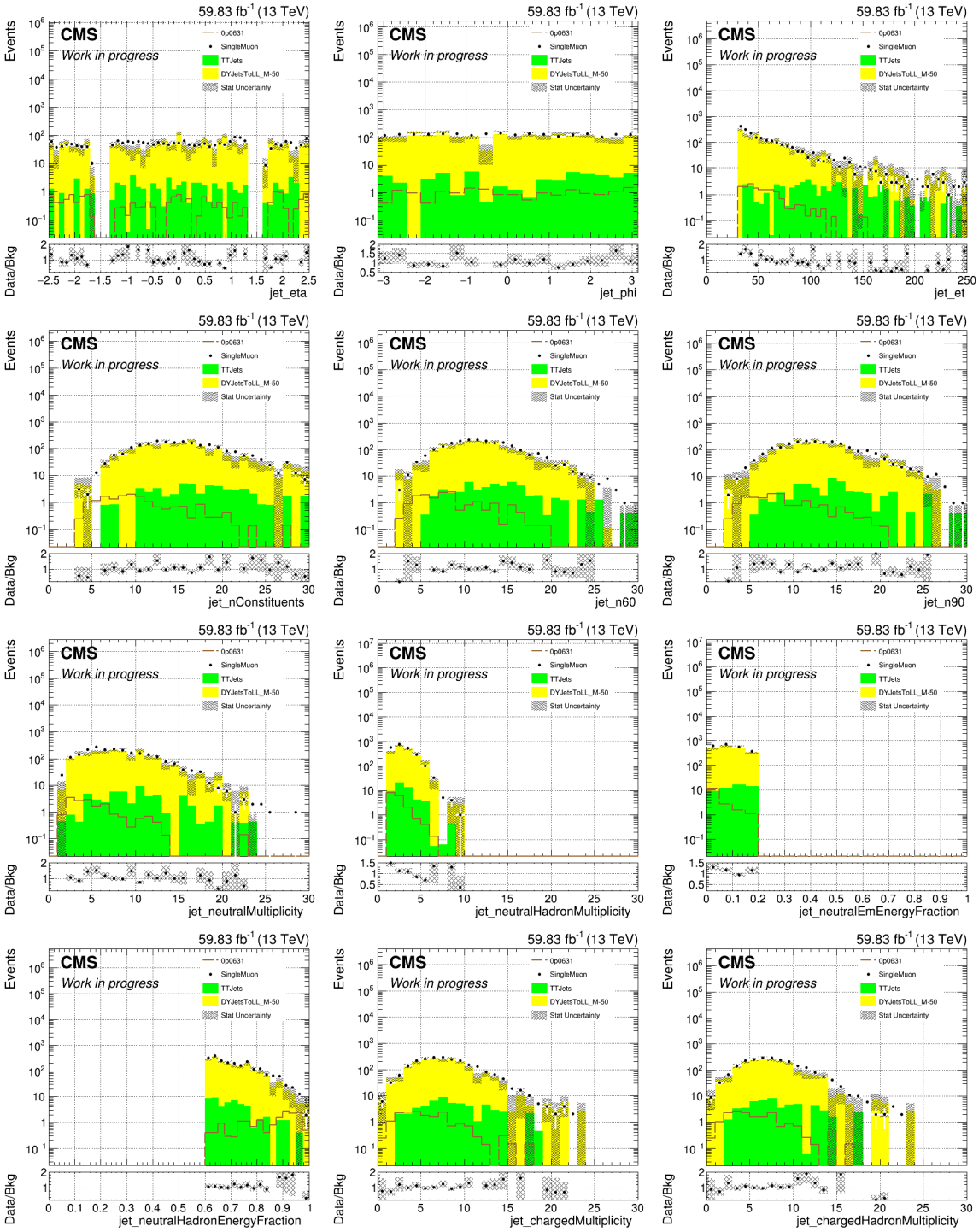
5.1 The SM processes

A control region CR_0 is defined with very similar selections as the SR that is described in Section 4, except that we select events with $M_{\mu\mu j}$ higher than 140 GeV. An event in the control region is selected if:

- The event passes an isolated muon trigger and passes `HBHENoiseIsoFilter`.
- There exists a pair of oppositely charged muons as described in Section 4.1.
- $M_{\mu\mu}$ is between 55 GeV to 80 GeV.
- There does not exist a jet such that its $M_{\mu\mu j}$ is between 85 GeV to 120 GeV.
- There exists a jet that makes $M_{\mu\mu j}$ higher than 140 GeV.
- All other selections on jets are the same as the signal region as described in Section 4.2 and 4.3.

The selections are listed in Table 5.1. The validation of control region focuses on basic kinematics of the muons and jets, as well as the jets energy fractions. The distributions employed in the selection process are contained in Figure 5.1. In general, the simulations of SM processes are consistent with what we observe in the CR_0. Therefore, we can confirm that the dominant SM processes have been included in this analysis.


59.83 fb-1 (13 TeV)
CMS
Work in progress
0p0631
SingleMuon
TTJets
DYJetsToLL_M-50
Stat Uncertainty
Events
Data/Bkg
leadingMuPt
leadingMuEta
leadingMuPhi
subleadingMuPt
subleadingMuEta
subleadingMuPhi
MuMuMass
MuMuJetMass
deltaR_leadingMu_jet
deltaR_subleadingMu_jet
deltaR_MuMu
jet_pt


59.83 fb-1 (13 TeV)
CMS
Work in progress
0p0631
SingleMuon
TTJets
DYJetsToLL_M-50
Stat Uncertainty
Events
Data/Bkg
jet_eta
jet_phi
jet_et
jet_nConstituents
jet_n60
jet_n90
jet_neutralMultiplicity
jet_neutralHadronMultiplicity
jet_neutralEmEnergyFraction
jet_neutralHadronEnergyFraction
jet_chargedMultiplicity
jet_chargedHadronMultiplicity

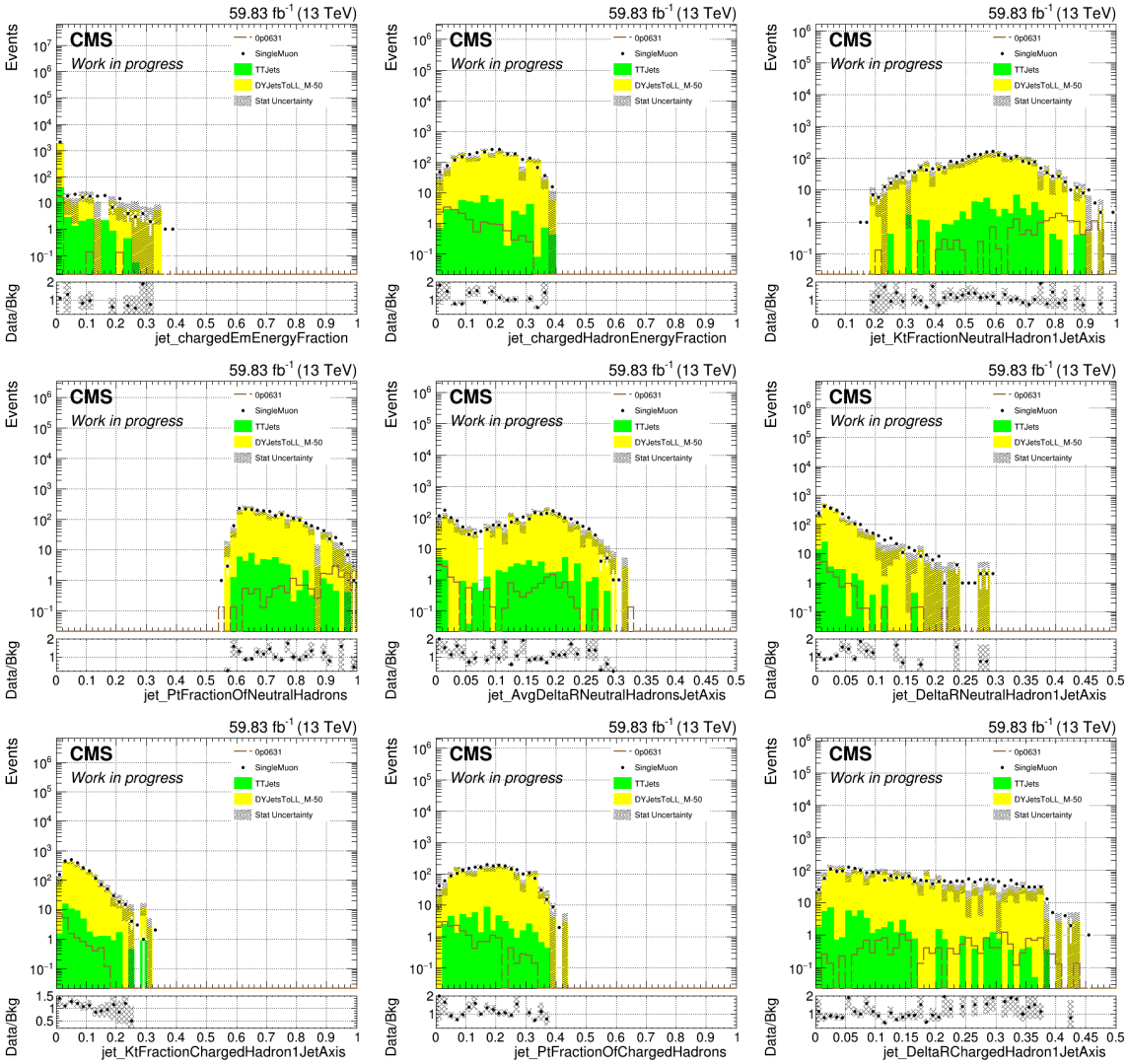


Figure 5.1: Distributions of CR_0 events when dark factor is 0.0631, including basic kinematics and the variables listed in Table 4.2 that will be used as inputs to training a multi-variate method.

5.2 Pileup

As described in Section 4.3, we applied a selection criterion (jet $p_T > 30$ GeV) to reduce the amount of pileup jets. Pileup is simulated and included in both `DYJetsToLL` and `TTJets` MC samples, so pileup can be identified as a background just as the SM processes in the BDT. The BDT scores of pileup jets are checked to make sure that the BDT behaves well in identifying pileup. A jet is identified as a pileup jet if there is no generated jet within $\Delta R < 0.4$ and its jet parton flavour is 0. The BDT score distribution is shown in Figure 5.2.

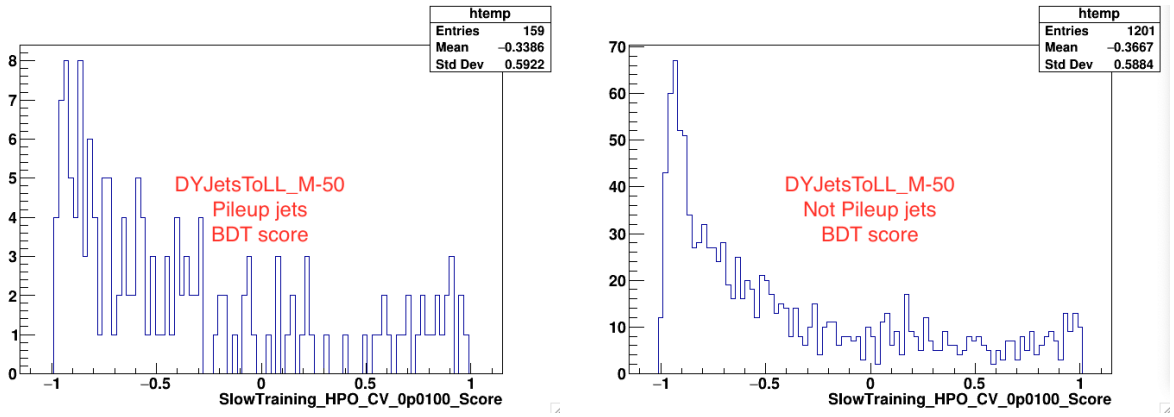


Figure 5.2: BDT Scores of pileup jets(left) and other jets(right) in the DYJetsToLL sample in the SR.

From the CR defined in Section 5.1, we didn't see a big discrepancy between data and MC. However, the CR we looked at in the previous section does not have the same jet kinematic distributions as the SR. The amount of pileup jets that pass the selections strongly depends on the jet's energy and η as we've seen in Figure 4.2. Another CR that has similar jet kinematics as the SR is defined. We call the region CR(B). Basically CR(B) events have

di-muon resonance around Z peak. The events in the CR(B) are selected with the following rules:

- The event passes an isolated muon trigger and passes HBHENoiseIsoFilter.
- There exists a pair of oppositely charged muons as described in Section 4.1.
- There does not exist a jet that makes $M_{\mu\mu j}$ between 85 GeV to 120 GeV, and make $M_{\mu\mu}$ between 55 GeV to 80 GeV.
- There exists a jet that $M_{\mu\mu j}$ is between 120 to 145 GeV, and $M_{\mu\mu}$ is between 80 GeV to 100 GeV.
- All other selections on the jets are the same as the signal region as described in Section 4.2 and 4.3.

Figure 5.3 compares the jet's energy, p_T and η distributions of the events that in the SR and the CR(B) in the DYJetsToLL sample without applying BDT selections.

After confirming that the CR(B) has similar jet kinematics as the SR, the data and MC in the CR(B) are compared to check if there is a discrepancy contributed by pileup. The jet's kinematics are compared and discrepancy is found when jet's $|\eta|$ is between 0.9 and 1.3, as shown in Figure 5.4 (right). In order to gain more statistics to demonstrate the discrepancy, and to check that it is not caused by BDT selections, Figure 5.4 (left) plots the distribution of jet's $|\eta|$ before applying BDT selections.

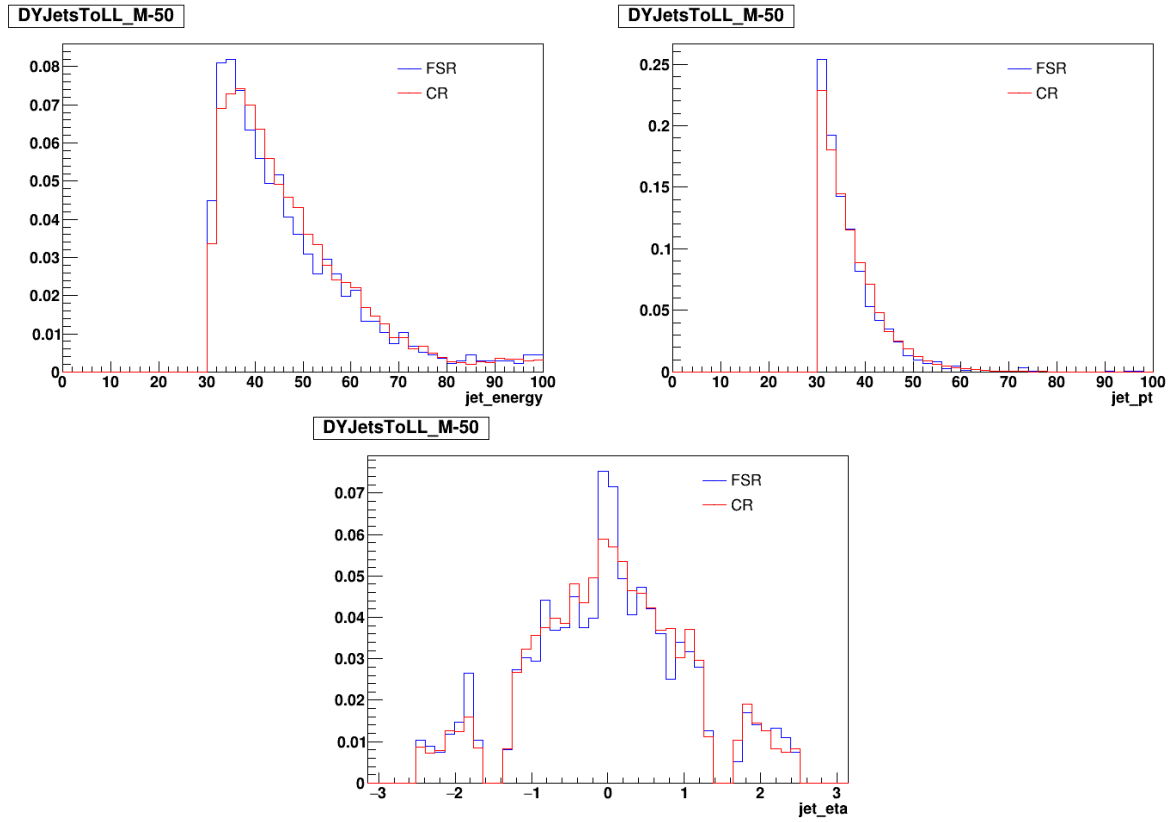


Figure 5.3: Jet's energy, p_t and η distributions of SR(FSR) and CR(B) using DYJetsToLL sample.

The discrepancy was suspected to be caused by the pileup that is not removed by the BDT selections, so the distributions of the number of pileup vertices (npvs) and the number of Good pileup vertices (npvsGood) are plotted before and after the BDT selections. In Figure 5.5, it can be seen that before BDT selections, npvs and npvsGood have mismatch due to the pileup. However, after BDT selections, npvs and npvsGood has a good match between data and MC. Therefore, it is not pileup that causes the two bumps in the data.

Since we do not yet have a clear answer to what causes the bumps in the data, the region

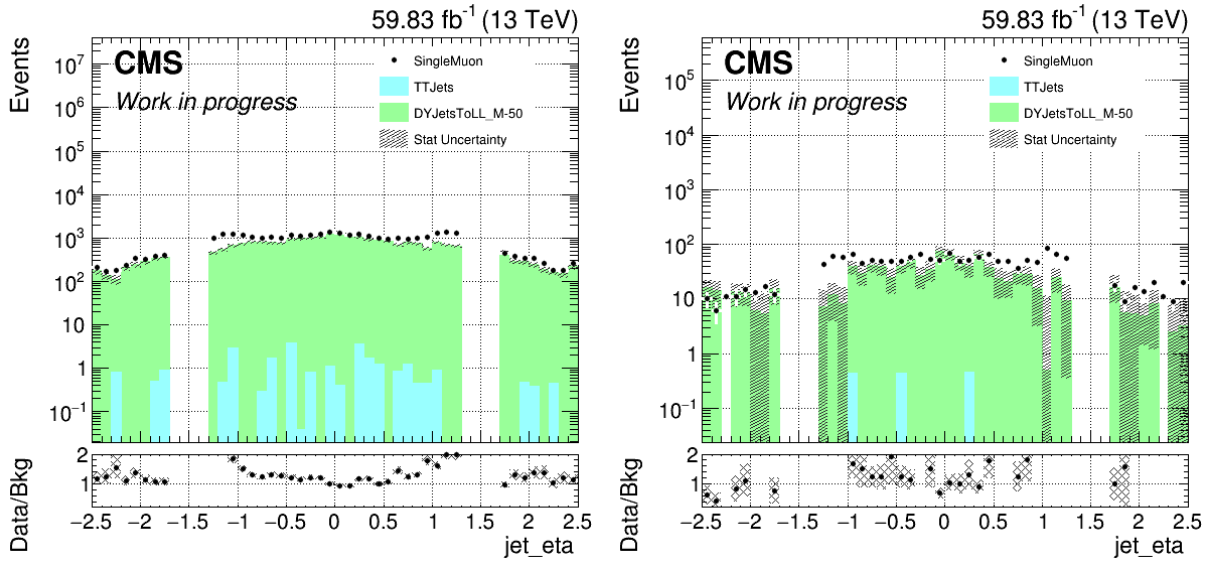


Figure 5.4: Jet's eta distribution of MC and data in CR(B) before(left) and after(right) applying BDT selections.

where jet's $|\eta|$ is between 0.9 and 1.3 is removed from the analysis. This leads to reduction of about 23% in the background estimate as well as about 20% reduction in the expected signal, as compared to the number of events listed in the Table 4.4 in Section 4.

