UNIVERSITY OF CALIFORNIA
Los Angeles

Search for Long-Lived Particles that Decay to Muon Pairs at 13 TeV and an Algorithm to Improve Resolution in the Muon System of the Compact Muon Solenoid

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Physics

by

William Arthur Nash

2022

ABSTRACT OF THE DISSERTATION

Search for Long-Lived Particles that Decay to Muon Pairs at 13 TeV
and an Algorithm to Improve Resolution in the
Muon System of the Compact Muon Solenoid

by

William Arthur Nash
Doctor of Philosophy in Physics
University of California, Los Angeles, 2022
Professor Jay Hauser, Chair

This thesis presents a search for new long-lived particles decaying to two muons in the Compact Muon Solenoid (CMS) detector using proton-proton collision data taken at $\sqrt{s} = 13$ TeV corresponding to 96.6 fb^{-1} of integrated luminosity during the Run 2 of the Large Hadron Collider (LHC). The search spans the full extent of the CMS detector and probes vertices displaced from the interaction point ranging from less than a millimeter to multiple meters. No evidence for beyond the standard model physics is observed. Results are interpreted in the context of both scalar and vector models for the long-lived particle. We also present a detailed description of a novel algorithm for the endcap muon system of CMS which improves muon resolution of the Level-1 Trigger to be implemented for the upcoming Run 3 data taking period.

The dissertation of William Arthur Nash is approved.

Michalis Bachtis

Zvi Bern

Huan Huang

Jay Hauser, Committee Chair

University of California, Los Angeles

2022

To my friends–

those who inspired me to go to graduate school,

and those that kept me here.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Thank you to Slava Valuev, who has continued to show me by example how to be a thorough and observant scientist. Your attention to detail has both challenged me and helped me grow as a researcher. You have relentlessly and admirably done the right thing even when it is not easy.

Thank you to my advisor Jay Hauser, who gave me the freedom to pursue my own study of Bremmstrahlung within the muon system, and was a source of both insightful feedback and physical understanding throughout my time at UCLA. I am also thankful for his hand in guiding the evolution of my thesis work.

Thank you to Robert Cousins, who provided sage advice and always made sure the complicated logistics of being a graduate student in a foreign country were taken care of. I am also indebted to him for introducing me to the wild world of statistics.

I also extend this thank you to the full displaced dimuons group who have held me to a standard of excellence and been great sources of both ideas and feedback. This group includes: Wolfgang Adam, Robert Cousins, Riju Dasgupta, Alberto Escalante, Alice Florent, Ansar Iqbal, Siddharth Hariprakash, Jay Hauser, Suchita Kulkarni, Ivan Mikulec, Mangesh Sonawane, Sebastian Templ, Slava Valuev and Claudia-Elisabeth Wulz. Together they have shown me the grit required to get something difficult done. The work presented in Chapter 4 was completed in collaboration with this group.

The work done on the CMS muon-system was performed together with Cameron Bravo, Jay Hauser, Nick McColl, and Andrew Peck. The group members were both patient with my questions and impressive in their breadth and depth of knowledge. I am lucky to have learned so much from them.

Thank you to my friends in Los Angeles who have helped me to make Los Angeles my home over the these last few years, including: Cameron Bravo, Riju Dasgupta, Robert Dorst,

Neal Donnelly, Alex Epner, Casey Fischl, David Hamilton, Ethan Julius, Tyler Lam, Alice Polyakov, Chris Schaible, Brent Stone, Maxx Tepper, and Grace Whang. I am indebted to them for maintaining my good spirits.

During my year abroad, I would like to thank the group of friends who made Geneva a welcoming place even in the middle of a pandemic. This group includes Steffie Ann, Sean Gasiorowski, Nicole Hartman, Stavros Mallios, Mason Proffitt, Melissa Quinnan, and Nathan Simpson who made my time in Geneva an enjoyable one - involving rock climbing, cross-country skiing and the "ROCs for Jocks" journal club. I feel very lucky to have met them all.

I am also thankful for those who inspired me to come here in the first place. I would like to thank my parents who helped to support me along the way (and bought me plenty of legos which may have helped). My high school teachers also helped to peak my interest in science which I am grateful for, including both Ms. Johnson and Ms. Prudholme. I would also like to thank my high school friends including Neal Donnelly, Mulu Fratkin, Thomas Hirsch, Sam Richardson, and Jake Winniman. Countless nights musing in the lair are what I believe really set off my interest in physics to begin with. I also owe a great deal of gratitude to my teachers and colleagues at Boston University without which I would not have pursued high energy physics, including: Robert Carey, Adi Foord, Joe Serigano, Larry Sulak, and Kolby Weisenburger.

This thesis includes material that is based upon work supported by the U.S. Department of Energy under Award Number DE-SC0009937.

VITA

2012	Undergraduate Research Assistant (Biophysics), Boston University.
2013	Undergraduate Research Assistant (CALICE), CERN.
2014	Undergraduate Research Assistant (ATLAS), Boston University.
2014	B.A. (Physics), Boston University.
2014–2015	Physics Assistant, Mevion Medical Systems.
2015–2016	Software Engineer I: Physics and Algorithms, Mevion Medical Systems.
2016–2018	Teaching Assistant, Physics Department, UCLA.
2017–present	Graduate Student Researcher, Physics Department, UCLA.
2017	M.S. (Physics), UCLA.

CHAPTER 1

Introduction

What is it that we humans depend on? We depend on our words... Our task is to communicate experience and ideas to others. We must strive continually to extend the scope of our description, but in such a way that our messages do not thereby lose their objective or unambiguous character... We are suspended in language in such a way that we cannot say what is up and what is down. The word "reality" is also a word, a word which we must learn to use correctly.

— Niels Bohr (1885 - 1962)

Four fundamental forces are believed to govern the full body of interactions and dynamics within our universe; the weak force, the strong force, electromagnetism, and gravity. So far, three of these forces (with the exception of gravity) have been successfully described by what is known as the standard model (SM) of physics [Tul11]. The SM itself has stood strong to ever finer inspection from the physics community for nearly the last 50 years, providing predictions that continue to fall in line with experimental results. Nevertheless, there are still many unsolved mysteries in the realm of physics that do not fit into the modern rendition of the SM − leaving room for experimenters to probe unexplored spaces looking for direction towards their resolution.

One such mystery is the nature of a substance which makes up nearly 85% of the matter density of our universe [Zyl20]. Since the 1930s [Zwi33, Smi36], discrepancies have appeared between the velocities of galaxies in galaxy clusters when they are measured either by their motion or calculated using Kepler's law after their mass is inferred through the light they emit. Galaxies were found to be moving faster than would be predicted by the observable distribution of mass alone. In the 1970s, researchers [OPY74, Ein74] proposed that this discrepancy could be due to matter which was not seen in the emitted light spectrum of

the galaxy. This unobserved matter is today referred to as 'dark matter' and is believed to hardly interact with any of the forces described by the SM. Supporting evidence for the existence of dark matter has also been seen within galaxies themselves and when accounting for the gravitational lensing effect of general relativity [Zyl20]. A widely held view of dark matter ascribes its origin to a massive 'dark' particle which lies outside the current list of particles predicted by the SM [Sal19].

The origin of the observed matter-antimatter asymmetry of our universe also lies outside the current scope of the SM. One of the most promising ways to account for this asymmetry is through the process of electroweak baryogenesis (EWBG) [MR12]. Baryons, a composite subatomic particle composed of an odd number of valence quarks (three or more), have been observed in far greater numbers than their anti-particle counterparts (anti-baryons) [Cli06, Pla20, Zyl20]. EWBG proposes that as the Universe cooled ($T \lesssim 100$ GeV) at the end of the electroweak epoch, the electroweak symmetry is broken and results in an excess of baryonic matter. The observed excess of baryons in this context motivates the need for new physics at the electroweak scale. This new physics is also thought to have potential ties to dark matter. A leading theory of this kind proposes new scalar (spin-0) particles with significant coupling to the Higgs [MR12].

The Higgs boson was first theorized in 1964 [EB64, Hig64] and found at the Large Hadron Collider (LHC) at CERN in 2012 [Col12b, Col12a]. Through the mechanism of electroweak symmetry breaking, the Higgs field gives many fundamental particles their mass – a discovery that resulted in the Nobel prize awarded in 2013. The experimental discovery of an excitation of this field, the Higgs boson, was achieved in part by reconstructing muons resulting from one of its decay channels. Muons are fundamental particles that provide a particularly clean and unambiguous signature useful for understanding and reconstructing their progenitors.

This thesis presents a search for long-lived neutral particles which decay into a pair of muons. The search is intended to be as independent of theoretical models as is possible to both reduce potential biases and maximize the covered kinematic space. Results are interpreted in the context of two benchmark models which propose solutions to some of

the outstanding questions left unanswered by the SM, such as the origins of dark matter and the matter-antimatter asymmetry. Chapter 2 contains an overview of the SM and the theoretical background and motivation for such long-lived particles. Chapter 3 then describes the Compact Muon Solenoid (CMS) detector which is used for the search. The search itself is covered in Chapter 4. Work towards an upgraded muon detector which will be used during the next data-taking run at the LHC, having potential implications for future long-lived analyses is then described in Chapter 5.

CHAPTER 2

Theoretical Motivations for Long-Lived Particles

If you have built castles in the air, your work need not be lost; that is where they should be. Now put the foundations under them.

— Henry David Thoreau (1817 - 1862)

The SM accurately describes three of the four fundamental forces believed to be the bedrock of our universe. Gravity, the exception, is the weakest of the four and therefore plays a negligible role at the level of subatomic particles [Zyl20]. This chapter details the theoretical origins of long-lived particles primarily in the context of dark matter as an extension to the SM. This chapter and those that follow use the convention that $\hbar = c = 1$.

2.1 Overview of the Standard Model

The SM posits that all matter in the universe is formed out of combinations of specific breeds of fundamental particles; leptons, quarks, and mediators. Each known fundamental particle is shown Figure 2.1. All particles are understood to be synonymous with discrete excitations of a quantum field.

2.1.1 Quantum Field Theory

The marriage of classical field theory, special relativity, and quantum mechanics was made possible with the development of Quantum Field Theory (QFT). Within the framework of QFT, observed symmetries of a physical system are translated into a Lagrangian via Noether's Theorem. The entirety of the SM is famously invariant under local gauge trans-

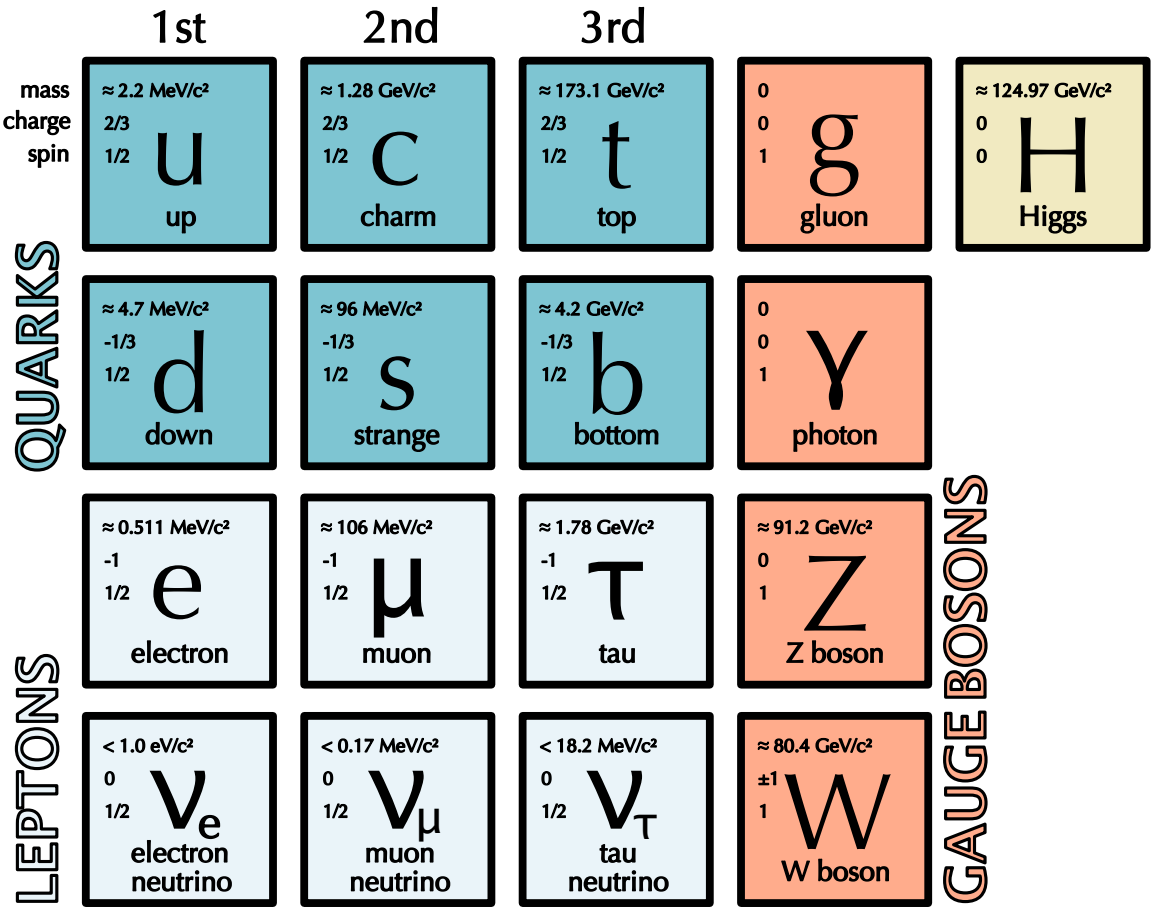


Figure 2.1: Elementary particles in the standard model grouped by type. Quarks are shown in blue and leptons in gray. Mediators are shown in orange and yellow, signifying the gauge bosons and Higgs boson respectively. Each labeled column represents the generation of the corresponding particles. Cited values are obtained from [Zyl20].

formations of the group

$$\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y. \tag{2.1}$$

Together, these three factors give rise to the three fundamental interactions: strong, weak, and electromagnetic. The first term represents the fact that the action of the SM Lagrangian is invariant under transformations of color charge belonging to the special unitary group of degree three, SU(3). Similarly, weak isospin can be freely transformed by elements of the SU(2) group. Finally, weak hypercharge Y can be transformed by elements of the U(1) group.

2.1.2 Electroweak Symmetry Breaking

Each factor in Equation 2.1 corresponds to its own gauge field tensor and results in a gauge boson for each group generator associated with the field. Each field itself is characterized by a coupling constant that measures its strength [EF20]. Electroweak symmetry breaking involves the gauge fields associated with the latter two components of Equation 2.1.

The weak isospin field contains three gauge bosons $W^i_\mu,\ i \in (1,2,3)$ with a coupling constant g, while the weak hypercharge field contains just one, B_μ, with a coupling constant g'. Each of these bosons are *vector* bosons, which is synonymous with the fact that they have spin-1. The μ subscript, therefore, corresponds to the components of the four-potential required to describe them.

Spin-1 particles are described by the Proca equation which does not allow for massive solutions that maintain the local gauge invariance needed to construct the SM. Electroweak symmetry breaking introduces an additional scalar (spin-0) field, called the Higgs field, that couples to the vector fields thereby allowing them to have non-zero masses consistent with experimental results.

However, the initial vector fields W^i_μ, B_μ become *rotated* in the presence of the Higgs field

via a process known as spontaneous symmetry breaking. The fields are rotated according to the weak mixing angle $\theta_W \equiv \tan^{-1}(g'/g)$, resulting in physical gauge fields consisting of two charged fields, $W^{\pm}$, and two neutral gauge boson fields Z and γ:

$$W_\mu^\pm \equiv \frac{1}{\sqrt{2}}\Big(W_\mu^1 \mp iW_\mu^2\Big), \tag{2.2}$$

$$Z_\mu \equiv -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W, \tag{2.3}$$

$$A_\mu \equiv B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W. \tag{2.4}$$

Three of these new fields notably *obtain a non-zero mass* with

$$M_W \propto g, \qquad M_Z \propto \sqrt{g^2 + g'^2}, \qquad M_\gamma = 0. \tag{2.5}$$

2.1.3 Particle Lifetime

The survival probability for a particle to a time t is given by an exponential probability distribution parameterized by its mean lifetime τ, with

$$\mathcal{P}(t|\tau) = e^{-t/\tau}/\tau. \tag{2.6}$$

The mean lifetime of a particle is the inverse of its decay rate Γ,

$$\tau = \frac{1}{\Gamma}. \tag{2.7}$$

A particle's decay rate, often called its *width*, is the probability per unit time that the particle will decay. Each individual channel to which the particle could decay, i, has its decay rate Γ_i summed to yield the total decay rate:

$$\Gamma = \sum_{i=1}^{n} \Gamma_i. \tag{2.8}$$

For a particle of mass M, individual decay rates can be calculated via Fermi's Golden Rule [Lee19]

$$\Gamma_i = \frac{1}{2M} \int d\Pi |\mathcal{M}(M \to \{p_f\})|^2, \tag{2.9}$$

where $\mathcal{M}$ is the matrix element which connects the initial state of the particle to that of its decay products $\{p_f\}$, and $d\Pi_f$ is the allowed Lorentz-invariant phase space for the decay. The matrix element $\mathcal{M}$ is calculated using the Feynman rules and includes the coupling strength between initial and final states.

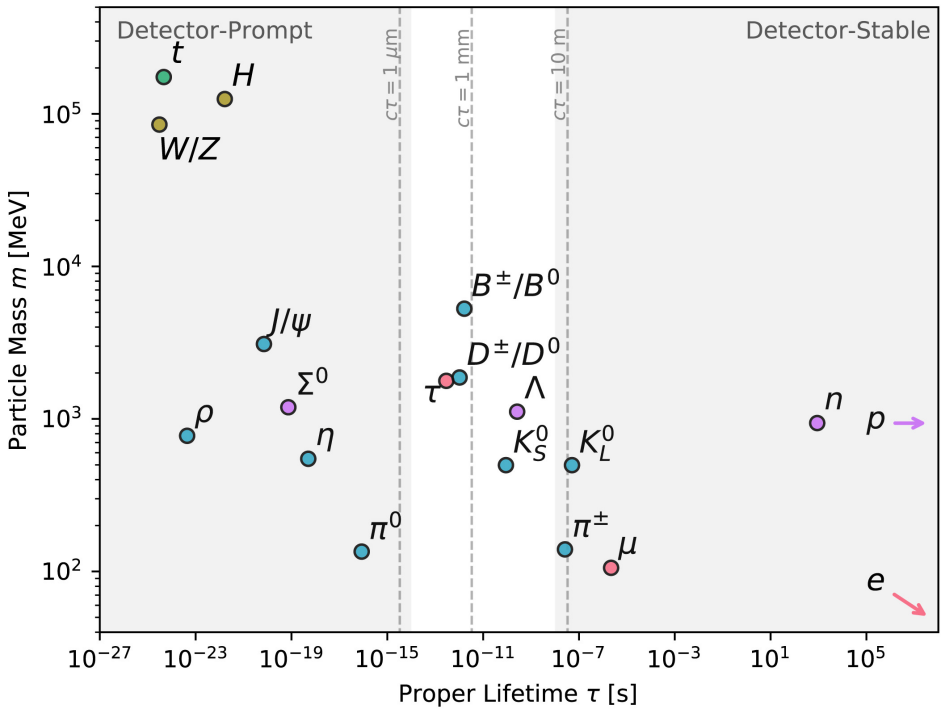


Figure 2.2: A selection of the SM particle spectrum is shown as a function of mass and proper lifetime. Shaded regions roughly represent the detector-prompt ($\tau < 1$ μm) and detector-stable ($\tau > 10$ m) regions of lifetime space for a particle moving at close to the speed of light. Reproduced from [Lee19].

Particles in the SM have lifetimes which span many orders of magnitude as shown in Figure 2.2. Lifetimes of individual particles are predominately determined by the largest individual decay rate Γ_i among their available decay channels. Particles that decay via

the strong, electromagnetic, and weak force typically have lifetimes of 10^{-23}, 10^{-16}, and $10^{-13} - 10^{3}$ seconds respectively. In general, a particle can become long-lived if there is a small matrix element and/or there is limited phase space in which the decay can take place.

2.2 Motivation for Long-Lived Particles

The high p_{T} physics program at the LHC has yet to see a clear sign of physics beyond the standard model (BSM) that could explain any of the outstanding mysteries described in Chapter 1. However, BSM long-lived particles (LLPs) remain a well-motivated and understudied area of research with signatures that can be probed at the LHC. Specifically, searches for LLPs with masses above 10 GeV, and lifetimes above 1 μm are particularly attractive, as they have no counterparts predicted by the SM (Figure 2.2).

2.2.1 Hidden Abelian Higgs Model

Hidden or "dark" sector models near the weak scale could provide explanations for both dark matter and electroweak baryogenesis while representing a general expectation of BSM physics [CEG15]. Hidden sectors in general do not interact directly with SM particles, but do so only indirectly via mixing effects. In this way, they reproduce features similar to those of dark matter.

The Hidden Abelian Higgs Model (HAHM) considers a spontaneously broken "dark" $\mathrm{U}(1)_D$ gauge symmetry which mixes with the SM and allows for particles with macroscopic lifetimes. The corresponding new field is mediated by a vector boson called the "dark photon", Z_{D}, which interacts with the SM only through a mixing with the SM hypercharge gauge boson, B. A kinetic mixing parameter ϵ connects the two fields through what is known as the *hypercharge portal*. When one adds a dark Higgs s to account for the spontaneous symmetry breaking of the $\mathrm{U}(1)_D$ group, one finds an analogous mixing to the SM Higgs h described in Section 2.1.2. Both scalar bosons are mixed via coupling parameter κ through what is known as the *Higgs portal*, and the dark photon obtains a non-zero mass $m_{Z_{\mathrm{D}}}$. These

relationships are shown in the Feynman diagram in Figure 2.3.

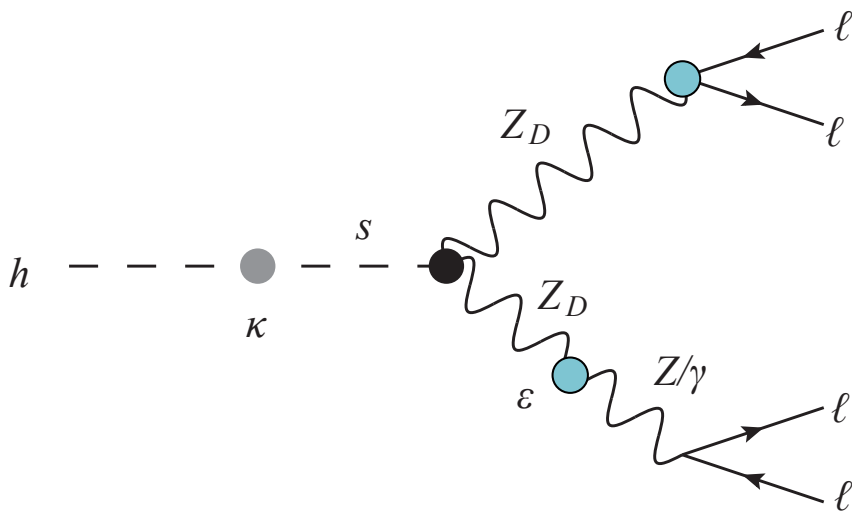


Figure 2.3: Feynman diagram of the HAHM model. The standard model Higgs h mixes with a dark Higgs s via coupling parameter κ. The dark Higgs then decays to a pair of dark photons Z_{D} which couple to the SM photon γ and Z boson via kinetic mixing parameter ϵ. The strength of ϵ determines the lifetime of Z_{D} which then decays to SM particles, here shown as pairs of leptons l.

If there are no hidden-sector states with masses smaller than $m_{Z_{\mathrm{D}}}$, the dark photon must decay exclusively to SM particles, with a significant branching ratio to leptons. Because the matrix element connecting these states is proportional to ϵ, Equation 2.9 forces the lifetime of the dark photon $\tau_{Z_{\mathrm{D}}}$ to be proportional to ϵ^{-2}. When 10 GeV $\lesssim m_{Z_{\mathrm{D}}} < m_h/2$, the dark photon is expected to have macroscopically large mean decay lengths $c\tau_{Z_{\mathrm{D}}} \gtrsim \mathcal{O}(100\ \mu\mathrm{m})$ for $\epsilon < \mathcal{O}(10^{-6})$ [Val19b]. Both the mass of the dark Higgs m_s and its mixing parameter κ do not affect the lifetime of the dark photon, only their production rate.

For $m_{Z_{\mathrm{D}}} > \mathcal{O}(1\ \mathrm{GeV})$, the Higgs portal provides the most sensitive search to the hidden sector in the long-lived regime ($\epsilon < \mathcal{O}(10^{-6})$). To date, there is only one experimental constraint on the dark photon in this regime set by the ATLAS experiment [Aab19], due to the challenging nature of the search.

2.2.2 BSM Higgs Model

LLPs could also be produced via heavy mediators. The BSM Higgs model considers an exotic scalar Higgs boson Φ which serves as a window into a new sector of particles uncharged under the SM [SZ08] and unconstrained in their lifetimes. The model adds an additional scalar boson X to the SM which may decay to muon pairs, with resulting potential having the form

$$V = -\mu^2 H^2 + \lambda H^4 + M^2 X^2 + \kappa X^4 + \zeta X^2 H^2 \tag{2.10}$$

$$+ aX + bX^3 + cXH^2. \tag{2.11}$$

If one takes $a = b = c = 0$, the potential is symmetric under the cyclic group of order two, $\mathbf{Z}_2$, with $X \to -X$. Furthermore, if $\langle X \rangle = 0$, the symmetry will not be spontaneously broken, and the X itself will be stable. This results in the hypothetical decay $\Phi \to XX$ to be invisible. If the parameters a, b, c are instead nonzero, but small, the lifetime of the X becomes finite due to the softly broken $\mathbf{Z}_2$ symmetry. If the eigenstate with the lowest mass is of the form $|X\rangle + \epsilon|h\rangle$, the decay rate for X becomes $\Gamma_X = \epsilon^2 \Gamma_h$ with branching fractions identical to those of the Higgs. The parameter ϵ itself is unconstrained, allowing for lifetimes that could span the body of the particle detectors used at the LHC. Consequently, this model was used in the Run 1 searches for displaced dimuons [Col15, Kha15].

CHAPTER 3

The CMS Detector at the LHC

That is the essence of science: ask an impertinent question, and you are on your way to the pertinent answer.

— Jacob Bronowski (1908 - 1974)

The LHC accelerates two counter-rotating beams of protons around a 27.6 km ring roughly 100 m underground at CERN in Geneva, Switzerland. These beams collide at four interaction points, home to large detectors designed to capture and decipher the collisions. CMS is one of two general-purpose detectors designed to investigate such collisions: built to find and gain an understanding of the Higgs boson and to search for exotic physics processes outside our current understanding, such as Supersymmetry, extra dimensions, and dark matter.

3.1 The Large Hadron Collider

Protons within the LHC are accelerated to a speed thousands of times their rest energy and collide in one of four locations along the ring. The LHC was designed to accelerate protons capable of colliding at a center-of-mass energy of 14 TeV [EB08] for the purpose of perturbing underlying quantum fields enough to produce their associated particles. The resulting particles are then studied to gain an understanding of the composition of our universe. The most recent Run 2 of the LHC achieved an operating center-of-mass energy of 13 TeV from 2015 to 2018. Complementary to proton acceleration, the LHC is also capable of accelerating heavy ions used in the study of quark-gluon plasma.

3.1.1 Accelerator Complex and Design

The high energies of the protons used at the LHC are achieved through a multi-stage transfer process as they are passed to progressively higher and higher energy accelerators. Protons are first produced via ionization of hydrogen gas and are then injected into the Proton Synchrotron Booster (PSB) at 50 MeV from the LINear ACcelerator 2 (LINAC 2). Within the PSB, they are further accelerated and increase their kinetic energy by nearly a factor of 30 to 1.4 GeV. The protons are then fed into the Proton Synchrotron (PS), followed by the Super Proton Synchrotron (SPS), where they increase their energy to 25 GeV then 450 GeV, respectively. From the SPS, they are finally transferred to the LHC where they are split among two separate counter-rotating rings, which are both accelerated to 6.5 TeV. The full layout of the accelerator complex at CERN is shown in Figure 3.1.

Acceleration at each stage is achieved using radio frequency (RF) cavities that are pulsed at regular intervals, synchronized with the location of the protons. This design packs protons into discrete *bunches* such that they follow the frequency pattern set by the accelerator. Within the LHC, proton bunches are separated in time by 25 ns, corresponding to a 40 MHz frequency with which a bunch will pass a single point along the accelerator.

The size of the proton bunches varies along the body of the accelerator; away from the collision points they can be a roughly 11 cm long and a millimeter wide, while close to the collision point they can shrink to about 8 cm long and 20 μm wide to increase the probability of a proton-proton collision [Bru04]. The beams themselves are steered and focused using superconducting dipole and quadrupole magnets, respectively.

3.1.2 Luminosity and Cross-Section

The intensity of the collisions of a particle beam is characterized by its instantaneous luminosity $\mathcal{L}$ (typically just called its luminosity), defined as the number of incident particles per unit area, per unit time. The number of outgoing particles produced as a result of a particle collision scales linearly with the luminosity according to

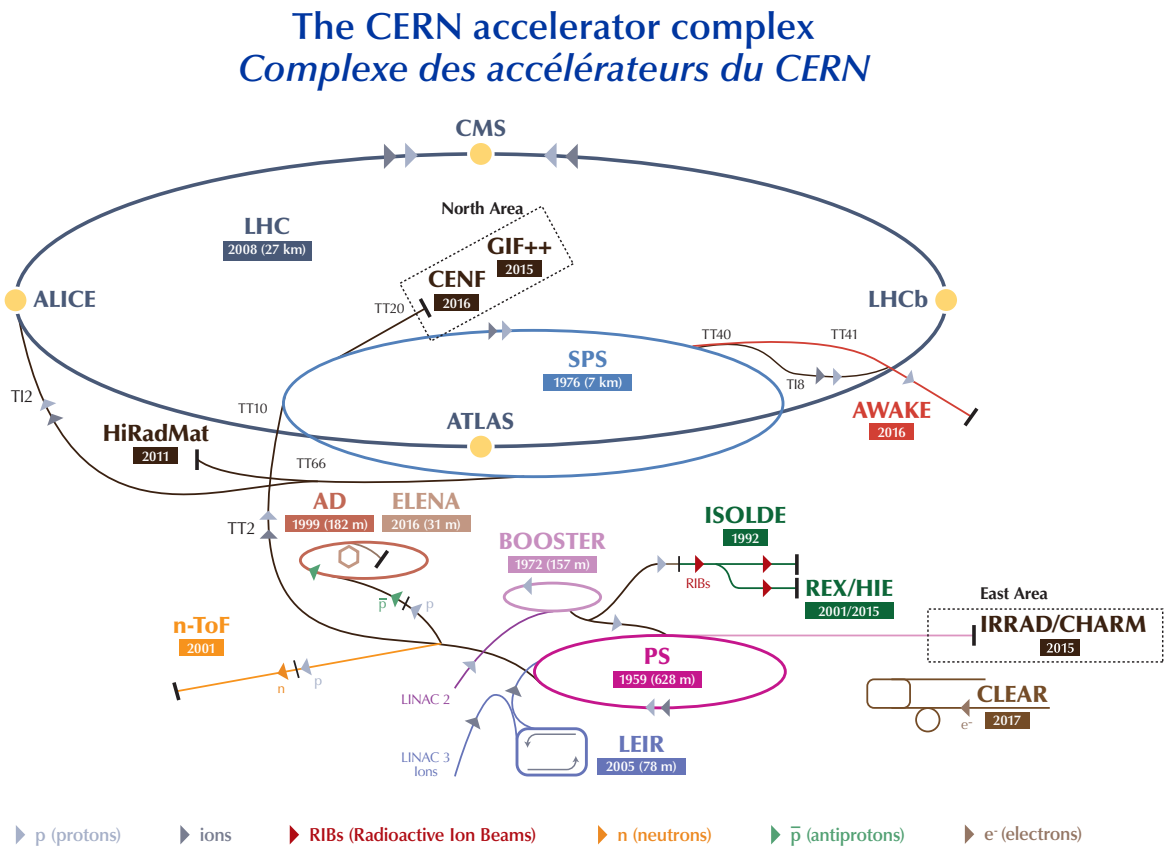


Figure 3.1: Overview of the CERN accelerator complex in 2018. Accelerator facilities are shown along with their completion date and length if applicable. Reproduced from [Mob18]

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \sigma\mathcal{L}, \tag{3.1}$$

where N is the number of outgoing particles, $\mathcal{L}$ is the instantaneous luminosity, and σ is a proportionality factor of the process called the cross-section. For proton beams at the LHC, the instantaneous luminosity is given by

$$\mathcal{L} = \frac{N_p^2 N_b f \gamma}{4\pi\epsilon\beta^*} F. \tag{3.2}$$

Variables making up this expression are listed and summarized in Table 3.1. This gives the instantaneous luminosity according to the LHC design specifications as $\mathcal{L} = 10^{-34}$ cm^{-2} s^{-1}. In practice, this corresponds to roughly 30 protons per bunch colliding at a frequency of 40 MHz. Each bunch collision is typically called an *event* and discretizes subsequent detector read-out.

Symbol	Description	Value
N_p	Number of protons per bunch	1.15×10^{11}
N_b	Number of bunches	2808
f	Revolution frequency	11.245 [kHz]
γ	Proton Lorentz factor	7461
ϵ	Transverse normalized beam emittance	3.75 [μm rad]
β^*	Optical beta function at the collision point	55 [cm]
F	Geometric luminosity reduction factor	0.836

Table 3.1: Nominal design parameters for calculating instantaneous luminosity of the LHC for collisions at the CMS detector. Values are obtained from Reference [Bru04].

The total number of outgoing particles may be calculated through the integration of the instantaneous luminosity

$$N = \sigma \int \mathcal{L}\ \mathrm{d}t = \sigma\mathcal{L}_{\mathrm{int}}, \tag{3.3}$$

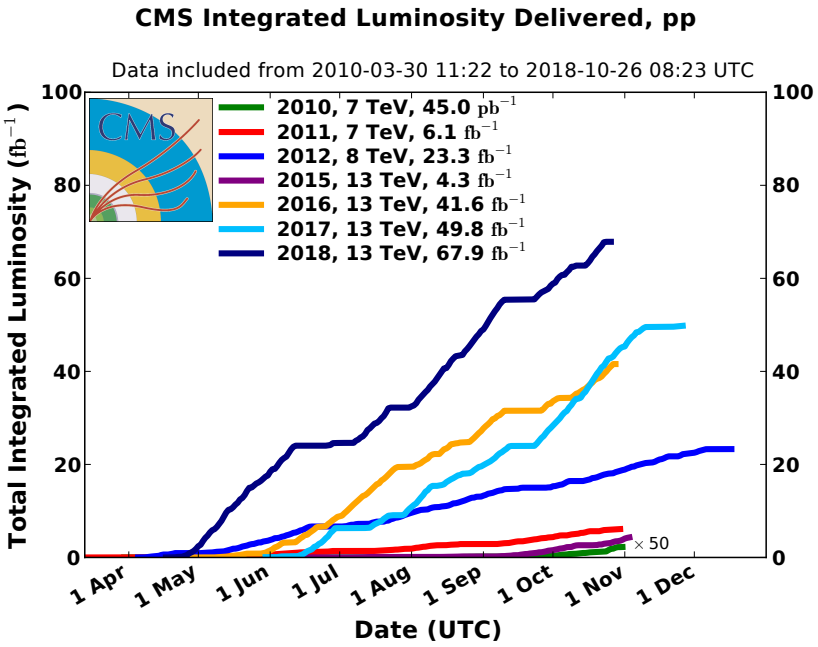


Figure 3.2: Integrated luminosities for proton-proton collisions delivered to CMS for each data taking year at the nominal center-of-mass energy. Reproduced from [Col21].

where $\mathcal{L}_{\text{int}}$ is called the *integrated* luminosity. Units of integrated luminosity are measured in inverse *barns* (1 b = 100 fm^2), which is roughly the cross-sectional area of a typical nucleus for nuclear interactions. Integrated luminosities for proton-proton data taking periods at the LHC are shown in Figure 3.2.

The cross-section of a process indicates how likely that process is to occur, and is measured in units of area. The larger the cross-section, the more scattered particles will be produced according to Equation 3.1. This value is experimentally calculated using the measured luminosity and integrating the observed events over all angles after the collisions take place.

3.2 The CMS Detector

The Compact Muon Solenoid (CMS) is one of four main detectors along the LHC ring and was built to probe the existence of the Higgs boson, as well as search for physics beyond the Standard Model. CMS gets its name from three of its defining features: its compactness relative to detectors of similar capabilities [Col08a]; its excellent muon resolution – the first

item listed in its physics design goals [Col08b]; and its strong solenoidal magnet it uses for charged particle momentum measurement. The detector is built in the shape of a cylinder, with a diameter of 14.6 m, length of 21.6 m, and a weight of over $12,500$ tons. The detector is housed roughly 100 m below the ground in Cessy, France, and encircles the LHC beam-pipe which runs parallel through the center of its flat faces. The experiment is managed by a collaboration of over $3,500$ scientists, engineers, and students who designed, built, and maintain the detector, and interact with the data that it produces.

The CMS detector is comprised of multiple sub-detectors which work in synchrony to reconstruct the particles resulting from each passing bunch collision. Secondary particles are identified by measuring their energy loss as they travel through the various sub-detectors in conjunction with the modulation in their trajectory from the presence of a strong magnetic field. The sub-detectors form a nested structure such that a particle that makes it outside the detector (e.g. a muon) has to pass through each of the individual types of sub-detectors. CMS is constructed in such a way that the vast majority of particles do *not* make it outside the detector, and the location of where they stop is used as a key variable used in their identification. Each of the sub-detectors is shown in Figure 3.3, overlaid with the characteristic signatures of particles CMS aims to distinguish. CMS uses a superconducting magnet placed inside this nested sub-detector structure to bend the trajectory of charged particles using a strong field of 3.8 T. This then allows reconstruction algorithms to distinguish the sign of their charge and measure their momentum. Both the sub-detectors and the superconducting magnet are described in more detail in the text that follows.

3.2.1 Geometry

The nominal collision point determines the center of the CMS detector, the y axis points upward, the x axis points radially towards the center of the LHC ring, and the z axis completes the right-handed coordinate system and points parallel to the beam-pipe, towards the direction of the Jura mountains. The azimuthal angle ϕ is defined with respect to the x-axis in the $x-y$ plane and is paired with the radial coordinate r, defined in that same

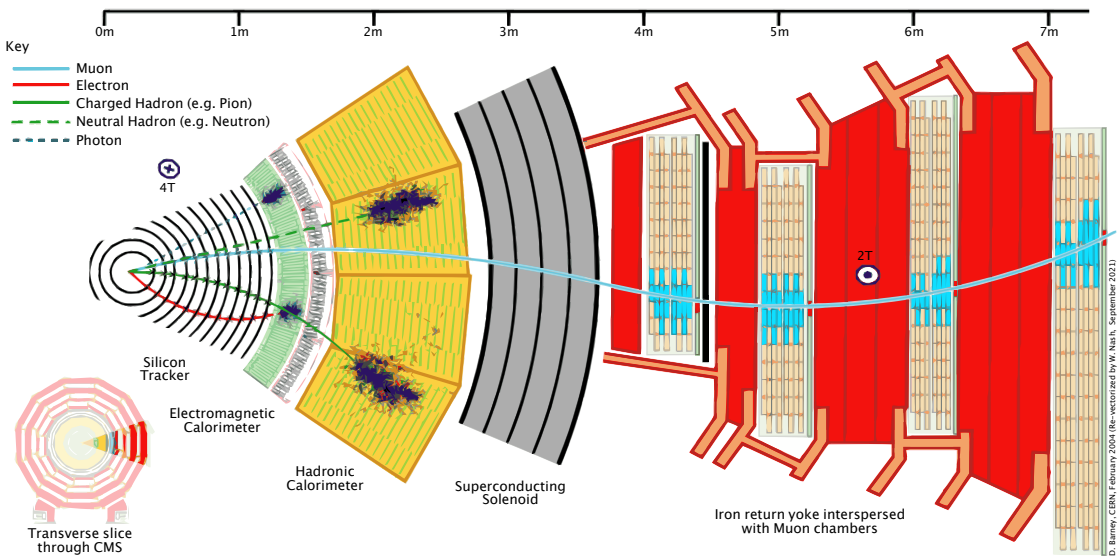


Figure 3.3: Transverse cross-section of the CMS detector. Graphic shows the scale of the detector and the characteristic signatures left by five separate particles within its subsystems. Reproduced from [Dav16].

plane. The polar angle θ is defined with respect to the z-axis, although it is rarely used in practice. More often the pseudorapidity η is used in its place, defined as

$$\eta = -\ln \tan(\theta/2). \tag{3.4}$$

The pseudorapidity gets its name from its similarity to the rapidity y_r, a particularly useful combination of kinematic quantities, defined as

$$y_r = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \tag{3.5}$$

where E is the particles energy, and p_z is the z component of the particles momentum $\mathbf{p}$. Differences in the rapidity of particles are invariant under Lorentz boosts along the z-axis, which make this an attractive quantity to use when reconstructing particles produced in a highly relativistic system such as the LHC. The pseudorapidity may be equivalently defined

as

$$\eta = \frac{1}{2} \ln \Big(\frac{|\mathbf{p}| + p_z}{|\mathbf{p}| - p_z} \Big), \tag{3.6}$$

which illuminates its relation to the rapidity more clearly – one sees that in the limit of a very relativistic particle (i.e. $|\mathbf{p}| \to E$), its pseudorapidity becomes equivalent to its rapidity. The choice of η over the θ has another advantage in that particles are produced roughly equally in slices of η, while the same is not true for the polar angle. The pseudorapidity spans the full set of real numbers from $-\infty \to \infty$ as it moves from pointing anti-parallel to parallel to the z axis, respectively.

3.2.2 Superconducting Magnet and Momentum Measurement

Momentum measurement and charge identification in CMS are enabled by a superconducting, solenoidal magnet. The magnet was designed such that the bending it provides gives a momentum resolution of muons in the muon system near 1 TeV of $\Delta p/p \approx 10\%$, where $p \equiv |\mathbf{p}|$ [Col06]. The magnetic field used to this end reaches nearly 4 T and is achieved using a current of 19.5 kA that travels through a reinforced NbTi superconductor cooled to 4 K with liquid helium. The magnet is installed between two sub-detector systems (the hadronic calorimeter and the muon system), and has a diameter and length of 6 m and 15 m, respectively. The field it produces runs along the beam axis in the center of the detector and curls around as it reaches the outer endcaps. Charged particles however are still bent along the ϕ directions in both cases due to the directionality of their momentum. In the center of the detector, the magnetic field runs along the beam-pipe, such that a particle produced with momentum in the transverse $x - y$ plane will move in a circle in that plane with a radius R according to the Lorentz force law

$$R = \frac{p_{\mathrm{T}}}{qB}, \tag{3.7}$$

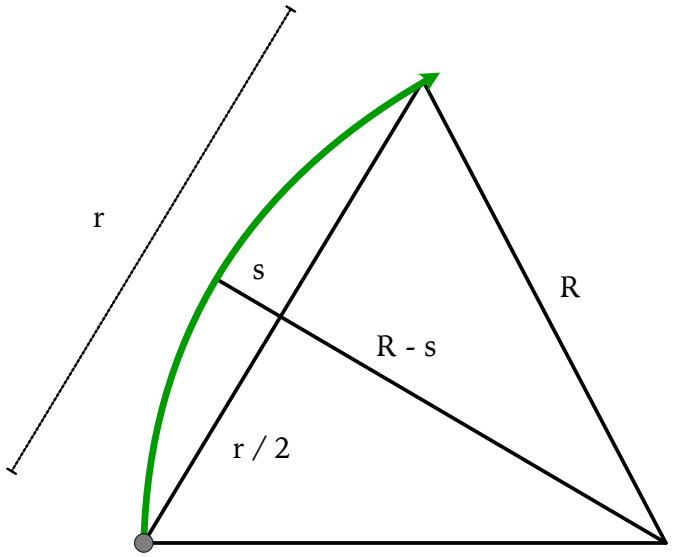


Figure 3.4: The trajectory of a charged particle (green) will have a curvature of radius R due to the presence of a magnetic field. The particle is tracked inside a circular detector in CMS with radius r. The sagitta s is the farthest the particle strays from a straight line directed radially from the center of the detector (gray dot). This geometry allows one to derive the radius of curvature R as a function of both r and s.

where p_{T} is the particle's transverse momentum, q is its charge, and B is the strength of the magnetic field. Track reconstruction algorithms calculate the p_{T} of charged particles using this relationship and infer the charge of the particle from the handedness of the curve (i.e. positive charges will move one way, while negative charges move the other). For particles of sufficient p_{T}, only a portion of the full circle is tracked, as the radius of the loop increases with transverse momentum. This requires one to use the sagitta s of the track to calculate its p_{T}, measured with respect to the radius of the active detector, defined as the distance between both curves at their center.

Figure 3.4 graphically shows the relationship between the radius of curvature R of the particle trajectory, the radius of the active detector r, and the sagitta of the track s. One can then geometrically derive the relationship between the three as

$$R = \frac{r^2}{8s} + \frac{s}{2} \approx \frac{r^2}{8s}. \tag{3.8}$$

This approximation is valid in the case that the sagitta is much smaller than the radius of the detector, as one should expect from particles of sufficient p_{T}. Measurement of the track within the detector, and its corresponding sagitta, therefore, lead to the measurement of the particles transverse momentum through

$$p_{\mathrm{T}} \approx \frac{qBr^2}{8s}. \tag{3.9}$$

The p_{T} resolution is primarily determined by the resolution on the sagitta, which arises from a collection of position measurements with errors distributed approximately normally. This translates into the measurement of q/p_{T} to vary about the true value with approximate Gaussian errors, *not* p_{T} itself [Das19] since by Equation 3.9, we have

$$\frac{q}{p_{\mathrm{T}}} \approx \frac{8s}{Br^2}. \tag{3.10}$$

When measuring tracking p_{T} resolution, it is therefore important to keep this in mind. Additionally, because the sagitta s shrinks as p_{T} increases, with the resolution on s is constant, transverse momentum resolution becomes worse with increasing particle p_{T}.

3.2.3 Inner Tracking System

Protons first collide within the beam-pipe kept at 1.9 K under a pressure of roughly 10^{-10} mbar. Secondary particles are not measured until they have traversed the 2.2 cm radius of the beam-pipe to reach the inner tracking system, the first system used in the reconstruction of the event. It is designed to provide precise tracking information capable of operating at the 40 MHz rate set by the LHC. The choice of a silicon-based detector was determined to be optimal in satisfying these criteria together with those on cooling efficiency, radiation hard-

ness, and material density. The inner tracker sub-system is composed of two parts; the pixel detector and the strip tracker [Col08b].

The inner-most portion of the tracker sees the densest flux of secondary particles anywhere in the detector. Therefore to distinguish between the multitude of tracks that are produced, pixel detectors are used, for which each pixel is roughly 100×150 μm^2. In total there are 66 million pixels in the pixel detector that cover a sensitive area of only ≈ 1 m^2. Individual pixels are distributed among $1,440$ pixel detector modules, each installed in one of three barrel layers, at $r = 4.4$ cm, 7.3 cm, and 10.2 cm, respectively, or on one of four end disks, which mirror each other at $|z| = 34.5$ cm and 46.5 cm.

Particles then pass through the strip detector, which uses a larger silicon technology allowed by the reduced particle flux density. Each strip has a thickness and pitch which varies with its placement, ranging from $80 - 180$ μm and $320 - 500$ μm, respectively. In total there are 9.6 million strips, that cover a sensitive area of 200 m^2 and extend to a radius of 110 cm $-$ substantially larger than that of the pixel detector. The strip detector is similarly split into $15,400$ modules which are distributed between barrel and endcap regions within $r < 1.2$ m and $|z| < 2.8$ m, excluding $|\eta| > 2.4$.

Both sub-detectors rely on electron-hole pair production via silicon ionization which occurs as charged particles traverse the detector. Each pixel/strip uses an electric current to collect and amplify these charge signals. This translates into a matrix of *hits* across the detector which then can be passed to reconstruction algorithms and form particle tracks.

3.2.4 Calorimeters

The remaining particles then enter the CMS calorimeters, where the vast majority are expected to stop entirely. Both charged and neutral particles interact with dense material within the detectors which typically leads to most particles losing all of their energy. This lost energy is then recorded by the calorimeter from which the initial energy of the particle is inferred. The electromagnetic calorimeter (ECAL) immediately follows the inner tracker and is designed to measure particles that lose energy according to the density of charge they see.

The remaining particles then enter the hadronic calorimeter (HCAL), which uses a material with a high density of nuclei, and aims to stop particles that interact strongly in addition to those which interact electromagnetically.

3.2.4.1 Electromagnetic Calorimeter

The ECAL surrounds the inner tracker and is made using scintillating lead tungstate ($PbWO_4$) crystals. The ECAL is primarily designed to measured electrons and photons which interact with the lead tungstate molecules to produce a cascading series of events involving bremsstrahlung, e^+e^- pair production, Compton scattering, and ionization – collectively known as an electromagnetic shower. The resulting shower is subsequently measured by the detector and is used to estimate the incident particles energy.

Materials can be characterized by the shape and size that these showers take inside them. The depth of an electromagnetic shower is characterized by the material's radiation length X_0, defined as the distance over which an electron will have its energy reduced to just $1/e$ of its initial energy, where e is Euler's number. The radiation length is strongly dependent on the charge density of the material, or equivalently its atomic number Z. The transverse width of a shower can be similarly characterized by a related quantity known as the Molière length. Lead tungstate was chosen for the ECAL for its small radiation length (0.89 cm) and Molière length (2.2 cm), which respectively allow for a compact design and good angular separation of nearby particles.

Conservation of energy requires that if an incident particle is stopped as a result of electromagnetic showering, the shower itself must contain its lost energy. Photodetectors installed on the ECAL take advantage of this principle and are designed to collect photons with total energy proportional to that of the incident particle, from which the initial energy is inferred. In total there are $61,200$ crystals in the barrel region and $7,324$ in each endcap. In the barrel region, each crystal is 22 cm long and has a cross-sectional area of 2.2×2.2 cm^2 to match the Molière length. This is done such that a typical electromagnetic shower is captured within a 2×2 square grid of crystals that pinpoint the location of the incident

particle.

The fractional energy resolution for particles contained in the ECAL improves with increasing particle energy. This is because the energy is measured through the counting of large numbers of photons in the photodetectors, which therefore takes errors according to the Gaussian limit of the Poisson distribution

$$\frac{\delta E}{E} \sim \frac{1}{\sqrt{E}}, \tag{3.11}$$

where E is the energy of the incident particle. This behavior is unlike that of tracking, which performs less well with increasing energy due to decreased trajectory curvature coupled with radiative losses that obscure the track during reconstruction.

3.2.4.2 Hadronic Calorimeter

The HCAL surrounds the ECAL and is designed to detect particles that interact via the strong force in addition to those that interact electromagnetically. Unlike the ECAL, the HCAL is a sampling calorimeter and is composed of alternating layers of absorber material and active scintillators, inside of which the particles interact and then have their energy loss measured, respectively.

The nuclear interaction length λ_I is a quantity analogous to the radiation length for particles that interact strongly and is defined as the mean distance traveled by a hadronic particle before undergoing an inelastic nuclear interaction. Materials with short nuclear interaction lengths have a high density of nuclei, equivalent to a large atomic mass number A. Materials with large A are therefore attractive when selecting an absorber to be used for a compact HCAL.

In CMS, absorber layers are composed of either brass or stainless steel because of their non-magnetic properties: brass is used in the barrel and endcap regions of the detector, while steel is used in the forward region ($3 < |\eta| < 5.2$). Two stainless steel absorbers are used as

structural support in the barrel region of the detector. Both absorber materials have a short nuclear interaction length of roughly $\lambda_n \approx 16.5$ cm [Col08b, CER97].

Particle showers produced in absorber layers then pass through read-out layers which contain scintillators that translate the showers into light. This light is then read out using wavelength-shifting fibers. In total, the HCAL contains roughly $70,000$ scintillator tiles it uses to reconstruct the location of incident particles. In the forward region of the HCAL, the intense radiation requires the use of quartz fibers, which generate a signal only when electrons above the Cherenkov threshold pass through them. This results in the detector becoming primarily sensitive to only the electromagnetic component of the shower which suppresses noise caused by the underlying radiation and therefore makes the read-out chamber more radiation hard [Col08b].

3.2.5 Muon System

The muon system of the CMS detector is the outermost detector subsystem and makes up the majority of the detector by volume. Muons play a crucial role in the reconstruction of an event due to their long lifetime (2.2 μs) and the ways in which they lose their energy [Gri04, Zyl20]. Because of these two peculiarities, muons are effectively the only charged particles that reach the muon system, which was foremost in mind as CMS was designed – good muon identification and momentum resolution were the first listed design goals in the CMS technical design report [Col06]. Sharp momentum resolution, coupled with radiative losses less than the electron when inside the silicon tracker, gave the $\mathrm{H} \to \mathrm{ZZ}^* \to 4\mu$ "gold-plated" decay channel its name [Col08b].

Muon momentum is measured via tracking as is done in the innermost silicon detector. Bending is achieved utilizing the return solenoidal field that loops through the muon system: the field bends to point radially inside the endcaps and then points opposite the field direction inside the solenoid once it reaches the barrel. This bends the muons in either the positive or negative ϕ direction depending on their charge. Notably, in the entire barrel region of the detector, this direction is opposite to that inside the silicon tracker.

Multiple detector technologies are used inside the muon system to satisfy operational constraints, each of which uses gas ionization as a means of tracking [Col08b]. The endcap region uses cathode strip chambers (CSCs), which are capable of handling high rates associated with large $|\eta|$, and irregular magnetic fields due to the placement of the solenoid. The barrel region sees a much lower muon rate than the endcaps, and therefore can use drift tubes (DTs), a less-specialized technology that can cover a significantly larger area ($18,000$ m^2 vs $7,000$ m^2) at a lesser cost [Sev18b]. Both barrel and endcap regions are complemented by resistive plate chambers (RPCs), which are less precise than their counterparts, and are used as redundant detectors for reconstructing muons at the 40 MHz collision rate set by the LHC. The layout of the three detector technologies used during the Run 2 data taking period is shown in Figure 3.5.

3.2.5.1 Cathode Strip Chambers

Precise measurements of muons in the endcap region of the CMS muon system are made using cathode strip chambers (CSCs). CSCs contain cathode strips that run radially to measure muon ϕ, paired with perpendicular anode wires that measure their pseudorapidity η. Strips have a width of $3.15 - 16$ mm, while wires are separated by distances of $2.5 - 3.16$ mm. In total 540 individual chambers cover an effective area of $7,000$ m^2.

Each CSC chamber contains six active layers of anode-cathode pairings that make largely independent measurements along the muon trajectory [Sev18b]. Each layer is held at high voltage ($2,900 - 3,600$ V) and filled with a gas mixture that is ionized by passing muons. The mixture is composed of: Argon (40%), which is ionized by the passing muons; CO_2 (50%), a quenching gas, which absorbs UV light that would otherwise cause continuous discharge; and CF_4 (10%), which is used to slow the aging of the detector.

Resultant ions and electrons are pulled apart by the large electric field and travel towards the cathode strips and anode wires respectively. Electrons, due to their lighter mass, accelerate more quickly and ionize additional gas molecules near the anode wires. This cascading ionization results in a macroscopic charge measurable on the anode, followed by an induced

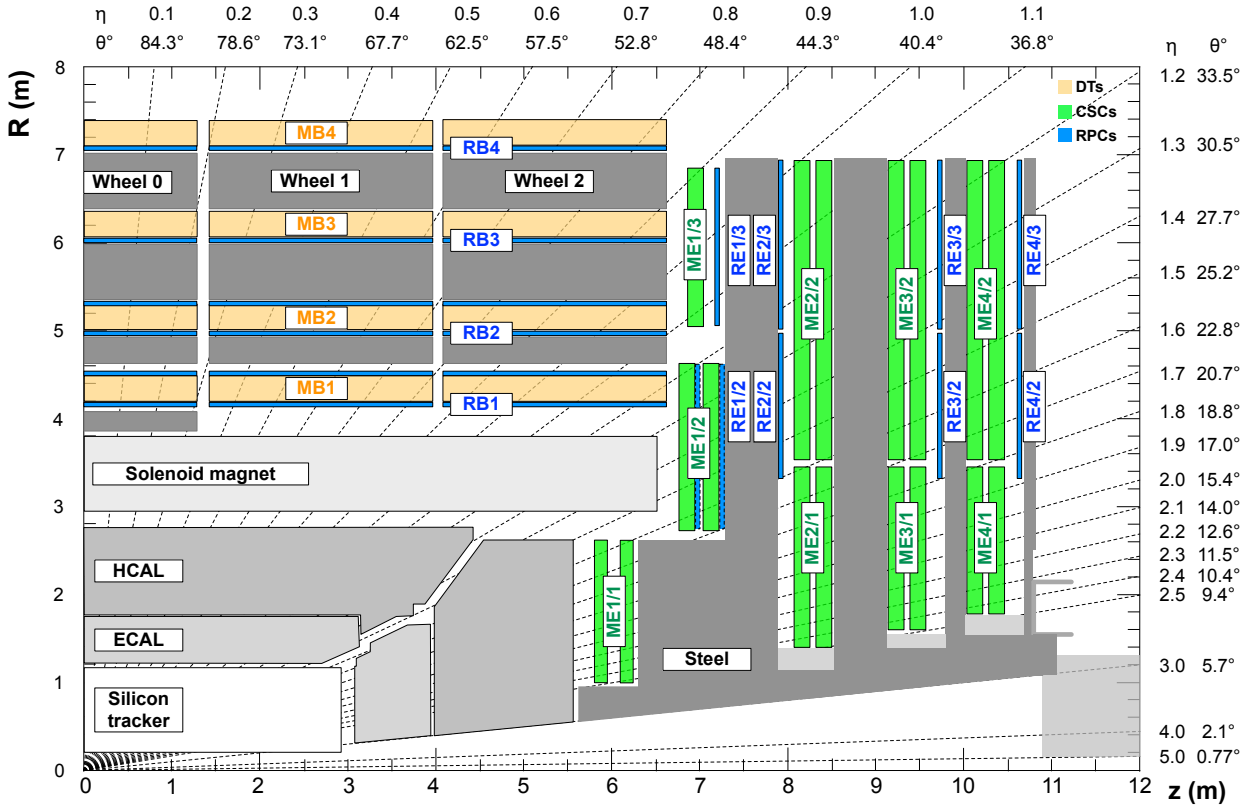


Figure 3.5: Cross-section of the CMS detector in the $r-z$ plane during the Run 2 data-taking period. The barrel (yellow) and endcap (green) muon chambers provide precise tracking information, often complemented by the silicon tracker. Drift tubes (DTs) are used in the barrel, while cathode strip chambers (CSCs) are used in the endcap. Four layers of chambers are present in both technologies, which sample the muon trajectory. Resistive plate chambers (RPCs) complement both sets of detectors and provide a fast response time used to determine if an event is read out. Reproduced from [cms18].

charge on the cathode provided by the slower moving ions. This process is shown graphically in Figure 3.6. Both analog charge measurements are then digitized using on-chamber electronics. Additional details of the endcap muon system are provided in Section 5.1.

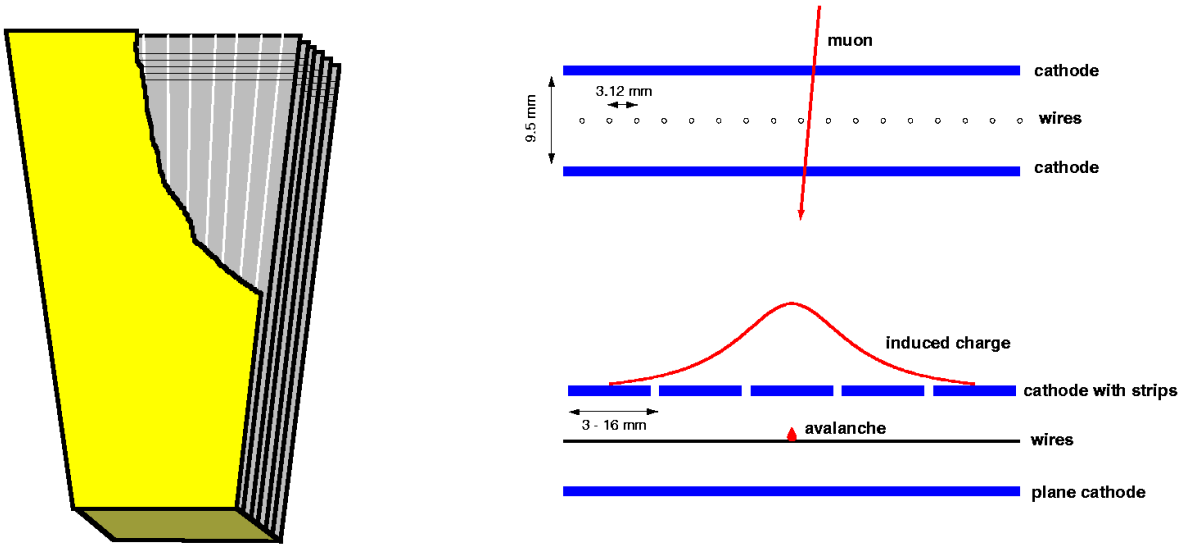


Figure 3.6: A CSC detector with six layers, containing vertical cathode strips (light gray) and radial anode wires (dark gray). A muon traverses the six planes of the detector, inducing a charge on both the strips and wires measured by nearby electronics. The measurements are then put together to reconstruct a portion of the muon's trajectory. Reproduced from [bru19].

3.2.5.2 Drift Tubes

Precise measurements of muons in the barrel region of the CMS muon system are made using drift tubes (DTs). Similar to the CSCs, DTs use anode-cathode pairings held at a multi-kilo-volt potential difference to measure the muon trajectory. The gas mixture (85% Argon, 15% CO_2) also operates via the production of ions that generate electromagnetic cascades which are read-out by onboard electronics [Cer07]. However, for each DT, anodes and cathodes run parallel to each other, unlike in the CSCs where they are orthogonal. This enforces a fixed distance between the anode and cathode (21 mm) and is used when inferring the location of the muon as it passes through the detector. Specifically, the time it takes for

the signal to reach the anode is used to localize the muon by using knowledge of the speed of the ions along the electric field, and the distance they must travel. Figure 3.7 shows a diagram of a DT cell, each of which has a cross-sectional area of 13×42 mm^2.

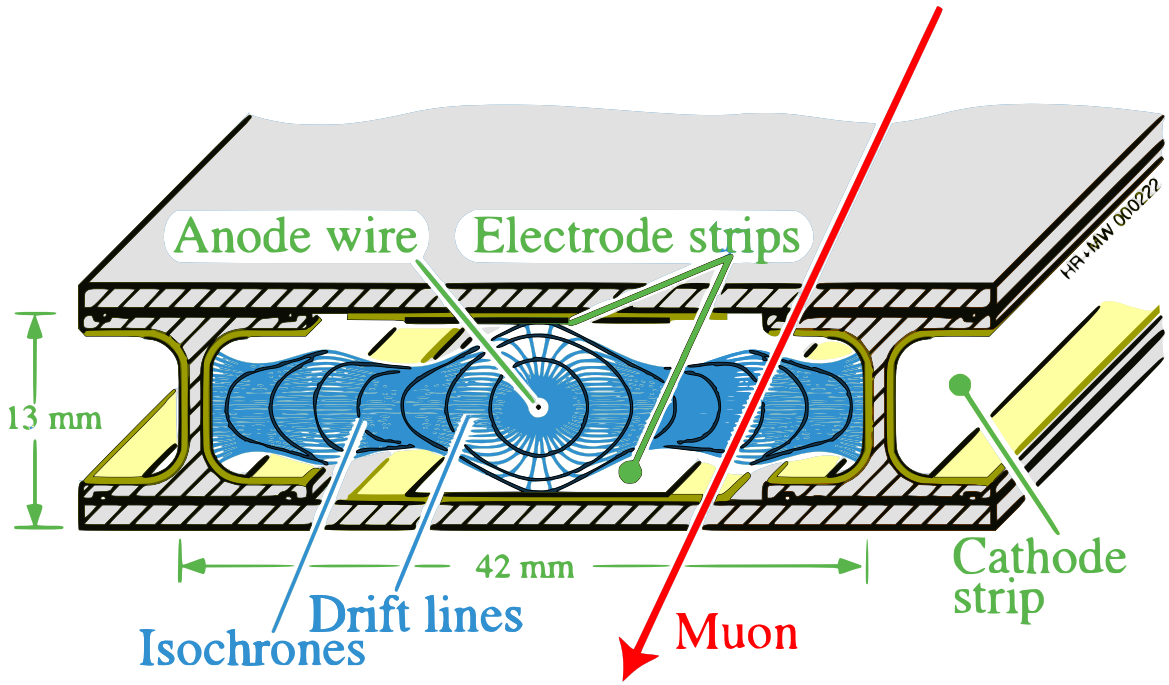


Figure 3.7: A drift tube cell overlain with electric field lines (blue) and a passing muon (red). The passing muon produces electrons through the ionization of the gas inside the cell. Resultant electrons then take a known amount of time to reach the anode and are used to infer the location of the passing muon. Isochrone contours (black) indicate locations at which initial electrons take an equal amount of time to travel from to the anode wire. Reproduced from [Col08b].

Within each chamber, drift tubes are layered together in stacks of four to create what is known as a superlayer (SL). A full chamber contains two to three SLs, which are stacked perpendicular to each other, with both outermost SLs oriented such that its wires run parallel to the beam-line. These SLs, and their complement which run orthogonally to the beam-line, measure the $r - \phi$ and z of the muon trajectory, respectively. In total there are 250 DT chambers in CMS that cover an area of $18,000$ m^2 [Sev18b].

3.2.5.3 Resistive Plate Chambers

The extraordinary bunch crossing frequency of the LHC requires a time resolution shorter than the 25 ns between collisions to distinguish muons from separate events. This is made possible with the addition of resistive plate chambers (RPCs), which are capable of resolving the time at which a muon passes through the detector to within roughly 1 ns [SC81].

Each RPC chamber maintains a uniform electric field between two parallel electrode plates 2 mm apart. The sensitive volume is filled with a gas mixture (95% $C_2H_2F_4$, 4.5% iso-C_4H_{10}, and 0.5% SF_6) which is ionized and whose charge is subsequently read out by onboard electronics [Abb04].