5.3 HBHENoiseIsoFilter leakage

The HCAL noise has three types of electronics failures as its source: Ion feedback noise, which affects one or few HPD pixels; HPD Noise, which affects most or all pixels in a given HPD; RBX Noise, which affects nearly all 72 channels in a given RBX. The dark photon shower in the HCAL will look similar to ion feedback noise, since it is a narrow shower that could be contained within a single HPD pixel. The HCAL noise study group developed an

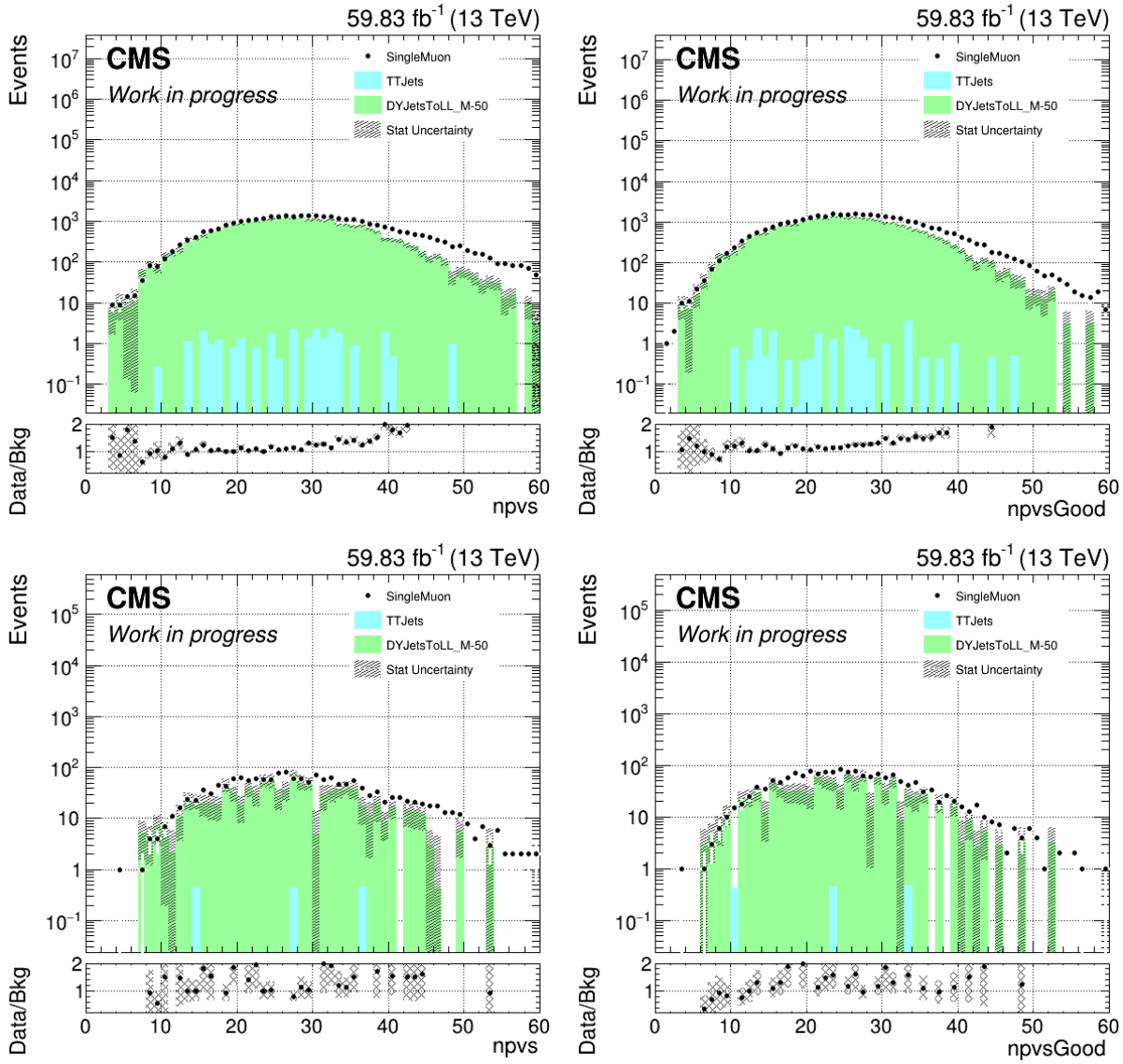


Figure 5.5: The distributions of number of primary vertices (left) and number of good primary vertices (right) of MC and data in CR(B) before(top) and after(bottom) applying the BDT selections.

Isolation-based Noise Filter to identify the ion feedback noise. The noise filter is claimed to be close to 100% efficient for high energy jets, however, for our analysis a large fraction of the jets that we select have p_T between 30 and 50 GeV. According to the noise filter efficiency

measured by the HCAL noise study group in Figure 5.6, the efficiency of the filter tagging an isolated noise is about 53% for jets E_T around 30 to 50 GeV when they measure it with NoBPTX datasets. However, it cannot be applied directly to our analysis because we have made selections that prefer a noise-like signature including the jet energy fractions and the BDT selections. Therefore, we use a data-driven method to measure the noise that leaks through the filter and passes the BDT selection.

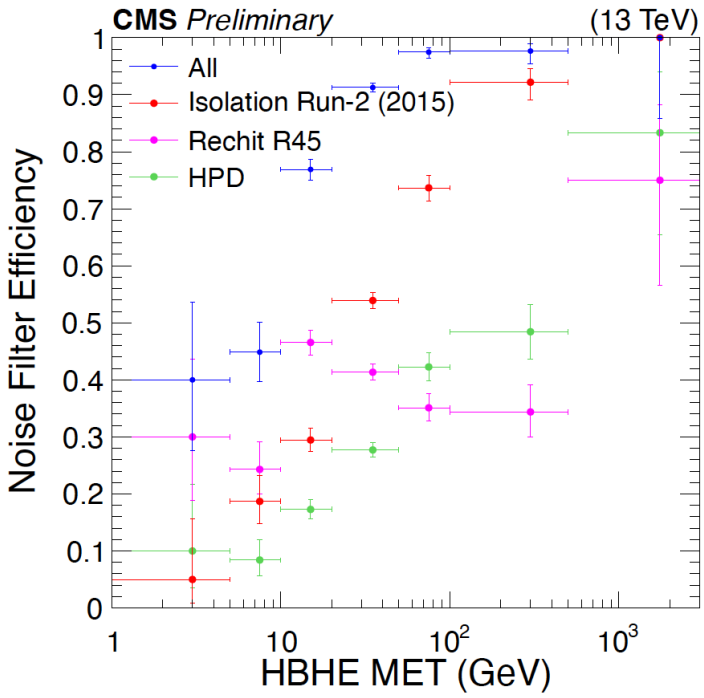


Figure 5.6: Refer to [69]. Noise filter efficiency with respect to the HBHE MET (which is equivalent to the HCAL noise transverse energy) in 25ns NoBPTX dataset that records the detector reactions between bunch crossings. The "Isolation" means tagging the isolated HCAL noise.

The definitions of the regions are shown in Figure 5.7. The CR(B) in the ABCD method has the same selections as described in Section 5.2. The CR(C) and CR(D) regions contain

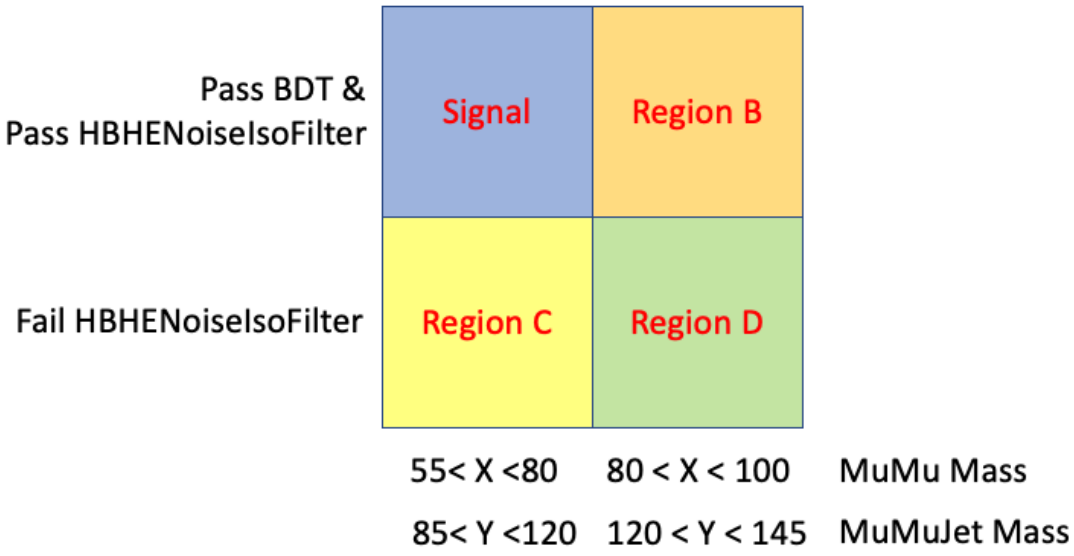


Figure 5.7: Definition of Signal, CR(B), CR(C), CR(D) regions for the ABCD method.

events that fail the HBHENoiseIsoFilter, before BDT selections. An event passing or failing the noise filter is independent of $M_{\mu\mu}$. We assume that the excess number of events from data with respect to MC is from the leakage of the noise filter. After extracting the number of simulated events out of the number of observed events, the ratio between region C and D provides an extrapolation factor for region A and B.

A closure test is performed to validate the ABCD method. We split the CR(B) and CR(D) further into 4 regions by placing a cut on $M_{\mu\mu}$ at 90 GeV. This selection is indicated in Figure 5.8. Details of this calculation are listed in Table 5.2. The estimated A' from the B' C' and D' region is about 143 $\pm$ 82 events, while the observed HCAL noise in A' is 165 $\pm$ 44 events. This estimate matches the observation in region A' with statistic uncertainties. Hence, it proves that the design of the ABCD method is valid.

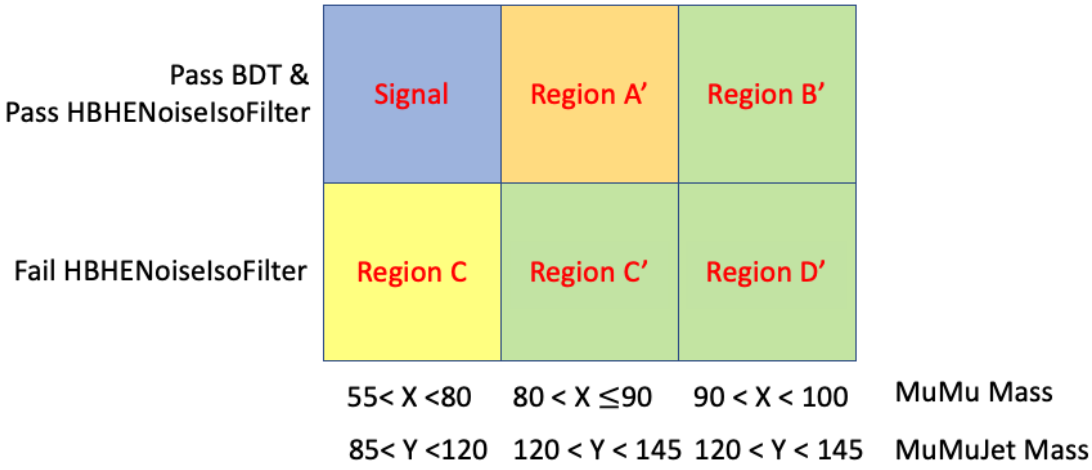


Figure 5.8: Definition of ABCD method validation regions.

Validation regions	#Observed events	#Expected events	#Exceeded events
A'	457	292	165
B'	125	52	73
C'	707	557	150
D'	218	142	76

Table 5.2: Result of the ABCD validation. A', B', C' and D' region correspond to the definition in the Figure 5.8. It lists the number of observed events from 2018UL data and number of expected events from 2018UL MCs that are scaled to the 2018 integrated luminosity 58.93 fb^{-1}.

The number of observed and expected events in CR(B) (C) and (D) are listed in Table 5.3. As a result, we expect about 37 $\pm$ 22 events in the signal region from HCAL noise leakage, before we consider any signal contamination.

Signal contamination depends on the dark factor. It is negligible in CR(B) due the large number of observed events around the Z boson mass resonance from the SM. However, CR(C) is contaminated with signals strongly if the dark factor is above 0.01. In Section 3.5.6 we have observed that the `HBHENoiseIsoFilter` fails for about 40% of the signal events, and

when dark factor is above 0.01, we expect at least 95 events to fails the `HBHENoiseIsoFilter` before BDT selections. This is far beyond what we observe in the CR(C). We treat the dark factor range between 0.01 and 0.1 as not useful for fitting with systematic and statistic uncertainties, due to the obvious fact that it violates the observation. Examining the CR(C) is still necessary for the analysis strategy because we have to take HCAL noise leakage into consideration and have to know how many events are tagged as HCAL noise by the noise filter.

CR	#Observed events	#Expected events	#Exceeded events
B	1163	849	314
C	27	13.5	13.5
D	312	195	117

Table 5.3: Result of the ABCD method. It lists the number of observed events from 2018UL data and number of expected events from 2018UL MCs that are scaled to the 2018 integrated luminosity 58.93 fb^{-1}.

5.4 Scale Factors and Systematic Uncertainties

We consider the following sources of scale factors (SFs) that cover the differences between MC and data and affect the expected number of signal events once we examine the data, post application of all selection criteria. The SFs are applied on the MC samples and the results in Table 4.4 are obtained after applying all the SFs under consideration. Several sources of systematic uncertainties are also calculated. They are summarized in Table 5.4. For 2018 data, an uncertainty of 2.5% is measured on the integrated luminosity, as provided

by Luminosity Physics Object Group.

Source Uncertainties(%)	Signal	DYJetsToLL_M-50	TTJets
Integrated Luminosity	± 2.5	± 2.5	± 2.5
PDF re-weighting	± 0.04	± 0.1	± 0.18
Muon SF stat	± 0.26	± 0.32	± 0.52
Muon SF syst	± 0.78	± 0.74	± 0.37
Pileup re-weighting	$^{+0.5}_{-0.0}$	$^{+0.02}_{-1.18}$	$^{-18.3}_{-12.3}$
Jet energy scale	$^{+8.1}_{-10.3}$	$^{+27.8}_{-29.7}$	$^{+12.4}_{-24.4}$
Jet energy resolution	$^{+2.3}_{-3.7}$	$^{+1.8}_{-12.0}$	$^{-0.0}_{-0.0}$

Table 5.4: Summary of systematic uncertainties.

5.4.1 Pileup

MC samples are produced with a distribution of number of pileup interactions that is roughly but not exactly matched to the real-time condition in the data taking period. To accommodate this effect, we re-weight the MC events by their number of pileup interactions from the simulation truth (nTrueInt). The distributions of nTrueInt from data and MC are normalized to unity, and the ratio between data and MC's normalized distributions is taken as the SF on MC as a function of nTureInt. The systematic uncertainties on the SFs are calculated with the up and down distributions of nTrueInt on data.

5.4.2 Partonic distribution function (PDF)

The selection of a parton distribution function affects the hadronic cross sections of processes in pp collisions. Therefore, we measured the change in backgrounds that are estimated through MC with 103 alternative sets of PDF. For each PDF set, an event has a corresponding

LHE event weight. By re-weighting the event with its LHE event weight, we obtain 103 different results and calculate the $\pm 1\sigma$ error from this sample.

5.4.3 Muon trigger and reconstructed muon efficiencies

The muon trigger that we use is a single muon trigger IsoMu24 for 2018 data, and its trigger efficiency is provided by Muon POG[70]. However, we are selecting events with 2 muons, and there is a small probability that the sub-leading muon satisfies the trigger requirement, instead of the leading muon. The probability of neither of the two muons satisfying the trigger is calculated as in Equation 5.1

$$P = (1 - \epsilon_1)(1 - \epsilon_2) \tag{5.1}$$

where ϵ_1 and ϵ_2 are the trigger efficiencies for the two muons. The SF of muon trigger on di-muon events is calculated as $(1 - P_{data})/(1 - P_{MC})$.

Both of the muons in the analysis are required to pass tight Id, and the leading muon requires to pass tight PF Isolation. There is no PF Isolation requirement on the subleading muon. Based on the SFs and efficiencies that are provided by the Muon POG, the di-muon total efficiency is calculated in three scenarios:

- For events in which the subleading muon passes the tight PF Isolation, the di-muon total efficiency is calculated as in the Equation 5.2.

$$\epsilon_{total} = \epsilon 1_{TRK} * \epsilon 1_{ID/TRK} * \epsilon 1_{ISO/ID} * \epsilon 2_{TRK} * \epsilon 2_{ID/TRK} * \epsilon 2_{ISO/ID} * \epsilon_{Trigger/ISO} \tag{5.2}$$

where ϵ_{TRK} is the muon track efficiency, $\epsilon_{ID/TRK}$ is the muon ID efficiency given a certain track requirement and $\epsilon_{ISO/ID}$ is the muon isolation efficiency given a certain ID requirement. In this scenario, we quote for both muons:

ϵ_{TRK} from `NUM_TrackerMuons_DEN_genTracks`,

$\epsilon_{ID/TRK}$ from `NUM_TightID_DEN_TrackerMuons`,

$\epsilon_{ISO/ID}$ from `NUM_TightRelIso_DEN_TightIDandIPCut`,

in the json file provided by the Muon POG in [71]. Trigger efficiency for the subleading muon quotes: `NUM_IsoMu24_DEN_CutBasedIdTight_and_PFIsoTight`.

- For the events that the subleading muon passes the medium PF Isolation, the di-muon total efficiency is calculated as in the Equation 5.2, but with subleading muon's $\epsilon_{ISO/ID}$ quoted from `NUM_LooseRelIso_DEN_TightIDandIPCut`, and trigger efficiency quoted from `NUM_IsoMu24_DEN_CutBasedIdMedium_and_PFIsoMedium`.

- For events in which the subleading muon does not pass the medium PF Isolation, the subleading muon is considered as not likely to fire the trigger, therefore, the trigger efficiency calculation considers only the leading muon. The $\epsilon_{ISO/ID}$ is not also considered in the calculation for the di-muon total efficiency.

5.4.4 Jet energy scale and resolution

The energy of a jet needs to be corrected due to the fact that the detector's response to particles is not linear. A set of tools is developed to properly match a jet energy deposition

in the detector to the particle's true energy. The corrections are categorized into 1) L1 correction that removes the energy added by pileup; 2) L2L3 MC-truth correction that compares the reconstructed jet p_T to its particle level gen jet, as determined by a QCD di-jet sample; 3) L2L3 residuals correction that corrects the remaining small differences between data and MC based on their jets' p_T and η distributions. It is calculated with Z($\mu\bar{\mu}$, $e\bar{e}$)+jet, photon+jet and multi-jet events for barrel jets.

The jet energy corrections are applied to both signal and background simulations. This step is performed automatically in the customized NanoAODv9 samples that are used in the analysis. The jet energy correction version being used in NanoAODv9 for 2018 UL datasets is `Summer19UL18 V5`.

A jet's energy needs to be smeared when a jet's energy resolution is worse in data than in MC. We follow the "hybrid" jet energy smearing procedure that is recommended by the jet POG, i.e., when a matching gen-level jet is found, we use the scaling method that takes the p_T difference between reco jet and gen jet into consideration. Otherwise, we use a stochastic smearing method, which randomly smears the jet based on its p_T resolution.

Jet energy corrections and jet energy smearing generate a SF respectively. The uncertainties caused by them are calculated based on a simultaneous shift up and shift down on the jet energy, jet p_T and jet mass for all events.

5.4.5 Jet veto in hot/cold calorimeter regions

Some regions of the calorimeters produce anomalously high or low jet rates, due to somewhat flawed calorimeter calibrations for jet measurement performance. The jet mass and energy (JME) POG provides a map that records the hot/cold regions of the calorimeter regions. For this analysis, we discard the event if the selected dark photon candidate jet in the signal region, or the selected jet in the control region, falls into any of the jet veto regions. It reduces the samples by less than about 4% for both data and MC simulations. Applying the jet veto map is recommended by the JME POG to debug the excess events in the $|\eta|$ region between 0.9 and 1.3, as discussed in Section 5.2. It does not reduce the discrepancy that we have observed, as indicated in Figure 5.9, but we decide to apply it to the whole analysis to avoid any slight mis-matching between data and simulations.

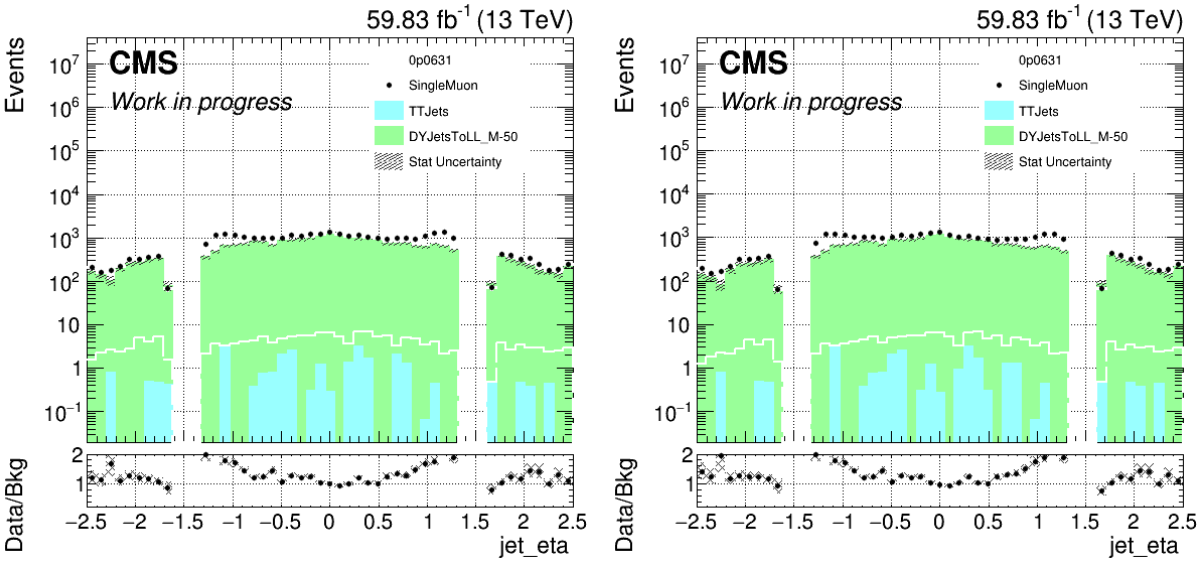


Figure 5.9: The jet η distribution of the events in the CR(B). The left plot is before applying the jet veto map, and the right plot is after applying the jet veto map. There is very tiny difference can be seen, and it does not reduce the discrepancy we saw in $|\eta|$ between 0.9 and 1.3.

Chapter 6

Results

This analysis consists of a single bin counting experiment. We are examining the data to count the number of events that survive all of our selection criteria, and compare that number to what is expected from our extensive simulations of backgrounds. In the case that no significant excess about background expectations is observed, an upper limit is to be placed on the dark factor associated with a dark photon.

The Higgs Combine toolbox[72] [73] [74] is used in the analysis. It is a software package used in the CMS experiment for statistical interpretation of data analysis. It takes into consideration the statistic and systematic uncertainties, and uses hypothesis testing and likelihood-based techniques to compare theoretical predictions to experimental observations.

An asymptotic limit is fit using the Higgs Combine toolbox and the fit is performed simultaneously on the SR bin and the 3 CR bins for the ABCD method.

6.1 Fitting result with Asimov dataset

We first perform a fit with an Asimov dataset before unblinding the signal region. The Higgs Combine toolbox is used with the option `--run blind` and `--expectSignal 0`.

The Asimov dataset is a constructed dataset that simulates observed events for the fittings. It is not real observed data. In the Asimov dataset, the observed data is in agreement with the assumed theoretical model or hypothesis. The Asimov dataset is used to test the statistical models of fits and calculate the maximum signal significance before looking into real data. When we calculate the exclusion limits, we expect that the signal does not exist, and the number of observed events in the Asimov dataset matches the expected number of events from backgrounds. When we test the signal significance, we expect the signal to exist, and the number of observed events in the Asimov dataset matches the expected number of signal and background events.

A 95% upper limit is provided as shown in Figure 6.1 on the cross sections of the $pp \rightarrow \mu^+\mu^-\gamma_D$ process as a function of kinetic mixing (square root of dark factor) from 0.032 to 0.1.

The fitting procedure considers the signal contamination in all control regions for the ABCD method for measuring the HCAL noise filter leakage. For the cases when the signal contamination is larger than the observation in a bin, a lower limit 0.02 is given to the `rateParam` in the data card to make the fit work as desired. The higher limit for `rateParam` is set to 1.5 times the observed data subtracting expected events from SM processes without

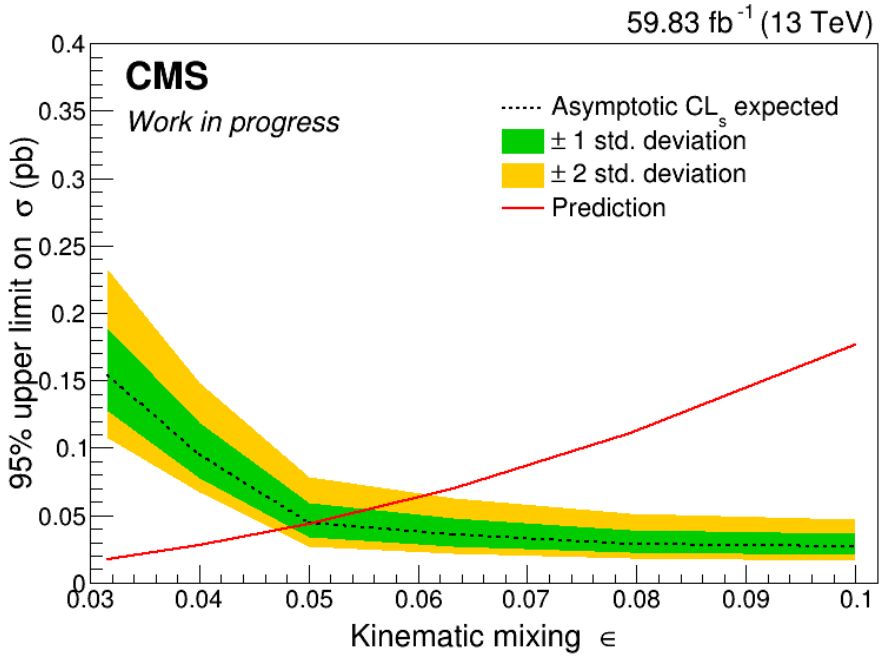


Figure 6.1: Asimov-fit 95% upper limit on the cross section of the $pp \rightarrow \mu^+\mu^-\gamma_D$ process as a function of kinetic mixing from 0.031 to 0.1. The prediction comes from the MADGRAPH calculation of the process's cross section.

considering signal.

Data cards with kinetic mixing at 0.032 and at 0.1 are provided as screenshots in Figures 6.8 and 6.9. The "OT" or "other" represents the HCAL noise filter leakage that needs to be calculated with the ABCD method. The upper and lower 1σ bounds for each type of systematic uncertainties on each bin are listed in the data card as "lower bound / upper bound". The systematic uncertainties are fit with log-normal distribution. The statistic uncertainties for each bin for each process are fit with gmN distribution. The impact of nuisance parameters, shown in Figure 6.2 to Figure 6.7, is that the kinetic mixing is from 0.032 to 0.1. The -1σ impact pull sometimes is not shown for C_OT (the noise leakage in

region C) because the fitting finds a mathematically negative value for C_OT corresponding to the signal strength's upper bound, but it is physically impossible to have a negative number of noise events. A threshold is placed in the data cards to prevent the fit from looking into negative C_OT. In this case, C_OT's lower bound is not found and is not shown in the impact plots. The $+1\sigma$ impact pull is sometimes not shown for C_OT because the fit finds a positive value that is more than number of observed events, which is also physically impossible.

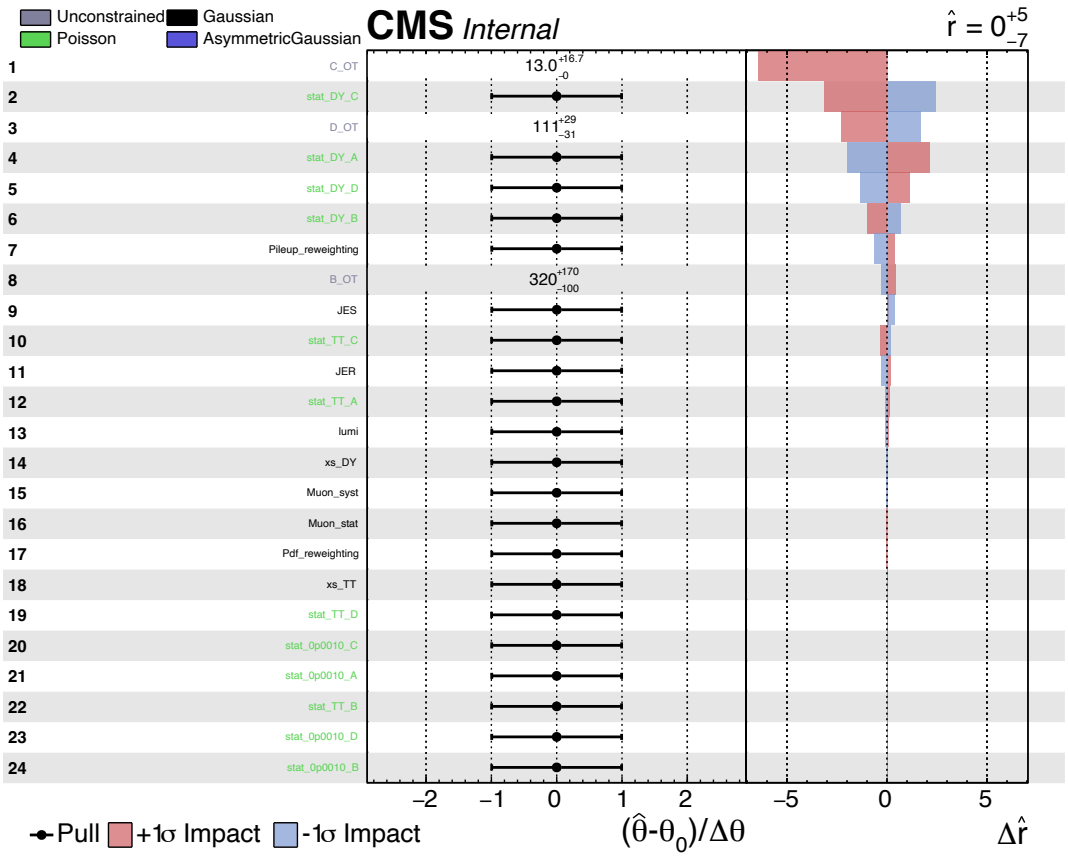


Figure 6.2: Impact of nuisance parameters to the parameter of interest (cross section of the process $pp \to \mu^+\mu^-\gamma_D$)when kinetic mixing is at 0.032. The number of observed events in the data card is set to be equal to the expected number of events.

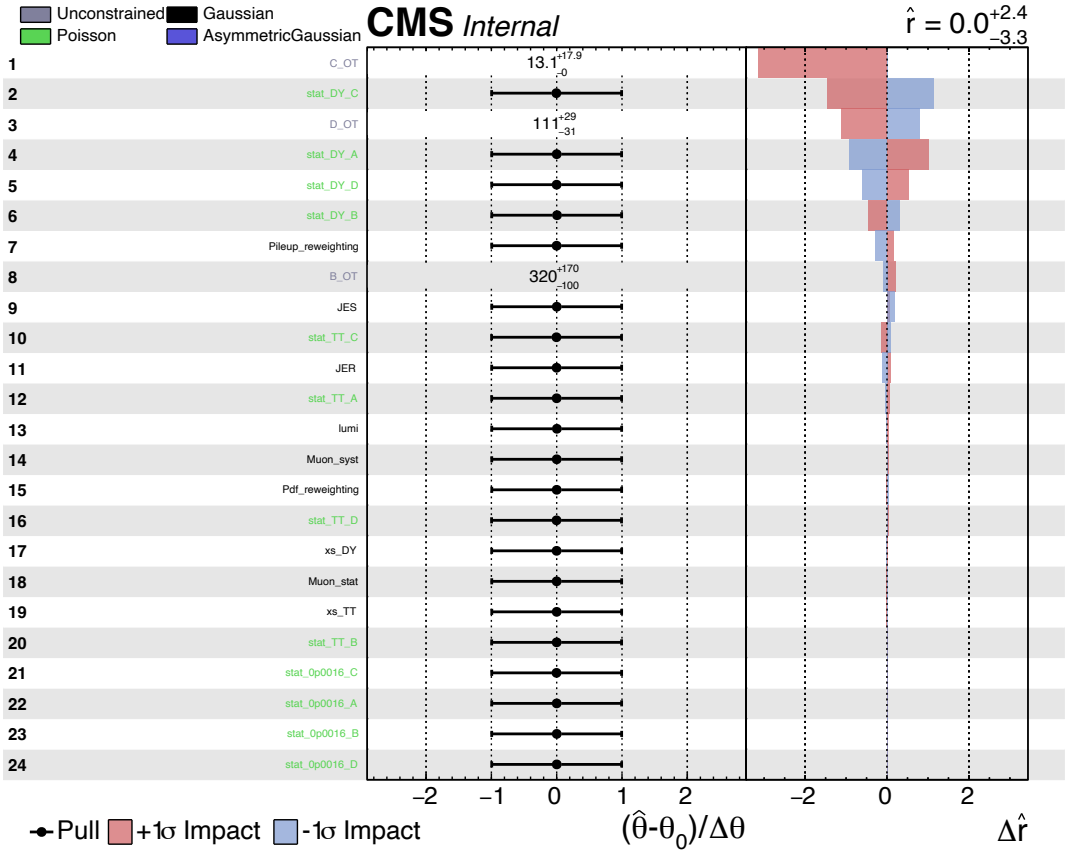


Figure 6.3: Impact of nuisance parameters to the parameter of interest (cross section of the process $pp \to \mu^+\mu^-\gamma_D$) when kinetic mixing is at 0.04. The number of observed events in the data card is set to be equal to the expected number of events.