In total, there are 576 and 480 RPCs installed in the endcap and barrel regions of the muon system, respectively. Each of these detectors is paired with a complimentary DT or CSC chamber, both of which have much better spatial resolution than that of the RPCs (roughly 10x) and are used for offline muon reconstruction.

3.2.6 Trigger

The high collision rate at the LHC comes into direct conflict with bandwidth limitations set by the detectors. Each recorded event holds over one megabyte of data which implies that if CMS was to fully record every event at the nominal 40 MHz rate, more than 40 terabytes of data would be produced per second – far beyond modern hardware capabilities. CMS is instead designed around this fast rate in such a way that read-out is *triggered* only for certain events. The rate reduction is achieved using algorithms that operate on coarse grain primitives with a latency compatible with that of the collision rate. In parallel, high precision data used offline is kept in a rolling buffer to be permanently stored only if the event passes this preliminary selection. CMS uses a two-stage trigger system to this end: The Level-1 (L1) Trigger makes decisions at the hardware level, reducing the event rate from 40 MHz to 100 kHz, then the High-Level Trigger (HLT) further reduces the rate to roughly 1 kHz using a processor farm [CMS06].

3.2.6.1 Level-1 Trigger

The L1 Trigger uses only information from the muon system and calorimeters and fires according to the results of two sub-triggers: the Global Muon Trigger and Global Calorimeter Trigger (GCT) [Col]. The architecture of the L1 trigger is explained below and shown graphically in Figure 3.8.

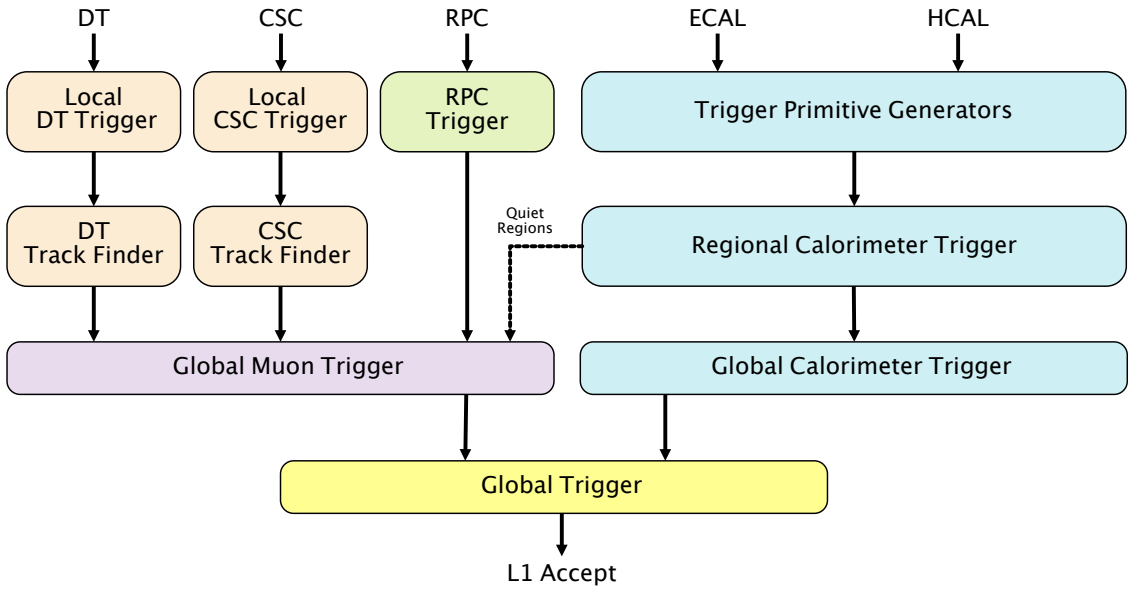


Figure 3.8: Structure of the L1 Trigger. Drift tubes (DT), cathode strip chambers (CSC), and resistive plate chambers (RPC) feed primitives to the Global Muon Trigger, while information from the electromagnetic (ECAL) and hadronic (HCAL) calorimeters are used by the Global Calorimeter Trigger. Primitive objects are then evaluated by the Global Trigger in the determination of a Level-1 (L1) accept. The diagram is based on Figure 8.1 of [Col08b].

The L1 Global Muon Trigger forms imprecise muon tracks using individual primitives provided by the DTs, CSCs, and RPCs. In both the DTs and CSCs, individual *hits* first must pass through a local trigger, during which chamber-sized track segments are formed. These local tracks are then passed to track finders which form full muon trajectories. The Global Muon Trigger then sorts and corroborates the tracks provided by the DTs, CSCs, and RPCs, and sends the top four muon candidates based on their quality to the Global Trigger. The L1 Muon trigger also uses a grid (η, ϕ) of quiet regions provided by the calorimeters used

to select for isolated muons (i.e. those with limited nearby activity). Additional information on primitives used within the CSCs is provided in Section 5.1.

The L1 Global Calorimeter Trigger (GCT) loosely identifies particle candidates such as electrons, photons, taus, and jets in addition to their energies. Local circuits generate trigger primitives which provide individual calorimeter cell energies and are then fed to the Regional Calorimeter Trigger, which constructs particle candidates. These candidates are then sent to the GCT, while quiet regions of the detector are sent to the Global Muon Trigger. The GCT then sends the top four of each particle type (electrons, photons, taus, and jets) sorted by their transverse energy to the Global Trigger.

The Global Trigger combines information provided by both the muon and calorimeter triggers to determine if the event is above the read-out threshold. The trigger itself sorts particle candidates and makes read-out decisions based on logical operators applied to the sorted results. For example `L1_DoubleMu_11_4` requires the first and second muons in the event have $p_{\mathrm{T}} > 11$ GeV and 4 GeV, respectively. An event may then be selected if it satisfies at least one of a specified number of acceptable *paths*, after which the event is further scrutinized by the HLT.

3.2.6.2 High-Level Trigger

Events passing the L1 Trigger have precision data stored in a rolling cache sent to the HLT, along with regions of interest indicating where to begin reconstruction. The HLT works at a software level and consists of a conventional CPU farm. This data is the same format (and precision) as that used by offline algorithms run on events which have been selected for permanent storage. Algorithms used at the HLT level are designed such that they operate as close as possible to those used offline. The HLT is tasked with further reducing the event rate to roughly 1 kHz [CMS06] and does by so using a streamlined reconstruction chain that rejects undesirable events as quickly as possible while minimizing overall CPU usage [Tro14].

As is the case for the L1 Trigger, the HLT operates using trigger paths used for different physics analyses. For example, there are single muon, double photon, and high missing

transverse energy ($\not\!\!E_{\mathrm{T}}$) triggers. Whether or not an event passes one of these triggers is used in determining if it is permanently recorded for offline analysis. Each path requires a set of physics objects which are constructed in a sequence of increasing complexity such that if the path fails at any stage, the rest of the path is skipped. For example, when identifying electrons, only calorimeter information is initially used, a positive result is then corroborated with hits in the pixel detector, and only then is the full track reconstruction performed.

To further reduce the amount of computation necessary for each incoming event, only part of each event is reconstructed. Where to look in each event is provided by the L1 Trigger and allows the HLT to operate at speeds orders of magnitude faster than that needed offline. Events passing the HLT are then permanently recorded in full, allowing the entire catalog of offline reconstruction algorithms to operate on them.

3.2.7 Event Reconstruction

Event reconstruction converts raw detector output into collections of high-level physics objects. Reconstruction takes place at three stages: the L1 Trigger, the HLT, and offline, each of which builds a progressively more intricate picture of the full event.

During both HLT and offline reconstruction, algorithms first use detector primitives to form nascent particle-like objects: charged-particle tracks, calorimeter clusters, and muon tracks [CMS09]. These elements are then used as input to a higher-level reconstruction algorithm known as Particle-flow (PF). The PF algorithm aims to reconstruct and identify all stable particles in the event and uses the full breadth of the CMS detector to that end. The output of the PF algorithm is a list of particles, known as `PFCandidates`, similar to those provided during Monte-Carlo event generation. `PFCandidates` are then passed to still higher-level reconstruction algorithms, such as those used to identify jets; high-level objects used to represent the collimated shower of particles produced by hadronic scattering.

The search presented in Chapter 4 looks for displaced particles which are not optimally reconstructed as `PFCandidates`. The analysis instead uses only lower-level objects – specifically the charged-particle tracks and muon tracks. These objects are therefore explained in

more detail in the following two sections.

3.2.7.1 Charged-Particle Reconstruction

Within the silicon tracker, charged-particle tracks are formed from combinations of pixel hits and silicon strip hits [SFS06]. These tracks, referred to as *tracker tracks*, are the product of a four-stage modular procedure:

1. Trajectory *seeds* are constructed using combinations of two or three hits which define the initial estimate of track parameters and their uncertainties.

2. Seeds are then extrapolated along the expected flight path of the particle. Hits that are found to be consistent with the current trajectory then form new track candidates. This process is run recursively to the end of the tracker volume.

3. A subset of compatible candidates is selected to resolve ambiguities in the ownership of the hits.

4. Final track fits are performed. Those failing certain criteria are discarded.

Tracker tracks are generated under the paradigm of *inside-out* tracking, where seeds are generated using the innermost region of the detector and the track is progressively built outwards. Seeds typically use the center of the detector (often called the beam-spot) as a constraint, which allows a trajectory seed to be formed from just two hits instead of three.

Individual seeds are then propagated as tracks according to the combinatorial Kalman filter, which acts equivalently to a global least-squares minimization for the track [Fru87]. In general, five parameters are required to describe a track within CMS: x_0, y_0, z_0, η, and p_{T} [Col14]. Together, x_0, y_0, and z_0 describe the coordinate in space called the impact point, defined as the point of closest of the track to the beam-spot. The angle η describes the pseudorapidity of the track, and p_{T} is the transverse momentum of the track. The Kalman filter aims to reconstruct the parameters of the track $\Theta(x_0, y_0, z_0, \eta, p_{\mathrm{T}})$ through propagation

of a discrete recursive system equation

$$\Theta_k = f_{k-1}(\Theta_{k-1}) + w_{k-1}, \tag{3.12}$$

where f_{k-1} is the track propagator from detector $k-1$ to k and w_{k-1} describes the process noise (e.g. multiple scattering). In practice, one builds this track through measurements, which are taken as having a linear relationship:

$$m_k = \mathbf{H}_k \Theta_k + \epsilon_k, \tag{3.13}$$

where m_k are the measured outputs of detector k, $\mathbf{H}_k$ is a matrix parametrizing the dependency of measurements on the track, and ϵ_k represents a noise term present on the detector. The Kalman filter treats the track Θ (sometimes called the *state vector*) as a random variable and exploits properties of linear models to arrive at estimates of both the track parameters and their associated errors. To this end, both noise terms (w and ϵ) are assumed to be normally distributed with a mean of zero.

Given our distribution of possible tracks on detector $k-1$, we wish to estimate what we should find on detector k. We do this by calculating the expectation value of Equation 3.12. The noise term does not contribute, while the track propagator, f, is a non-linear function due to the magnetic field and therefore requires special treatment in the context of the Kalman filter. To maintain linearity, the track propagator is first Taylor expanded around the expected value of the track at detector k, with

$$f_k(\Theta_k) = f_k(\hat{\Theta}_k) + \mathbf{J}_k(\Theta_k - \hat{\Theta}_k) + \ldots \tag{3.14}$$

Where $\hat{\Theta}_k \equiv E[\Theta_k]$, the expected value of the track on detector k, and $\mathbf{J}_k \equiv \partial f_k / \partial \Theta_k|_{\hat{\Theta}_k}$, the Jacobian [Ter]. Higher order terms are neglected. Because $E[\Theta_k - \hat{\Theta}_k] = 0$, we have that

$$\hat{\Theta}_k = f_{k-1}(\hat{\Theta}_{k-1}). \tag{3.15}$$

This approach is called the Extended Kalman Filter. The corresponding errors are calculated using the covariance matrix $\mathbf{C}_k \equiv E[(\Theta_k - \hat{\Theta}_k)^2]$. Equations 3.12 and 3.14 together enforce that Θ_k is linearly dependent on Θ_{k-1}, which allows us to write the propagated covariance matrix as

$$\mathbf{C}_k = \mathbf{J}_{k-1}\mathbf{C}_{k-1}\mathbf{J}_{k-1}^T + \mathbf{Q}_{k-1}, \tag{3.16}$$

where $\mathbf{Q}_{k-1}$ is the covariance matrix of w_{k-1}, and the T superscript denotes the matrix transpose.

However, this tells us nothing about how to update our track estimate after making a measurement. This is done by first comparing the difference between the observed measurements on detector k with their expectation values as calculated by Equation 3.13, written as

$$r_k = m_k - \mathbf{H}_k\hat{\Theta}_k, \tag{3.17}$$

where r_k are called the residuals. The corresponding covariance matrix of this quantity is denoted as $\mathbf{R}_k$ and can be expressed as

$$\mathbf{R}_k = \mathbf{H}_k\mathbf{C}_k\mathbf{H}_k^T + \mathbf{V}_k, \tag{3.18}$$

where $\mathbf{V}_k$ is the covariance matrix of the measurement noise ϵ_k. We then take an ansatz for our *updated* track estimate $\hat{\Theta}'$ to be linearly dependent on these residuals

$$\hat{\Theta}'_k = \hat{\Theta}_k + \mathbf{K}_k(m_k - \mathbf{H}_k\hat{\Theta}_k), \tag{3.19}$$

where $\mathbf{K}_k$ is called the Kalman gain matrix. The optimal Kalman gain is determined by minimizing the trace of the covariance matrix of the new track estimate $\mathbf{C}'_k = E[(\Theta'_k - \hat{\Theta}'_k)^2]$

with respect to $\mathbf{K}_k$, yielding

$$\mathbf{K}_k = \mathbf{C}_k \mathbf{H}_k^T (\mathbf{H}_k \mathbf{C}_k \mathbf{H}_k^T + \mathbf{V}_k)^{-1}. \tag{3.20}$$

This formalism ultimately enables the Kalman filter to behave as an additive χ^2 fit as one continues to recursively add measurements from the detectors.

In this way, charged-particle candidates can be built from initial trajectory seeds – each added hit provides a new trajectory estimate and a new χ^2, allowing for pruning of those with poor quality. Once initial candidates have been produced, a subset of those with compatible hits are selected to avoid overlapping tracks. The remaining tracks then have their parameters at each detector recalculated using the full set of measurements. These tracks are then assigned as the reconstructed tracker tracks of the event.

3.2.7.2 Muon Reconstruction

Muons are first independently reconstructed in two subsystems of the CMS detector – the inner tracker and the muon system [Abb08]. The inner tracker initially reconstructs muons as tracker tracks described in the previous section, while the muon system reconstructs them as *standalone-muon tracks*, created by a similar seeding and Kalman filtering scheme applied to track primitives found there. The muon system produces standalone-muon tracks using muon segments – small, nearly straight trajectories built using reconstructed hits (`recHits`) from a single detector. Information in both the inner tracker and muon system is then combined to construct two distinct varieties of muon objects:

- *Global Muons* are formed by finding a matching tracker track for each standalone-muon track. Measurements from both are then globally refitted using the Kalman filter technique.

- *Tracker Muons* are formed by extrapolating tracker tracks outward to the muon system while checking for signatures compatible with muons in the calorimeters and muon system. No combined (silicon + muon system) fit is performed. Tracker Muons require

at least one muon segment to match the extrapolated tracker track [Col12c]. The algorithm is intended to complement Global Muon reconstruction, recovering low p_{T} muons (below 7 GeV) which may not make it to the muon system.

Muons produced within the geometrical acceptance of the muon system and having sufficient p_{T} can be constructed as either a Global Muon, Tracker Muon, or very often both. In total, about 99% of these muons are reconstructed [Col12c]. Muons reconstructed as neither may still be reconstructed as only standalone-muons. Candidates found as both a Tracker Muon and Global Muon are merged into a single muon candidate, and typically have their combined momentum assigned according to the "sigma switch" algorithm described in Reference [Col12c].

CMS has developed additional specialized algorithms which improve muon momentum resolution for those with high p_{T} [Col12c]. As muons increase in energy, so too does their probability to create an electromagnetic shower. These showers confuse tracking algorithms and reduce the accuracy of their momentum measurement. Several algorithms have been developed to improve muon momentum resolution in these cases:

- *Tracker-Plus-First-Muon-Station (TPFMS)* refits the global-muon track using only hits from the innermost muon station, which prevents electromagnetic showering which develops later in the trajectory from entering the track fit.

- *Picky* looks at the hits in each muon chamber and identifies chambers that appear to have an electromagnetic shower. The global fit is then performed again using only chambers without the showers.

- *Tune P* chooses between tracker-only, TPFMS, and Picky results according to their goodness of fit.

Muons objects reconstructed by the PF algorithm (*Particle-flow Muons*) apply selection criteria to muon candidates found by the Global and Tracker Muon algorithms as described above. Selection criteria use information from other sub-detectors (such as the calorimeters)

to optimize the identification of muons produced in jets while maintaining a low fake-rate of hadrons misidentified as muons.

3.2.7.3 Vertex Reconstruction

The common ancestry of reconstructed tracks is deduced through vertex reconstruction. The location of the initial proton-proton collision is referred to as the primary vertex. As mentioned in Section 3.1.2, there are $\mathcal{O}(30)$ primary vertices in each event. Reconstruction of primary vertices [CMS10] uses prompt tracks selected by requiring:

1. A maximum transverse impact parameter significance d_0/σ_{d_0}, where $d_0 = \sqrt{x_0^2 + y_0^2}$ and is obtained from the tracks impact point as described in Section 3.2.6.1.

2. A minimum number of strip and pixel hits used to reconstruct the track.

3. A maximum normalized χ^2 of the track fit.

Tracks are then clustered using a *deterministic annealing* (DA) algorithm, which finds the global minimum for a problem with many degrees of freedom [Col14, Ros98]. The optimization borrows ideas from statistical mechanics and treats the system as if it has a temperature T which is gradually reduced, analogous to an annealing process used on materials.

The process begins by assigning the tracks to some unknown number of primary vertices distributed along the z-axis. The z-coordinate of each vertex k is referred to as z_k^V, while that of each track i is labeled as z_i^t. The uncertainty on z_i^t is labeled as σ_i^z. The analog of the Helmholtz free energy of the system is then given by

$$F = -T \sum_{i}^{\#\text{ tracks}} p_i \log\Big(\sum_{k}^{\#\text{ vertices}} \rho_k \exp\Big[-\frac{1}{T} \frac{(z_i^t - z_k^V)^2}{\sigma_i^{z\,2}} \Big] \Big). \tag{3.21}$$

Each track is given a constant weight p_i which reflects their consistency with coming from the primary vertex, and ρ_k is a vertex weight, only constrained such that $\sum_k \rho_k = 1$. The assignment of any track to a given vertex is then given by a probability p_{ik} calculated by

minimizing F with respect to z_k^V, which follows the Boltzmann distribution

$$p_{ik} = \frac{\rho_k \exp\left[-\frac{1}{T}\frac{(z_i^t - z_k^V)^2}{\sigma_i^{z2}}\right]}{\sum_{k'} \rho_{k'} \exp\left[-\frac{1}{T}\frac{(z_i^t - z_{k'}^V)^2}{\sigma_i^{z2}}\right]}. \tag{3.22}$$

Following the standard rules of statistical mechanics, tracks are compatible with many vertex locations at high temperatures, while only one vertex per track is allowed as $T \to 0$. The DA algorithm is initialized at a high temperature and recalculated iteratively as T is lowered, thereby tracing the global minimum of F. As the temperature falls, vertices are combined and ρ_k is renormalized. Therefore the process finds not only position and track assignments but also the optimal number of vertices. The 'annealing' process is continued to a temperature of $T_{\min} = 4$, where one reaches a compromise between resolving power and the possibility of separating tracks from the same true vertex.

Resulting vertices with at least two tracks are then fit, using the Kalman filter applied to vertex fitting [Fru96], which iteratively adds tracks to the vertex analogously to what is done in Section 3.2.7.1 for detector measurements. Both primary and secondary vertices are reconstructed using variants of this procedure, described in Reference [Ada07]. Each iteration updates the vertex location in addition to the momenta of the comprising tracks. The χ^2 of the vertex fit is also updated after each iteration and can be used for the identification of tracks not associated with the true vertex.

The primary vertex specifically is fit using an *adaptive vertex fitter* [Fru07], which additionally assigns a weight w_i to each track, reflecting the likelihood that the track belongs to the vertex in question. This fitting is done using a Kalman filter implementation capable of handling these weights. The tracks are then fit to a vertex position, which finds the minimum sum of weighted squares. The vertex position is consequently changed as the result of the fit, which requires reassignment of the track weights. After the weights are reassigned, the vertex fit is performed again. This process is run iteratively until reaching a predefined stopping condition detailed in Reference [Col14].

One often speaks of *the* primary vertex of the event, which during the Run 1 data-

taking period was simply the primary vertex found to have the maximum squared transverse momentum sum over reconstructed tracks. During the Run 2 data-taking period, the primary vertex is selected as the candidate vertex with the largest value of summed physics-object p_{T}^2. The physics objects are the jets, clustered using the jet finding algorithm [CGS08, CPS12] with the tracks assigned to candidate vertices as inputs, and the associated missing transverse momentum, taken as the negative vector sum of the p_{T} of those jets.

CHAPTER 4

Search for a Long-Lived Displaced Dimuon Resonance

However beautiful the strategy, you should occasionally look at the results.

— Sir Winston Churchill (1874 - 1965)

This chapter presents a search for long-lived neutral particles decaying to a pair of muons in the CMS detector [Val19b]. The search uses data taken in 2016 and 2018 at a center-of-mass energy of 13 TeV corresponding to an integrated luminosity of 97.6 fb^{-1}. The analysis extends two CMS analyses [Phy15, Col15] performed using data taken at a center-of-mass energy of 8 TeV . The experimental signature is a pair of oppositely charged muons with an invariant mass above 10 GeV that originate from a common vertex that is spatially separated from the proton interaction point by distances ranging from several hundred microns to several meters. Results are interpreted in the context of the HAHM and BSM Higgs models as described in Section 2.2.

4.1 Main Discriminating Variables

The primary feature used to distinguish our dimuon signal from the background is its displacement. Event reconstruction algorithms define the central-most location of the event as the primary vertex (PV) as described in Section 3.2.7.3, which aims to describe the point in space where the protons collided to produce particles of the highest energy. For the signal we are searching for, this point should be distinct from the origin of the muon pair. In the analysis, muons are paired and fit to a common vertex (CV) using a Kalman filter which is then measured against the location of the PV. The distance between the CV and PV in the transverse plane is called the transverse decay length, or L_{xy}.

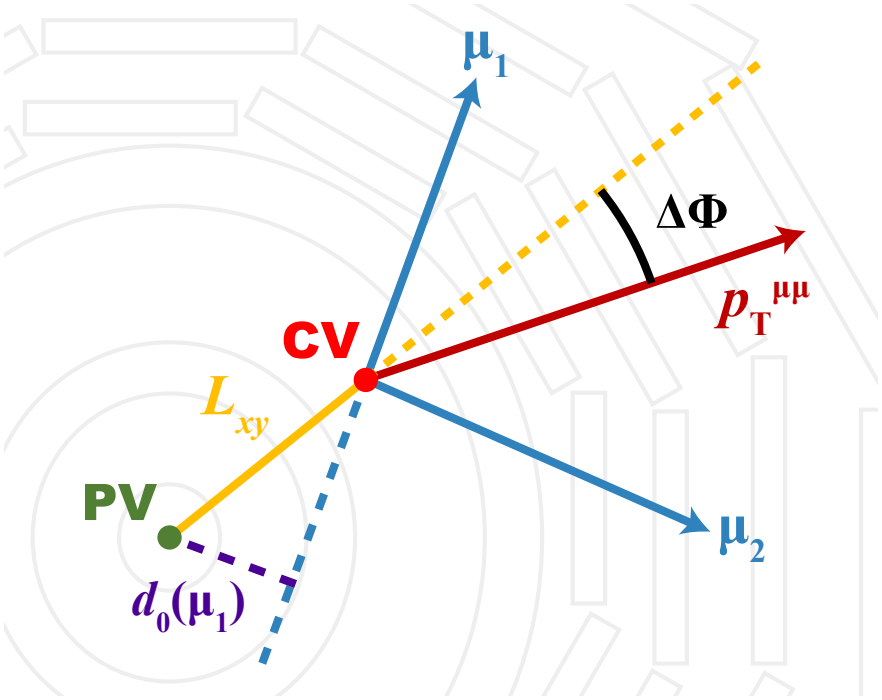


Figure 4.1: Main variables used in distinguishing dimuon signal from background. Dimuon displacement is measured via L_{xy}, the transverse distance between the location of the dimuon common vertex (CV) fit and the primary vertex (PV). The angle between the L_{xy} vector and the dimuon transverse momentum vector $p_{\mathrm{T}}^{\mu\mu}$ is labeled as $|\Delta\Phi|$. This variable is expected to be small for signal and symmetric for most backgrounds. Each individual muon provides an indication of the dimuon displacement via its transverse impact parameter, d_0.

For a genuine signal, the transverse momentum of the dimuon should point in a direction parallel to the L_{xy} vector, as the neutral particle that decayed to the dimuon is assumed to have been produced at and traveled away from the PV. The momentum vector of the dimuon is measured using the four-vectors of the two constituent muons. The angle between the transverse component of the dimuon momentum vector ($p_{\mathrm{T}}^{\mu\mu}$) and the L_{xy} vector is called the collinearity angle, or $|\Delta\Phi|$, and is depicted in Figure 4.1. For signal, $|\Delta\Phi|$ is expected to be narrowly peaked at zero.

Unlike signal, the majority of the expected background is distributed symmetrically in $|\Delta\Phi|$ around $\pi/2$. For most backgrounds, measured L_{xy} is the result of resolution effects and its direction is a random variable around the PV. This is similarly true for $p_{\mathrm{T}}^{\mu\mu}$ when the dimuon is produced at rest and causes the background to fall equally almost everywhere. The symmetry in $|\Delta\Phi|$ arises due to a correlation between the transverse momentum resolution and its direction. For a given L_{xy}, momentum is most accurately measured using tracks that travel radially from the beam-spot and progressively worsens as $|\Delta\Phi|$ approaches $\pi/2$. The effect is depicted in Figure 4.2, with more details given in Section 4.5.2 of Reference [Das19].

One may also use quantities associated with individual muons to assess the displacement of reconstructed dimuons. The transverse impact parameter d_0 is the closest the muon track ever gets to the PV in the transverse plane. This value is obtained from the impact point of the track described in Section 3.2.7.1. In general, a larger d_0 is associated with a dimuon with a larger displacement. The muon impact parameter itself is particularly useful because it is measured *twice* per dimuon, thereby reducing the chances of a prompt dimuon being mismeasured as displaced. The complementary longitudinal impact parameter and its significance are denoted as d_z and d_z/σ_{d_z}, respectively.

4.2 Main Backgrounds

The SM predicts no dimuon resonances which are measurably long-lived in the mass range used for this analysis (see Figure 2.2). Therefore the expected background in our signal region

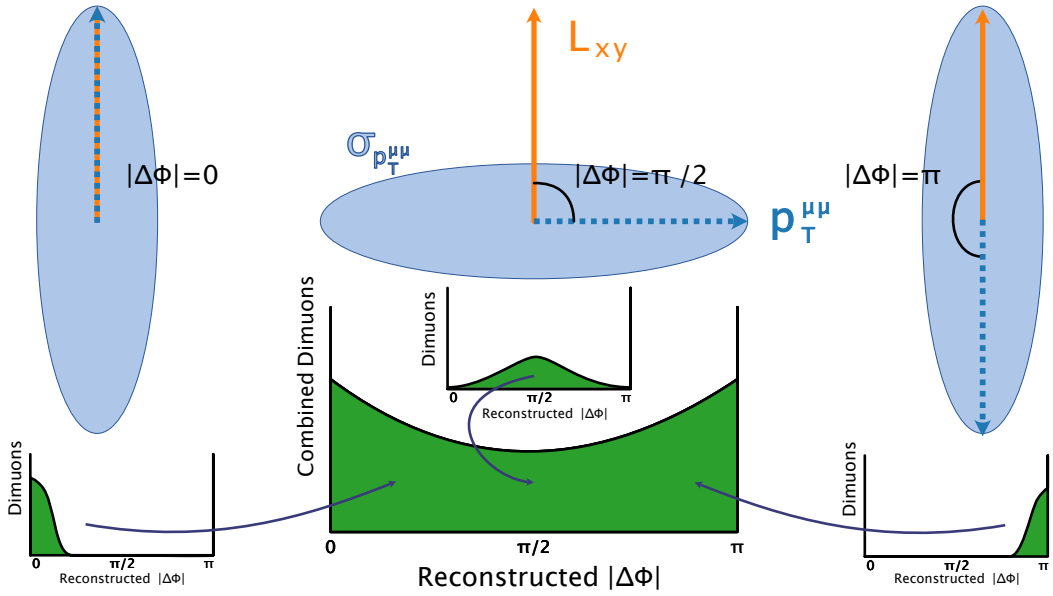


Figure 4.2: Distribution of background dimuons which are symmetric in $|\Delta\Phi|$ (bottom middle) and related diagrams. Uncertainty on the dimuon transverse momentum $p_{\mathrm{T}}^{\mu\mu}$ takes the form of an ellipse $\sigma_{p_{\mathrm{T}}^{\mu\mu}}$. Dimuons with $|\Delta\Phi| \sim 0$ (left) and π (right) have their transverse momenta measured more accurately in the direction of the L_{xy} vector and are thus measured with relatively sharp precision around their true values. Dimuons with $|\Delta\Phi| \sim \pi/2$ upper middle) however have their momenta poorly measured relative to the L_{xy} vector, and become smeared when reconstructed. This results in a reconstructed $|\Delta\Phi|$ distribution that is symmetric with respect to $\pi/2$, with dimuons clustering equally at the boundaries.

is purely due to instrumental mistakes, reconstruction mistakes, and/or mispairings of muons originating from separate physical processes. This section covers the general characteristics of the three main backgrounds present in the analysis and outlines methods used for their suppression. The full list of backgrounds studied using Monte Carlo simulation is given in Section 4.3.4.

4.2.1 Drell-Yan

Inelastic scattering of quark-antiquark pairs from initial proton-proton collisions can annihilate to form lepton pairs through what is known as the Drell-Yan process [Ken82]. This occurs through an exchange of a virtual photon γ^* or Z boson, as is shown in Figure 4.3. The resulting dilepton pairs are produced promptly along a mass continuum through what is one of the most well-studied processes throughout all of particle physics.

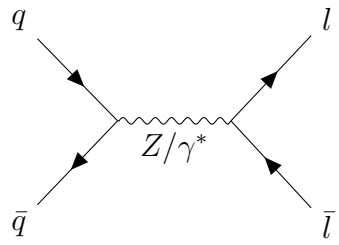


Figure 4.3: Feynman diagram of the Drell-Yan process. Quarks ($q\bar{q}$) from the initial proton-proton collisions annihilate and produce a virtual photon (γ^*) or Z boson, which then decay into a pair of oppositely charged leptons ($l\bar{l}$).

Dimuons produced via Drell-Yan that have an invariant mass that falls within our signal region are built from muons that are both oppositely charged and typically isolated from other event activity. These features mirror those we look for in our signal, however, these dimuons are produced at the PV and are strongly suppressed using displacement requirements. Drell-Yan dimuons that fall within our signal region are often a result of fitting mistakes when identifying the CV location. Drell-Yan dimuons tend to have low transverse momentum, which allows one to take advantage of the $|\Delta\Phi|$ symmetry covered in Section 4.1.

Ditaus can also be produced through the Drell-Yan process, with each tau having a

significant branching ratio to muons, $\mathcal{B}(\tau \rightarrow \mu\ \bar{\nu}_\mu\ \nu_\tau) = 17.4\%$ [Zyl20]. The tau itself is measurably long-lived with a lifetime of roughly 3×10^{-13} s [Zyl20] (Figure 2.2), and results in muon pairs that are produced at locations distinct from the primary vertex. In this case, the muons are not produced at the same location and can be suppressed using requirements on the CV fit. Ditau events make up roughly half the remaining Drell-Yan background after the full signal selection is applied.

4.2.2 QCD

Muons produced by J/Ψ decays, B-meson cascades, and in association with dijets / multijets also form a significant background for the search. Backgrounds resulting from strong interaction processes are labeled as QCD, and generally produce muons surrounded by other particles. Therefore QCD-like background are suppressed using requirements on the muon isolation, demanding that tracks around the muon are of low p_{T} and/or few in number.

The J/Ψ meson decays to two oppositely charged muons with a branching ratio of $\mathcal{B}(\mathrm{J}/\Psi \rightarrow \mu\bar{\mu}) = 5.7\%$ [Zyl20]. Its relatively low-mass (3.1 GeV) compared to the collision center-of-mass energy enables it to have a significant relativistic boost which, when combined with the the long lifetime of the B meson which it is often produced from, allow it to become measurably displaced. Additionally, these decays are not expected to be $|\Delta\Phi|$ symmetric because they are genuinely displaced. These decays, and other similar low-mass resonances (such as the ϕ), are further suppressed using an invariant mass requirement, which is placed significantly above where the bulk of these events would be found.

Muons resulting from B-meson cascades or dijet/multijet events often do not originate from the same vertex. An example B-meson cascade event is shown in Figure 4.4, where a common vertex is found at a location distinct from the origin of either muon. These events are suppressed using vertex fit requirements in addition to those listed above. Cascading decay chains do not always produce muons of opposite charge and allow for the estimation of such backgrounds from the number of same-sign events.