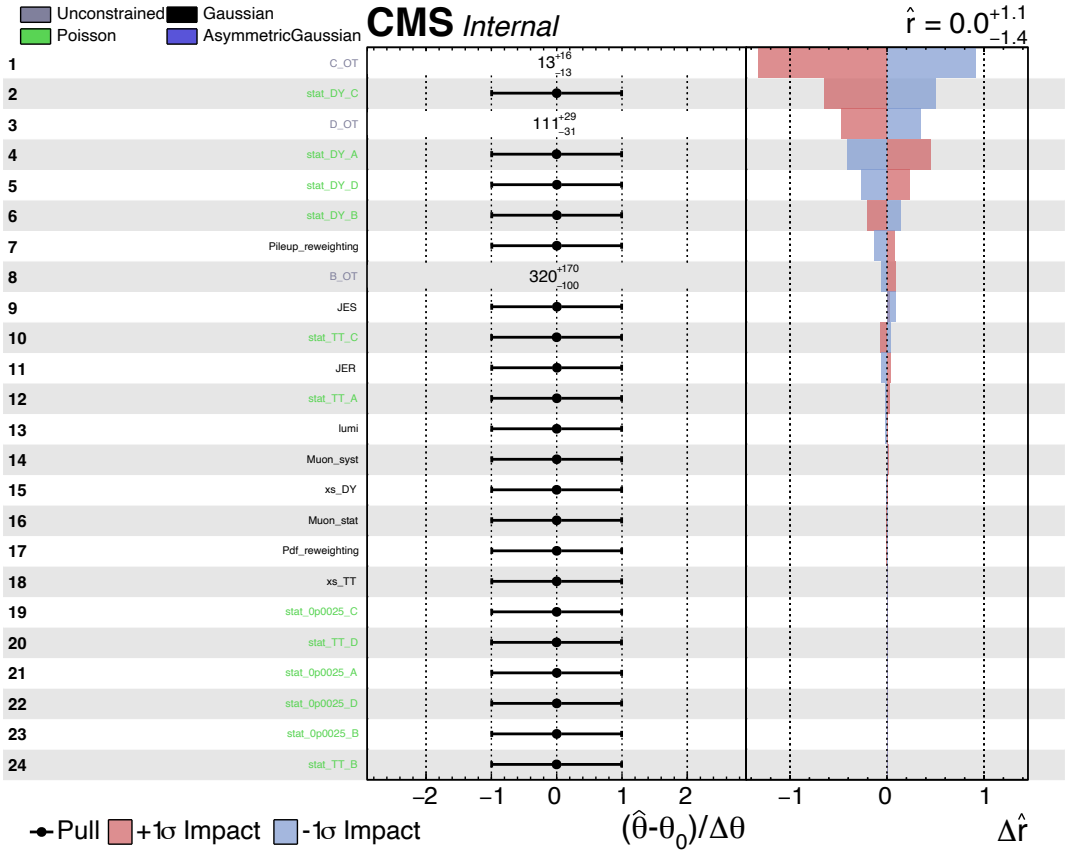


Figure 6.4: Impact of nuisance parameters to the parameter of interest (cross section of the process $pp \to \mu^+\mu^-\gamma_D$) when kinetic mixing is at 0.05. The number of observed events in the data card is set to be equal to the expected number of events.

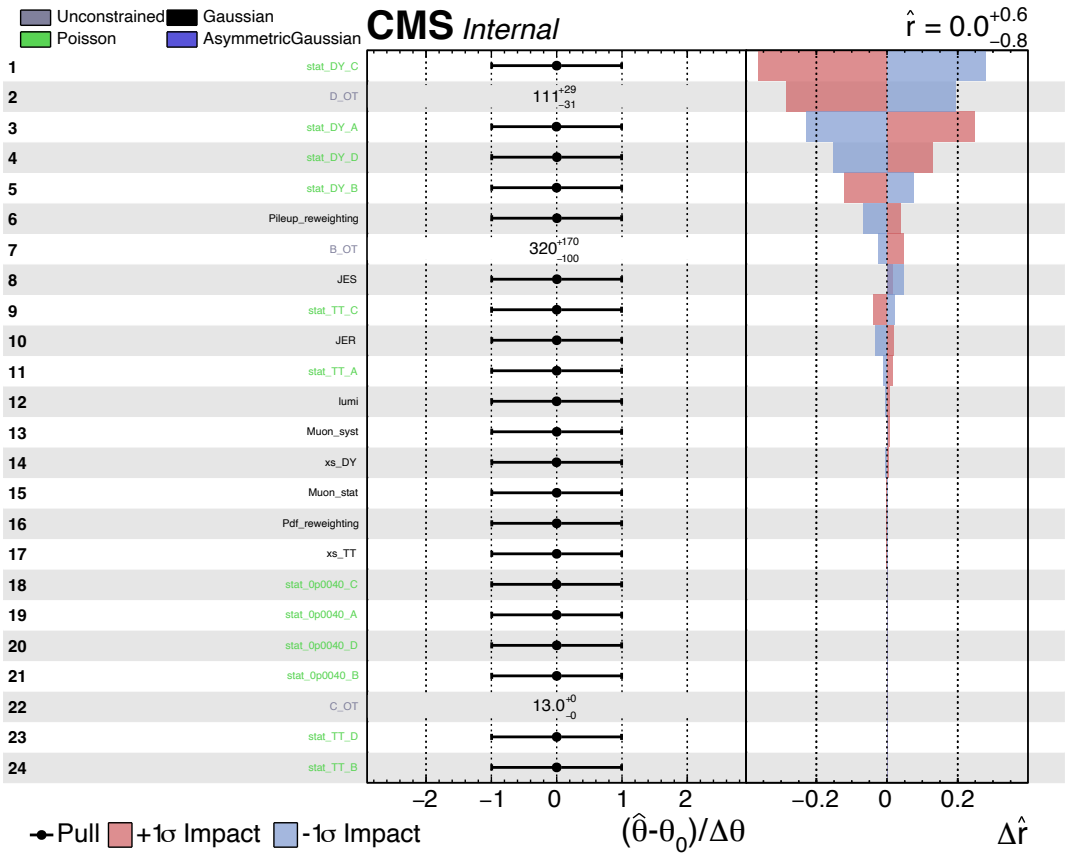


Figure 6.5: Impact of nuisance parameters to the parameter of interest (cross section of the process $pp \to \mu^+\mu^-\gamma_D$) when kinetic mixing is at 0.063. The number of observed events in the data card is set to be equal to the expected number of events.

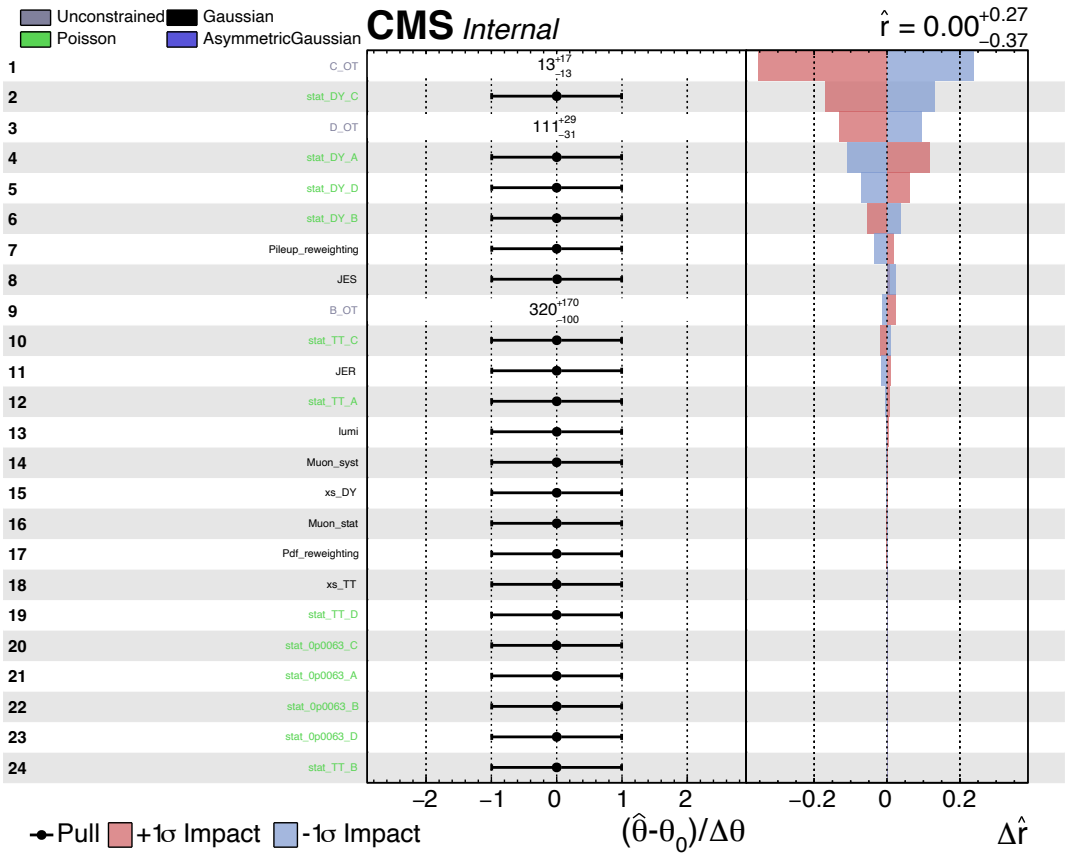


Figure 6.6: Impact of nuisance parameters to the parameter of interest (cross section of the process $pp \to \mu^+\mu^-\gamma_D$) when kinetic mixing is at 0.079. The number of observed events in the data card is set to be equal to the expected number of events.

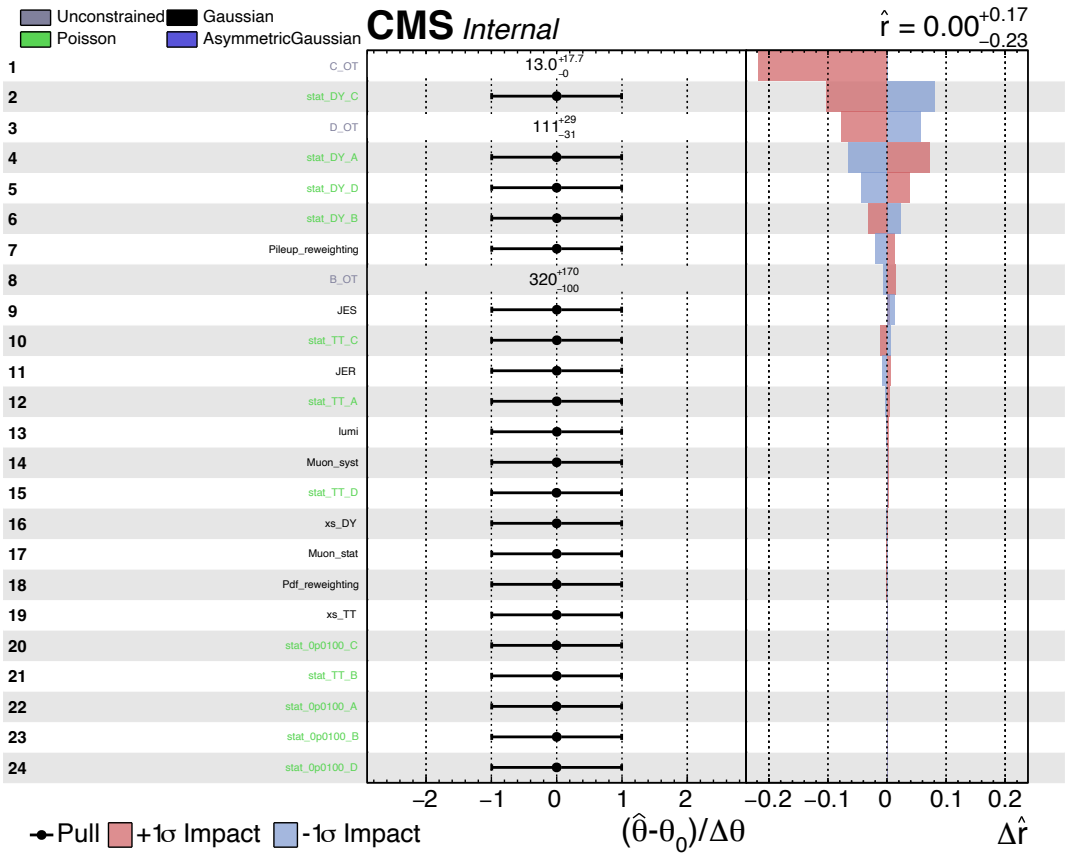


Figure 6.7: Impact of nuisance parameters to the parameter of interest (cross section of the process $pp \to \mu^+\mu^-\gamma_D$) when kinetic mixing is at 0.1. The number of observed events in the data card is set to be equal to the expected number of events.

```
# Simple counting experiment, with one signal and a few background processes
# An ABCD method is used to estimate backgrounds from HCAL Noise
imax 4 number of channels
jmax 3 number of backgrounds
kmax * number of nuisance parameters (sources of systematical uncertainties)
shapes * * FAKE
-------------------
bin             A     B     C     D
observation     114   1133  27    305
-------------------
bin                          A                 B                 C                 D
process                      0p0010            0p0010            0p0010            0p0010
process                      0                 0                 0                 0
rate                         3.270             0.4148            3.024             0.5337
-------------------
Muon_stat lnN                0.9976/1.002      0.9989/1.001      0.9971/1.003      0.9985/1.002
Muon_syst lnN                0.9929/1.007      0.9962/1.004      0.9924/1.008      0.9961/1.004
Pdf_reweighting lnN          0.9997/1.000      0.9995/1.000      0.9998/1.000      0.9992/1.001
Pileup_reweighting lnN       0.9940/1.007      0.9887/1.004      0.9983/1.006      0.9866/1.012
JER lnN                      0.9431/1.032      0.9793/1.051      0.9980/1.004      1.000/1.006
JES lnN                      0.8274/1.128      0.8854/1.135      0.9761/1.021      0.9939/1.013
lumi lnN                     0.9750/1.025      0.9750/1.025      0.9750/1.025      0.9750/1.025
xs_DY lnN                    -                 -                 -                 -
xs_TT lnN                    -                 -                 -                 -
stat_0p0010_A gmN 963        0.003396          -                 -                 -
stat_0p0010_B gmN 122        -                 0.003400          -                 -
stat_0p0010_C gmN 891        -                 -                 0.003394          -
stat_0p0010_D gmN 156        -                 -                 -                 0.003421
stat_DY_A gmN 47             -                 -                 -                 -
stat_TT_A gmN 5              -                 -                 -                 -
stat_DY_B gmN 516            -                 -                 -                 -
stat_TT_B gmN 3              -                 -                 -                 -
stat_DY_C gmN 7              -                 -                 -                 -
stat_TT_C gmN 1              -                 -                 -                 -
stat_DY_D gmN 102            -                 -                 -                 -
stat_TT_D gmN 1              -                 -                 -                 -

A_OT rateParam A OT (@0*@1/@2) B_OT,C_OT,D_OT
B_OT rateParam B OT 318.352 [63.670,642]
C_OT rateParam C OT 10.106 [2.021,31]
D_OT rateParam D OT 110.777 [22.155,227]
```

bin	A	A	A	B	B	B
process	DY	TT	OT	DY	TT	OT
process	1	2	3	1	2	3
rate	75.317	1.287	1.000	813.856	0.3776	1.000
Muon_stat lnN	0.9975/1.002	0.9911/1.009	-	0.9993/1.001	0.9984/1.002	-
Muon_syst lnN	0.9950/1.005	0.9974/1.003	-	0.9970/1.003	0.9977/1.002	-
Pdf_reweighting lnN	0.9987/1.001	0.9983/1.002	-	0.9986/1.001	0.9941/1.006	-
Pileup_reweighting lnN	0.9275/1.062	0.9401/1.271	-	0.9806/1.020	0.001000/1.658	-
JER lnN	0.9297/0.9995	1.000/1.000	-	0.9625/1.001	1.000/1.000	-
JES lnN	0.8380/1.194	0.7439/1.000	-	0.8044/1.100	1.000/1.000	-
lumi lnN	0.9750/1.025	0.9750/1.025	-	0.9750/1.025	0.9750/1.025	-
xs_DY lnN	0.9798/1.020	-	-	0.9798/1.020	-	-
xs_TT lnN	-	0.9485/1.036	-	-	0.9485/1.036	-
stat_0p0010_A gmN 963	-	-	-	-	-	-
stat_0p0010_B gmN 122	-	-	-	-	-	-
stat_0p0010_C gmN 891	-	-	-	-	-	-
stat_0p0010_D gmN 156	-	-	-	-	-	-
stat_DY_A gmN 47	1.602	-	-	-	-	-
stat_TT_A gmN 5	-	0.2575	-	-	-	-
stat_DY_B gmN 516	-	-	-	1.577	-	-
stat_TT_B gmN 3	-	-	-	-	0.1259	-
stat_DY_C gmN 7	-	-	-	-	-	-
stat_TT_C gmN 1	-	-	-	-	-	-
stat_DY_D gmN 102	-	-	-	-	-	-
stat_TT_D gmN 1						

bin	C	C	C	D	D	D
process	DY	TT	OT	DY	TT	OT
process	1	2	3	1	2	3
rate	13.414	0.4557	1.000	193.257	0.4320	1.000
Muon_stat lnN	0.9938/1.006	0.9997/1.000	-	0.9983/1.002	0.9987/1.001	-
Muon_syst lnN	0.9968/1.003	0.9987/1.001	-	0.9964/1.004	0.9877/1.012	-
Pdf_reweighting lnN	0.9939/1.006	0.9962/1.004	-	0.9992/1.001	0.9977/1.002	-
Pileup_reweighting lnN	0.9119/1.057	0.8525/1.080	-	0.9544/1.013	0.9277/1.076	-
JER lnN	1.000/1.000	1.000/1.000	-	1.000/1.015	1.000/1.000	-
JES lnN	1.000/1.245	1.000/1.000	-	0.9707/1.059	1.000/1.000	-
lumi lnN	0.9750/1.025	0.9750/1.025	-	0.9750/1.025	0.9750/1.025	-
xs_DY lnN	0.9798/1.020	-	-	0.9798/1.020	-	-
xs_TT lnN	-	0.9485/1.036	-	-	0.9485/1.036	-
stat_0p0010_A gmN 963	-	-	-	-	-	-
stat_0p0010_B gmN 122	-	-	-	-	-	-
stat_0p0010_C gmN 891	-	-	-	-	-	-
stat_0p0010_D gmN 156	-	-	-	-	-	-
stat_DY_A gmN 47	-	-	-	-	-	-
stat_TT_A gmN 5	-	-	-	-	-	-
stat_DY_B gmN 516	-	-	-	-	-	-
stat_TT_B gmN 3	-	-	-	-	-	-
stat_DY_C gmN 7	1.916	-	-	-	-	-
stat_TT_C gmN 1	-	0.4557	-	-	-	-
stat_DY_D gmN 102	-	-	-	1.895	-	-
stat_TT_D gmN 1	-	-	-	-	0.4320	-

Figure 6.8: Higgs Combine data card when kinetic mixing is 0.032.

```
# Simple counting experiment, with one signal and a few background processes
# An ABCD method is used to estimate backgrounds from HCAL Noise
imax 4 number of channels
jmax 3 number of backgrounds
kmax * number of nuisance parameters (sources of systematical uncertainties)
shapes * * FAKE
-------------------
bin             A     B     C     D
observation     114   1133  27    305
-------------------
bin                         A                 B                 C                 D
process                     0p0100            0p0100            0p0100            0p0100
process                     0                 0                 0                 0
rate                        106.582           14.577            94.309            17.927
-------------------
Muon_stat lnN               0.9977/1.002      0.9991/1.001      0.9972/1.003      0.9986/1.001
Muon_syst lnN               0.9929/1.007      0.9965/1.003      0.9924/1.008      0.9964/1.004
Pdf_reweighting lnN         0.9998/1.000      0.9989/1.001      0.9996/1.000      0.9991/1.001
Pileup_reweighting lnN      0.9903/1.010      0.9881/1.010      0.9996/1.004      0.9847/1.022
JER lnN                     0.9563/1.037      0.9634/1.045      0.9973/1.003      1.000/1.005
JES lnN                     0.8557/1.128      0.9127/1.176      0.9819/1.019      0.9814/1.024
lumi lnN                    0.9750/1.025      0.9750/1.025      0.9750/1.025      0.9750/1.025
xs_DY lnN                   -                 -                 -                 -
xs_TT lnN                   -                 -                 -                 -
stat_DY_A gmN 47            -                 -                 -                 -
stat_TT_A gmN 5             -                 -                 -                 -
stat_DY_B gmN 516           -                 -                 -                 -
stat_TT_B gmN 3             -                 -                 -                 -
stat_DY_C gmN 7             -                 -                 -                 -
stat_TT_C gmN 1             -                 -                 -                 -
stat_DY_D gmN 102           -                 -                 -                 -
stat_TT_D gmN 1             -                 -                 -                 -
stat_0p0100_A gmN 1265      0.08425           -                 -                 -
stat_0p0100_B gmN 172       -                 0.08475           -                 -
stat_0p0100_C gmN 1121      -                 -                 0.08413           -
stat_0p0100_D gmN 211       -                 -                 -                 0.08496

A_OT rateParam A OT (@0*@1/@2) B_OT,C_OT,D_OT
B_OT rateParam B OT 304.189 [60.838,642]
C_OT rateParam C OT 0.1000 [0.02000,31]
D_OT rateParam D OT 93.384 [18.677,227]
bin                         A                 A                 A                 B                 B                 B
process                     DY                TT                OT                DY                TT                OT
process                     1                 2                 3                 1                 2                 3
rate                        75.317            1.287             1.000             813.856           0.3776            1.000
-------------------
Muon_stat lnN               0.9975/1.002      0.9911/1.009      -                 0.9993/1.001      0.9984/1.002      -
Muon_syst lnN               0.9950/1.005      0.9974/1.003      -                 0.9970/1.003      0.9977/1.002      -
Pdf_reweighting lnN         0.9987/1.001      0.9983/1.002      -                 0.9986/1.001      0.9941/1.006      -
Pileup_reweighting lnN      0.9275/1.062      0.9401/1.271      -                 0.9806/1.020      0.001000/1.658    -
JER lnN                     0.9297/0.9995     1.000/1.000       -                 0.9625/1.001      1.000/1.000       -
JES lnN                     0.8380/1.194      0.7439/1.000      -                 0.8044/1.100      1.000/1.000       -
lumi lnN                    0.9750/1.025      0.9750/1.025      -                 0.9750/1.025      0.9750/1.025      -
xs_DY lnN                   0.9798/1.020      -                 -                 0.9798/1.020      -                 -
xs_TT lnN                   -                 0.9485/1.036      -                 -                 0.9485/1.036      -
stat_DY_A gmN 47            1.602             -                 -                 -                 -                 -
stat_TT_A gmN 5             -                 0.2575            -                 -                 -                 -
stat_DY_B gmN 516           -                 -                 -                 1.577             -                 -
stat_TT_B gmN 3             -                 -                 -                 -                 0.1259            -
stat_DY_C gmN 7             -                 -                 -                 -                 -                 -
stat_TT_C gmN 1             -                 -                 -                 -                 -                 -
stat_DY_D gmN 102           -                 -                 -                 -                 -                 -
stat_TT_D gmN 1             -                 -                 -                 -                 -                 -
stat_0p0100_A gmN 1265      -                 -                 -                 -                 -                 -
stat_0p0100_B gmN 172       -                 -                 -                 -                 -                 -
stat_0p0100_C gmN 1121      -                 -                 -                 -                 -                 -
stat_0p0100_D gmN 211       -                 -                 -                 -                 -                 -
bin                         C                 C                 C                 D                 D                 D
process                     DY                TT                OT                DY                TT                OT
process                     1                 2                 3                 1                 2                 3
rate                        13.414            0.4557            1.000             193.257           0.4320            1.000
-------------------
Muon_stat lnN               0.9938/1.006      0.9997/1.000      -                 0.9983/1.002      0.9987/1.001      -
Muon_syst lnN               0.9968/1.003      0.9987/1.001      -                 0.9964/1.004      0.9877/1.012      -
Pdf_reweighting lnN         0.9939/1.006      0.9962/1.004      -                 0.9992/1.001      0.9977/1.002      -
Pileup_reweighting lnN      0.9119/1.057      0.8525/1.080      -                 0.9544/1.013      0.9277/1.076      -
JER lnN                     1.000/1.000       1.000/1.000       -                 1.000/1.015       1.000/1.000       -
JES lnN                     1.000/1.245       1.000/1.000       -                 0.9707/1.059      1.000/1.000       -
lumi lnN                    0.9750/1.025      0.9750/1.025      -                 0.9750/1.025      0.9750/1.025      -
xs_DY lnN                   0.9798/1.020      -                 -                 0.9798/1.020      -                 -
xs_TT lnN                   -                 0.9485/1.036      -                 -                 0.9485/1.036      -
stat_DY_A gmN 47            -                 -                 -                 -                 -                 -
stat_TT_A gmN 5             -                 -                 -                 -                 -                 -
stat_DY_B gmN 516           -                 -                 -                 -                 -                 -
stat_TT_B gmN 3             -                 -                 -                 -                 -                 -
stat_DY_C gmN 7             1.916             -                 -                 -                 -                 -
stat_TT_C gmN 1             -                 0.4557            -                 -                 -                 -
stat_DY_D gmN 102           -                 -                 -                 1.895             -                 -
stat_TT_D gmN 1             -                 -                 -                 -                 0.4320            -
stat_0p0100_A gmN 1265      -                 -                 -                 -                 -                 -
stat_0p0100_B gmN 172       -                 -                 -                 -                 -                 -
stat_0p0100_C gmN 1121      -                 -                 -                 -                 -                 -
stat_0p0100_D gmN 211       -                 -                 -                 -                 -                 -
```

Figure 6.9: Higgs Combine data card when kinetic mixing is 0.1.

6.2 Observed limits

The expected number of events in the SR is 114 ± 26(stat) events, among which about 75 events are contributed by Drell-Yan process by simulation and about 39 events are contributed by HCAL noise leakage that is estimated by the ABCD method. We observe 124 events in the SR, hence no obvious deviation from expected backgrounds is seen.

A 95% upper limit is provided as shown in Figure 6.10 on the cross sections of the $pp \to \mu^+\mu^-\gamma_D$ process as a function of kinetic mixing parameter (square root of dark factor) from 0.032 to 0.1. Based on the observed limit and post-fit expected limits, we are able to exclude dark photon kinetic mixing down to 0.063 with 95%CL.

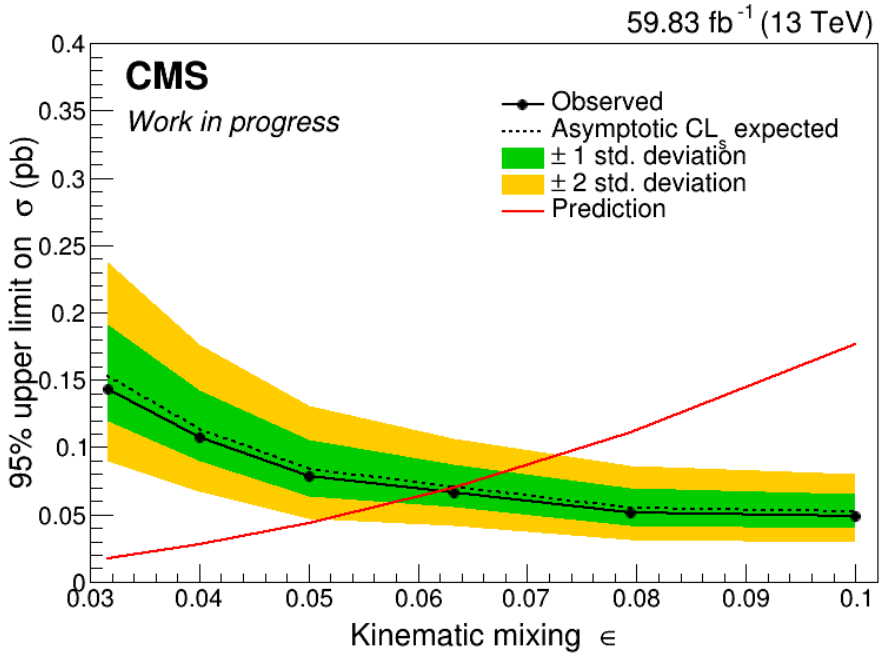


Figure 6.10: Post-fit 95% upper limit on the cross section of the $pp \to \mu^+\mu^-\gamma_D$ process as a function of kinetic mixing from 0.031 to 0.1. The observed limit is very close to the medium value of the expected limits.

The impact of nuisance parameters is shown in Figure 6.11 to Figure 6.16 that the kinetic mixing is from 0.032 to 0.1. The fits are performed with observed number of events in the signal region.

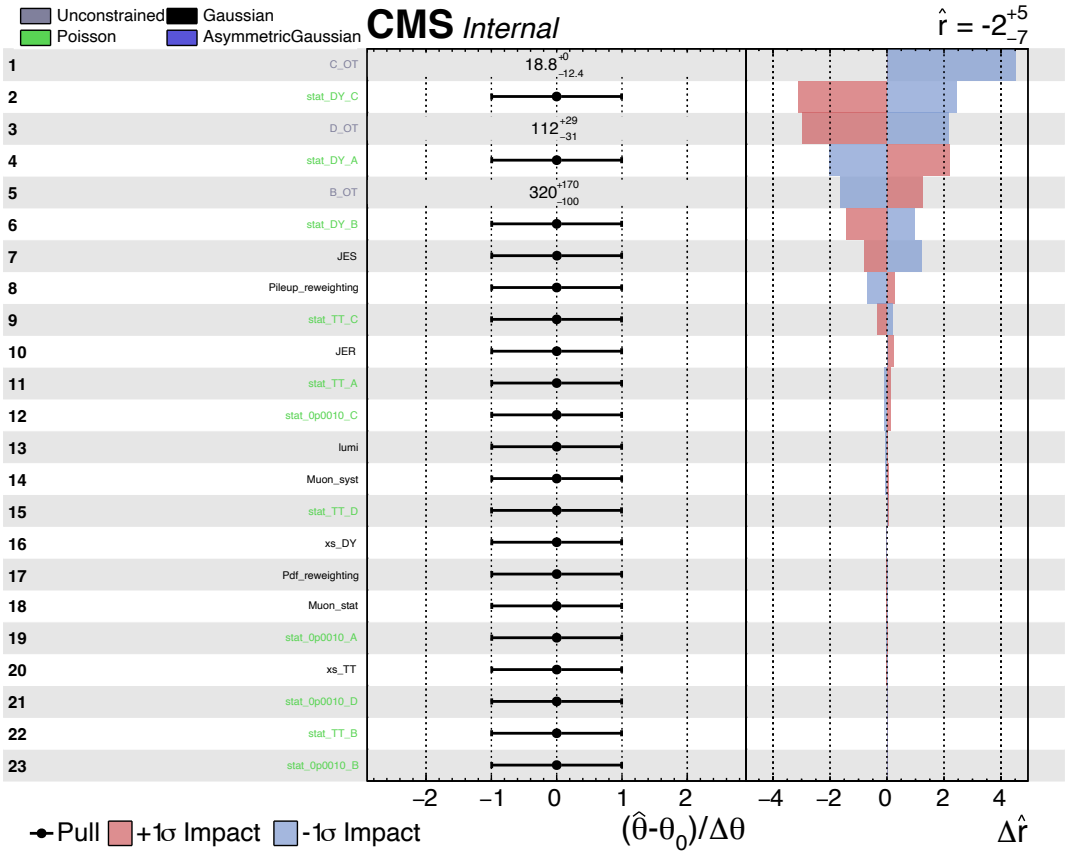


Figure 6.11: Post-fit Impact of nuisance parameters to the parameter of interest (cross section of the process $pp \to \mu^+\mu^-\gamma_D$) when kinetic mixing is at 0.032.

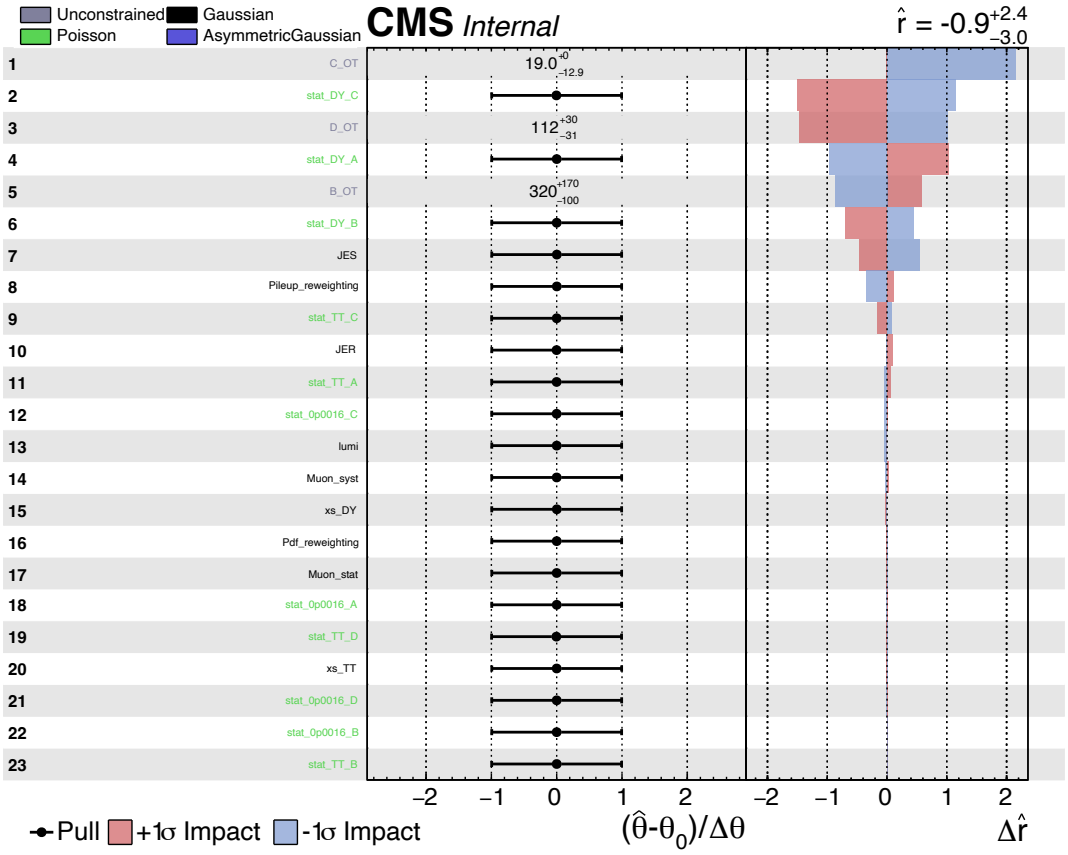


Figure 6.12: Post-fit Impact of nuisance parameters to the parameter of interest (cross section of the process $pp \to \mu^+\mu^-\gamma_D$) when kinetic mixing is at 0.04.

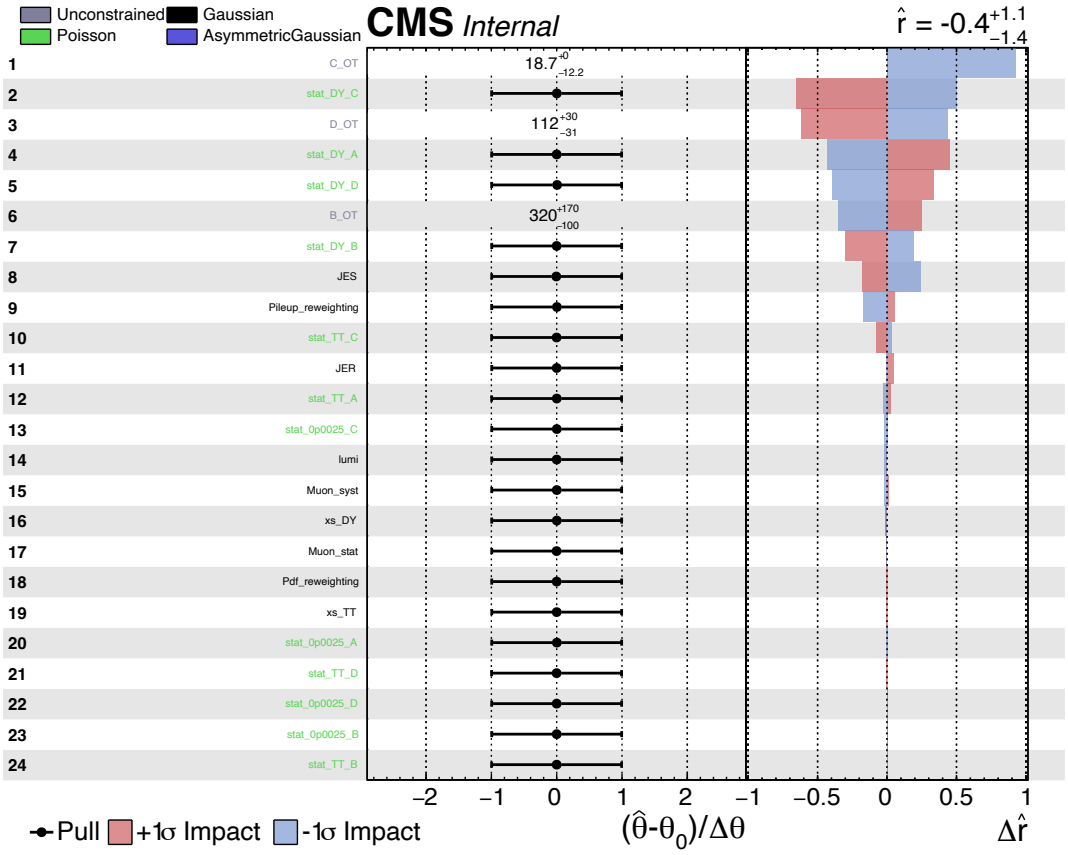


Figure 6.13: Post-fit Impact of nuisance parameters to the parameter of interest (cross section of the process $pp \to \mu^+\mu^-\gamma_D$) when kinetic mixing is at 0.05.