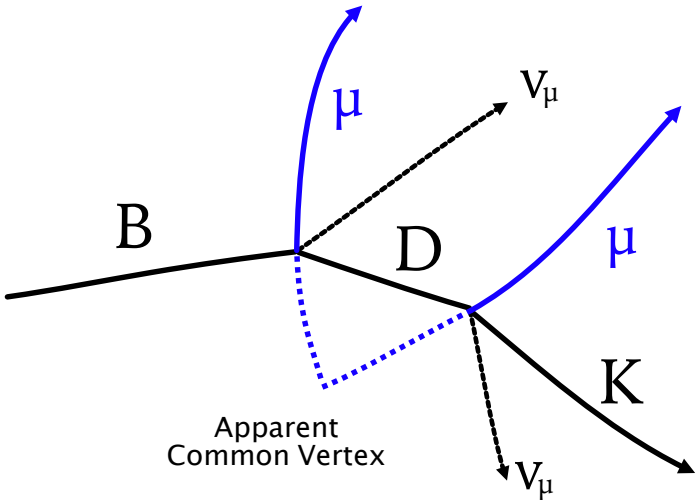


Figure 4.4: Typical QCD background dimuon. Muons originate from two distinct locations as a product of a B-meson cascading decay chain. Muons are incidentally combined to form a common vertex.

4.2.3 Cosmic Rays

The earth's atmosphere is subject to a nearly uniform flux of incoming high energy particles – often from sources outside our solar system [Zyl20]. These particles are known as cosmic rays and are typically composed of protons and other atomic nuclei capable of living long enough to reach the earth. When cosmic rays interact with our atmosphere, they produce showers of secondary particles. Muons, which originate from the mesons produced in these interactions, are the most numerous charged particle that reach sea level due to their long lifetime and penetrating nature. Muons from cosmic rays frequently pass through the CMS detector and are often used for calibrating detectors when the accelerator is turned off.

A cosmic muon that passes through the CMS detector can sometimes be reconstructed as *two* muons. An example of such a case is shown in Figure 4.5, where the vertex of the two reconstructed muons lies away from the PV, giving the appearance of a truly displaced dimuon. Such muons arrive at random times with respect to the bunch crossing, and can therefore be suppressed using timing information of the reconstructed muons. Furthermore, one of the muons will have passed through the detectors in the wrong order, which can be

similarly used to suppress such events. Muons that cannot be suppressed using the timing information can be identified using other characteristics of these events – both muons will be back to back from each other. One may suppress cosmic muons in this way by requiring the 3-d opening angle α between the two muons to not be near π.

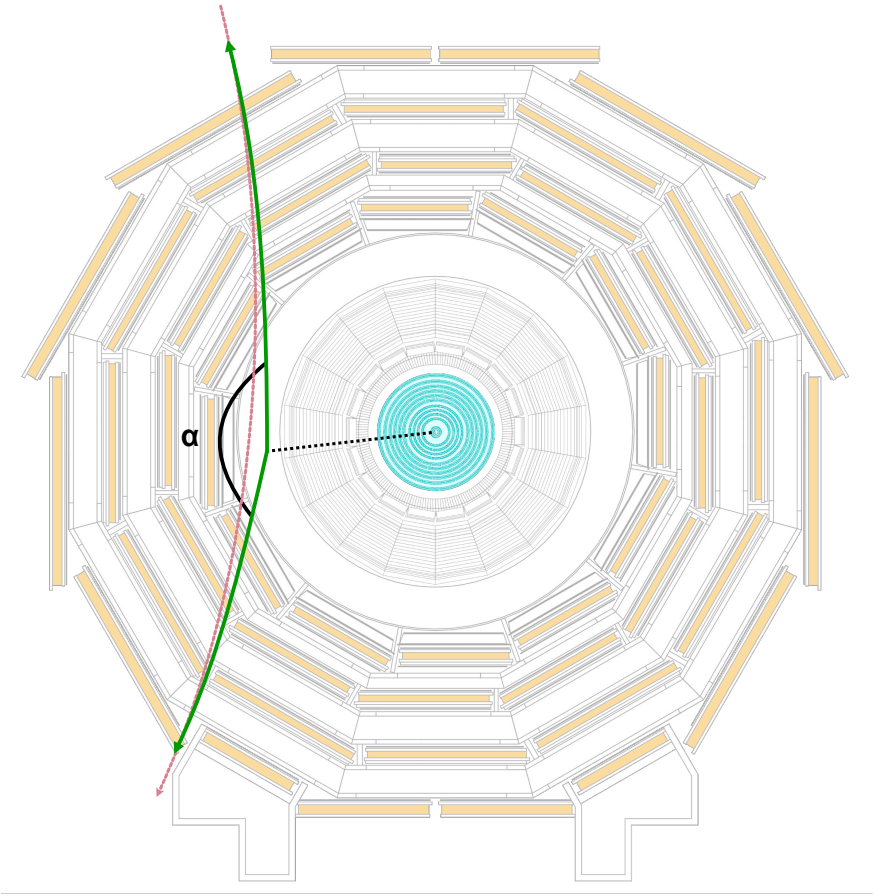


Figure 4.5: Cosmic ray muon (pink) is reconstructed as two muons in the barrel muon system (green). The reconstructed dimuon appears as displaced, forming a vertex away from the PV. Cosmics can be suppressed using a cut on the reconstructed 3-d opening angle α.

4.3 Dataset Production

The search looks at pp collisions taken during 2016 and 2018 at a center-of-mass energy of 13 TeV. Events are selected from a dedicated Level-2 trigger described in Section 4.3.1 and were recorded in the `DoubleMuon` datasets listed in Tables 4.1 and 4.2. Complementary

Monte Carlo simulations are produced for both the HAHM and BSM Higgs models, as well as expected background to aid in the development of selection criteria. Studies using cosmic ray muons are primarily documented in Reference [Val19a]. The analysis was performed using `CMSSW_8_0_31` and `CMSSW_10_2_5` for 2016 and 2018, respectively. The software developed for this analysis is stored in Reference [Tem].

4.3.1 Trigger

The triggers used in the analysis aim to select events containing two muons of modest p_{T} that will be reconstructed as a displaced dimuon offline. No relevant HLT path was maintained for 2017, which accounts for the non-usage of data from this period. The trigger path used in 2018 is an updated version of that used in 2016 that was designed to improve our signal efficiency [Val19a].

For both 2016 and 2018 datasets, selected events first pass a Level-1 (L1) muon trigger. Both years utilize double muon triggers that require two muons of relatively low p_{T} to be reconstructed using coarse-grain primitives found in the muon system, as described in Reference [AT13]. For example, the `L1_DoubleMu_11_4` trigger path used in 2016 requires the p_{T} of the leading and subleading muons to be above 11 and 4 GeV, respectively. Thresholds are set low to compensate for the fact that reconstruction assumes both that the muon originates from the beam-spot, and that they are built out of inherently imprecise information. Events are required to pass a logical or of multiple L1 triggers as documented in Reference [Val19b]. Selected events in 2018 may also be triggered by a triple muon L1 trigger that probes even lower values of p_{T}.

The subsequent HLT path used in 2016 was

`HLT_L2DoubleMu28_NoVertex_2Cha_Angle2p5_Mass10`

The trigger selects events in which two muons have p_{T} above 28 GeV (`L2DoubleMu28`), both of which have no beam-spot constraint used in their track fits (`NoVertex`). Both muons tracks must also be formed using at least two CSC or DT stations (`2Cha`) to suppress those of low quality. As described in Section 4.2.3, cosmic muons can occasionally mimic a dimuon

formed by back-to-back muons. To suppress this type of background, the 3-D angle α between both reconstructed muons must be less than 2.5 rad (`Angle2p5`). This requirement is equivalent to $\cos\alpha > -0.8$. Finally, the invariant mass $m_{\mu\mu}$ of the two muons is required to be above 10 GeV (`Mass10`), which serves to keep the event rate within an acceptable level and to suppress events dominated by QCD processes outlined in Section 4.2.2.

The HLT used in 2018 was modified and includes two independent triggers based on that used in 2016

`HLT_DoubleL2Mu23NoVtx_2Cha`

`HLT_DoubleL2Mu23NoVtx_2Cha_CosmicSeed`

Similar to 2016, both triggers select for events which have two muons with p_{T} above 23 GeV (`DoubleL2Mu23`) formed without a beam-spot constraint (`NoVtx`). It was found that both the requirements on the muon angle (`Angle2p5`) and dimuon mass (`Mass10`) gave only a marginal rate decrease and were therefore removed in 2018. The two paths used in 2018 differ in the initial seeding that is used for fitting the muon trajectory as described in Section 3.2.7.1. The first listed uses "pp-seeding", which assumes the muons are produced in the vicinity of the beam-spot, while the second generates the track according to "cosmic-seeding" as if they originated from cosmic rays (`CosmicSeed`). The HLT itself is a logical OR of these two triggers. The muons used in 2016 HLT are purely pp-seeded. Additional details on the track seeding can be found in Reference [Val19a].

4.3.2 Data Samples

The analysis itself searches the `DoubleMuon` datasets in 2016 and 2018 among events that pass the HLT paths described above. Good quality data requires the synchronicity of many different sub-detectors, each of which must be performing within specified operating conditions. Quality assurance of data is managed by the Physics Data and Monte Carlo Validation (PdmV) group at CMS [Mey21], who provide recommendations for reprocessing and configuration files towards these ends, the most recent of which have been used for this analysis. Specifically, our analysis does not use any calorimeter information, therefore we require

only that detectors used to measure muons are in proper working conditions by using the `MuonPhys` JSON file provided by the PdmV group. The total integrated luminosities for 2016 and 2018 are listed in Tables 4.1 and 4.2, respectively.

Table 4.1: 2016 data sets used in the main analysis. Listed are the data set names, run ranges as certified by CMS for physics analysis, and corresponding integrated luminosities.

Data set	Run range	$\mathcal{L}_{\text{int}}[\text{fb}^{-1}]$
`/DoubleMuon/Run2016B-07Aug17_ver2-v1/AOD`	273150–275376	5.8
`/DoubleMuon/Run2016C-07Aug17-v1/AOD`	275657–276283	2.6
`/DoubleMuon/Run2016D-07Aug17-v1/AOD`	276315–276811	4.3
`/DoubleMuon/Run2016E-07Aug17-v1/AOD`	276831–277420	4.0
`/DoubleMuon/Run2016F-07Aug17-v1/AOD`	277981–278808	3.1
`/DoubleMuon/Run2016G-07Aug17-v1/AOD`	278820–280385	7.7
`/DoubleMuon/Run2016H-07Aug17-v1/AOD`	281613–284044	8.8
Total		36.3

Table 4.2: 2018 data sets used in the main analysis. Listed are the data set names, run ranges as certified by CMS for physics analysis, and corresponding integrated luminosities.

Data set	Run range	$\mathcal{L}_{\text{int}}[\text{fb}^{-1}]$
`/DoubleMuon/Run2018A-17Sep2018-v2/AOD`	315257–316995	14.7
`/DoubleMuon/Run2018B-17Sep2018-v1/AOD`	317080–319077	7.2
`/DoubleMuon/Run2018C-17Sep2018-v1/AOD`	319337–320065	6.9
`/DoubleMuon/Run2018D-PromptReco-v2/AOD`	320673–325172	32.5
Total		61.3

4.3.3 Signal Monte Carlo Simulation

The search itself is intended to be as model-independent as possible, therefore signal samples have been generated for a wide range of signal kinematics. Samples were simulated for both the HAHM and BSM Higgs models, described in Section 2.2, which cover the range of kinematics listed in Table 4.3 and 4.4, respectively.

The HAHM samples are parametrized by the mass of the dark photon, m_{Z_D}, which determines $\mathcal{B}(Z_\text{D} \to \mu\mu)$, as calculated in Equation 2.15 and enumerated in Table 2 of Reference [CEG15]. Generated masses vary from 10 GeV, where the Z_D is fairly boosted ($\gamma \sim$

6.25) to 60 GeV, where the dark photon is produced with very little transverse momentum and is therefore relatively slow-moving ($\gamma \sim 1$). As explained in Section 2.2, ϵ controls the dark photon lifetime, scaling as ϵ^{-2}. The transverse decay length, L_{xy}, increases with the dark photon lifetime and boost.

The BSM Higgs Model includes Higgs bosons with masses consistent with that in the SM, as well as exotic Higgs bosons with significantly larger masses. Larger mass Higgs bosons allow for larger mass LLPs and are used to probe additional regions of phase space.

Table 4.3: Characteristics of the HAHM signal samples used in the search.

m_{Z_D} [GeV]	$\mathcal{B}(Z_\mathrm{D} \to \mu\mu)$	ϵ	$c\tau$[mm]	$\langle L_{xy} \rangle$[cm]
10	0.1538	$1 \cdot 10^{-6}$	1.14	0.77
		$5 \cdot 10^{-7}$	4.55	2.89
		$1 \cdot 10^{-7}$	114	72.0
		$3 \cdot 10^{-8}$	1264	800
20	0.1476	$5 \cdot 10^{-7}$	2.17	0.72
		$2 \cdot 10^{-7}$	13.6	4.14
		$5 \cdot 10^{-8}$	217	66.4
		$1 \cdot 10^{-8}$	5425	1665
30	0.1437	$3 \cdot 10^{-7}$	3.90	0.80
		$1 \cdot 10^{-7}$	35.1	6.73
		$3 \cdot 10^{-8}$	390	74.7
		$7 \cdot 10^{-9}$	7165	1374
40	0.1462	$2 \cdot 10^{-7}$	6.21	0.86
		$8 \cdot 10^{-8}$	38.8	5.10
		$2 \cdot 10^{-8}$	621	81.4
		$5 \cdot 10^{-9}$	9937	1301
50	0.1257	$2 \cdot 10^{-7}$	4.42	0.47
		$6 \cdot 10^{-8}$	49.1	4.37
		$1 \cdot 10^{-8}$	1768	158
		$4 \cdot 10^{-9}$	11049	982
60	0.1069	$1 \cdot 10^{-7}$	11.8	0.70
		$4 \cdot 10^{-8}$	73.7	3.93
		$7 \cdot 10^{-9}$	2405	129
		$2 \cdot 10^{-9}$	29464	1591

Samples are produced through a pipeline of simulation stages. Event generation is done using Pythia [S15] for the BSM Higgs samples, while the HAHM samples use a combination of both Pythia [S15] and MadGraph [AFF14]. Subsequent interactions of resulting

Table 4.4: Simulated $\Phi \to XX$ signal samples used in the analysis. Both $\Phi \to 2X \to 2\mu 2j$ and $\Phi \to 2X \to 4\mu$ samples were generated at each combination of m_Φ, m_X, and $c\tau(X)$.

m_Φ [GeV]	m_X [GeV]	$c\tau(X)$ [mm]
125	20	13, 130, 1300
125	50	50, 500, 5000
200	20	7, 70, 700
200	50	20, 200, 2000
400	20	4, 40, 400
400	50	8, 80, 800
400	150	40, 400, 4000
1000	20	2, 20, 200
1000	50	4, 40, 400
1000	150	10, 100, 1000
1000	350	35, 350, 3500

particles with the detector are carried out using GEANT4 [Ago03]. The pipeline thereafter treats the simulated detector signals as if they were real data. Simulated collisions are superimposed over the signal to mimic additional inelastic pp interactions that would typically occur during the same bunch crossing, known as pileup.

Both sets of signal samples are required to pass through the triggers described in Section 4.3.1. Individual $\Phi \to 2X \to 2\mu 2j$ and $\Phi \to 2X \to 4\mu$ signal samples have their trigger efficiencies shown in Figure 4.6. In general, more events pass the trigger as m_Φ increases, as this increases the p_T of the individual muons which pass the associated trigger requirement more often. Larger m_X for fixed m_Φ also increases efficiency, as heavy objects travel less far and our efficiency drops with displacement. Larger $c\tau$ corresponds to a drop in efficiency as this increases the mean displacement and the L1 trigger uses the beam-spot as a constraint on the muon fits. One also notices the efficiency increase for the $\Phi \to 2X \to 4\mu$ samples over $\Phi \to 2X \to 2\mu 2j$, which occurs not only because there are two dimuons, but also due to advantageous mispairings of muons occurring at the trigger level. The combined L1 and HLT efficiency for all $\Phi \to 2X \to 2\mu 2j$ signal samples is shown plotted as a function of generated L_{xy} in Figure 4.7. One observes a strong efficiency dependence on displacement, which is due to the inability of the trigger to properly measure muon p_T for those that are not produced promptly. One also notes that the efficiency is substantially higher in 2018,

due to the loosened requirements covered in Section 4.3.1. The trigger is the largest source of efficiency loss in the analysis for both the BSM Higgs and HAHM models after the full set of selection requirements are imposed.

4.3.4 Background Monte Carlo Simulation

As stated in Section 4.2, the SM predicts no dimuon resonances which are measurably long-lived with a mass above 10 GeV, the threshold used in this analysis. The expected background is entirely due to either reconstruction and/or instrumental failures, for which one does not expect reliable fidelity of Monte Carlo simulation. Therefore background simulation is used only to gain intuition into the general characteristics of events that are expected to fall in our signal region.

Background samples which were used in this analysis are listed in Tables 4.5 and 4.6 for 2016 and 2018 respectively. Each sample is produced following a similar pipeline as the one described for the signal samples and is documented in detail in Reference [Val19b].

Table 4.5: Names of simulated background samples corresponding to the 2016 data. The `RunIISummer16*` string stands for `RunIISummer16DR80Premix-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6`.

Process	Data set
$Z/\gamma^* \to ll$	/DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/RunIISummer16*-v1
	/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/RunIISummer16*_ext2-v1
$WW \to ll\nu\nu$	/WWTo2L2Nu_13TeV-powheg-herwigpp/RunIISummer16*-v1
WZ	/WZ_TuneCUETP8M1_13TeV-pythia8/RunIISummer16*-v1
	/WZ_TuneCUETP8M1_13TeV-pythia8/RunIISummer16*_ext1-v1
ZZ	/ZZ_TuneCUETP8M1_13TeV-pythia8/RunIISummer16*-v1
	/ZZ_TuneCUETP8M1_13TeV-pythia8/RunIISummer16*_ext1-v1
$t\bar{t} \to b\bar{b}ll\nu\nu$	/TTTo2L2Nu_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8/RunIISummer16*-v1
tW	/ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/RunIISummer16*_ext1-v1
$\bar{t}W$	/ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/RunIISummer16*_ext1-v1
$W \to l\nu$+jets	/WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16*-v1
QCD	/QCD_Pt-20toInf_MuEnrichedPt15_TuneCUETP8M1_13TeV_pythia8/RunIISummer16*-v1

Individual background samples are not produced with the same number of events as are expected in data. Therefore simulated background events are weighted such that they have an equivalent integrated luminosity as that expected in data. Each event weight (w) is defined as the ratio of the integrated luminosity of the corresponding LHC data set ($\mathcal{L}_{\text{int}}$)

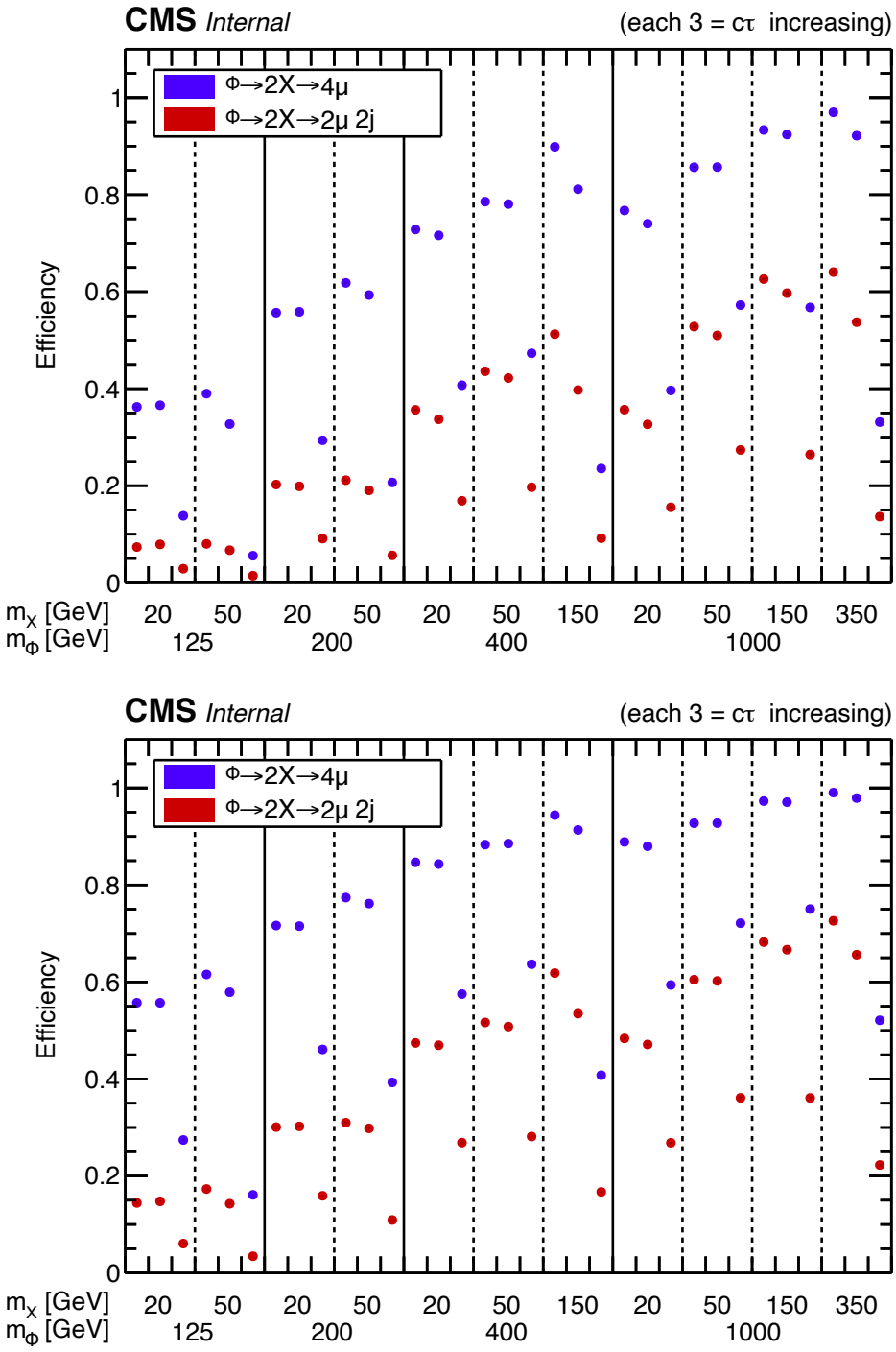


Figure 4.6: Trigger efficiencies for all $\Phi \to 2\text{X} \to 2\mu 2j$ (red) and $\Phi \to 2\text{X} \to 4\mu$ (blue) signal samples. Efficiency is defined as the fraction of events that pass the trigger paths for 2016 (top) and 2018 (bottom). The $c\tau$ values (not labeled) correspond to those listed in Table 4.4 and are shown in increasing order (i.e. shortest to longest).

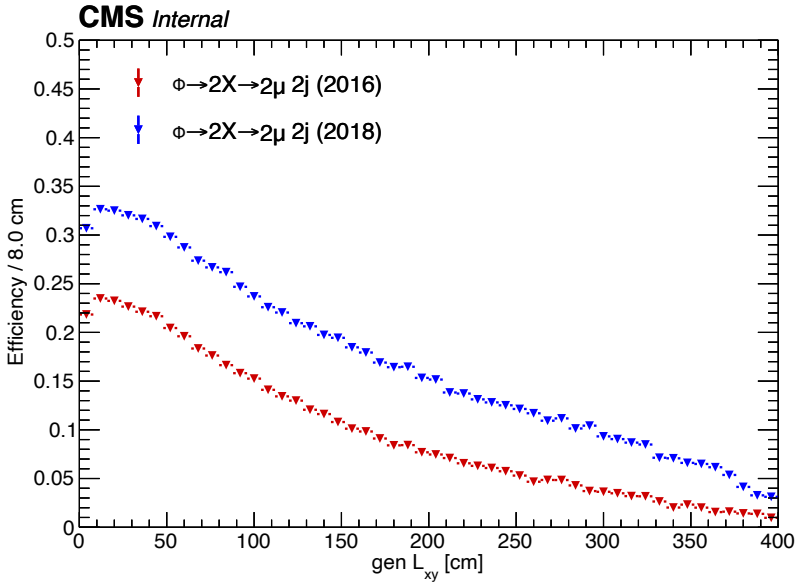


Figure 4.7: Trigger efficiency as a function of generated L_{xy} combining all $\Phi \to 2\text{X} \to 2\mu 2j$ signal samples. Efficiency is defined as the fraction of signal events that fired the 2016 (red) and the 2018 (blue) trigger paths described in Section 4.3.1.

Table 4.6: Names of simulated background samples corresponding to the 2018 data. The `RunIIAutumn18*` string stands for `RunIIAutumn18DRPremix-102X_upgrade2018_realistic_v15`.

Process	Data set
$\text{Z}/\gamma^* \to \text{ll}$	`/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18*-v2`
	`/DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18*-v1`
$\text{WW} \to \text{ll}\nu\nu$	`/WWTo2L2Nu_NNPDF31_TuneCP5_13TeV-powheg-pythia8/RunIIAutumn18*-v1`
WZ	`/WZ_TuneCP5_13TeV-pythia8/RunIIAutumn18*-v3`
ZZ	`/ZZ_TuneCP5_13TeV-pythia8/RunIIAutumn18*-v2`
$\text{t}\bar{\text{t}} \to \text{b}\bar{\text{b}}\text{ll}\nu\nu$	`/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/RunIIAutumn18*-v1`
tW	`/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIIAutumn18*_ext1-v1`
$\bar{\text{t}}\text{W}$	`/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIIAutumn18*_ext1-v1`
$\text{W} \to \text{l}\nu$+jets	`/WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18*-v2`
QCD	`/QCD_Pt-20toInf_MuEnrichedPt15_TuneCP5_13TeV_pythia8/RunIIAutumn18*-v1`

divided by the equivalent luminosity ($\mathcal{L}_{\text{int}}^{\text{eq}}$) of the simulated sample

$$w = \frac{\mathcal{L}_{\text{int}}}{\mathcal{L}_{\text{int}}^{\text{eq}}}. \tag{4.1}$$

The equivalent luminosity is obtained using the number of generated events (N_{gen}) and the process cross section (σ). Certain event generators such as MadGraph5_aMC@NLO 2.2.2 assign negative weights to events before this reweighting procedure takes place [AFF14]. This is done to account for the signs of contributing diagrams. Integrated luminosities for these samples are calculated looking at the number of positive (N_+) and negative (N_-) generated events

$$N_{\text{gen}} = N_+ + N_-, \qquad N_{\text{evts}} = N_+ - N_-. \tag{4.2}$$

Defining the fraction of negative events $f = N_-/N_{\text{gen}}$, one can then express the total number of events as

$$N_{\text{evts}} = (1 - 2f)N_{\text{gen}}. \tag{4.3}$$

This quantity is used when calculating the equivalent luminosity and reduces to the form used for non-negatively weighted events when $f = 0$. Only the Drell-Yan samples in 2016 include negatively weighted events, with $f = 0.136$ for `DYJetsToLL_M-10to50` and $f = 0.165$ for `DYJetsToLL_M-50`. Simulated samples are listed in Table 4.7, including their cross-section, number of generated events, and equivalent integrated luminosity. Some samples include kinematic selections applied at the generator level to provide simulated events more like that expected in the analysis signal region.

4.4 Event and Object Selection

The analysis targets dimuon decays reconstructed in the inner tracker and/or the muon system of CMS. The probed phase space includes both heavily boosted and relatively slow

Table 4.7: Simulated background samples used in the analysis. Listed are the simulated processes with their generator-level selection (if any), cross section (σ), number of generated events (N_{gen}), and equivalent luminosity ($\mathcal{L}_{\text{int}}^{\text{eq}}$) for 2016 and 2018. The equivalent luminosity is obtained using Equations 3.3 and 4.3.

			2016		2018	
Process	Selection	σ (pb)	N_{gen}	$\mathcal{L}_{\text{int}}^{\text{eq}}$ (fb^{-1})	N_{gen}	$\mathcal{L}_{\text{int}}^{\text{eq}}$ (fb^{-1})
$Z/\gamma^* \to ll$	$10 < m_{ll} < 50$ GeV	18445	30,935,823	1.22	39,433,275	2.14
	$m_{ll} > 50$ GeV	6077	122,547,040	13.5	100,194,597	16.5
$WW \to ll\nu\nu$		12.2	1,999,000	164	7,881,400	646
WZ		47.1	3,997,571	85	3,885,000	82
ZZ		16.5	1,988,098	120	1,979,000	120
$t\bar{t} \to b\bar{b}ll\nu\nu$		87.3	79,140,880	906	64,330,000	737
tW, $\bar{t}W$		35.8	13,885,924	388	17,335,000	484
$W \to l\nu$+jets		61527	29,804,825	0.48	71,122,242	1.15
QCD	$\hat{p}_T > 20$ GeV , $\lvert\eta(\mu)\rvert < 2.5$, $p_T(\mu) > 15$ GeV	302672	22,094,081	0.07	22,165,320	0.07

dimuons ($\gamma \in (0 - 25)$) in a large range of both masses ($10^1 - 10^3$ GeV) and displacements ($10^{-4} - 10^3$ m). Combinations of two types of reconstructed muons are used to form signal dimuon candidates which are covered in the following sections. Selected events are a subset of those that pass the trigger described in Section 4.3.1. Subsequent selection criteria were developed aiming to minimize SM background while maintaining sensitivity to this broad range of signal topologies. A preselection is first applied to the muons and the event. Selection for the muons is then applied which includes a match-and-replace procedure designed to replace muons reconstructed in the muon system with those found in the tracker. Muons are then paired and selection criteria are applied to the dimuons, yielding up to two dimuons per event.

4.4.1 Muon Preselection

Various algorithms have been developed for muon reconstruction as covered in Section 3.2.7. However, no single algorithm gives optimal performance over the full range of displacements used by this analysis.

The probed dimuon lifetimes allow for muons production outside of the tracker volume,

and consequently, one variety of muons used in this analysis are formed using only hits in the muon system. Muons reconstructed in this way are referred to as *standalone-muons* (STA) and can be reconstructed multiple meters away from the interaction point. The standard standalone reconstruction algorithm can be run without a beam-spot constraint, an attractive feature for a displaced analysis, however previous studies found that the p_{T} measurement of highly displaced tracks it produced was biased to be anomalously low for generator level quantities [Mat11]. Recent attempts at mitigating this bias led to the production of the displaced standalone (DSA) reconstructor [AC15]. This algorithm replaces the default seed generator with that used for cosmic muon reconstruction, which forms seeds from muon segments with the highest y and builds the tracks downward. This led to an improved p_{T} resolution for displaced muon tracks over previous reconstruction algorithms observed by both the authors as well as our own analysis group on our generated signal samples [Das19]. The collection of STA muons built using the DSA reconstruction algorithm are therefore used at the first stage of the dimuon reconstruction.

The initial list of STA muons must first pass two quality requirements ensuring they are sufficiently well measured. Specifically, the muons must have more than 12 combined CSC and/or DT recHits, and have muon segments measured at more than one station, as listed in Table 4.8. Both complementary requirements enforce that there are sufficient measurement points needed to accurately recover the muon p_{T}. Comparisons of STA muons which pass or fail the $N(\text{CSC+DT hits}) > 12$ requirement are shown in Figure 4.8. For muons failing this requirement, the p_{T} is very often underestimated and the correlated charge measurement is found to be incorrect nearly 40% of the time. Applying these requirements removes no more than 2% of our reconstructed signal muons. Preselected STA muons are then used as input to the remainder of the analysis.

4.4.2 Event Preselection

Only events passing both the L1 and HLT triggers described in Section 4.3.1 are available for offline analysis. This section describes preselection requirements imposed on each event.

STA preselection	Description
N(CSC+DT hits) > 12	Combined number of CSC/DT hits used in the fit must exceed 12
N(CSC+DT stations) > 1	More than one CSC/DT station used to measure the track

Table 4.8: Preselection requirements for STA muons before they are potentially replaced by their matching counterparts in the tracker.