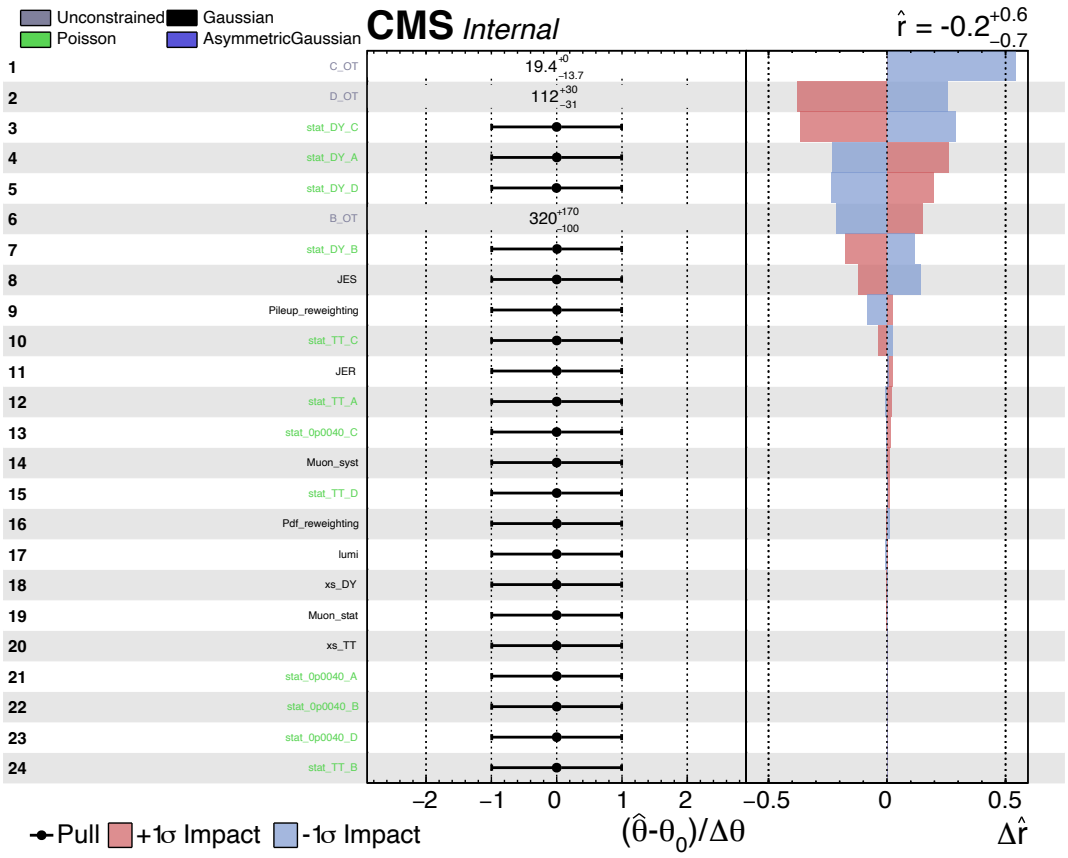


Figure 6.14: Post-fit Impact of nuisance parameters to the parameter of interest (cross section of the process $pp \to \mu^+\mu^-\gamma_D$) when kinetic mixing is at 0.063.

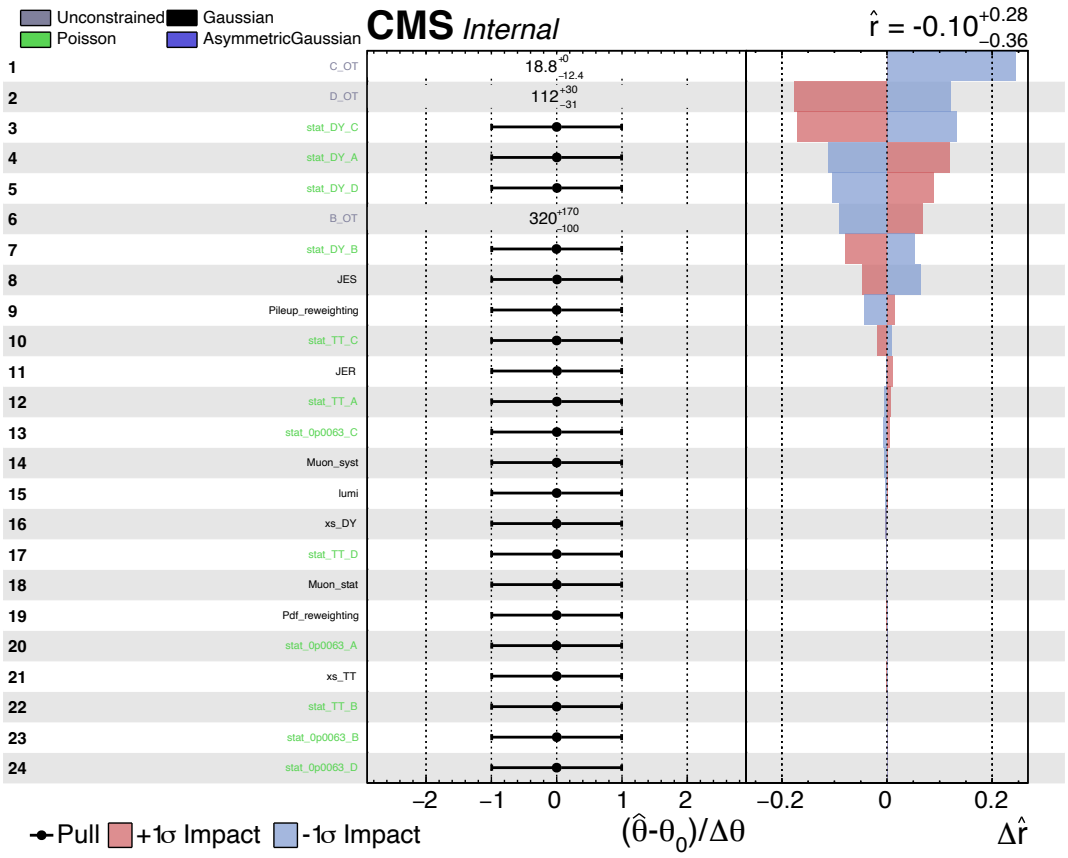


Figure 6.15: Post-fit Impact of nuisance parameters to the parameter of interest (cross section of the process $pp \to \mu^+\mu^-\gamma_D$) when kinetic mixing is at 0.079.

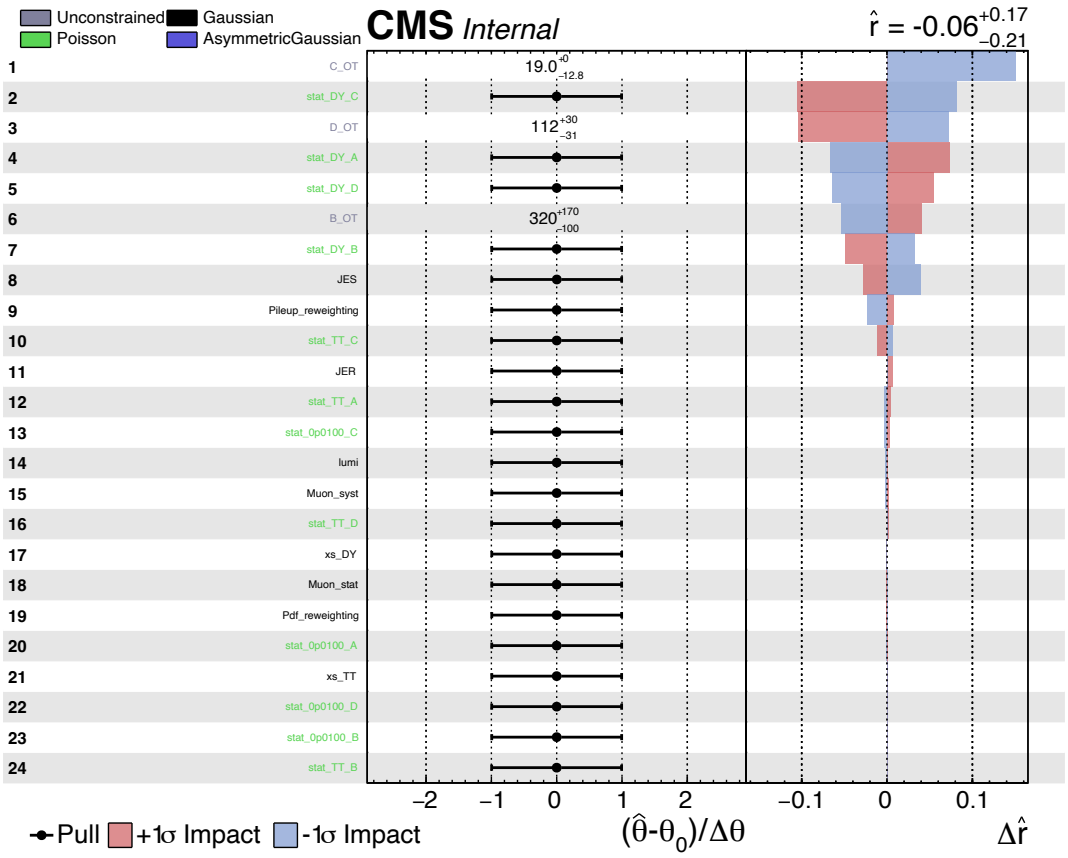


Figure 6.16: Post-fit Impact of nuisance parameters to the parameter of interest (cross section of the process $pp \to \mu^+\mu^-\gamma_D$) when kinetic mixing is at 0.1.

6.3 Conclusions

This is the first search for an ultra-low mass dark photon ($M_{\gamma_D} < 1$MeV) in proton-proton collisions using the CMS detector. A dark photon kinetically mixes with a SM photon with a small mixing ϵ, so a dark photon weakly couples to charged SM particles. A dark photon with mass lower than 1 MeV cannot decay to a pair of oppositely charged electrons, so it is stable. Such a stable dark photon can interact with detector materials through a virtual pair production that has a much smaller cross section compared to a SM photon's pair production, affected by ϵ^2 which we define as dark factor. It implies that a dark photon can penetrate farther in the detector before it showers in calorimeters, and once it has pair-produced starts, the shower develops in a manner which is the same as a SM photon's shower.

In this search, we look for a dark photon radiated from a muon. We select events with at least two muons, and the two muons and the radiated dark photon should has an invariant mass close to a Z resonance. The kinematics of the final state objects is independent on the mass of the dark photon since a dark photon is at least 100 times lighter than the muon that it is radiated from. Therefore, the search applies to dark photon mass lower than 1 MeV. On the other hand, the range of kinetic mixing that we can search for is restricted by the cross section of the dark photon radiated from a muon as well as a dark photon's virtual pair production cross section. After studying simulated signal events, we concluded that the sensitive range of kinetic mixing will be between 0.032 to 0.32, which corresponds to dark factor between 0.001 and 0.1. This is the range in which we could still find some

dark photons that shower in the HCAL with the integrated luminosity in Run II. Compared to the other experiments that also search for a dark photon as shown in Figure 1.4 and 1.5, we are the first collider experiment that has placed a limit on the mixing parameter, for a dark photon mass under 1 GeV, and we do not have a lower limit on the mass in the kinetic mixing range that we are searching for. Note that if a dark photon has mass below 10^{-22} eV, it is not massive enough to explain all the dark matter mass in the universe.

A dark photon is reconstructed as neutral energy in the HCAL and is clustered in a `AK4PFCHS` jet. Therefore, we look for a jet that is dominated by a neutron hadron together with two muons. The contribution to background events is mostly from SM processes that contain two muons and extra jets or pileup jets. HCAL electronic noise from HPD boxes is the second major source of background events. A forest of Boosted Decision Trees was trained to distinguish a dark photon jet from light flavored hadronic jets and pileup jets. The HCAL noise was estimated with a data-driven ABCD method, which was validated in a side band region. When applying the ABCD method, we reveal the control region that has the same selections as the signal region but fails the `HBHENoiseIsoFilter`. This region has very small amount of observed events and is much smaller than the signal contamination when kinetic mixing is higher than 0.1. Therefore, the fits are only performed when kinetic mixing is smaller than 0.1 for this study.

The systematic uncertainties from integrated luminosity, PDF distributions, muon reconstruction and triggers, pileup, jet energy scale and resolution are considered. We fit the signal region simultaneously with the control regions for the HCAL noise estimation, and

it is indicated in Section 6.1. Due to the signal contamination in the control region in the ABCD method, part of the kinetic mixing range is already excluded before opening the box for the signal region.

After opening the box, we observed 124 events in the SR, which agreed with the expected number of events within its statistical uncertainty. The post-fit results exclude dark photon kinetic mixing down to 0.063 with 95%CL, which corresponds to dark factor of 0.004. This result is independent of the is dark photon mass. The effect of dark photon mass on the cross section of the process is negligible due to the big center-of-mass energy of proton collisions compared to the dark photon mass. The effect of dark photon mass on the kinematics of the FSR process is also negligible due to the fact that the transverse momentum of the dark photon is much larger than the dark photon mass. Therefore, the limit we set applies to all dark photon masses below 1 MeV.

Figure 6.17 displays our limit (blue band) in the context of worldwide results. While this exclusion is not competitive in the mass range above 10^{-18} eV, it does extend the investigation into mass ranges below that value. Further, this work has established a unique method for searching for dark photons at hadron colliders. Future searches, while enjoying the absence of HCAL noise in the data, will find this work to be a highly useful resource in their efforts.

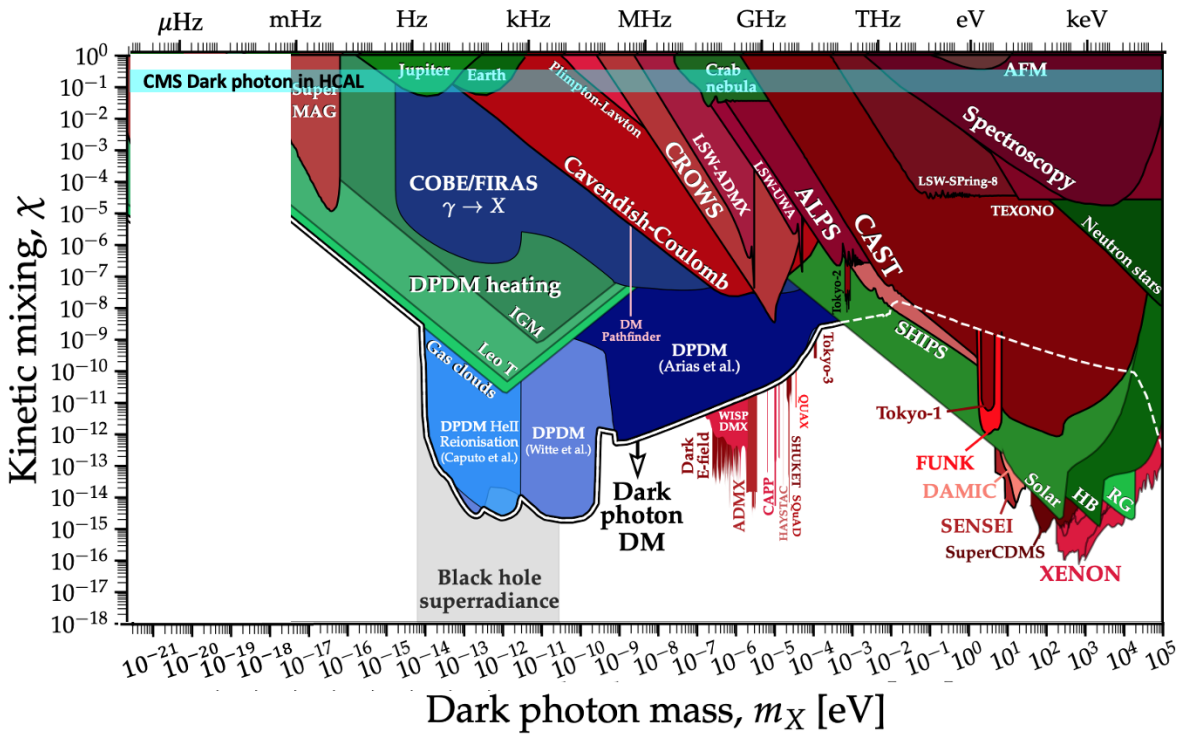
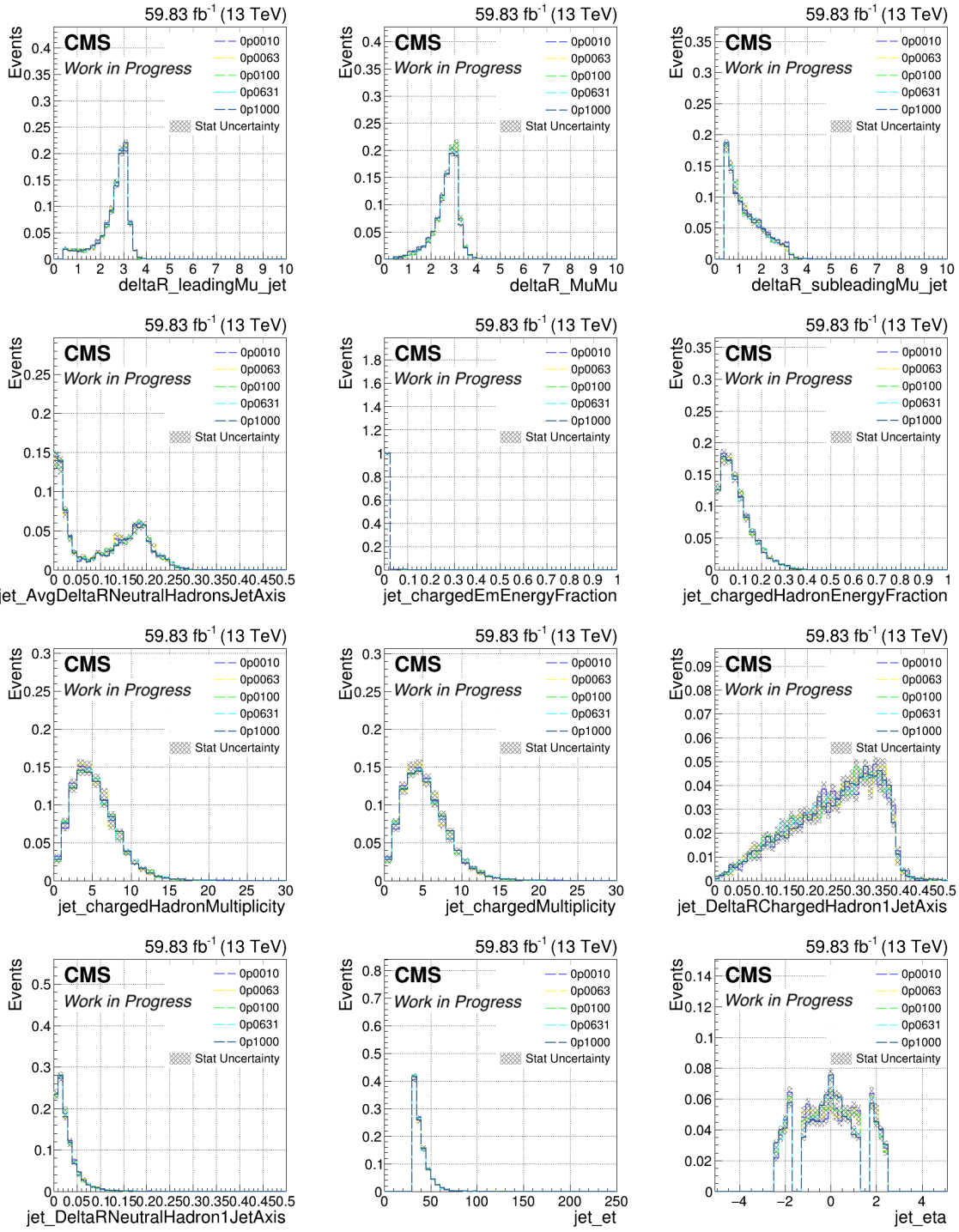


Figure 6.17: The search set a limit on dark photon with kinetic mixing up to 0.32 and down to 0.063, and mass lower than 1 MeV.

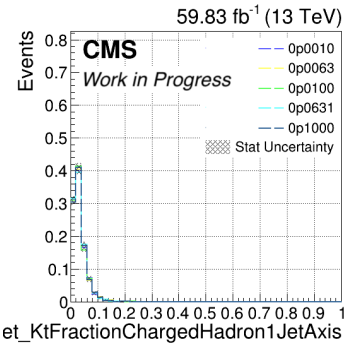
Appendix A

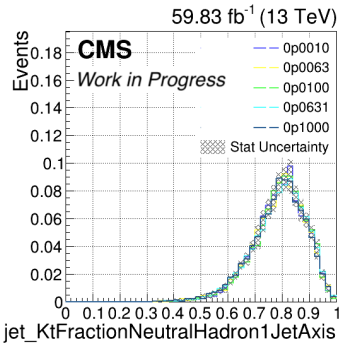
Reconstructed Dark Photon Jet Properties

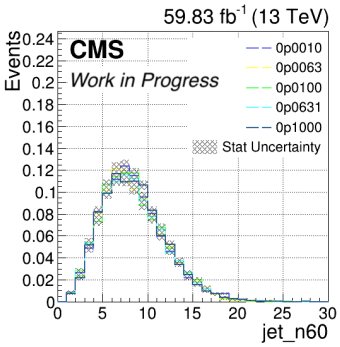
A. RECONSTRUCTED DARK PHOTON JET PROPERTIES

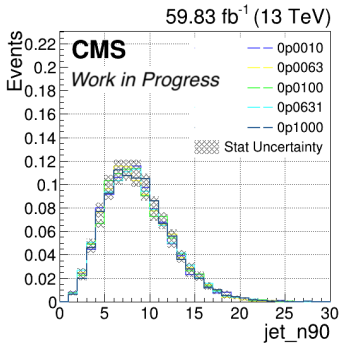


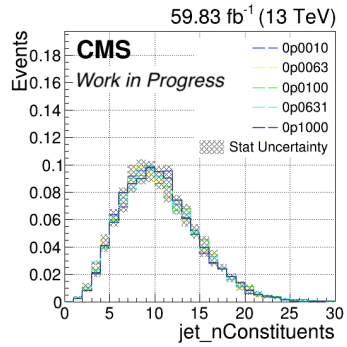
A. RECONSTRUCTED DARK PHOTON JET PROPERTIES

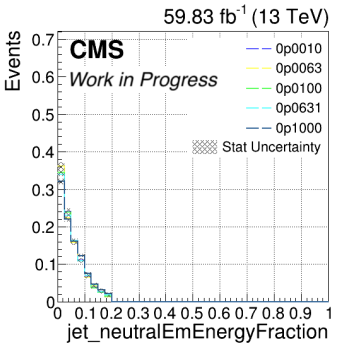


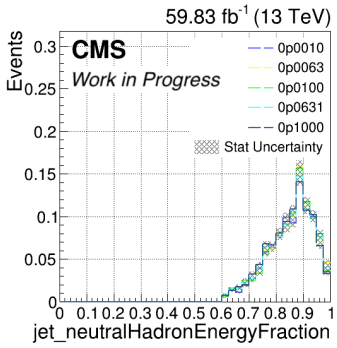


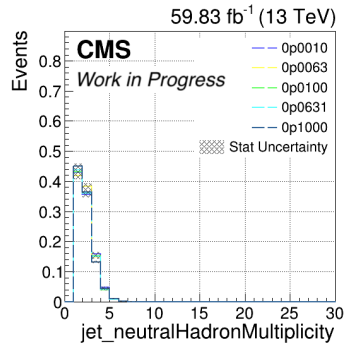


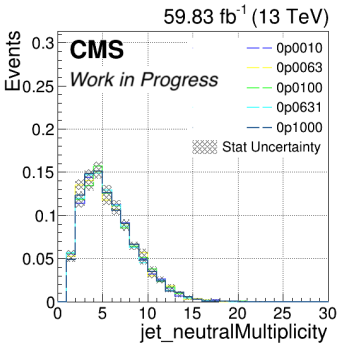


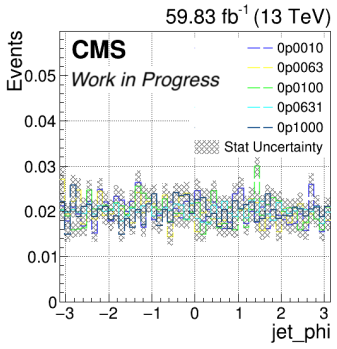


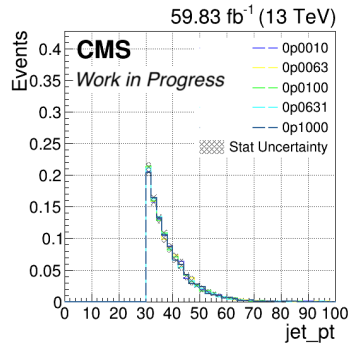


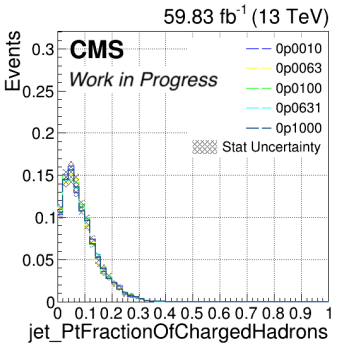












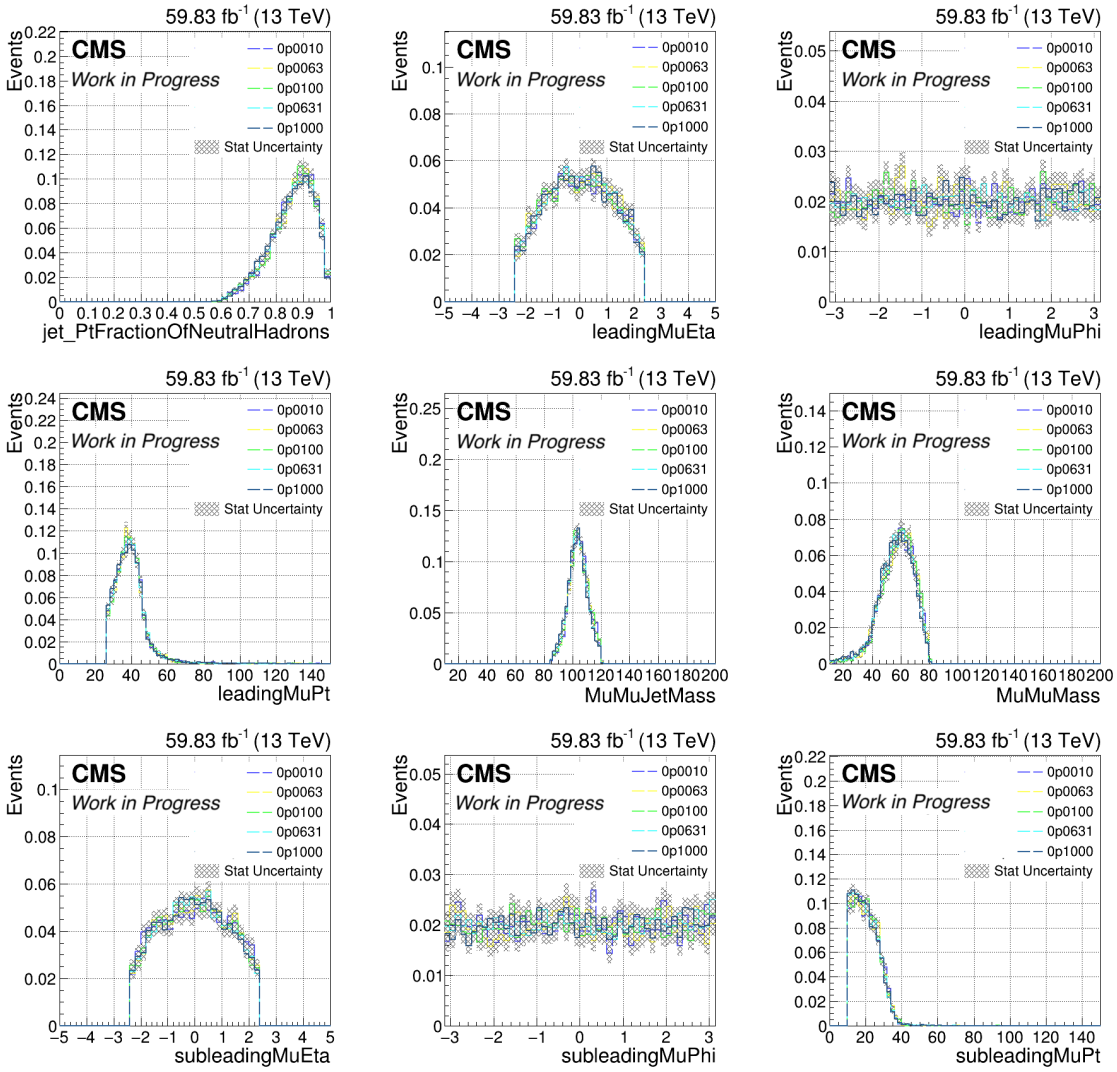
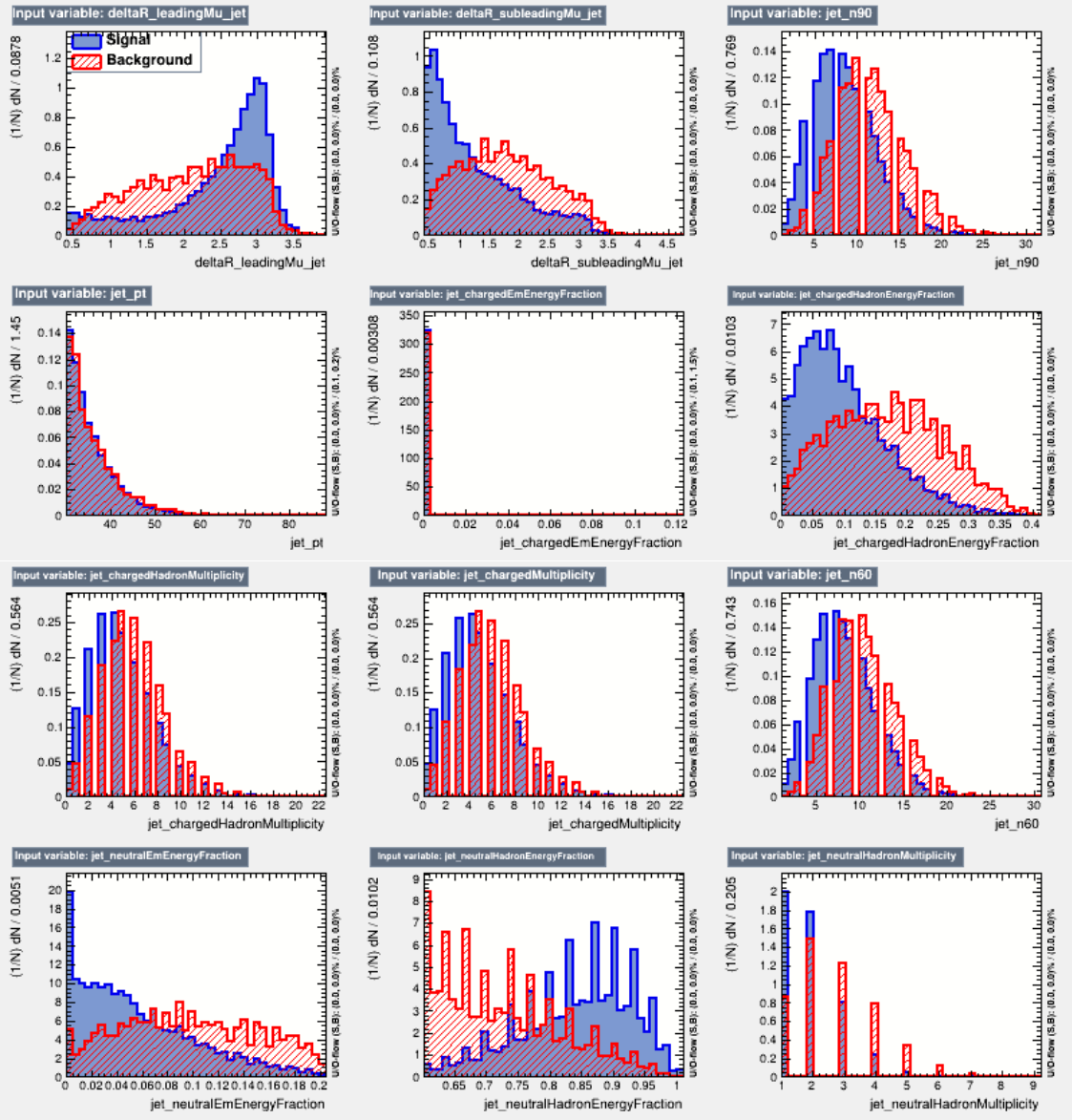


Figure A.1: Dark photon jets basic kinematics and input variables for BDT training listed in Table 4.2. The events have been selected as described in Section 4.5 and Table 4.5 as BDT_{Signal}. It proves that the kinematic of dark photon jets are similar with different dark factors.

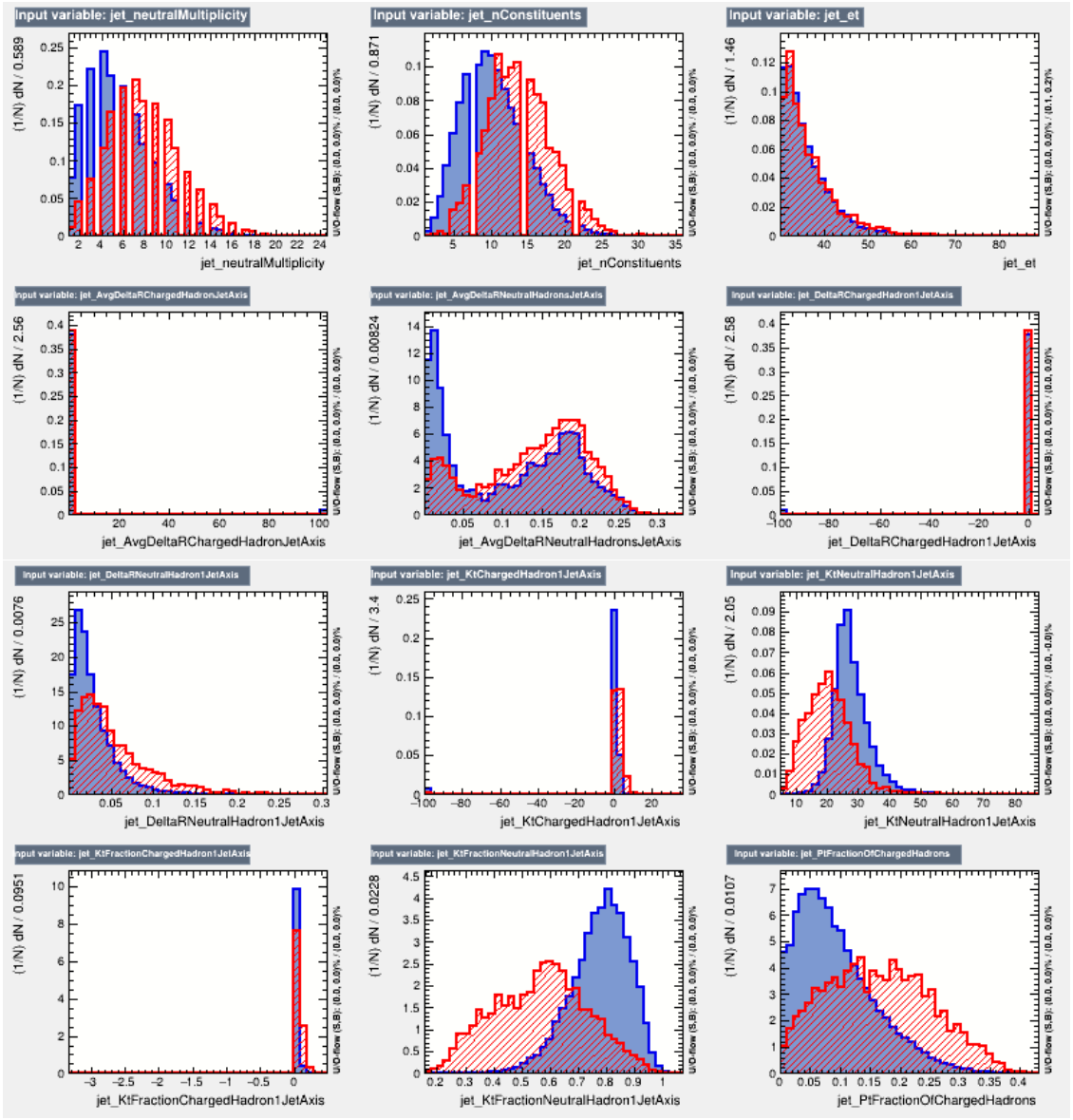
Appendix B

BDT Input Variables

B. BDT INPUT VARIABLES



B. BDT INPUT VARIABLES



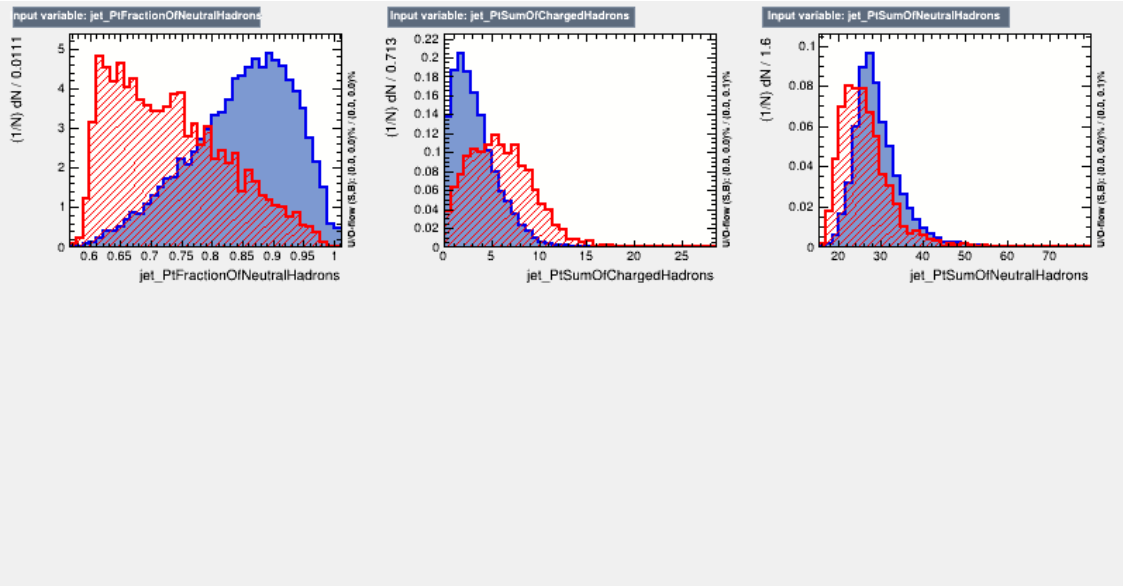


Figure B.1: BDT input variables' distributions used for the BDT training. Signal events are plotted as blue, while background events are plotted as red. The corresponding selections are listed in Table 4.5.