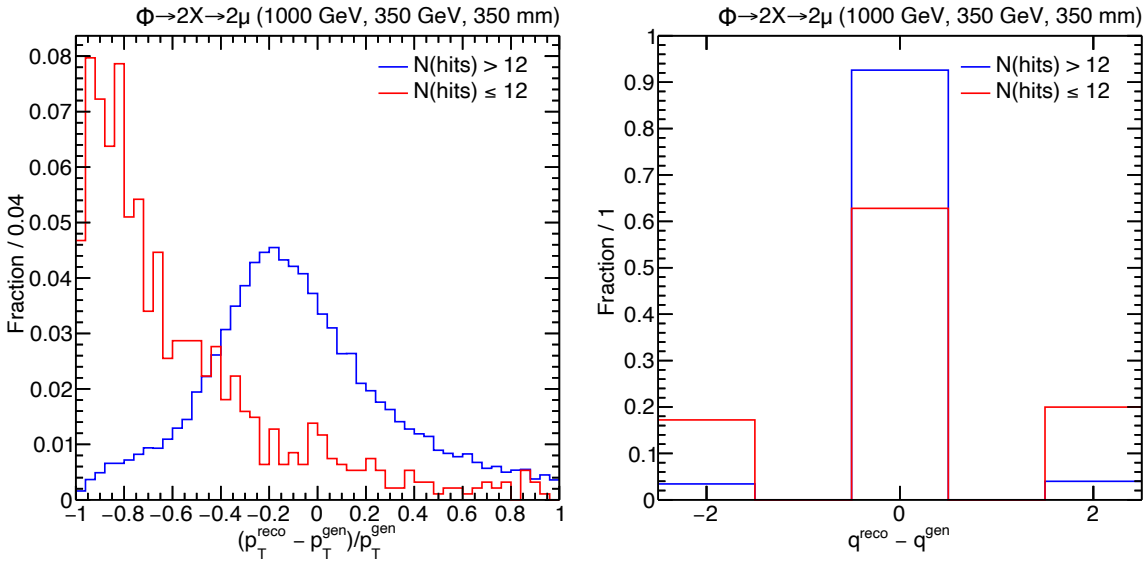


Figure 4.8: Distributions showing the effect of the N(CSC+DT hits) > 12 requirement on preselected STA muons. Both muon p_{T} resolution (left) and the difference between the reconstructed and true charge (right) are shown for the $\Phi \to 2\mathrm{X} \to 2\mu 2j$ signal sample with $m_{\Phi} = 1000$ GeV, $m_{\mathrm{X}} = 350$ GeV, and $c\tau = 350$ mm. Both plots compare the population of muons that pass (blue) and do not pass (red) the N(CSC+DT hits) > 12 requirement. All are normalized to unit area. The overflow bin is not used.

It is typical to require that objects reconstructed offline include a subset corresponding to trigger level objects which induced the event readout. This is done to avoid pathologies associated with studying "volunteer" events from other triggers which have event topologies unlike those which one is aiming to probe. Therefore the preselected STA muons are matched to the Level-2 (HLT) muons that caused the event to be stored. Matching of objects within CMS is often done using the quantity

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}, \tag{4.4}$$

where $\Delta\eta$ and $\Delta\phi$ are the differences between the two objects pseudorapidities and azimuthal angles, respectively. We take this approach, requiring that each Level-2 muon that triggered the event has an associated STA muon found within a cone of $\Delta R < 0.4$. Only a subset of the preselected STA muons are allowed to be matched to: those with $p_{\mathrm{T}} > 10$ GeV, and $|\eta| < 2.5$. These additional requirements are imposed to suppress accidental matches from poor-quality STA muons. The HLT matching requirement removes between $1-5\%$ of signal dimuons, with losses primarily in low-mass samples where one of the two STA muons has $p_{\mathrm{T}} < 10$ GeV.

To suppress cosmic ray muons which arrive asynchronously with the proton-proton collisions, we require that each event have at least one sufficiently well-reconstructed primary vertex. Specifically, a primary vertex must be formed from at least four associated tracks and be within $|z| < 24$ mm and $\sqrt{x^2+y^2} < 2$ mm of the interaction point. Within CMS, this requirement is known as the `PrimaryVertexFilter` and suppresses cosmic events recorded when there was no proton-proton collision.

Although a single cosmic muon can be mistakenly reconstructed as two STA muons, multiple cosmic muons provide an even richer breeding ground for combinatorial failure modes. Cosmic rays can occasionally provide a multitude of muons that pass through the CMS detector within a time window of a single bunch crossing. An example of this type of event is given in Figure 4.9. In these cases, dimuons can be found not only back-to-back from each other (where they are rejected by requirements on α) but can also be paired such that they are parallel. Dimuons formed in this way should be found to have an invariant mass near zero, however anomalous behavior of the vertex fitter gives them large masses above our requirement [Val19b]. These events are suppressed by globally looking at the number of parallel pairs that could be formed from the preselected STA muons. Specifically, we require $N(\text{parallel pairs}) < 6$ among preselected STA muons with $p_{\mathrm{T}} > 5$ GeV. A pair is considered parallel if $\cos\alpha > 0.99$. The p_{T} requirement is imposed to suppress efficiency loss from pileup STA muons present in real collisions. Only 0.2% of the 2μ and 0.4% of the 4μ signal samples fail the requirement imposed on the number of parallel pairs. The event

preselection requirements are summarized in Table 4.9.

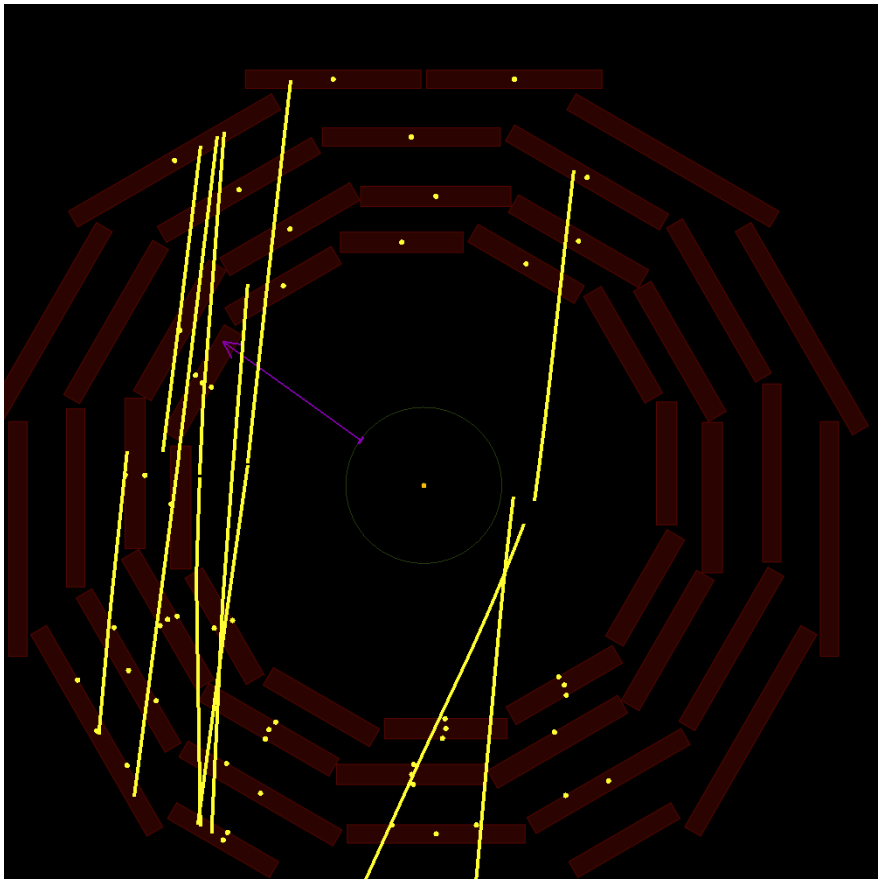


Figure 4.9: Display of $r - \phi$ for a cosmic shower event identified with event, run, and lumi numbers of 283283, 2947743777, and 1734, respectively. Most reconstructed STA muons (yellow) are roughly parallel to each other, a feature used for the suppression of this type of background.

4.4.3 STA Muon Replacement

The inclusion of tracker information substantially improves measurements of both the muon position and momentum. Therefore upgrading STA muon objects to those that also include tracker information, if possible, strongly increases signal sensitivity. For each preselected STA muon described in Section 4.4.1, we look for a matching counterpart reconstructed with tracker information. Specifically, we look for global and/or tracker muon objects that correspond to the STA muon. The collection of tracker muons used in this analysis are "arbitrated", requiring that no two tracker muons are formed using overlapping muon seg-

Event preselection	Description
HLT matching	HLT muons must match preselected STA muons with $p_{\mathrm{T}} > 10$ GeV, $\lvert\eta\rvert < 2.5$
`PrimaryVertexFilter`	Event must have a well measured primary vertex
N(parallel pairs) < 6	Less than six parallel pairs are formed by preselected STA muons with $p_{\mathrm{T}} > 5$ GeV

Table 4.9: Preselection event requirements.

ments [Mul08]. The analysis uses these lower-level muon objects as opposed to PF muons after it was found that the additional PF requirement degraded performance with muon displacement [Nas21a].

Global/tracker muons which are to be matched to the STA muons are collectively known as *tracker muon system* (TMS) muons. The STA to TMS matching algorithm was developed using simulated signal and background in addition to signal-free regions of 2016 data. TMS muons can be matched to in one of two ways:

- **Segment-based match**: More than two thirds of muon segments used to build the STA muon are also used in the TMS muon.
- **Proximity match**: $\Delta R = \sqrt{(\eta_{\mathrm{hit}} - \eta_{\mathrm{pca}})^2 + (\phi_{\mathrm{hit}} - \phi_{\mathrm{pca}})^2} < 0.1$.
  - $\eta_{\mathrm{hit}}/\phi_{\mathrm{hit}}$ correspond to the positions of the innermost hit of the STA muon.
  - $\eta_{\mathrm{pca}}/\phi_{\mathrm{pca}}$ correspond to the positions of the point of closest approach (pca) of the TMS muon with respect to the innermost hit of the STA muon.

Additional details of the matching procedure are given in References [Val19b] and [Das19]. STA muons which are matched to TMS muons are then replaced. Because a substantial portion of our background is truly prompt (e.g. Drell-Yan), replacing poorly measured STA muons with well measured TMS muons acts as a strong background suppressant. In the entire 2016 data set, simulated proton-proton background dimuons formed from pairs of STA muons becomes suppressed by over four orders of magnitude through the replacement procedure [Val19b].

In contrast, simulated signal dimuons formed in this way are relatively unharmed. Figure 4.10 shows the population of signal dimuons that remain after the replacement procedure

takes place. One sees that the population of signal dimuons decaying outside the active tracker region ($L_{xy} > 65$ cm) in large part remain as they were. Outside the tracker, roughly 90% of signal dimuons remain intact, indicating that the probability of accidentally matching to a TMS muon (and improperly removing an STA muon) is roughly $1 - \sqrt{0.9} \sim 5\%$.

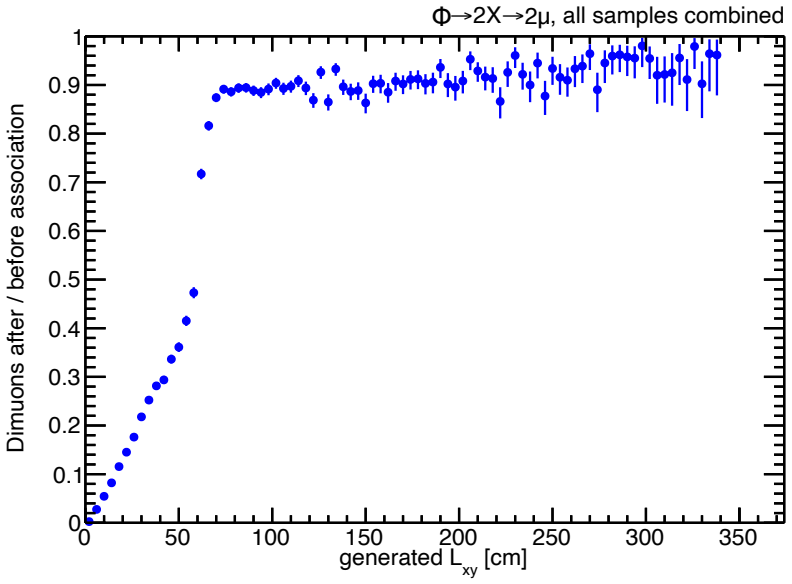


Figure 4.10: Ratio of signal matched STA-STA dimuons found after and before the STA replacement procedure is performed as a function of L_{xy}. STA muons are considered matched to signal dimuons if both reconstructed muons are found within a cone of $\Delta R < 0.2$ of the generator level muons. The ratio is calculated combining all generated $\Phi \to 2\text{X} \to 2\mu 2j$ signal samples.

4.4.4 Muon Selection

Both STA muons resulting from the displaced standalone algorithm and TMS muons used for displaced searches did not have a standard set of requirements in place prior to this search. Therefore a custom set of selection criteria were developed for these objects. Both STA and TMS muons differ widely in their resolving power and are therefore optimized independently. For example Figure 4.11 compares the difference in their transverse momentum resolution.

Some selection criteria, however, remain the same irrespective of the muon type, namely two requirements on their transverse momentum measurement. Both types of muons must

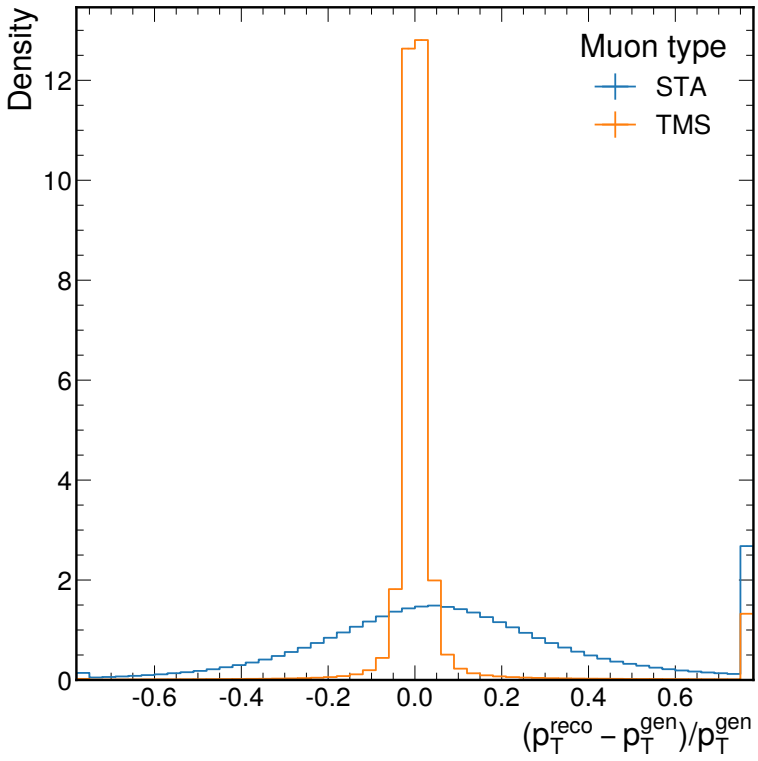


Figure 4.11: STA (blue) and TMS (orange) muon momentum resolution calculated combining all generated $\Phi \to 2X \to 2\mu 2j$ signal samples. Includes all signal-matched reconstructed muons before the replacement procedure is applied. Muons are matched to generated muons requiring $\Delta R < 0.2$. Overflows are shown in the outermost bins.

have $p_{\mathrm{T}} > 10$ GeV and $\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} < 1$, where $\sigma_{p_{\mathrm{T}}}$ is the uncertainty on the transverse momentum provided by the track fit. The former requirement suppresses low mass background and synchronizes the analysis with the thresholds set by the HLT (23 GeV and 28 GeV for 2016 and 2018, respectively). The latter requirement on the relative uncertainty of the track momentum prohibits grossly mismeasured muons from being used in the analysis. The requirement primarily affects STA muons – removing less than 1% of those with low p_{T} and up to 10% of those with p_{T} of a few hundred GeV. Figure 4.12 compares muons passing and failing this requirement measured in our signal sample containing the largest proportion of high p_{T} muons, for which the track's charge is incorrectly assigned roughly 25% of the time. Less than 1% of TMS from any signal sample are lost due to this requirement.

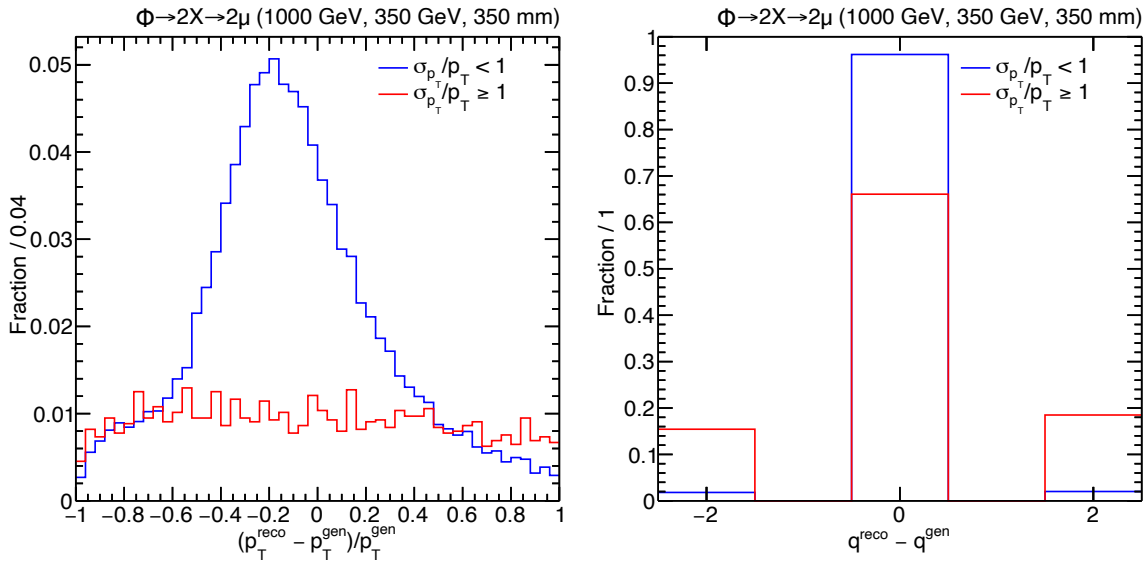


Figure 4.12: Track p_{T} resolution (left) and charge mis-measurement (right) of signal matched preselected STA muons. Tracks with $\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} < 1$ are shown in blue and those which do not pass the requirement are shown in red. Each distribution shows the 2016 $\Phi \to 2\mathrm{X} \to 2\mu 2j$ signal sample with $m_{\Phi} = 1000$ GeV, $m_{\mathrm{X}} = 350$ GeV, and $c\tau = 350$ mm normalized to unit area.

4.4.4.1 STA Muon Selection

Preselected STA muons that are not replaced following the matching procedure described in Section 4.4.3 then undergo further selection described here. Selection criteria for STA muon

were developed using simulated signal and background samples in addition to background-free regions of data. The criteria developed for STA muon identification using the displaced standalone algorithm now serve as the standard recommended by the muon POG for these objects [Valc]. The development of these criteria is described in References [Val19b] and [Das19].

Poorly measured STA muons are further suppressed using more stringent track recHit requirements for muons measured in the barrel region. Specifically, we require $N(\text{DT hits}) > 18$ if the muon is measured only in the barrel region. The quality of selected muons is further improved by requiring $\chi^2_{\text{track}}/\text{dof} < 2.5$, calculated from the track fit. Both quality requirements individually lose roughly 5% of integrated signal yield.

The $r - z$ cross-sectional area of the muon system is over an order of magnitude larger than that of the tracker. This causes the cosmic muon background to be significantly larger for unreplaced STA muons than for TMS muons with a component measured in the tracker. As mentioned in Section 4.2.3, muons of this nature can be suppressed by using both timing and their direction of travel. Each muon chamber hit is assigned a time that corresponds to the difference from the nominal arrival time of a muon traveling at the speed of light from the interaction point to the muon system. Muons built from these hits are then assigned a time $t_{\text{in-out}}$ corresponding to their weighted sum. The timing distribution of signal muons, shown in Figure 4.13, has a mean value near 0 ns with a resolution of $1.4 - 16$ ns. A requirement of $|t_{\text{in-out}}| < 12$ ns maintains 99.9% of signal dimuons, while suppressing handfuls of events found in various control regions passing all other selection criteria.

The timing information of the muon is also correlated among chambers to determine its direction of flight. If a dimuon is formed by a single cosmic muon, one of the reconstructed "muons" will be found to be going the wrong way. A muon that travels from the muon system to the interaction point will have muon hits recorded at times differing from those which travel the other way. The complementary time $t_{\text{out-in}}$ describes the time relative to hits recorded from a muon traveling outwards-in. The muon direction is determined by comparing the uncertainties in both $t_{\text{in-out}}$ and $t_{\text{out-in}}$, denoted as $\sigma_{t_{\text{in-out}}}$ and $\sigma_{t_{\text{out-in}}}$, respectively. Muons

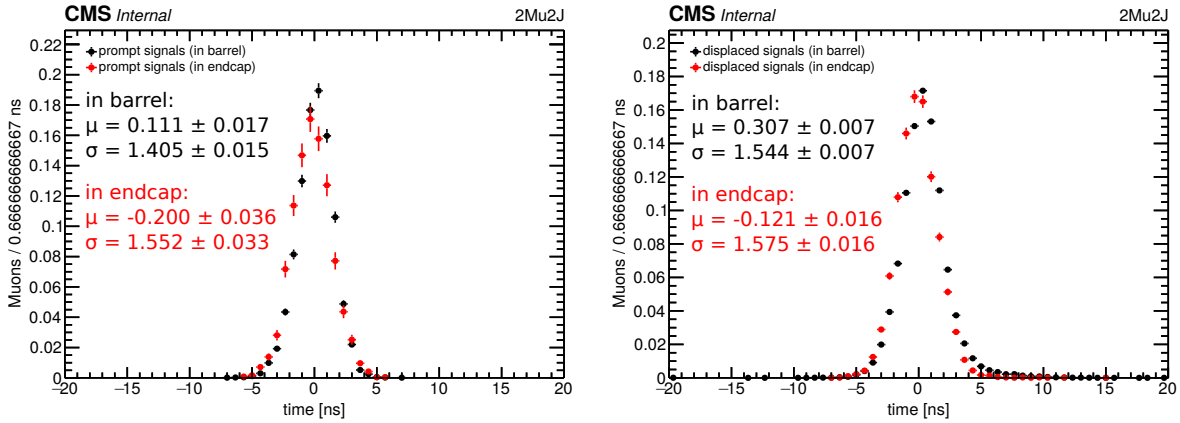


Figure 4.13: Distributions of $t_{\text{in-out}}$ for STA muons found in the 2018 $\Phi \to 2\text{X} \to 2\mu 2j$ signal samples with the shortest (left) and longest (right) lifetimes and combining all mass points. Signal samples are listed in Table 4.4. The legend provides the parameters of a Gaussian fit for distributions found in the barrel (black) and endcap (red).

with $\sigma_{t_{\text{in-out}}} < \sigma_{t_{\text{out-in}}}$ are assigned as outward going, while muons with $\sigma_{t_{\text{in-out}}} > \sigma_{t_{\text{out-in}}}$ are assigned as inward going. If less than two timing measurements are used in calculating the uncertainties, the direction is left unassigned. This last case is very rare in signal samples and occurs at a rate of roughly one per 10^5 muons. Each STA muon is required to have traveled outward from the interaction point as measured in this way. This requirement removes no more than 1% of STA muons from signal events, and strongly suppresses cosmic muons found in control regions enriched in such events described in Section 4.4.6.1.

Timing information can also be used to reject events for which the L1 trigger had pre-fired. In some cases, noise present in the electronics used in the L1 trigger can cause it to fire prematurely. Corresponding offline muon system information is also read out, however, it is paired with tracker information from the earlier bunch crossing. Because the time measured at the collision point is early, offline muon hits are assigned a value of $t_{\text{in-out}}$ corresponding to that of one bunch crossing (25 ns). Additionally, the tracker component of these STA muons is missing, which causes them to not be replaced according to the procedure described in Section 4.4.3. Timing requirements, namely $|t_{\text{in-out}}| < 12$, suppress this type of instrumental background in addition to that of cosmic muons. A summary of requirements placed on STA muons is given in Table 4.10. Notably, we do not use an isolation requirement for STA

muons, however, this variable is used when evaluating the character of events that pass our full selection shown in Section 4.7.

STA selection	Description
$p_{\mathrm{T}} > 10$ GeV	Muon transverse momentum must exceed 10 GeV
$\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} < 1$	Fractional error on the transverse momentum must be less than 100%
$N(\mathrm{DT\ hits}) > 18$	Muons measured only in the barrel must be fit using more than 18 hits
$\chi^2_{\mathrm{track}}/\mathrm{dof} < 2.5$	Normalized chi-squared of the track fit must be less than 2.5
$\lvert t_{\text{in-out}}\rvert < 12$ ns	Timing information must be consistent within 12 ns of an outward going muon
outward going	Muons must be reconstructed as travelling outward from the interaction point

Table 4.10: STA muon selection. Selection is applied to preselected STA muons which are not replaced following the procedure described in Section 4.4.3.

4.4.4.2 TMS Muon Selection

STA muons with a matching tracker component undergo a parallel selection procedure to that described in Section 4.4.4.1. TMS muons first undergo identical $p_{\mathrm{T}} > 10$ GeV and $\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} < 1$ requirements applied to the momentum obtained by the "Tune P" algorithm described in Section 3.2.7.2. Muons are then subjected to quality requirements analogous to those used for unmatched STA muons.

There are many standard identification (ID) criteria that have been developed for global and tracker muons, however, each was developed under the assumption that the muon is produced promptly. Therefore it was necessary for our analysis to investigate the performance of these commonly used ID criteria for our displaced signal. A summary of performance for various IDs is shown in Figure 4.14 as a function of d_0 and d_z. Our analysis determined the tracker muon ID requirement was optimal, as it performs the best as a function of displacement. We supplement this criterion by additionally requiring that at least two muon segments found in the muon system are matched to the tracker muon. The requirement that the muon object is a tracker muon, and that $N(\mathrm{matched\ segments}) > 1$ removes less than 1% of the combined 2016 $\Phi \to 2\mathrm{X} \to 2\mu 2j$ signal samples, while suppressing roughly 10% of mismeasured background measured in a control region. Notably the requirement that the

muon is reconstructed as a tracker muon does not preclude it from also being a global muon. Components of various IDs were also investigated, such as the segment compatibility and the global muon fit χ^2, each of which gave marginal signal-over-background discrimination power and therefore were not used.

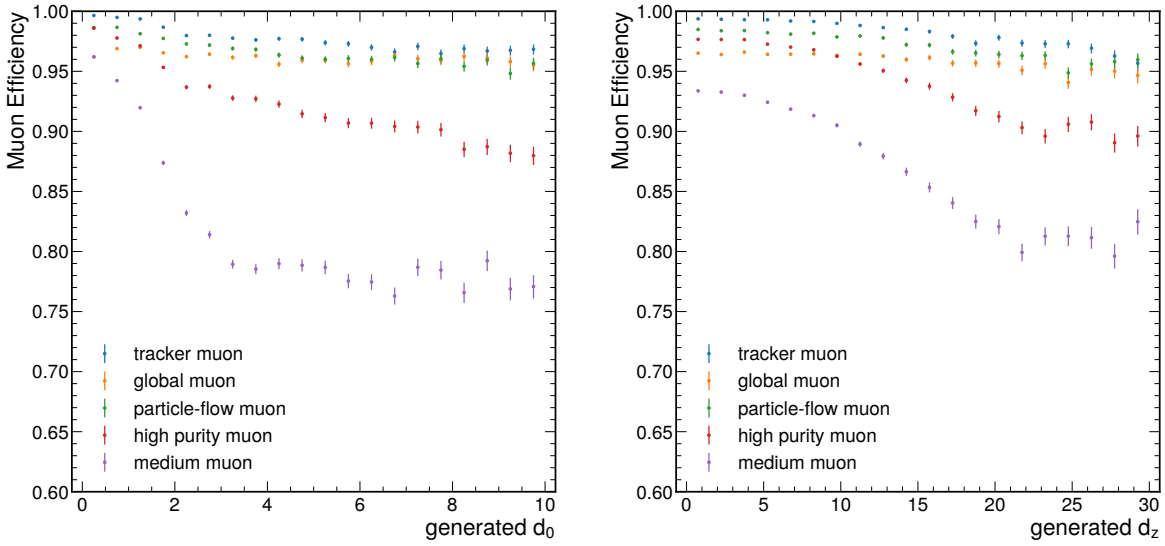


Figure 4.14: Muon identification efficiencies for combined 2016 $\Phi \rightarrow 2\mathrm{X} \rightarrow 2\mu 2j$ signal samples as a function of generated d_0 (left) and d_z (right). The denominator is defined as all generator-matched TMS muons found within a cone of $\Delta R < 0.2$. The last bin includes all events in the overflow region.

As described in Section 4.2.2, QCD makes up a substantial portion of the background of the search. Our analysis suppresses many of these events using a relative isolation variable, which quantifies the magnitude of activity around each muon. This is done by first identifying the number of tracks and/or particles that neighbor the muon within a cone of some size $\Delta R < \mathcal{R}$ (excluding the muon itself). Each object in the cone then has either its transverse momentum or transverse energy summed together. This sum is called the *isolation variable*. Because the activity of nearby particles is less relevant as the p_T of the muon increases, it is typical to use a *relative isolation variable*, which is this sum normalized by the muon p_T.

We have studied several variations of relative isolation to find that which optimally covers the range of topologies used in this analysis [Nas20a]. Isolation variables based on calorimeter

information were found to perform worse for signals with overlapping displaced signal dijets than isolation calculated using only tracker tracks. Figure 4.15 shows a diagram of this topology present in the $\Phi \rightarrow 2X \rightarrow 2\mu 2j$ signal samples, where each event contains both a displaced dimuon and dijet. Low p_T samples (e.g. the $\Phi \rightarrow 2X \rightarrow 2\mu 2j$ sample with $m_\Phi = 125$ GeV and $m_X = 50$ GeV) often only pass the trigger if the Higgs is produced with non-zero p_T. This causes the long-lived X particles to be collimated and results in energy deposits from the displaced jet in the calorimeters which overlap with the displaced muon trajectory. These deposits are added to the isolation sum which contributes to the rejection of the signal muon. On the other hand tracker isolation uses only tracks near the primary vertex so the energy from the neighboring jets are not included in the isolation sums.

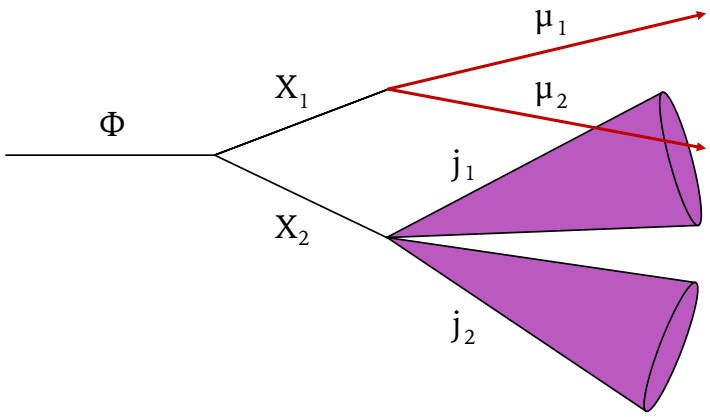


Figure 4.15: Diagram depicting a signal jet j_1 overlapping with a signal muon μ_2.

The generated dimuon efficiency of a particle-flow relative isolation requirement of < 0.15 is shown in Figure 4.16. The BSM sample with $m_\Phi = 125$ GeV and $m_X = 50$ GeV is most sensitive to this requirement, which was also found to contain the highest population of signal muons with nearby jets after the trigger is applied. The efficiency dependence on nearby jets is quantified in Figure 4.17, which shows the dimuon isolation efficiency as a function of the closest ΔR signal muon-jet pair. Particle-flow isolation efficiency drops substantially when a signal jet approaches the size of the cone used for the calculation. This same behavior is observed for ECAL and HCAL only isolation variables, as well as individual components

used for calculating the particle-flow isolation variable.

The larger efficiency obtained using tracker isolation is a result of additional geometric constraints placed on tracks that enter the isolation sum, which prevents decay products of the second dimuon from being included. By default, tracker isolation includes tracks satisfying:

- $|\Delta z| < 2$ mm between the reference point of the track and that of the muon;
- $d_0 < 1$ mm of the track with respect to the interaction point.

Following the discovery of the source of this efficiency loss, these parameters were adjusted to improve signal-over-background rejection. The optimized region extends the default region, and includes tracks satisfying:

- $|\Delta z| < 5$ mm between the reference point of the track and that of the muon;
- $d_0 < 2$ mm of the track with respect to the interaction point.

This allows for significantly stronger background rejection (roughly 40% of data in a control region), while only decreasing signal efficiency by roughly 1% – still significantly higher than that obtained using particle-flow isolation.

Tracker isolation is further adjusted by the vetoing of the paired signal muon from the isolation sum. Muons that are reconstructed and form a dimuon as discussed in Section 4.4.5 have the muon they are paired with removed from the isolation sum if they would otherwise contribute. This primarily affects the signal efficiency of signal samples with highly boosted muons. A comparison of the dimuon efficiency using the default tracker isolation and that using the fully optimized isolation variable is shown in Figure 4.18.

The analysis selects for isolated TMS muons using the adjusted relative tracker isolation variable ($I^{\text{rel}}_{\text{trk}}$) discussed in this section. The strong sensitivity of our signal efficiency to the isolation variable requires figure of merit optimization as discussed in Section 4.4.6.3. The optimal selection was determined to be $I^{\text{rel}}_{\text{trk}} < 0.075$.

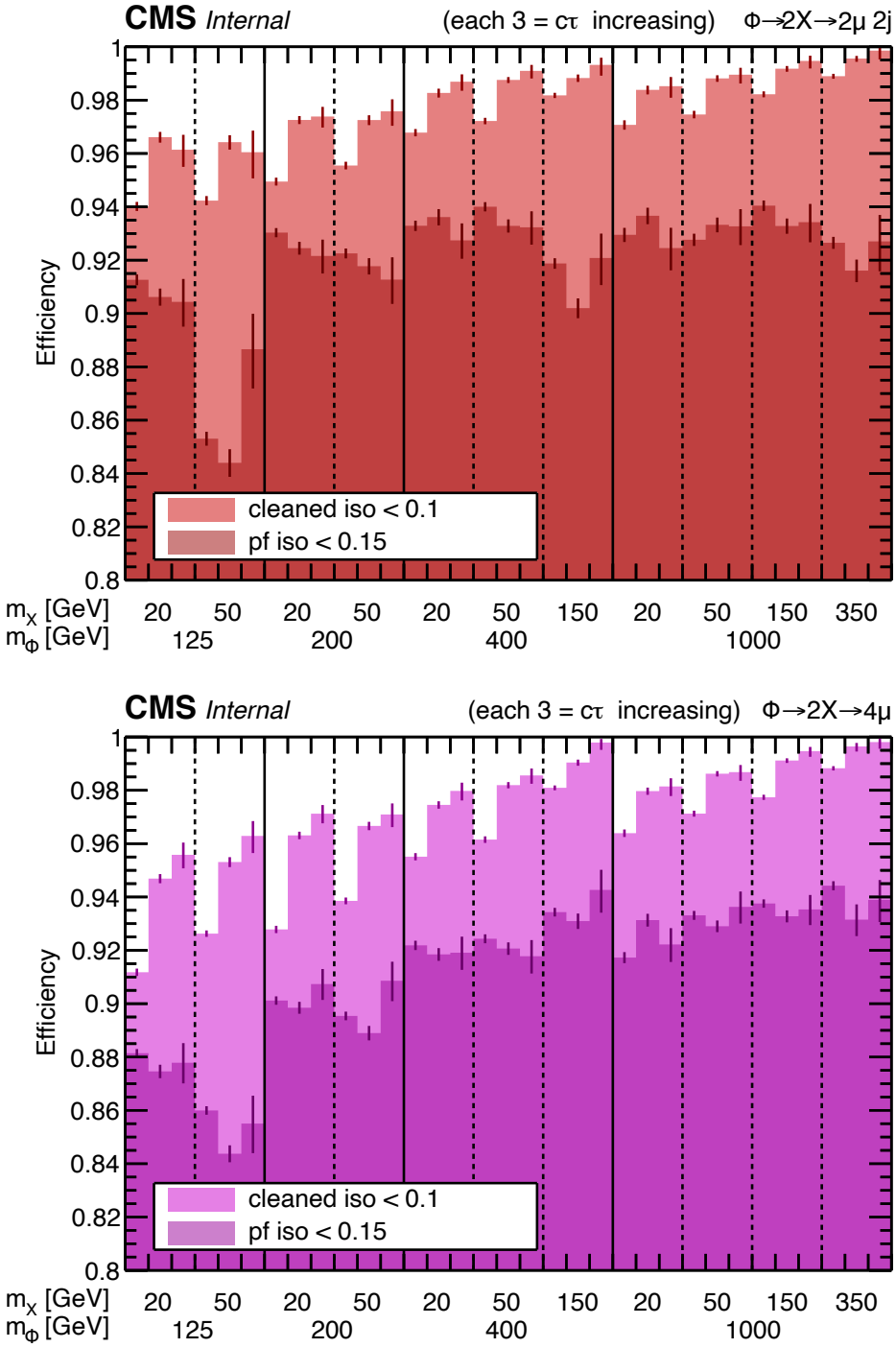


Figure 4.16: Generated dimuon isolation efficiencies in the $\Phi \rightarrow 2X \rightarrow 2\mu 2j$ (top) and $\Phi \rightarrow 2X \rightarrow 4\mu$ (bottom) signal samples. Both TMS muons belonging to the generated dimuon are required to have either a relative particle-flow isolation of < 0.15 (dark red and purple), or a relative tracker isolation of < 0.1 (light red and purple). The tracker isolation has the paired signal muon removed from the sum ("cleaned") and is calculated using the adjusted parameters detailed in Section 4.4.4.2.

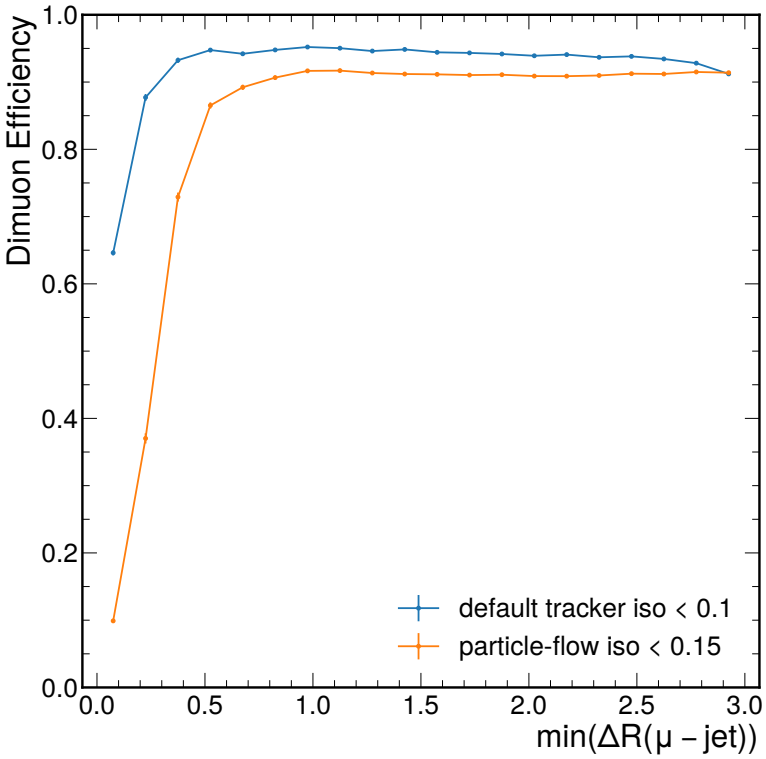


Figure 4.17: Generated dimuon efficiency for all $\Phi \to 2\mathrm{X} \to 2\mu 2j$ combined as a function of the closest ΔR muon-jet pairs. Both requirements: relative tracker isolation < 0.1 (blue), and relative particle-flow isolation < 0.15 (orange), remove approximately (within 1%) the same background from a control region in data. Both isolation variables sum in a cone of $\Delta R < 0.3$. Tracker isolation performs better overall, but most notably in regions containing overlapping signal jets.

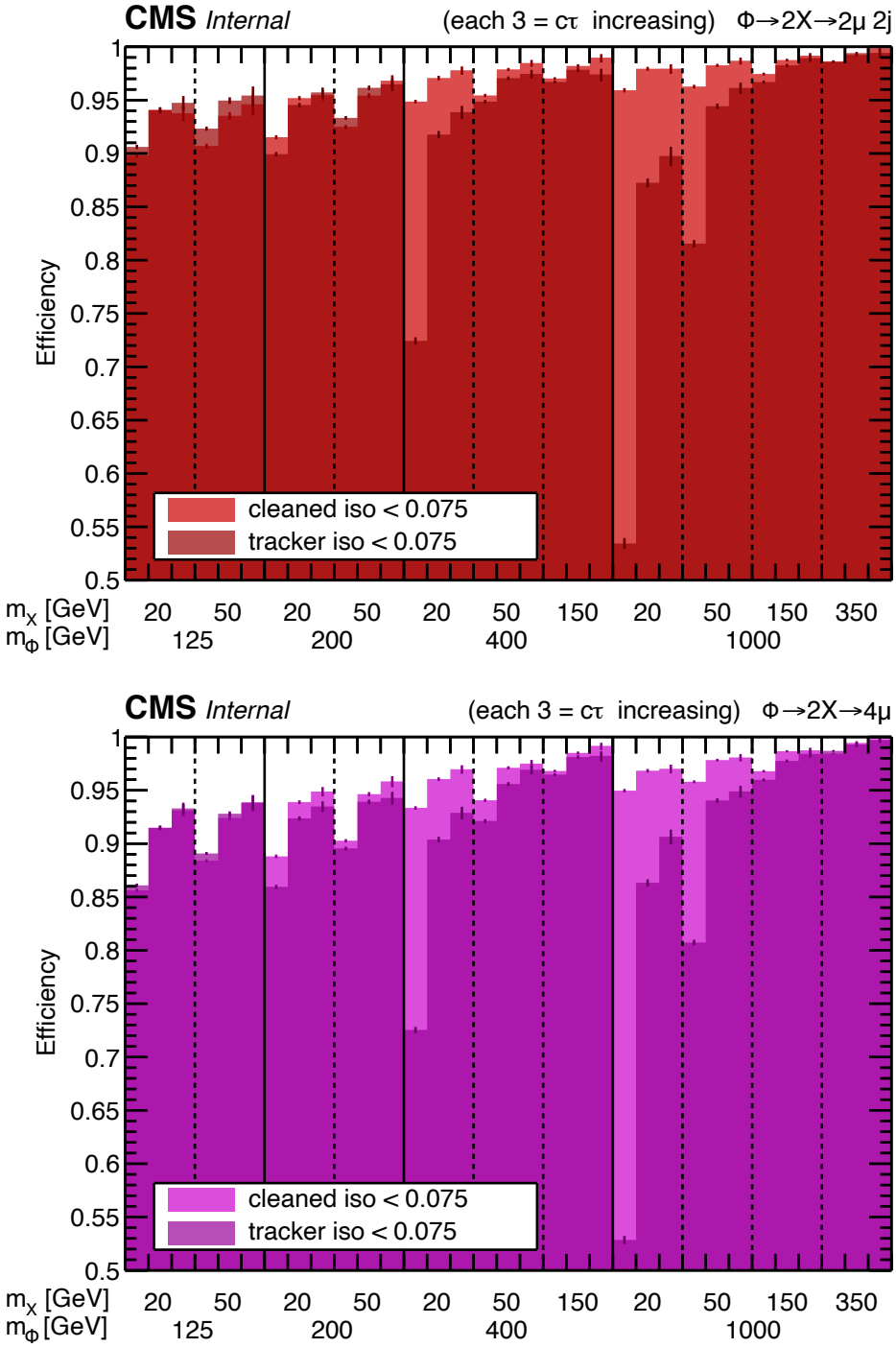


Figure 4.18: Generated dimuon isolation efficiencies of the $\Phi \rightarrow 2\text{X} \rightarrow 2\mu 2j$ (top) and $\Phi \rightarrow 2\text{X} \rightarrow 4\mu$ (bottom) signal samples. Default tracker isolation (dark red and purple) performs worse than cleaned tracker isolation (light red and purple), which removes the paired muon from the isolation sum, in addition to adjusting the geometric parameters discussed in Section 4.4.4.2. This gain is primarily evident in signal samples which are highly collimated. Both isolation sums are calculated in a cone of $\Delta R < 0.3$.

As discussed in Section 4.1, individual muons can be used to select for dimuon displacement through their impact parameter d_0. Our analysis uses the transverse impact significance d_0/σ_{d_0}, which is the measured d_0 divided by its measurement error. Due to the range of large lifetimes probed by the analysis, a figure of merit is also used here to select the optimal requirement. This requirement is found to be $d_0/\sigma_{d_0} > 6$. The resolution of the muon impact parameter is $\mathcal{O}(10\ \mu\text{m})$, which translates to a minimum d_0 requirement of roughly 60 μm. A summary of requirements placed on TMS muons is given in Table 4.11.

TMS selection	**Description**
$p_{\text{T}} > 10$ GeV	Muon transverse momentum must exceed 10 GeV
$\sigma_{p_{\text{T}}}/p_{\text{T}} < 1$	Fractional error on the transverse momentum must be less than 100%
is a tracker muon	Muon must be reconstructed as a tracker muon
$N(\text{matched segments}) > 1$	The number of segments matched to the tracker muon must exceed one
$I^{\text{rel}}_{\text{trk}} < 0.075$	Relative tracker isolation must be less than 0.075
$d_0/\sigma_{d_0} > 6$	The impact parameter significance must be greater than six

Table 4.11: TMS muon selection. Selection is applied to TMS muons which replaced preselected STA muons following the procedure described in Section 4.4.3.

4.4.5 Dimuon Selection

We allow combinations of both types of muons to form reconstructed signal candidates. This yields three distinct types of dimuons: STA-STA, STA-TMS, and TMS-TMS. Each dimuon category is depicted in Figure 4.19. Figure 4.20 shows where each of the categories are relevant as a function of generated L_{xy}. One sees that each dimuon category has a region in L_{xy} over which it is used to reconstruct signal most often. The tracker itself ends at around 60 cm and accounts for the sharp drop there seen in both the TMS-TMS and STA-TMS categories. Dimuons found in the aforementioned categories outside the tracker volume are caused by accidental matching to muons originating from the underlying event. The greatly differing resolution between constituent muons necessitates individual optimization of each dimuon category. This section covers the formation of candidate dimuons and the selection applied to those that fall within our signal region.

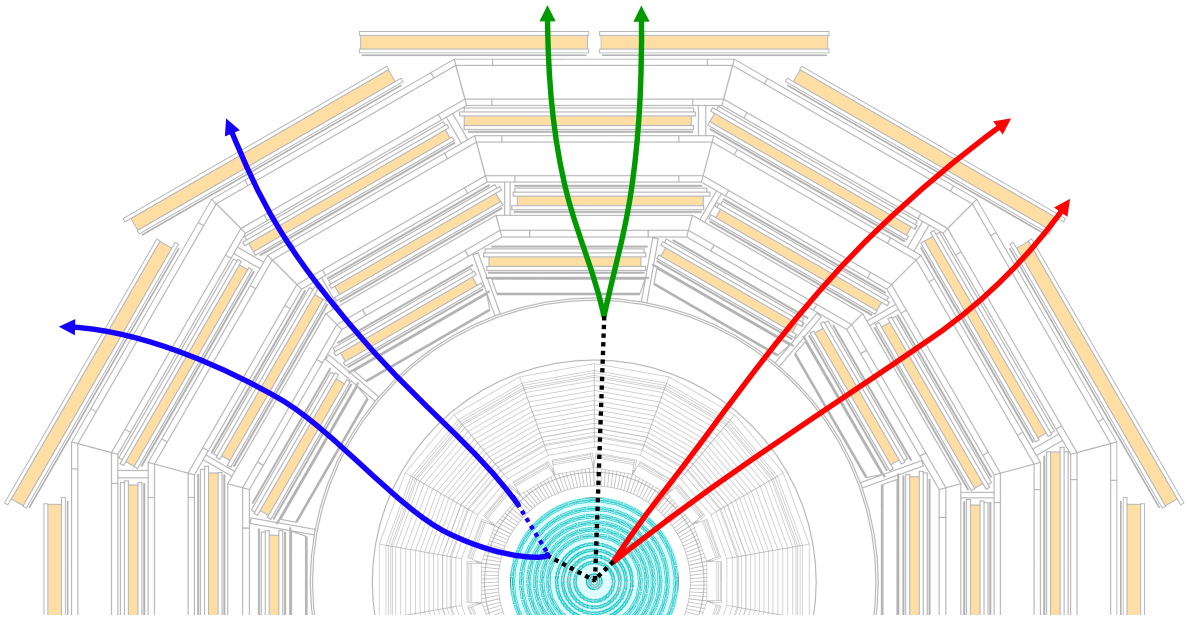


Figure 4.19: Transverse cross-section of the CMS detector showcasing the three types of reconstructed dimuons. STA-TMS (blue), STA-STA (green) and TMS-TMS (red).

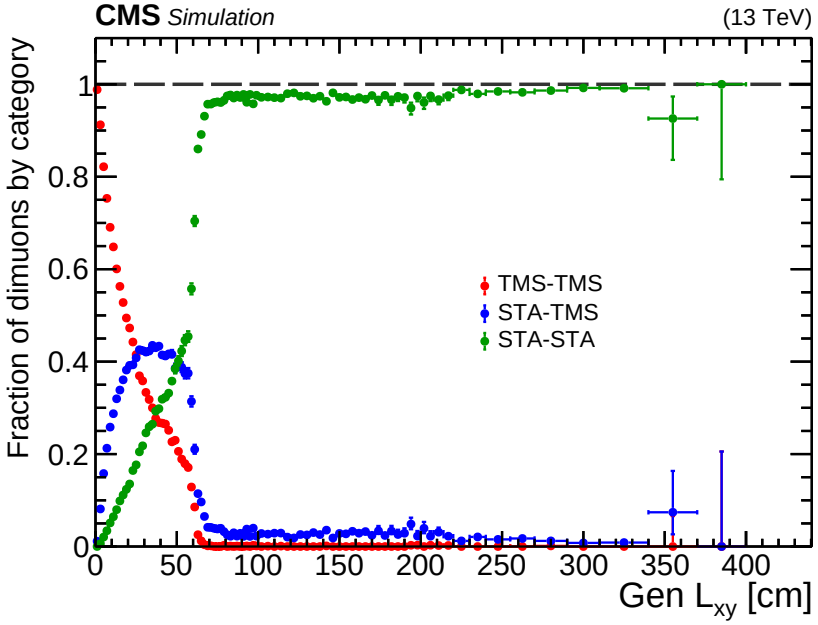


Figure 4.20: Fraction of reconstructed signal event dimuons in the three categories: TMS-TMS (red), STA-STA (green), and TMS-STA (blue). Dimuons are shown taken from the $\Phi \to 2\mathrm{X} \to 2\mu 2j$ sample and are shown as a function of generated L_{xy}.

4.4.5.1 Muon Pairing

The analysis reconstructs at most two signal dimuons, therefore an algorithm is required to find the two leading candidates in each event. This is done by first selecting the top two candidates in each category, and then reducing these dimuons to the top two overall. Each of the three dimuon categories uses an identical algorithm to select signal candidates. The method begins with $n > 1$ muons which pass the selection described in Section 4.4.4 that belong to the specified category. These muons are paired into all $n(n-1)/2$ combinations to form candidate dimuons. Pairs of unrelated muons are first suppressed by a requirement on the distance of closest approach (DCA) between their extrapolated tracks in 3-d space. A loose requirement of DCA < 15 cm reduces the combinatorics involved in selecting signal candidates and remains relatively efficient for signal dimuons. Specifically, this requirement removes no more than 8% of signal dimuons reconstructed in the STA-STA category and less than 1% in both the STA-TMS and TMS-TMS categories.

The remaining pairs are then fit to a common vertex, using the Kalman fitter described in Section 3.2.7.3. Our implementation follows that of the Run 1 analysis [Col15] and does not restrict the common vertex to be within the silicon tracker as is required by the standard implementation in CMSSW. This fit is required to converge since the dimuon object is later used to define variables that are also selected for. The convergence of the common vertex fit is nearly 100% throughout most of the active detector volume [Val19b]. Muon track parameters are modified by the common vertex fitting procedure, and kinematic quantities associated with the refitted tracks are used throughout the remainder of the analysis. The number of dimuon candidates is then reduced by keeping only those formed from the four highest p_{T} muons. This leaves at most $4(4-1)/2 = 6$ dimuon candidates to be selected from.

The top dimuon candidate(s) are chosen based on the χ^2 of the vertex fits. We first look to pair each dimuon with another that is built using a disjoint set of muons. If there are any such pairs, the pair of dimuons with the lowest summed χ^2 is selected as the signal candidates. If no such pairs exist, we then look for the single dimuon with the lowest χ^2

which is then assigned as the signal candidate. If no dimuon exists, the algorithm terminates and no dimuon is selected. Dimuons are then collected among all categories and up to two are selected which have the lowest χ^2, which are used for the remainder of the analysis. Selection efficiency for recovering the correct dimuons in the event is nearly 100% for the 2μ samples, and between $85-95\%$ in the 4μ samples.

4.4.5.2 Common Dimuon Selection

Dimuons formed by the pairing criteria undergo additional selection done independently and in parallel for each of the three dimuon categories. The selection applied to each can differ substantially. However, as is true for the reconstructed muon categories, some commonalities remain between all three. These selection criteria are listed in Table 4.12, which include the pairing criteria selection described in the previous section, in addition to those that will be further elaborated here.

Similar to the p_{T} requirement put in place for both muon categories, a minimum invariant mass of 10 GeV is set as the threshold for all three dimuon categories. This suppresses QCD-like events in addition to those that would occur due to material interactions inside the detector. Both muons are also required to have opposite charges consistent with a search for a neutral particle. The last common requirement is on the collinearity angle $|\Delta\Phi|$, whose distribution for all $\Phi \to 2\mathrm{X} \to 2\mu 2j$ signal types is shown in Figure 4.21. A requirement of $|\Delta\Phi| < \pi/4$ is selected, which is intentionally kept loose to maintain sensitivity to LLPs which decay to more particles than just the two muons.

Each dimuon category also has a requirement on the opening angle α between the muons, the χ^2 of the vertex fit, and the transverse decay length significance $L_{xy}/\sigma_{L_{xy}}$. However, the optimized requirements are not identical due to differences in their resolution. In 2016, we place a requirement on the opening angle α between the two muons to be less than 2.5 rad to match that set by our trigger, equivalent to a requirement of $\cos\alpha > -0.8$. In 2018, this trigger requirement was removed, which allowed us to re-optimize the corresponding offline requirement for the three categories. The study resulted in loosened requirements of

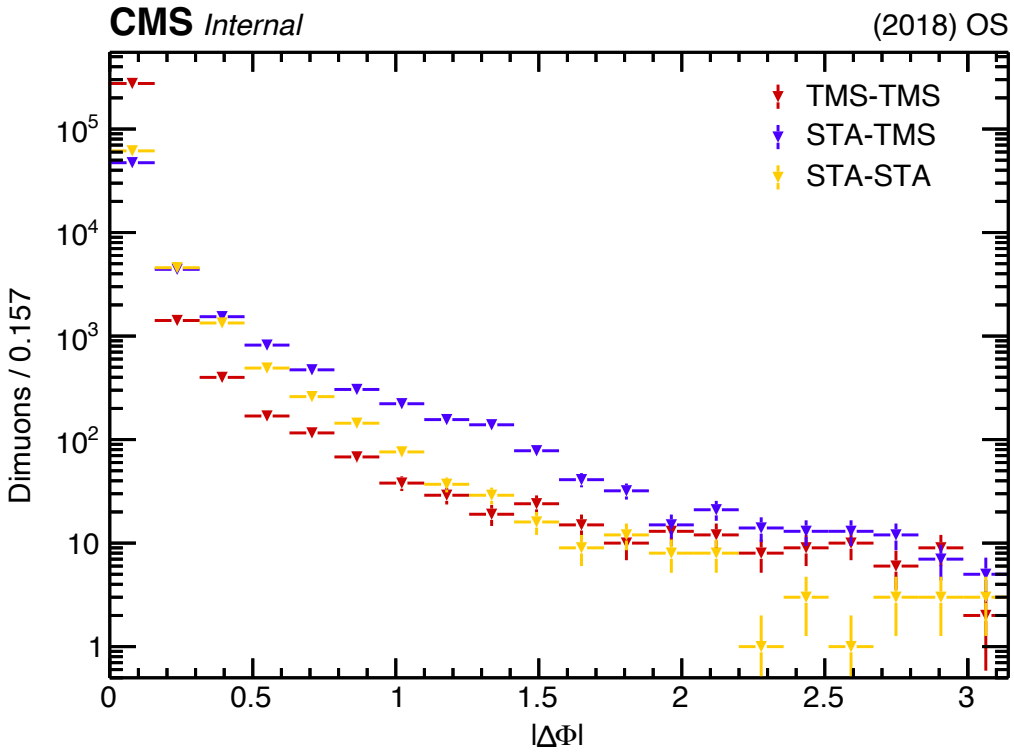


Figure 4.21: Distribution of $|\Delta\Phi|$ for TMS-TMS (red), STA-TMS (blue), and STA-STA (yellow) dimuons passing the full selection except for a requirement on $|\Delta\Phi|$ in all $\Phi \to 2\mathrm{X} \to 2\mu 2j$ signal samples combined.

Common dimuon selection	Description
DCA < 15 cm	Paired muon distance of closest approach is less than 15 cm
vertex fit converged	The common vertex fit applied to the dimuons must converged
selected pair(s)	Dimuon is ranked among the top two of the event
$m_{\mu\mu} > 10$ GeV	Dimuon invariant mass must be larger than 10 GeV
opposite-sign muons	Muons must be oppositely charged
$\lvert\Delta\Phi\rvert < \pi/4$	Collinearity angle of the dimuon must be smaller than $\pi/4$

Table 4.12: Dimuon selection applied independently of candidate dimuon type. These requirements are imposed in order, and before those that are category-specific.

$\cos\alpha > -0.9$, -0.9, and -0.99 for dimuons in the STA-STA, STA-TMS, and TMS-TMS categories, respectively.

Backgrounds from pairs of unassociated muons are suppressed with a requirement on the χ^2 of the vertex fit. A requirement of this type offers a strong handle on background suppression, however, the chosen requirement errs on the loose side to keep signal efficiency high, since this variable is known to be difficult to reproduce in simulation. Ultimately requirements of $\chi^2_{\text{vertex}} < 10, 20$, and 10 are selected for dimuons in the STA-STA, STA-TMS, and TMS-TMS categories.

Dimuons from each category are required to be displaced. This is done by selecting dimuons with a sufficiently large transverse decay length significance $L_{xy}/\sigma_{L_{xy}}$, in addition to requirements imposed on the d_0/σ_{d_0} of constituent muons in TMS-based categories. The $L_{xy}/\sigma_{L_{xy}}$ requirements were chosen following the figure of merit based procedure described in Section 4.4.6.3, which selects $L_{xy}/\sigma_{L_{xy}} > 6, 3$ and 6 for the for dimuons in the STA-STA, STA-TMS, and TMS-TMS categories, respectively. The L_{xy} of each dimuon is measured with respect to the primary vertex of the event. Studies revealed switching this reference point to the beam-spot did not improve signal-to-background discriminating power. Similarly, removal of the dimuon object from the primary vertex (if it is present), as was done by the Run 1 analysis [Kha15], also gave no clear improvement. A summary of these related dimuon requirements is given in Table 4.13. These include an invariant mass requirement that is described in Section 4.4.6.4.

Related Dimuon Selection

Requirement	Dimuon category		
	STA-STA	STA-TMS	TMS-TMS
$\cos\alpha$ [2016, 2018]	$> -0.8, -0.9$	$> -0.8, -0.9$	$> -0.8, -0.99$
χ^2_{vertex}	< 10	< 20	< 10
$L_{xy}/\sigma_{L_{xy}}$	> 6	> 3	> 6
$m_{\mu\mu}^{\text{min}} < m_{\mu\mu} < m_{\mu\mu}^{\text{max}}$	see Section 4.4.6.4		

Table 4.13: Dimuon selection applied to the three categories which vary in their cutoff values due to category resolution. These requirements are applied after those described in Table 4.12, and commute with those which are category-specific.

4.4.5.3 STA-STA Dimuon Selection

The strongest sensitivity to long lifetime signals is obtained by the STA-STA dimuon category. Further selection requirements improve the relative dimuon quality by imposing requirements on the sum of muon segments used for both muon track fits, specifically $N(\text{dimuon segments}) > 4$.

Additional background was found in the 2018 dataset which was larger than expected from the increased luminosity. By inverting the STA to TMS muon association logic we identified the *true* dimuon quantities as measured by TMS-TMS dimuons. Studies of this kind revealed the new background was primarily due to dimuons with a true mass below 10 GeV which was mismeasured in the STA-STA category. STA muons associated with these pairs were found to have very similar η values. Low mass background of this nature was therefore suppressed using a requirement on the difference in η between the two muons, requiring $|\Delta\eta_{\mu\mu}| > 0.1$.

Cosmic ray background is most prevalent in the STA-STA dimuon category – where the dimuon handles are relatively few, and the active area for a passing cosmic muon is the largest. Therefore, in addition to timing and angle requirements on the paired muons, we also look at additional muons which are back-to-back from those in the pair. Specifically, we look to reject background formed using one leg of a cosmic muon which becomes paired to a separate muon produced from the proton-proton collision or the associated cosmic ray. For

both muons in the STA-STA pair, we looked for an addition back-to-back (b2b) STA muon defined as having $\cos\alpha < -0.9$. We also require these back-to-back muons have $p_{\mathrm{T}} > 10$ GeV to minimize accidental pairings to low momentum muons produced in the underlying event. If one such muon is found which has a back-to-back time difference of $|\Delta t_{\mathrm{b2b}}| > 20$, the dimuon is vetoed, as this difference is consistent with a cosmic muon being read in opposite-halves of the detector. This additional requirement has a minor effect on signal efficiency, removing at most 2% of any individual signal sample. Each STA-STA dimuon specific requirement is listed in Table 4.14, with more details on the selection provided in Section 5 of Reference [Val19b].

STA-STA dimuon selection	**Description**
N(dimuon segments) > 4	Summed muon segments must exceed four
N(dimuon segments) > 5 or $\lvert\Delta\eta_{\mu\mu}\rvert > 0.1$	Summed muon segments must exceed five or the difference in muon η is greater than 0.1
N(DT hits) > 24 or $\lvert\Delta\eta_{\mu\mu}\rvert > 0.1$	Muons track fits in the barrel only use more than 24 hits or the difference in muon η is greater than 0.1
back-to-back muon veto	No back-to-back muon with $p_{\mathrm{T}} > 10$ GeV and $\lvert\Delta t_{\mathrm{b2b}}\rvert > 20$ ns

Table 4.14: STA-STA dimuon selection applied after those listed in Tables 4.12 and 4.13.

4.4.5.4 STA-TMS Dimuon Selection

Dimuons reconstructed in the STA-TMS dimuon category bridge the gap in signal efficiency left between those reconstructed in the STA-STA category to those in the TMS-TMS category, i.e. for L_{xy} between $20 - 50$ cm. The pairing of dissimilar muons creates a unique topology not seen in either of the two complementary dimuon categories. Because the resolution of the TMS muon is orders of magnitude greater than that of the STA muon, the location of the dimuon vertex almost always lies along the reconstructed TMS track. Consequentially, we can derive a variable that takes advantage of this characteristic to reject our background. The muon collinearity angle $\Delta\phi_{\mu}$ is such a variable and is defined as the angle between the TMS muon p_{T} vector and the L_{xy} vector of the STA-TMS dimuon. The variable, depicted in Figure 4.22, can be used to distinguish dimuons formed with a CV

location consistent with being displaced (i.e. $|\Delta\phi_\mu| < \pi/2$) from those that would have the TMS muon traveling in a direction inconsistent with the initial transverse momentum of the LLP (i.e. $|\Delta\phi_\mu| > \pi/2$). A requirement of $|\Delta\phi_\mu| < 2.9$ was found to suppress over 50% of remaining background in various control regions in data while having a negligible effect on signal efficiency.

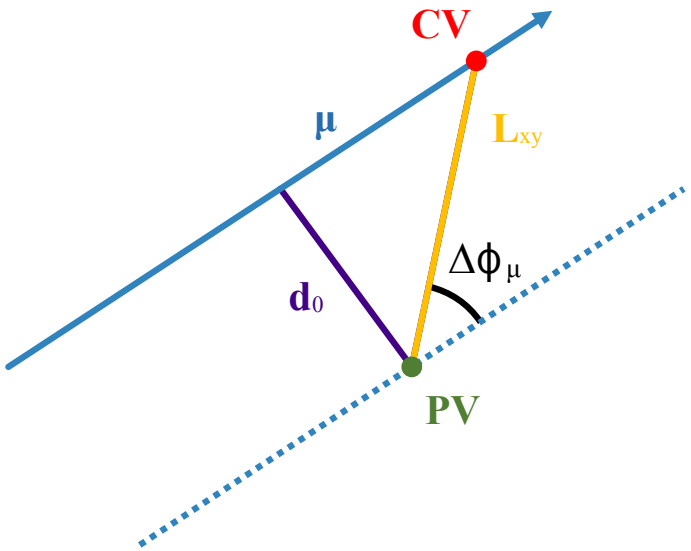


Figure 4.22: Diagram depicting the muon collinearity angle $|\Delta\phi_\mu|$. The location of the common vertex (red) typically lies along the TMS muon in a STA-TMS dimuon (blue). The d_0 of the TMS muon (purple) is used together with the L_{xy} of the reconstructed STA-TMS (yellow) to derive $|\Delta\phi_\mu| = \arcsin(d_0/L_{xy})$ (black). This variable is used similarly to the collinearity angle $|\Delta\Phi|$ and rejects background reconstructed misaligned with the primary vertex (green).

Dimuon quality cuts are also used to distinguish dimuons with features consistent with background from those that have characteristics like that expected of our signal. Occasionally TMS muons can be found with tracker hits closer to the primary vertex than the location of the CV fit. This pathology is associated with mismeasured muons and therefore a requirement of $N(\text{hits before vertex}) < 6$ is imposed on the TMS muon in the STA-TMS pair. The analysis also imposes a requirement on the number of tracker layers used in the TMS fit. However, because an LLP with a longer lifetime is associated with a TMS muon with fewer

tracker layers, this requirement is made a function of the measured dimuon L_{xy}. Specifically, we require N(tracker layers) + floor(L_{xy} [cm]/15) > 5 – derived to maximize our signal efficiency (less than 1% loss), and remove a handful of events left in the $3\pi/4 < |\Delta\Phi| < \pi$ control region described in Section 4.4.6.1 when all other selection criteria are applied. The floor operator rounds the floating point number down to the closest integer. More details on the selection placed on STA-TMS dimuons can be found in Section 5 of Reference [Val19b], and a summary of the selection is given in Table 4.15.