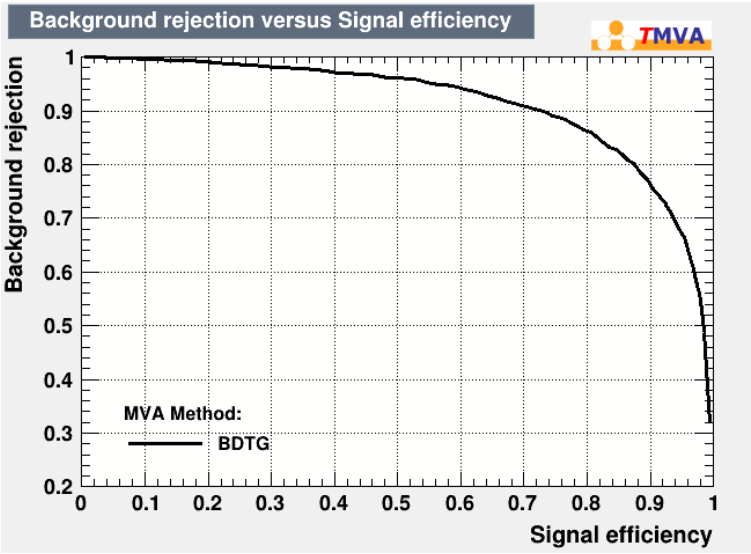


Figure B.2: BDT ROC curve.

```
Ranking input variables (method unspecific)...
       : Ranking result (top variable is best ranked)
-----------------------------------------------------------
Rank : Variable                              : Separation
-----------------------------------------------------------
   1 : jet_KtFractionNeutralHadron1JetAxis : 4.025e-01
   2 : jet_PtFractionOfNeutralHadrons      : 3.055e-01
   3 : jet_KtNeutralHadron1JetAxis         : 3.054e-01
   4 : jet_neutralHadronEnergyFraction     : 2.940e-01
   5 : jet_PtSumOfChargedHadrons           : 1.883e-01
   6 : jet_PtFractionOfChargedHadrons      : 1.823e-01
   7 : jet_chargedHadronEnergyFraction     : 1.708e-01
   8 : jet_KtChargedHadron1JetAxis         : 1.592e-01
   9 : jet_n90                             : 1.515e-01
  10 : jet_nConstituents                   : 1.496e-01
  11 : jet_neutralEmEnergyFraction         : 1.490e-01
  12 : jet_KtFractionChargedHadron1JetAxis : 1.431e-01
  13 : deltaR_subleadingMu_jet             : 1.416e-01
  14 : jet_neutralMultiplicity             : 1.338e-01
  15 : jet_DeltaRNeutralHadron1JetAxis     : 1.335e-01
  16 : jet_neutralHadronMultiplicity       : 1.234e-01
  17 : jet_PtSumOfNeutralHadrons           : 1.183e-01
  18 : jet_n60                             : 1.054e-01
  19 : deltaR_leadingMu_jet                : 1.040e-01
  20 : jet_AvgDeltaRNeutralHadronsJetAxis  : 7.751e-02
  21 : jet_chargedMultiplicity             : 4.945e-02
  22 : jet_chargedHadronMultiplicity       : 4.794e-02
  23 : jet_chargedEmEnergyFraction         : 4.032e-02
  24 : jet_DeltaRChargedHadron1JetAxis     : 7.263e-03
  25 : jet_AvgDeltaRChargedHadronJetAxis   : 7.190e-03
  26 : jet_pt                              : 6.543e-03
  27 : jet_et                              : 6.268e-03
```

Figure B.3: BDT input variables importance ranking.

Bibliography

[1] E. Corbelli and P. Salucci. "The extended rotation curve and the dark matter halo of M33". In: *Monthly Notices of the Royal Astronomical Society* 311.2 (Jan. 2000). Available at https://doi.org/10.1046%2Fj.1365-8711.2000.03075.x, pp. 441–447.

[2] P. Young et al. "The double quasar Q0957+561 A, B: a gravitational lens image formed by a galaxy at $z = 0.39$". In: *The Astrophysical Journal* 241 (Oct. 1980), pp. 507–520.

[3] *Discoveries - Highlights — Shining a Light on Dark Matter*. Available at https://www.nasa.gov/content/discoveries-highlights-shining-a-light-on-dark-matter.

[4] C. L. Bennett et al. "Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results". In: *The Astrophysical Journal Supplement Series* 208.2 (Sept. 2013). Available at https://doi.org/10.1088%2F0067-0049%2F208%2F2%2F20, p. 20.

[5] Y. Akrami et al. "Planck 2018 results. VII. Isotropy and Statistics of the CMB". In: *Astronomy and Astrophysics* 641 (Sept. 2020). Available at https://doi.org/10.1051%2F0004-6361%2F201935201, A7.

[6] Gary Steigman. "CMB constraints on the thermal WIMP mass and annihilation cross section". In: *Physical Review D* 91.8 (Apr. 2015). Available at https://doi.org/10.1103%2Fphysrevd.91.083538.

[7] Alexander Kusenko. "Sterile neutrinos: The dark side of the light fermions". In: *Physics Reports* 481.1-2 (Sept. 2009). Available at https://doi.org/10.1016%2Fj.physrep.2009.07.004, pp. 1–28.

[8] Carlos A. Argüelles, Matheus Hostert, and Yu-Dai Tsai. "Testing New Physics Explanations of the MiniBooNE Anomaly at Neutrino Scattering Experiments". In: *Phys-*

ical Review Letters 123.26 (Dec. 2019). Available at https://doi.org/10.1103%2Fphysrevlett.123.261801.

[9] J Billard et al. "Coherent neutrino scattering with low temperature bolometers at Chooz reactor complex". In: *Journal of Physics G: Nuclear and Particle Physics* 44.10 (Aug. 2017). Available at https://doi.org/10.1088%2F1361-6471%2Faa83d0, p. 105101.

[10] M. Yoshimura, N. Sasao, and M. Tanaka. "Experimental method of detecting relic neutrino by atomic de-excitation". In: *Physical Review D* 91.6 (Mar. 2015). Available at https://doi.org/10.1103%2Fphysrevd.91.063516.

[11] A. Tumasyan et al. "Search for chargino-neutralino production in events with Higgs and W bosons using 137 fb^{-1} of proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2021.10 (Oct. 2021). Available at https://doi.org/10.1007%2Fjhep10%282021%29045.

[12] B. Arsioli and G. Polenta. "A complete sample of LSP blazars fully described in γ rays". In: *Astronomy and Astrophysics* 616 (Aug. 2018). Available at https://doi.org/10.1051%2F0004-6361%2F201832786, A20.

[13] Manimala Chakraborti, Sven Heinemeyer, and Ipsita Saha. *Improved $(g-2)_\mu$ Measurements and Supersymmetry : Implications for e^+e^- colliders*. Available at https://arxiv.org/abs/2105.06408. 2021.

[14] Francesca Chadha-Day, John Ellis, and David J. E. Marsh. *Axion Dark Matter: What is it and Why Now?* Available at https://arxiv.org/abs/2105.01406. 2021.

[15] Lev B. Leinson. "Axion mass limit from observations of the neutron star in Cassiopeia A". In: *Journal of Cosmology and Astroparticle Physics* 2014.08 (Aug. 2014). Available at https://doi.org/10.1088%2F1475-7516%2F2014%2F08%2F031, pp. 031–031.

[16] Joerg Jaeckel, Gray Rybka, and Lindley Winslow. *Report of the Topical Group on Wave Dark Matter for Snowmass 2021*. Available at https://arxiv.org/abs/2209.08125. 2022.

[17] David Curtin et al. "Illuminating Dark Photons with High-Energy Colliders". In: *JHEP* 02 (2015), p. 157. arXiv: 1412.0018 [hep-ph].

[18] Andrea Caputo et al. "Dark photon limits: a cookbook". In: (May 2021). arXiv: 2105.04565 [hep-ph].

[19] Armen Tumasyan et al. "Search for long-lived particles decaying into muon pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV collected with a dedicated high-rate data stream". In: *JHEP* 04 (2022), p. 062. arXiv: 2112.13769 [hep-ex].

[20] Oliver Sim Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Available at https://cds.cern.ch/record/782076. Geneva: CERN, 2004.

[21] *CMS Luminosity Public Results*. Available at https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[22] *Linear accelerator 2*. Available at https://home.cern/science/accelerators/linear-accelerator-2.

[23] *Proton Synchrotron Booster*. Available at https://home.cern/science/accelerators/proton-synchrotron-booster.

[24] *Proton Synchrotron*. Available at https://home.cern/science/accelerators/proton-synchrotron.

[25] G. Arnison et al. "Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 GeV". In: *Physics Letters B* 122.1 (1983). Available at https://www.sciencedirect.com/science/article/pii/0370269383911772, pp. 103–116.

[26] M. Banner et al. "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider". In: *Physics Letters B* 122.5 (1983). Available at https://www.sciencedirect.com/science/article/pii/0370269383916052, pp. 476–485.

[27] G. Arnison et al. "Experimental observation of lepton pairs of invariant mass around 95 GeV at the CERN SPS collider". In: *Physics Letters B* 126.5 (1983). Available at https://www.sciencedirect.com/science/article/pii/0370269383901880, pp. 398–410.

[28] P. Bagnaia et al. "Evidence for Z0 to e+e– at the CERN pp collider". In: *Physics Letters B* 129.1 (1983). Available at https://www.sciencedirect.com/science/article/pii/037026938390744X, pp. 130–140.

[29] *Super Proton Synchrotron*. Available at https://home.cern/science/accelerators/super-proton-synchrotron.

[30] *CERN accelerator complex layout*. Available at https://cds.cern.ch/record/2693837/files/Poster-2019-858.pdf.

[31] Mark Thomson. *Modern particle physics*. New York: Cambridge University Press, 2013.

[32] *The Large Hadron Collider*. Available at https://home.cern/science/accelerators/large-hadron-collider.

[33] *The CMS Detector*. Available at https://cms.cern/detector.

[34] *Silicon Pixels*. Available at https://cms.cern/detector/identifying-tracks/silicon-pixels.

[35] *Silicon Strips*. Available at https://cms.cern/detector/identifying-tracks/silicon-strips.

[36] Paolo Azzurri. "The CMS Silicon Strip Tracker". In: *Journal of Physics: Conference Series* 41.1 (May 2006). Available at https://dx.doi.org/10.1088/1742-6596/41/1/011, p. 127.

[37] A.M. Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector". In: *Journal of Instrumentation* 12.10 (Oct. 2017). Available at http://dx.doi.org/10.1088/1748-0221/12/10/P10003, P10003–P10003.

[38] The Tracker Group of the CMS Collaboration. *The CMS Phase-1 Pixel Detector Upgrade*. 2020. arXiv: 2012.14304 [physics.ins-det].

[39] *CMS Tracker Detector Performance Results*. Available at https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK.

[40] M. Lozano et al. "Comparison of radiation hardness of P-in-N, N-in-N, and N-in-P silicon pad detectors". In: *IEEE Transactions on Nuclear Science* 52.5 (2005), pp. 1468–1473.

[41] Walaa Elmetenawee. *CMS track reconstruction performance during Run 2 and developments for Run 3*. 2020. arXiv: 2012.07035 [physics.ins-det].

[42] "CMS: The electromagnetic calorimeter. Technical design report". In: (Dec. 1997).

[43] *The CMS hadron calorimeter project: Technical Design Report*. Technical design report. CMS. Available at https://cds.cern.ch/record/357153. Geneva: CERN, 1997.

[44] "CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter". In: (Sept. 2012). Ed. by J. Mans et al.

[45] Federico De Guio and on behalf of the CMS Collaboration. "First results from the CMS SiPM-based hadronic endcap calorimeter". In: *Journal of Physics: Conference Series* 1162.1 (Jan. 2019). Available at https://dx.doi.org/10.1088/1742-6596/1162/1/012009, p. 012009.

[46] *The Phase-2 Upgrade of the CMS Muon Detectors*. Tech. rep. Available at https://cds.cern.ch/record/2283189. Geneva: CERN, 2017.

[47] J. G. Layter. *The CMS muon project: Technical Design Report*. Technical design report. CMS. Available at https://cds.cern.ch/record/343814. Geneva: CERN, 1997.

[48] V. Khachatryan et al. "The CMS trigger system". In: *Journal of Instrumentation* 12.01 (Jan. 2017). Available at https://dx.doi.org/10.1088/1748-0221/12/01/P01020, P01020.

[49] Hale Sert. *CMS Run 2 High Level Trigger Performance. CMS High Level Trigger performance in Run 2*. Tech. rep. Available at https://cds.cern.ch/record/2797787. Geneva: CERN, 2020.

[50] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07 (2014), p. 079. arXiv: 1405.0301 [hep-ph].

[51] T. Sjöstrand, et al, and Comput. Phys.Commun. "An Introduction to PYTHIA 8.2". In: (Oct. 2015). arXiv: 1410.3012 [hep-ph].

[52] T. Sjöstrand, S. Mrenna, and P. Skands. "PYTHIA 6.4 Physics and Manual". In: *JHEP* 05 (2006), p. 026. arXiv: hep-ph/0603175v2 [hep-ph].

[53] "Parton distributions for the LHC Run II". In: *JHEP* 04 (2015), p. 040. arXiv: 1410.8849 [hep-ph].

[54] Andy Buckley et al. "LHAPDF6: parton density access in the LHC precision era". In: *The European Physical Journal C* 75.3 (Mar. 2015). Available at http://dx.doi.org/10.1140/epjc/s10052-015-3318-8.

[55] S. Agostinelli et al. "GEANTfour — a simulation toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), p. 250.

[56] M. Tanabashi et al. "Review of Particle Physics". In: *Phys. Rev. D* 98 (3 Aug. 2018). Available at https://link.aps.org/doi/10.1103/PhysRevD.98.030001, p. 030001.

[57] The CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *Journal of Instrumentation* 9.10 (Oct. 2014). Available at http://dx.doi.org/10.1088/1748-0221/9/10/P10009, P10009–P10009.

[58] A.M. Sirunyan et al. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$ =13 TeV". In: *Journal of Instrumentation* 13.06 (June 2018). Available at http://dx.doi.org/10.1088/1748-0221/13/06/P06015, P06015–P06015.

[59] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet user manual". In: *The European Physical Journal C* 72.3 (Mar. 2012). Available at http://dx.doi.org/10.1140/epjc/s10052-012-1896-2.

[60] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. "The anti-kt jet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (Apr. 2008). Available at http://dx.doi.org/10.1088/1126-6708/2008/04/063, pp. 063–063.

[61] V. Khachatryan et al. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *Journal of Instrumentation* 12.02 (Feb. 2017). Available at http://dx.doi.org/10.1088/1748-0221/12/02/P02014, P02014–P02014.

[62] V.V. Abramov et al. "Studies of the response of the prototype CMS hadron calorimeter, including magnetic field effects, to pion, electron, and muon beams". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 457.1-2 (Jan. 2001). Available at http://dx.doi.org/10.1016/S0168-9002(00)00711-7, pp. 75–100.

[63] *HCAL Noise and Noise Filters in 2016*. Available at https://indico.cern.ch/event/534040/contributions/2178680/attachments/1280427/1901837/HCALnoise_JamboreeMeeting_27May2016.pdf.

[64] "Muon HLT Performance with 2018 Data". In: (June 2018). Available at https://cds.cern.ch/record/2627469.

[65] *Pileup Jet Identification*. Tech. rep. Available at https://cds.cern.ch/record/1581583. Geneva: CERN, 2013.

[66] *Heavy flavour tagging for 13 TeV data in 2018 Ultra-Legacy reprocessing and 10_6_X MC*. Available at https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation106XUL18Additional_information.

[67] Henning Kirschenmann. *Jet performance in CMS*. Tech. rep. Available at https://cds.cern.ch/record/1627818. Geneva: CERN, Oct. 2013.

[68] A. Hoecker et al. *TMVA - Toolkit for Multivariate Data Analysis*. Available at https://arxiv.org/abs/physics/0703039. 2007.

[69] Halil Saka. "HCAL Noise Filter Algorithms in Run-2". In: (Feb. 2016).

[70] *Muon Physics Object Group*. Available at https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonPOG.

[71] *Muon recommendations for 2018 Ultra Legacy data and Monte Carlo*. Available at https://twiki.cern.ch/twiki/bin/view/CMS/MuonUL2018#Trigger_efficiency_AN1.

[72] *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep. Available at https://cds.cern.ch/record/1379837. Geneva: CERN, 2011.

[73] J. S. Conway. *Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra*. Available at https://arxiv.org/abs/1103.0354. 2011.

[74] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *The European Physical Journal C* 71.2 (Feb. 2011). Available at https://doi.org/10.1140%2Fepjc%2Fs10052-011-1554-0.