STA-TMS dimuon selection	**Description**
$\|\Delta\phi_\mu\| < 2.9$	Muon collinearity angle is less than 2.9
N(hits before vertex)< 6	Fewer than six TMS tracker hits are measured before the vertex
N(tracker layers) + floor(L_{xy} [cm]/15) > 5	Number of TMS muon tracker layers added to the rounded down result of the L_{xy} (in cm) divided by 15 is greater than five

Table 4.15: STA-TMS dimuon selection applied after those listed in Tables 4.12 and 4.13.

4.4.5.5 TMS-TMS Dimuon Selection

Dimuons formed by pairs of two TMS muons give the strongest signal sensitivity at smaller ($L_{xy} < 10$ cm) lifetimes, and meaningfully contribute to sensitivity throughout the entire range probed by this analysis. Similar quality requirements to those imposed on the STA-TMS muons are also imposed on the TMS-TMS muons. Namely, a tighter requirement of N(hits before vertex) < 3 is used on *both* TMS muons of the dimuon. We also require N(tracker layers) + floor(L_{xy} [cm]/15) > 5, an identical requirement to that used in the STA-TMS category, which is also applied to both muons.

TMS muons associated with TMS-TMS dimuons have hits measured in the pixel detector more often than those that are associated with STA-TMS dimuons. In general, the quality of a track improves with an increasing number of hits used in the fit. However, because we wish to reconstruct the dimuons produced anywhere inside the tracker, we cannot make an explicit requirement on the number of pixel hits used in an individual track fit, since some may be produced outside the pixel volume entirely. Instead, we enforce that both muons

that form the dimuon have nearly the same amount of pixel hits used in their track fit, requiring that $\Delta N(\text{pixel hits}) = |N(\text{pixel hits})_{\mu_1} - N(\text{pixel hits})_{\mu_2}| < 3$. Figure 4.23 shows the distributions of this quantity for signal and a control region in data, revealing that there is a significant tail of background events in 2018 which is removed with this quality requirement.

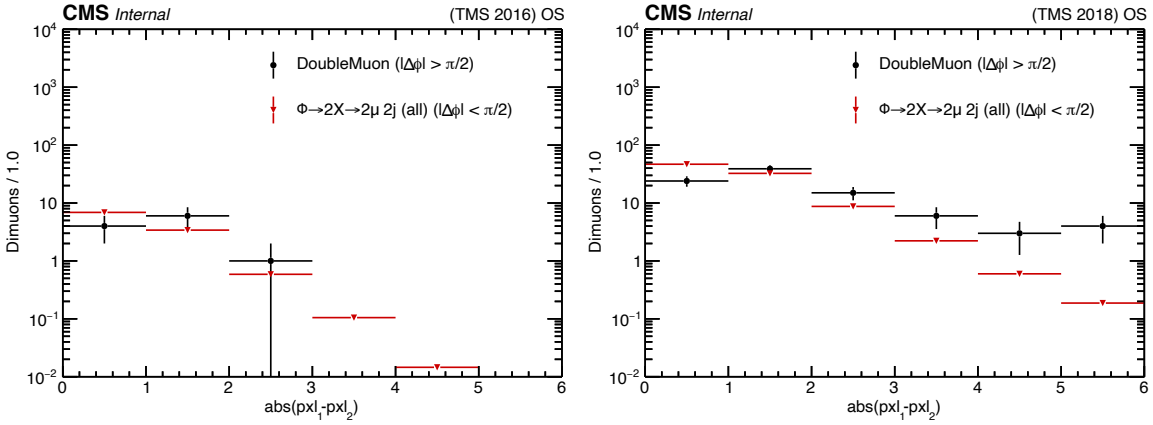


Figure 4.23: Distributions of the difference between the number of pixel hits on two muons in a TMS-TMS dimuon, $\Delta N(\text{pixel hits}) = |N(\text{pixel hits})_{\mu_1} - N(\text{pixel hits})_{\mu_2}|$, in 2016 (left) and 2018 (right). Signal simulation (red) includes all $\Phi \to 2\text{X} \to 2\mu 2j$ generated signal samples combined with $|\Delta\Phi| < \pi/2$ and is overlaid and normalized to remaining events in the $|\Delta\Phi| > \pi/2$ region in data (black). The overflow is added to the last bin in both plots.

The loosened trigger in 2018 was found to be a source of additional low p_{T} background in the TMS-TMS category. Dimuons of this type are suppressed with a requirement placed on the maximum p_{T} of the dimuon, requiring that $\max(p_{\text{T}}) > 25$ GeV. The requirement removes 42% of opposite-sign background dimuons in data with $|\Delta\Phi| > \pi/2$ while losing no more than 6% of efficiency in any individual signal sample. For simplicity, the same requirement is also used for 2016 data.

TMS-TMS dimuons which pass the full selection are then placed into one of three $\min(d_0/\sigma_{d_0})$ bins to improve signal-to-background sensitivity. The selected bins are: $6 < \min(d_0/\sigma_{d_0}) < 10$, $10 \leq \min(d_0/\sigma_{d_0}) < 20$, and $\min(d_0/\sigma_{d_0}) \geq 10$. Further increasing the number of bins gave negligible improvement to signal sensitivity [Nas21b]. More details on the selection placed on TMS-TMS dimuons can be found in Section 5 of Reference [Val19b],

and a summary of the selection is given in Table 4.16.

TMS-TMS dimuon selection	Description
N(hits before vertex)< 3	Fewer than three hits measured before the vertex
N(tracker layers) + floor(L_{xy} [cm]/15) > 5	Number of muon tracker layers added to the rounded down result of the L_{xy} (in cm) divided by 15 is greater than five
ΔN(pixel hits) < 3	Difference in pixel hits used to fit the muons is less than three
p_{T} of the leading muon > 25 GeV	Larger p_{T} of paired muons must exceed 25 GeV

Table 4.16: TMS-TMS dimuon selection applied after those listed in Tables 4.12 and 4.13.

4.4.6 Selection Optimization

This section covers how requirements set at specified values were derived. In general, we look to minimize the loss of any individual generated signal sample while suppressing background to only a handful of events or less to maximize signal sensitivity. Section 4.5 describes the background estimation methods, which are derived using control regions in data defined in this section. Monte Carlo is used only to gain insight into the nature of the events and is otherwise not used in the optimization process.

4.4.6.1 Control Regions

As covered in Section 4.1, background from most sources are expected to be distributed symmetrically in $|\Delta\Phi|$. This allows us to define a control region representative of our signal region by taking events in the paired half of the $|\Delta\Phi|$ distribution. Put another way, dimuons with $|\Delta\Phi| > 3\pi/4$ should share characteristics with many having $|\Delta\Phi| < \pi/4$ in our signal region. Therefore opposite-sign (OS) dimuons in the $|\Delta\Phi| > 3\pi/4$ region are used to sample characteristics of dimuons in our signal region and contain contributions from all three primary backgrounds of the analysis (Drell-Yan, QCD, and cosmic rays). This region is occasionally extended to $|\Delta\Phi| > \pi/2$ (complemented by $|\Delta\Phi| < \pi/2$) to increase statistics.

However, OS dimuons with $|\Delta\Phi| > 3\pi/4$ are not representative of the full population of events expected in our signal region. QCD-like events are expected to break the $|\Delta\Phi|$

symmetry and more frequently populate the region with $|\Delta\Phi| < \pi/4$. Studies showed that QCD dimuons formed by pairs of same-sign (SS) muons have very similar properties to those with OS that fall in our signal region. Limited Monte Carlo simulation revealed that the bulk of these events originate from B-meson cascades for both cases as shown in Figure 4.24. Therefore SS dimuons are used as a proxy for understanding OS QCD dimuons expected in our signal region.

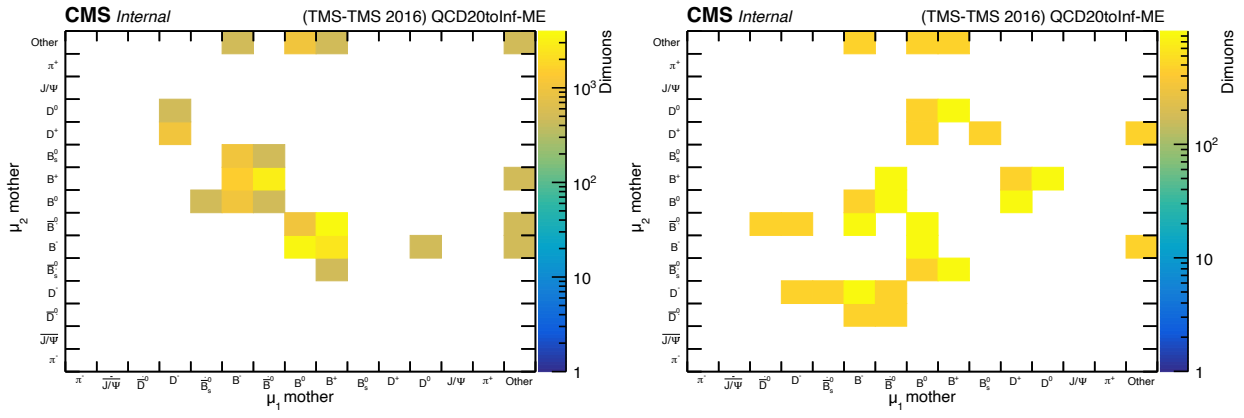


Figure 4.24: Muon parent particles of reconstructed OS (left) and SS (right) dimuons which pass our displacement and quality selection in the TMS-TMS dimuon category. Dimuons are obtained from QCD Monte Carlo simulation and the parent particle is extracted using generator level variables. No isolation or $|\Delta\Phi|$ requirement is imposed to increase statistics. μ_1 and μ_2 denote the leading and sub-leading p_{T} muons of the dimuon, respectively.

Non-isolated dimuons with both $|\Delta\Phi| < \pi/4$ and $|\Delta\Phi| > 3\pi/4$ are also used for control regions in the analysis. The number of OS dimuons was found to track well with the number of SS dimuons as a function of the isolation of the maximally non-isolated TMS muon, as shown in Figure 4.25. Only in the region containing very isolated muons (where there is additional contamination from Drell-Yan) are the two distributions in significant disagreement. Both OS and SS non-isolated regions, therefore, constitute additional controls used in dimuon categories with a TMS muon. A complementary region for dimuons with STA muons is created by initially requiring a matched, non-isolated ($I_{\mathrm{trk}}^{\mathrm{rel}} > 0.1$) TMS muon. A dimuon is then formed using the STA muon associated with the non-isolated TMS muon. Dimuons of this type can be formed in both the STA-STA and STA-TMS category, and are used to obtain events with significant QCD contributions. A similar procedure is done to

obtain Drell-Yan enriched events in the STA-STA and STA-TMS categories, requiring that the TMS-TMS dimuon was produced promptly ($L_{xy}/\sigma_{L_{xy}} < 1$ and/or $d_0/\sigma_{d_0} < 1.5$), and subsequently forming the dimuon using only the associated STA muons.

Regions with enhanced cosmic muon contributions are obtained by inverting the requirement on $\cos\alpha$ and/or the primary vertex filter. The analysis also uses a dedicated sample of cosmic muons to obtain signal-like cosmic events recorded when there was beam activity [Val19a].

4.4.6.2 Lifetime Reweighting

Selection efficiencies for signal samples with lifetimes differing from those assigned during production can be procured through a lifetime reweighting procedure. We first note that the decay time of each dimuon can be expressed as

$$t = \frac{m_{\mathrm{X}} \cdot L_{xy}}{p_{\mathrm{T}}^{\mathrm{X}}}, \tag{4.5}$$

which can be obtained through reconstructed quantities. We then note that a particle with lifetime τ will have its decay time distributed according to an exponential probability distribution through Equation 2.6. Furthermore, there are three and four simulated lifetimes for each unique combination of masses belonging to the BSM Higgs and HAHM samples, respectively. To take full advantage of the range of lifetimes we have generated, we define a meta-sample probability distribution

$$\mathcal{M}(t) = \frac{\sum_{i=1}^{n} \mathcal{P}(t|\tau_i) N_{\tau_i}}{\sum_{i=1}^{n} N_{\tau_i}}, \tag{4.6}$$

where $\mathcal{P}(t|\tau_i)$ and N_{τ_i} represent the decay probability distribution and number of generated dimuons for generated sample i among n total. The meta-sample can then be multiplied by a weight $w(t,\tau)$ such that it falls from a probability distribution with an arbitrary lifetime

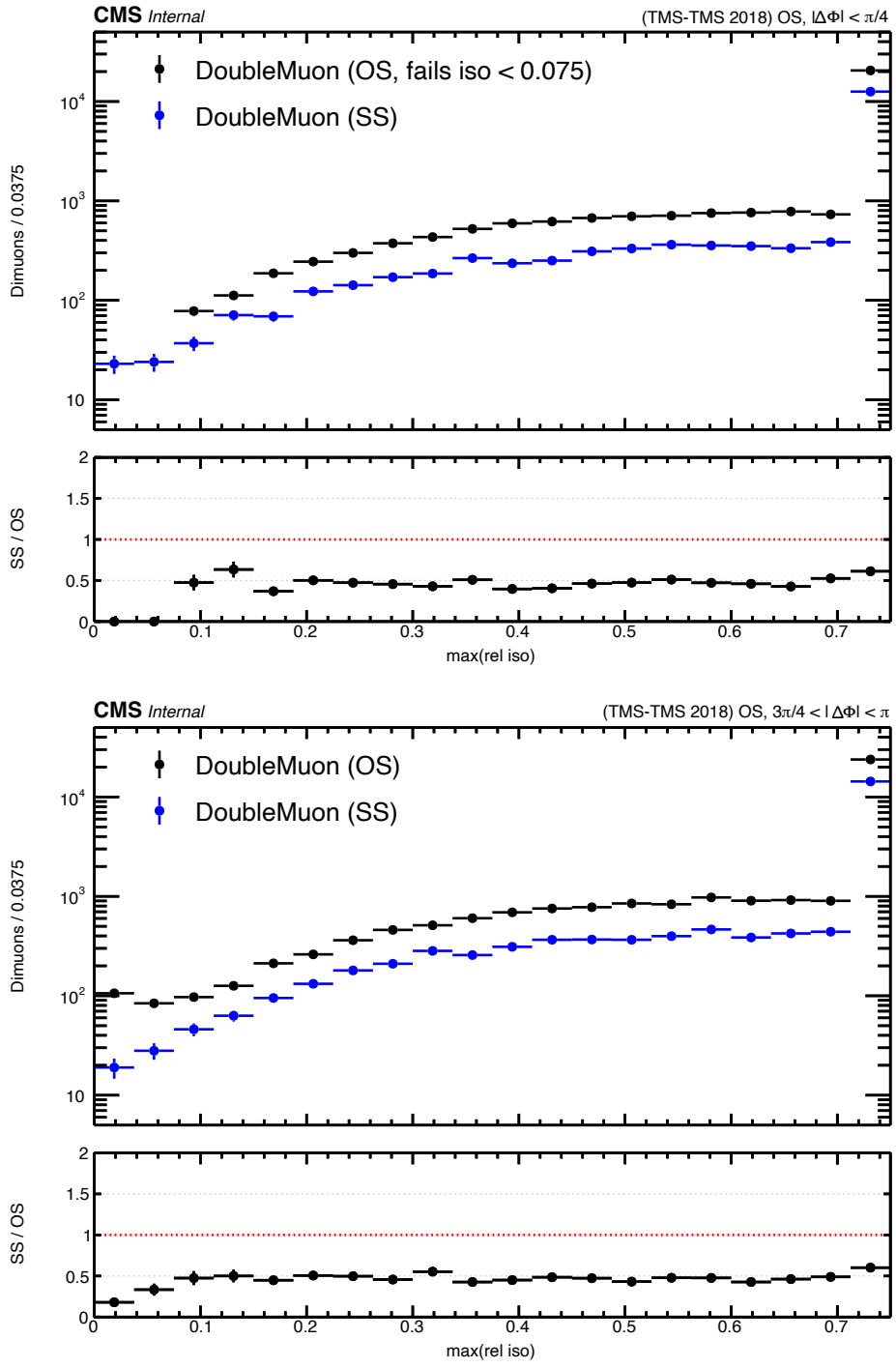


Figure 4.25: Distributions of the maximum relative isolation variable belonging to either OS (black) or SS (blue) TMS-TMS dimuons in 2018 data. Dimuons with $|\Delta\Phi| < \pi/4$ (top) and $|\Delta\Phi| > 3\pi/4$ (bottom) are shown which pass the remaining selection under a loosened displacement requirement ($\min(d_0/\sigma_{d_0} > 4)$). Dimuons in our signal region (OS, $|\Delta\Phi| < \pi/4$) remain blinded. The last bin includes events in the overflow.

$\mathcal{P}(t|\tau)$, defining

$$w(t,\tau) = \frac{\mathcal{P}(t|\tau)}{\mathcal{M}(t)}. \tag{4.7}$$

Our selection efficiency $\tilde{\epsilon}(t)$ is dependent only on the decay time t of the dimuon, not the generated lifetimes of the samples, which allows us to write the (non-normalized) distribution of selected dimuons $\mathcal{S}(t)$ as

$$\mathcal{S}(t) = \tilde{\epsilon}(t)\mathcal{M}(t). \tag{4.8}$$

We may easily obtain the selected distribution by passing the full meta-sample through the analysis framework. With the selected distribution and weight function fully defined, the selection efficiency for a sample with an arbitrary lifetime $\epsilon(\tau)$ can be obtained as

$$\int \tilde{\epsilon}(t)w(t,\tau)\mathcal{M}(t)\ dt = \int w(t,\tau)\mathcal{S}(t)\ dt = \epsilon(\tau). \tag{4.9}$$

This formulation allows us to derive the selection efficiency for lifetimes other than those generated for the analysis. Figure 4.26 shows an example of the reweighting procedure applied to one of the $\Phi \to 2\mathrm{X} \to 2\mu 2j$ samples. This procedure was found to significantly reduce statistical errors associated with reweighting just one of the generated samples [Nas20c].

4.4.6.3 Z_{Bi} Optimization

Selection criteria that strongly reject background but include non-negligible losses in signal efficiency are optimized using a figure of merit which is monotonic with the expected statistical significance. Our analysis uses Z_{Bi} [CLT08], which utilizes the number of events in the signal region (n_{on}) and a complementary "sideband" region (n_{off}) to calculate the Z value, or the number of standard deviations, the measurement is from that expected by the null (background-only) hypothesis. One first notes that the Poisson probability for observing

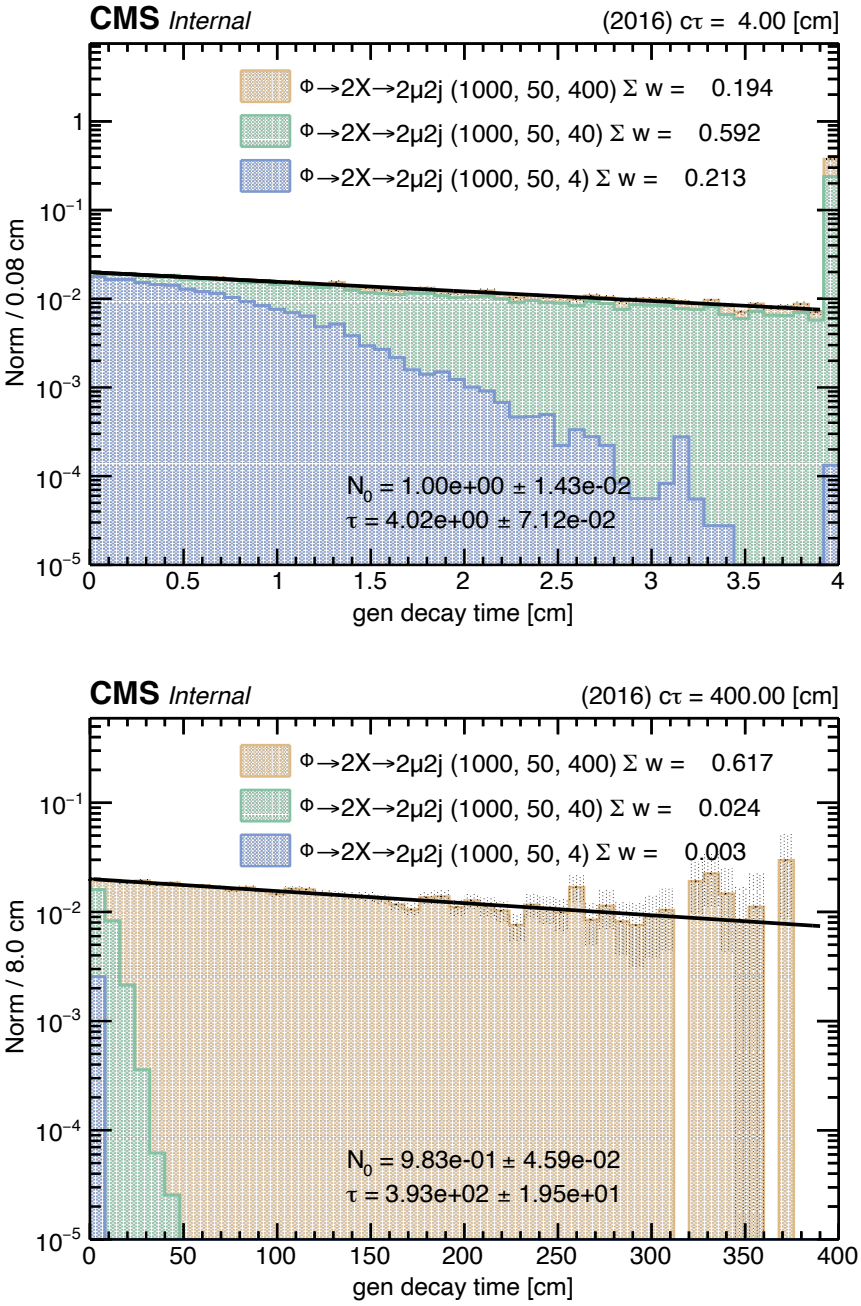


Figure 4.26: Meta-sample probability distribution for the $\Phi \to 2\mathrm{X} \to 2\mu 2j$ signal sample with $m_\Phi = 1000$ GeV and $m_\mathrm{X} = 50$ GeV. Each of the three generated lifetimes: 4 mm (blue), 40 mm (green), and 400 mm (orange); are combined according to Equation 4.6, which redistributes the dimuons according to an arbitrary lifetime, e.g. 4 cm (top) or 400 cm (bottom). To verify the reweighting procedure works as intended, the distribution is fit to an exponential itself, yielding the correct input parameters (N_0, τ). $\sum w$ denotes the contribution of the generated sample to the numerator of Equation 4.6. The last bin includes events in the overflow which is not included in the fit. The weights in the bottom distribution do not sum to one as a result of limited statistics.

both n_{on} and n_{off} events may be written as

$$P(n_{\text{on}}, n_{\text{off}}) = \frac{e^{-\mu_{\text{on}}}\mu_{\text{on}}^{n_{\text{on}}}}{n_{\text{on}}!} \cdot \frac{e^{-\mu_{\text{off}}}\mu_{\text{off}}^{n_{\text{off}}}}{n_{\text{off}}!} \tag{4.10}$$

$$= \frac{e^{-(\mu_{\text{on}}+\mu_{\text{off}})}(\mu_{\text{on}}+\mu_{\text{off}})^{n_{\text{tot}}}}{n_{\text{tot}}!} \cdot \frac{n_{\text{tot}}!}{n_{\text{on}}!(n_{\text{tot}}-n_{\text{on}})!}\rho^{n_{\text{on}}}(1-\rho)^{(n_{\text{tot}}-n_{\text{on}})}, \tag{4.11}$$

which transforms the two Poisson distributions into one with $n_{\text{tot}} = n_{\text{on}} + n_{\text{off}}$ events, which are distributed binomially between the two categories with parameter $\rho = \mu_{\text{on}} / (\mu_{\text{on}} + \mu_{\text{off}})$. All information concerning the ratios of events in the signal and sideband region is therefore contained in the conditional binomial probability $P(n_{\text{on}}|n_{\text{tot}};\rho)$. Therefore the Poisson from which the two categories of events are distributed plays no role in what follows. The expected ratio of events in the signal and sideband region is taken as fixed from some external measurement, equivalent to a fixed ρ. The p-value for finding the observed number of events in the signal region is therefore calculated from the one-tailed probability sum

$$p_{\text{Bi}} = \sum_{j=n_{\text{tot}}}^{n_{\text{tot}}} P(j|n_{\text{on}};\rho) = B(\rho, n_{\text{on}}, 1+n_{\text{off}})/B(n_{\text{on}}, 1+n_{\text{off}}), \tag{4.12}$$

where B are the incomplete and complete beta functions, distinguished by their number of arguments. The probability sum may then be easily translated into a Z-score (Z_{Bi}), which is used as our figure of merit.

For the $\min(d_0/\sigma_{d_0})$, $L_{xy}/\sigma_{L_{xy}}$, and $\max(I^{\text{rel}}_{\text{trk}})$ cuts, we set the optimum requirement using this score. The optimization of the $\min(d_0/\sigma_{d_0})$ requirement for the TMS-TMS category will be described as an example. The signal efficiency of this cut directly strongly depends on the lifetime of the sample in addition to its kinematics. To keep our search maximally model-independent, we optimize for the sample with the largest efficiency loss as a function of the $\min(d_0/\sigma_{d_0})$ requirement, found to be the $\Phi \to 2\text{X} \to 2\mu 2j$ signal sample with $m_\Phi = 1000$ GeV and $m_{\text{X}} = 20$ GeV [Nas20b]. Furthermore, we reweight the lifetime of this mass point to have a lifetime of 1 mm, roughly the smallest distance for which we require optimal performance since the tracker resolution is $\mathcal{O}(100\ \mu\text{m})$. To obtain the number of signal events, we scale this signal following the recipe described in Section 4.7, using the

previous limits obtained on the cross-section [Kha15].

The background estimation procedure described in Section 4.5 uses a transfer factor to predict the number of events in the signal region from a complementing sideband. This allows us to obtain the binomial parameter ρ under the background-only null hypothesis. The background and signal events are then combined, and the number of remaining events is counted as a function of the requirement. The number of events in the sideband region defines n_{off} and the sum of the signal and predicted background events define n_{on}. The Z_{Bi} score is then calculated at each point, and the requirement which yields the maximum is selected. Figure 4.27 shows the distributions described which led to a selection of $\min(d_0/\sigma_{d_0}) > 6$. The Z_{Bi} score calculated for the d_0/σ_{d_0} of the STA muon does not appreciably increase after the $L_{xy}/\sigma_{L_{xy}}$ requirements are applied and therefore is not used for our selection.

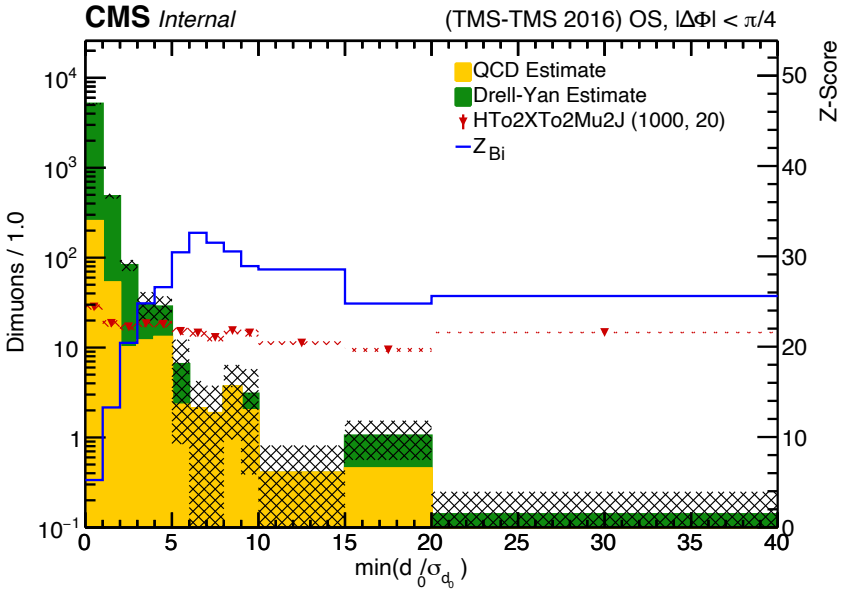


Figure 4.27: Distribution of estimated TMS-TMS background and signal dimuons from the $\Phi \to 2\text{X} \to 2\mu 2j$ signal sample with $m_\Phi = 1000$ GeV and $m_\text{X} = 20$, reweighted to a lifetime of 1 mm (red). In each bin, the Z_{Bi}score (blue) is calculated using all background and signal dimuons with $\min(d_0/\sigma_{d_0})$ larger than that bin. A selection requirement of $\min(d_0/\sigma_{d_0}) > 6$ is chosen which maximizes our figure of merit.

4.4.6.4 Mass Intervals

Signal sensitivity is further improved by using mass windows consistent with each signal. The mass resolution, shown in Figure 4.28, strongly vary between dimuon categories. Therefore the mass intervals are set separately for each category. Windows were chosen such that they typically contain $90 - 99\%$ of the signal in question. The windows used for generated signal samples are enumerated in Table 4.17.

Targeted mass [GeV]	Dimuon category interval [GeV]		
	STA-STA	STA-TMS	TMS-TMS
10	$10 - 32$	$10 - 14$	$10 - 12$
20	$10 - 32$	$14 - 26$	$18 - 22$
30	$15 - 60$	$20 - 40$	$27 - 33$
40	$20 - 80$	$28 - 52$	$37 - 43$
50	$20 - 80$	$36 - 64$	$46 - 54$
60	$35 - 120$	$42 - 78$	$55 - 65$
150	$35 - 245$	$95 - 205$	$140 - 160$
350	> 60	> 150	> 326

Table 4.17: Mass intervals used for selecting dimuons found in the generated signal samples for each dimuon category. The intervals are taken to be a function only of the LLP mass and the relatively small dependencies on both lifetime and parent mass seen in Figure 4.28 (right) are ignored.

Our analysis intends to be as model-independent as possible, and therefore we do not expect a signal to necessarily fall near one of the targeted masses. In the TMS-TMS category where the resolution is best, the mass interval is centered on the targeted mass, with a width that is parametrized as a function of the targeted mass through a simple linear relationship. The function: $\sigma_{\text{reco-gen}} = a + b \cdot m_{\mu\mu}$, is fit according to the method of least squares, yield $a = 0.704$ GeV and $b = 0.0635$. For each signal sample in the TMS-TMS category, windows were taken as the rounded value of this function at a given specified mass. Windows in the other categories were primarily found to be sufficiently wide such that no additional bins or parameterization were needed. The single exception was one STA-TMS bin which was added to target a mass of $m_{\mu\mu} = 80$ GeV with an interval of $56 < m_{\mu\mu} < 104$ GeV.

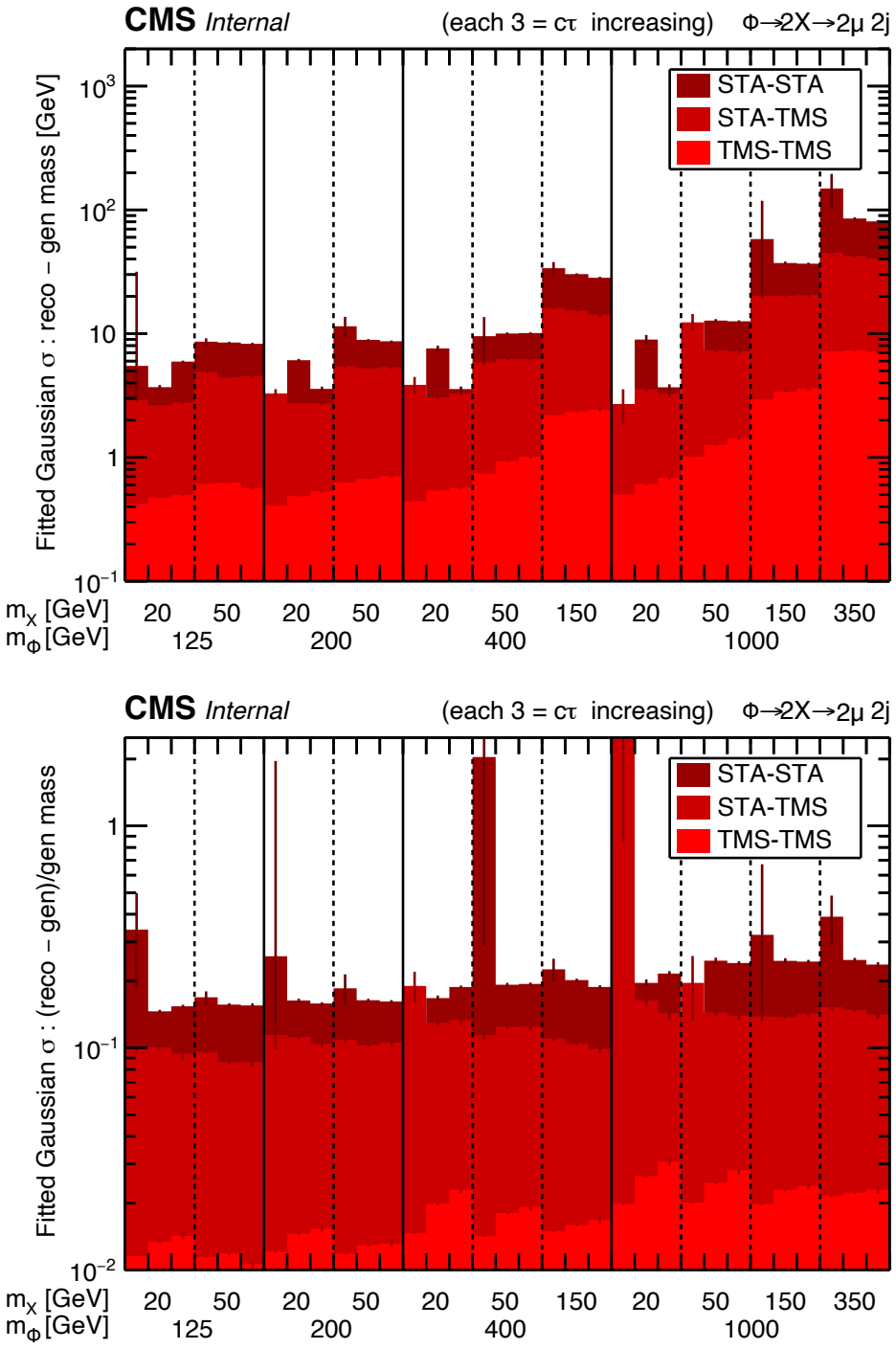


Figure 4.28: Width of a Gaussian fit applied to $m_{\mu\mu}^{\text{reco}} - m_{\mu\mu}^{\text{gen}}$ (top) and $(m_{\mu\mu}^{\text{reco}} - m_{\mu\mu}^{\text{gen}})/m_{\mu\mu}^{\text{gen}}$ (bottom) for each generated $\Phi \to 2\text{X} \to 2\mu 2j$ signal sample passing our selection. Resolution afforded by the TMS-TMS dimuons (light red), can be nearly an order of magnitude better than that of STA-TMS (red) or STA-STA (dark red) dimuons. The spikes in the distributions are caused by fit failures resulting from limited statistics.

4.4.7 Selection Summary

The full set of requirements used in the analysis are listed in Table 4.18. Figure 4.29 shows the breakdown of signal efficiencies for the $\Phi \to 2\text{X} \to 2\mu 2j$ signal samples. The overall signal efficiency (including acceptance and trigger effects) ranges from 1% for samples with small Φ and X masses and long lifetimes to roughly 50% for samples with large masses and short lifetimes. As mentioned earlier, the primary source of signal efficiency loss is the trigger.

4.5 Background Estimation

The analysis uses a purely data-driven approach to estimating the background found in our signal region. Estimates are derived from the amount of events found in various control regions covered in Section 4.4.6.1. Each dimuon category employs its own background estimation procedure due to the differing nature of their sources. For all categories, the number of events in the signal region was kept blinded until the full estimation procedure was developed and verified to avoid biasing the procedure. This section covers the estimation methods developed for the three primary backgrounds described in Section 4.2. All non-negligible sources of background covered in this section are summed to yield the total background estimate.

4.5.1 Drell-Yan Estimation

Background events arising from mis-measured Drell-Yan are expected in each dimuon category. These events are predicted using the observed events in the symmetric half of the $|\Delta\Phi|$ distribution (those with $|\Delta\Phi| > 3\pi/4$), and a transfer factor which relate the two

$$N^i_{\text{DY}}(\text{OS}; |\Delta\Phi| < \pi/4) = N^i_{\text{DY}}(\text{OS}; |\Delta\Phi| > 3\pi/4) \cdot R^i_{\text{DY}}, \tag{4.13}$$

Table 4.18: Summary of analysis selection criteria.

Muon preselection

Requirement	Muon type
	STA
N(CSC+DT hits)	> 12
N(CSC+DT stations)	> 1

Event preselection

Requirement	
HLT matching	yes
`PrimaryVertexFilter`	yes
N(parallel pairs)	< 6

Muon selection

Requirement	Muon type	
	STA	TMS
STA to TMS muon association	unreplaced	replaced
p_{T}	> 10 GeV	> 10 GeV
$\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}}$	< 1	< 1
N(DT hits) for muons in barrel	> 18	–
χ^2_{track}/dof	< 2.5	–
$\lvert t_{\text{in-out}}\rvert$	< 12 ns	–
muon direction	outward going	–
tracker muon	–	yes
N(matched segments)	–	> 1
$I^{\mathrm{rel}}_{\mathrm{trk}}$	–	< 0.075
d_0/σ_{d_0}	–	> 6

Dimuon selection

Requirement	Dimuon category		
	STA-STA	STA-TMS	TMS-TMS
DCA	< 15 cm	< 15 cm	< 15 cm
common vertex fit converged	yes	yes	yes
pairing criteria	best 1–2 ranked dimuons selected		
$m_{\mu\mu}$	> 10 GeV	> 10 GeV	> 10 GeV
opposite-sign muons	yes	yes	yes
$\lvert\Delta\Phi\rvert$	$< \pi/4$	$< \pi/4$	$< \pi/4$
$\cos\alpha$ [2016, 2018]	$> -0.8, -0.9$	$> -0.8, -0.9$	$> -0.8, -0.99$
χ^2_{vertex}	< 10	< 20	< 10
$L_{xy}/\sigma_{L_{xy}}$	> 6	> 3	> 6
$m^{\mathrm{min}}_{\mu\mu} < m_{\mu\mu} < m^{\mathrm{max}}_{\mu\mu}$	see Section 4.4.6		
N(dimuon segments)	> 4	–	–
if $\lvert\Delta\eta_{\mu\mu}\rvert < 0.1$			
- N(dimuon segments)	> 5	–	–
- N(DT hits) for muons in barrel	> 24	–	–
back-to-back muon veto	yes	–	–
$\lvert\Delta\phi_\mu\rvert$	–	< 2.9	–
N(hits before vertex)	–	< 6	< 3
N(tracker layers) + floor(L_{xy} [cm]/15)	–	> 5	> 5
ΔN(pixel hits)	–	–	< 3
p_{T} of the leading muon	–	–	> 25 GeV

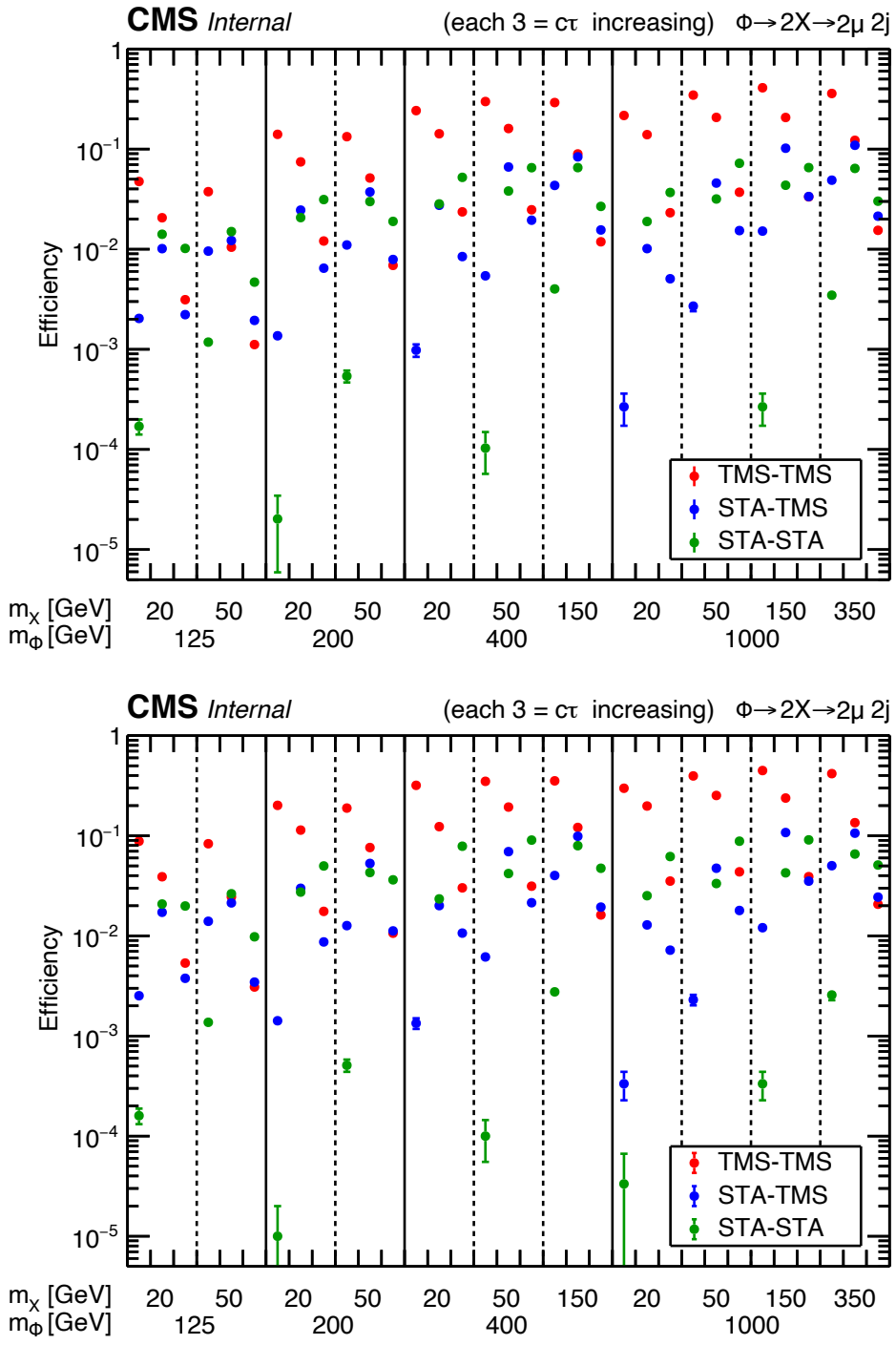


Figure 4.29: Signal efficiencies for the $\Phi \to 2X \to 2\mu 2j$ samples measured in 2016 (top) and 2018 (bottom) for the TMS-TMS (red), STA-TMS (blue), and STA-STA (green) dimuon categories. The efficiency is defined as the fraction of events that were selected by the trigger and offline requirements.

where $N^i_{\mathrm{DY}}(\mathrm{OS}; |\Delta\Phi| < \pi/4)$ and $N^i_{\mathrm{DY}}(\mathrm{OS}; |\Delta\Phi| > 3\pi/4)$ are the number of expected Drell-Yan events in the signal region and its symmetric $|\Delta\Phi|$ complement, respectively; R^i_{DY} is the transfer factor; and i denotes the dimuon category (STA-STA, STA-TMS, TMS-TMS). The number of Drell-Yan events in the control region, $N^i_{\mathrm{DY}}(\mathrm{OS}; |\Delta\Phi| > 3\pi/4)$, is taken as the number of observed events with $|\Delta\Phi| > 3\pi/4$) after the expected contribution of QCD events (discussed in Section 4.5.2) have been subtracted. The transfer factor R^i_{DY} is nominally unity but differs in some cases to account for residual asymmetries observed in certain categories.

Studies of the transfer factor for the three categories were performed using simulated Drell-Yan events and data [Val19b]. In the TMS-TMS category, we affirmed that $R^{\text{TMS-TMS}}_{\mathrm{DY}} = 1$ using simulated events and those in data that fail our vertex χ^2 requirement, which was found to be representative of Drell-Yan events in both $|\Delta\Phi|$ regions. However, in the STA-STA and STA-TMS categories, the transfer factor was found to be slightly smaller than unity. This feature was observed in both the dimuons formed from inverting prompt TMS-TMS dimuons (described in Section 4.4.6.1) and in Monte Carlo simulation. The transfer factor for these categories are taken to be $R^{\text{STA-TMS}}_{\mathrm{DY}} = 0.87$ and $R^{\text{STA-STA}}_{\mathrm{DY}} = 0.78$ [Val19b]. No transfer factor mass dependence was observed for any of the dimuon categories. Systematic uncertainties associated with the transfer factor are assigned values of $15\%, 20\%$, and 15% for the STA-STA, STA-TMS, and TMS-TMS categories, respectively.

4.5.2 QCD Estimation

Background from sources such as B-meson or J/Ψ decays discussed in Section 4.4.6.1 primarily populate the signal $|\Delta\Phi|$ region and are therefore evaluated using only dimuons that fall with the same corresponding $|\Delta\Phi|$. Dimuons are instead evaluated using an orthogonal control region, namely those formed with muons with like charges. Following a similar structure as is used for Drell-Yan estimation, our QCD prediction goes as

$$N^i_{\mathrm{QCD}}(\mathrm{OS}; |\Delta\Phi| < \pi/4) = N^i_{\mathrm{QCD}}(\mathrm{SS}; |\Delta\Phi| < \pi/4) \cdot R^i_{\mathrm{QCD}}, \tag{4.14}$$

where $N^i_{\text{QCD}}(\text{OS}; |\Delta\Phi| < \pi/4)$ and $N^i_{\text{QCD}}(\text{SS}; |\Delta\Phi| < \pi/4)$ denote the number of OS and SS QCD dimuon estimates, respectively; and R^i_{QCD} is the transfer factor. The transfer factor itself is evaluated taking the ratio of non-isolated events

$$R^i_{\text{QCD}} = \frac{M^i_{\text{QCD}}(\text{OS}; |\Delta\Phi| < \pi/4)}{M^i_{\text{QCD}}(\text{SS}; |\Delta\Phi| < \pi/4)}, \tag{4.15}$$

where $M^i_{\text{QCD}}(\text{OS}; |\Delta\Phi| < \pi/4)$ and $M^i_{\text{QCD}}(\text{SS}; |\Delta\Phi| < \pi/4)$ are the number of associated non-isolated OS and SS events, respectively. In the TMS-TMS category, these events are obtained by selecting for dimuons with at least one muon with $I^{\text{rel}}_{\text{trk}} > 0.075$ and both with $I^{\text{rel}}_{\text{trk}} < 0.5$. In both the STA-STA and STA-TMS categories, non-isolated dimuons are formed using STA muons associated to a non-isolated TMS muon, requiring $I^{\text{rel}}_{\text{trk}} > 0.1$ and $I^{\text{rel}}_{\text{trk}} > 0.125$ respectively for the two categories. The transfer factors obtained for these two categories exhibits a mass dependence [Val19b], and was therefore calculated using only events which fall in the corresponding mass intervals enumerated in Section 4.4.6.4. For dimuons in the TMS-TMS category, no such mass dependence was observed, and therefore events are combined among the full invariant mass spectrum to increase statistics. This is particularly important when one recognizes that the $R^{\text{TMS-TMS}}_{\text{QCD}}$ *is* d_0/σ_{d_0} dependent − requiring that the transfer factor be calculated independently among the three $\min(d_0/\sigma_{d_0})$ bins enumerated in Section 4.4.5.5. In practice, the QCD transfer factor typically sits around $1 - 2$ [Val19b].

QCD dimuons also populate the OS $|\Delta\Phi| > 3\pi/4$ control region used to estimate the Drell-Yan background. The control regions used for deriving QCD estimates (namely those containing SS and/or non-isolated dimuons) each include negligible contamination from Drell-Yan. Therefore an estimate of the QCD background is first calculated in the OS $|\Delta\Phi| > 3\pi/4$ control region, and the Drell-Yan contribution is taken as the remainder of observed events after the QCD estimate has been subtracted. The Drell-Yan estimate in our signal region is then calculated following the procedure described in Section 4.5.1. This procedure is done only in the STA-TMS and TMS-TMS categories after it was found that a negligible fraction of QCD events populates this region in the STA-STA category.

Systematic uncertainties associated with the QCD transfer factor of 30% and 15% are

assigned to dimuons in the STA-TMS and TMS-TMS categories, respectively. A series of auxiliary measurements made in the STA-STA category [Val19b] resulted in a systematic uncertainty associated with $R_{\mathrm{QCD}}^{\text{STA-STA}}$ which varies as a function of mass interval and year, and reaches values as large as 35%.

4.5.3 Cosmic Ray Estimation

In all categories, the expected background due to cosmic ray muons was found to be negligible after the full selection is applied. An upper bound in each category is calculated by first finding the efficiency for a cosmic ray muon to pass our selection. A dedicated sample of cosmic ray events [Val19a] is used to this end, from which the efficiency ϵ for a cosmic ray event to pass our selection is calculated in each category. Figure 4.30 shows the distribution of dimuons of this type formed from cosmic ray muons as a function of $\cos\alpha$, from which the efficiency is extracted. The corresponding $\cos\alpha$ selection requirements (listed in Table 4.18) are then tied to the corresponding dimuon category and data-taking period. These efficiencies rarely exceed 1%.

An upper bound for the number of cosmic muons expected to pass in collision data N_{p} is then calculated using the number of events that fail the $\cos\alpha$ requirement N_{f} together with the efficiency calculated in cosmic data as

$$N_{\mathrm{p}} = \frac{\epsilon}{1-\epsilon} N_{\mathrm{f}}. \tag{4.16}$$

Table 4.19 shows an upper bound on the number of cosmic ray events expected in data. In each category and data-taking period, the residual contribution of cosmic ray events is expected to be smaller than 0.1 events, and is therefore treated as negligible for the analysis.

4.5.4 Validation of Background Estimation Procedure

The full background estimation procedure is tested in several validation regions that are expected to contain a negligible fraction of signal events. Verifying the method correctly

Table 4.19: An upper bound on the number of cosmic ray muon events expected in data N_{p} in each dimuon category. Calculations are derived using Equation 4.16. For regions in which $N_{\mathrm{f}} = 0$, a Poisson upper bound of 1.8 is used.

	STA-STA		STA-TMS		TMS-TMS	
	2016	2018	2016	2018	2016	2018
$\cos\alpha$	> -0.8	> -0.9	> -0.8	> -0.9	> -0.8	> -0.99
ϵ	0.001	0.003	0.007	0.014	< 0.004	0.002
N_{f}	0	1	0	1	0	0
N_{p}	< 0.002	< 0.003	< 0.013	0.015	< 0.008	< 0.004

predicts the amount of Drell-Yan events is done using regions with relaxed or inverted displacement requirements. Figure 4.31 shows an example of such a prediction for the STA-STA dimuon category, in which we invert the displacement requirement, requiring that dimuons have $L_{xy}/\sigma_{L_{xy}} < 6$. The observed counts are consistent in each bin, and a larger background in 2016 is correctly predicted, consistent with a loss of tracking efficiency which occurred for a portion of 2016 data-taking [Sir19c], which led to more dimuons being found in the STA-STA category.

In the TMS-TMS category, which uses requirements on both $L_{xy}/\sigma_{L_{xy}}$ and $\min(d_0/\sigma_{d_0})$, its Drell-Yan estimation procedure can be evaluated still closer to the signal region, namely taking dimuons with $2 < \min(d_0/\sigma_{d_0}) < 6$. The result of the background estimation procedure in this region is shown in Figure 4.32 as a function of $L_{xy}/\sigma_{L_{xy}}$. The observed number of events fall in line with the expectation across the full $L_{xy}/\sigma_{L_{xy}}$ range. The distribution peaks away from zero due to the displacement requirement of $\min(d_0/\sigma_{d_0}) > 2$, which removes a substantial fraction of prompt dimuons.

Validation of background predictions from $|\Delta\Phi|$-asymmetric components of QCD are performed using regions enriched in such events. Background of this type is particularly important for the STA-STA category which is least able to distinguish low-mass events from those above our requirement of $m_{\mu\mu} > 10$ GeV. The STA-STA background estimation method is tested in two validation regions, selecting for dimuons with $6 < m_{\mu\mu} < 10$ GeV, and those with $|\Delta\eta_{\mu\mu}| < 0.1$ which would fail the subsequent requirements on DT hits and/or muon segments (referred to as 'small-$|\Delta\eta_{\mu\mu}|$'). These regions were verified to be

heavily populated by QCD-like events after inverting the STA to TMS matching logic as is described in Section 4.4.6.1 and looking at the associated TMS-TMS dimuons taken as proxies for the true dimuons. Figure 4.33 shows the observed number of events in these validation regions together with the prediction given by the estimation procedure. The low mass region is available only for 2018 data, as these types of events were rejected at the trigger level in 2016. Each bin contains an observed number of events consistent with that predicted by the background estimation procedure.

Backgrounds due to QCD-like processes are expected to be present in the tails of the displacement distributions in both the STA-TMS and TMS-TMS categories. Validation of the background estimation procedure in these regions is performed by looking at dimuons with $\pi/4 < |\Delta\Phi| < \pi/2$. The background estimation procedure is run by pairing this region with its symmetric half, namely, the region containing dimuons with $\pi/2 < |\Delta\Phi| < 3\pi/4$. The result is shown in Figure 4.34, in which the observed number of events are consistent with those predicted by the background estimation procedure.

4.6 Systematic Uncertainties and Scale Factors

Anticipated systematic differences between generated Monte Carlo signal simulation and data are corrected through the rescaling of simulation using scale factors (SFs), and by assigning uncertainties for cases in which an accurate correction is unavailable. The majority of these corrections and related uncertainties are evaluated separately between dimuon categories and data-taking periods. Unless otherwise stated, sources of uncertainties are considered to be uncorrelated in this regard.

The dominant systematic uncertainty of the analysis is due to the trigger. Evaluation of scale factors and corresponding uncertainties are calculated using muons from Z decays. For small displacements ($d_0 < 10$ cm), $p_{\mathrm{T}}-$ and $\eta-$dependent SFs associated with the trigger are measured using the tag-and-probe method [Flo]. They are calculated for both STA and TMS muons using single-muon versions of the double muon trigger used in the analysis,

covered in Section 4.3.1.

Trigger scale factors at large displacements ($d_0 > 10$ cm) are measured using cosmic data. The muon efficiencies as a function of d_0 are measured from L1 up to HLT to correct and/or validate the performance of our generated signal samples. The full results are documented in Section 6 of Reference [Val19a]. The scale factors are evaluated for each L1 p_{T} threshold and HLT used in the analysis and are shown as a function of d_0 in Figure 4.35, which includes both the cosmic and collision data results. No clear dependence is observed on d_0 is observed in any of the resulting distributions. The signal efficiencies derived from Monte Carlo are therefore corrected by the product of the corresponding single muon trigger SFs derived from collision data. A systematic uncertainty which increases linearly as a function of d_0 as shown in Figure 4.35 for TMS (pink band) and STA (orange band) muons is also assigned to account for the lack of knowledge we have in this region [Vala].

In categories that use STA muons, muon identification also plays a large role in assigned systematic uncertainties. Efficiencies for reconstructing and identifying STA muons are obtained by applying the tag-and-probe method to muons from J/Ψ and Z decays [Fra, Muoa, Muob]. These measurements provide prompt scale factors which are extrapolated to non-zero d_0 values using a d_0-dependent correction function derived using cosmic muon data and simulation [Valb, Val19a]. STA muon identification is found to have an SF linearly dependent on d_0, resulting in a 3.5% (7.0%) reduction in efficiency which is applied on top of prompt measurements at $d_0 = 100$ (200) cm. An associated systematic uncertainty of 2% is also assigned. Signal efficiencies in the STA-STA category are scaled by factors ranging from $0.83 - 0.88$ depending on kinematics. This correction is the largest applied to dimuons in the STA-STA category.

In the TMS-TMS category, the tracking efficiency plays the second largest role in assigned systematic uncertainties after those associated with the trigger. Tracking efficiency for displaced particles had already been measured in the context of a search for displaced muons and electrons [Ali17], which measured their efficiencies as a function of both d_0 and d_{z}. Identical efficiency measurements were made using the full suite of requirements listed in

Section 4.4.5.5, yielding measurements consistent between simulation and data. Therefore no additional scale factors are applied, although a 5% systematic is assigned per muon to account for inaccurate modeling of the TMS muon efficiency in simulation.

The remaining systematic uncertainties and scale factors used in the analysis are subdominant ($\lesssim 5\%$) and are covered in detail in Reference [Val19b]. These include sources associated with: the total integrated luminosity; the pileup distribution; the dimuon L_{xy} and vertex χ^2; the TMS isolation variable and d_0; and the STA to TMS replacement procedure.

4.7 Results

The observed number of events together with the predicted background yields for the STA-STA and STA-TMS categories are shown in representative $m_{\mu\mu}$ bins in Figures 4.36 and 4.37, respectively. The expected distribution predicted the majority of events at low mass and is similar to the observed distribution for which nearly all events lie below $m_{\mu\mu} < 20$ GeV. The number of observed events in both of these categories is consistent with predicted background yields. The observed events in the TMS-TMS category are shown with their expected background as a function of $m_{\mu\mu}$ and $\min(d_0/\sigma_{d_0})$ in Figures 4.38 and 4.39, respectively. The expected distribution predicts the majority of dimuons to have a $\min(d_0/\sigma_{d_0})$ near the cut of and fall with a mass near 70 GeV. The observed events exhibit this same behavior. Both figures show a deficit of observed events in the 2018 data-taking period. This deficit was the subject of a dedicated study documented in Appendix C of Reference [Val19b], which concluded it was a result of a statistical fluctuation. Elsewhere, the number of observed events is consistent with the predicted background. In total, no excess of events above those predicted by the standard model is observed.

4.7.1 Limit Setting Procedure

With no evidence of physics beyond the standard model having been observed, upper limits on signal cross-sections are then calculated. This is done by relating the number of signal

events *not* seen to the corresponding process cross-section. For the BSM Higgs model, the total number of signal events present in data N_{s} is calculated from Equation 3.3 and given by

$$N_{\mathrm{s}} = \sigma\mathcal{L}[2\mathcal{B}(1-\mathcal{B})\epsilon_1 + \mathcal{B}^2\epsilon_2], \tag{4.17}$$

where σ is the process cross-section; $\mathcal{L}$ is the integrated luminosity; $\mathcal{B}$ is the branching fraction for the LLP to decay to two muons; ϵ_1 and ϵ_2 are the efficiencies to reconstruct and event with one and two generated LLPs, respectively. If the efficiency to reconstruct each dimuon is independent for events with multiple dimuons, one can write $\epsilon_2 = 1-(1-\epsilon_1)^2 = 2\epsilon_1 - \epsilon_1^2$. In practice, these efficiencies are correlated and the presence of two dimuons increases the chances that either will be reconstructed. However, this underestimation only leads to less stringent limits and is taken as a simplification. One may then write the simplified version of Equation 4.17 as

$$N_{\mathrm{s}} = \sigma(\Phi \to \mathrm{XX})\mathcal{L}\mathcal{B}(\mathrm{X} \to \mu\mu)[2 - \mathcal{B}(\mathrm{X} \to \mu\mu)\epsilon_1]\epsilon_1. \tag{4.18}$$

This equation depends on the assumed value of $\mathcal{B}(\mathrm{X} \to \mu\mu)$, scaling as $1/[1 - \frac{1}{2}\mathcal{B}(\mathrm{X} \to \mu\mu)\epsilon_1]$. Since the signal efficiency is small in the majority of cases (Figure 4.29), we impose a conservative requirement of $\mathcal{B}(\mathrm{X} \to \mu\mu) = 1$ for the term in brackets, which gives us a linear relationship between the number of observed events and $\sigma(\Phi \to \mathrm{XX})\mathcal{B}(\mathrm{X} \to \mu\mu)$. For the HAHM model, the event yield is similarly given by

$$N_{\mathrm{s}} = \sigma(\mathrm{h} \to \mathrm{Z_D Z_D})\mathcal{L}\mathcal{B}(\mathrm{Z_D} \to \mu\mu)[2 - \mathcal{B}(\mathrm{Z_D} \to \mu\mu)]\epsilon_{\mathrm{Z_D}}. \tag{4.19}$$

No such simplification is performed for the HAHM model, since each value of $\mathcal{B}(\mathrm{Z_D} \to \mu\mu)$ is specified for a fixed mass according to Table 4.3. Signal efficiencies ϵ_1 and $\epsilon_{\mathrm{Z_D}}$ are calculated at arbitrary lifetimes via the lifetime reweighting procedure described in Section 4.4.6.2. The limit on the number of observed signal events is therefore easily translated to limits on $\sigma(\Phi \to \mathrm{XX})\mathcal{B}(\mathrm{X} \to \mu\mu)$ and $\sigma(\mathrm{h} \to \mathrm{Z_D Z_D})\mathcal{B}(\mathrm{Z_D} \to \mu\mu)$ for the two respective models.

The limit on the number of signal events is set using the `combine` tool [The], which was originally designed for statistically combining results of separate Higgs boson channels. Our analysis is binned which allows us to treat each bin as its own counting experiment. This builds the basis of our model: which posits that the full likelihood function $\mathcal{L}$ is the product of Poisson distributions for each respective bin. The rate of the Poisson distribution is assigned as the sum of rates of both signal and background processes in that bin. Bins that have associated systematic uncertainties are further multiplied by log-normal distributions parametrized by the size of the assigned uncertainty. Both the background rates and systematics uncertainties are taken as nuisances parameters θ while statistics on the signal rate μ are derived. A test statistic $\tilde{q}_\mu$ based on our observed data is then defined as

$$\tilde{q}_\mu = -2 \ln \frac{\mathcal{L}(\text{data}|\mu, \hat{\theta}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu}, \hat{\theta})}, \tag{4.20}$$

where $\hat{\mu}$ and $\hat{\theta}$ are the respective signal rate and nuisance parameters which globally maximize the likelihood function given the data. The parameter $\hat{\theta}_\mu$ is defined such that it maximizes the likelihood function conditional to a given μ. In `combine`, this is known as the `LHC-style` statistic. Furthermore, we require $0 \leq \hat{\mu} \leq \mu$. The lower constraint $0 \leq \hat{\mu}$ is simply a requirement that we cannot have a negative event yield, while the upper constraint $\hat{\mu} \leq \mu$ is enforced such that fluctuations in data giving $\hat{\mu} > \mu$ do not work against the signal hypothesis [com11].

Using the observed counts in each bin, one can perform a fit to find values of $\hat{\mu}^{\text{obs}}, \hat{\theta}^{\text{obs}}$ and $\hat{\theta}_\mu^{\text{obs}}$ which maximize the respective likelihoods. Plugging these values into Equation 4.20 then gives us our observed test statistic $\tilde{q}_\mu^{\text{obs}}$. The fitted values are then used to generate toy Monte Carlo pseudo-data, which treat the likelihood functions as probability density functions (pdfs). These pdfs are defined under both the background-only hypothesis and that which includes both signal and background, denoted as $\mathcal{P}(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\text{obs}})$ and $\mathcal{P}(\tilde{q}_\mu|0, \hat{\theta}_0^{\text{obs}})$, respectively. Notably these pdfs are defined according to the unblinded *post-fit* values, a point of some controversy [The].

One can then define two $p-$values associated with the observed counts: the probability that we would observe at least this many events given there is only background p_b; and the probability we would observe this many events if there is also a signal p_μ, with

$$1 - p_b = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}} | \text{background-only}) = \int_{\tilde{q}_0^{\text{obs}}}^{\infty} \mathcal{P}(\tilde{q}_\mu | 0, \hat{\theta}_\mu^{\text{obs}}) d\tilde{q}_\mu, \tag{4.21}$$

$$p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}} | \text{signal+background}) = \int_{\tilde{q}_\mu^{\text{obs}}}^{\infty} \mathcal{P}(\tilde{q}_\mu | \mu, \hat{\theta}_\mu^{\text{obs}}) d\tilde{q}_\mu. \tag{4.22}$$

One could use only the value of p_μ to compute corresponding limits, however, in the case of a downward fluctuation in the background from that which is expected, resulting upper limits can become computed as anomalously low [Jun99, Zyl20]. By also using the $p-$values associated with the background only hypothesis, one can define the modified frequentist confidence level[1] as

$$\text{CL}_\text{s}(\mu) = \frac{p_\mu}{1 - p_b}. \tag{4.23}$$

The 95% confidence level upper limit is therefore given by the value of μ at which $\text{CL}_\text{s}(\mu) = 0.05$. This value is then scaled to limits on the signal cross-section.

4.7.2 Upper Limits

The observed and expected 95% CL_s upper limits are shown as a function of $c\tau$ for the BSM Higgs model for $m_\Phi = 125$ GeV in Figure 4.40, $m_\Phi = 200$ GeV in Figure 4.41, $m_\Phi = 400$ GeV in Figure 4.42, and $m_\Phi = 1000$ GeV in Figure 4.43. All figures use the combined $2016 + 2018$ datasets and are overlaid with the contributions from the respective three categories. The most sensitive limits are achieved in regions ranging from 0.1 mm to 100 m, where $\sigma(\Phi \to \text{XX})\mathcal{B}(\text{X} \to \mu\mu)$ is excluded to cross-sections of less than 1 fb. In general limits at larger Φ mass are more stringent due to fewer background events present in the signal region.

[1]Strictly speaking, the value of CL_s is not a confidence level, but a ratio of two confidence levels [Rea02].

Corresponding HAHM limits are shown in Figure 4.44. Each set of limits are shown with the theoretical predictions given a specified branching fraction of $\mathcal{B}(\mathrm{h} \to \mathrm{Z_D Z_D})$, ranging from values of $0.001 - 1\%$. The results are also presented in two dimensions in Figure 4.45 as a function of both dark photon lifetime vs mass (left), and ϵ vs mass (right). Additional signal points ($m(Z_\mathrm{D}) = 13, 16, 25, 35, 45$, and 55 GeV) are used which have their efficiencies interpolated from neighboring generated samples to cover the gaps in mass listed in Table 4.17. The limits set by this analysis exclude branching fractions of $\mathcal{B}(\mathrm{h} \to \mathrm{Z_D Z_D}) = 0.1\%$ for dark photon masses ranging from 20 to 60 GeV over a range of lifetimes spanning roughly seven orders of magnitude. These results put constraints on rare SM Higgs boson decays which are tighter than those derived from searches for invisible Higgs boson decays [Sir19b] and those from indirect constraints obtained by measurements of the SM Higgs boson couplings [Sir19a].

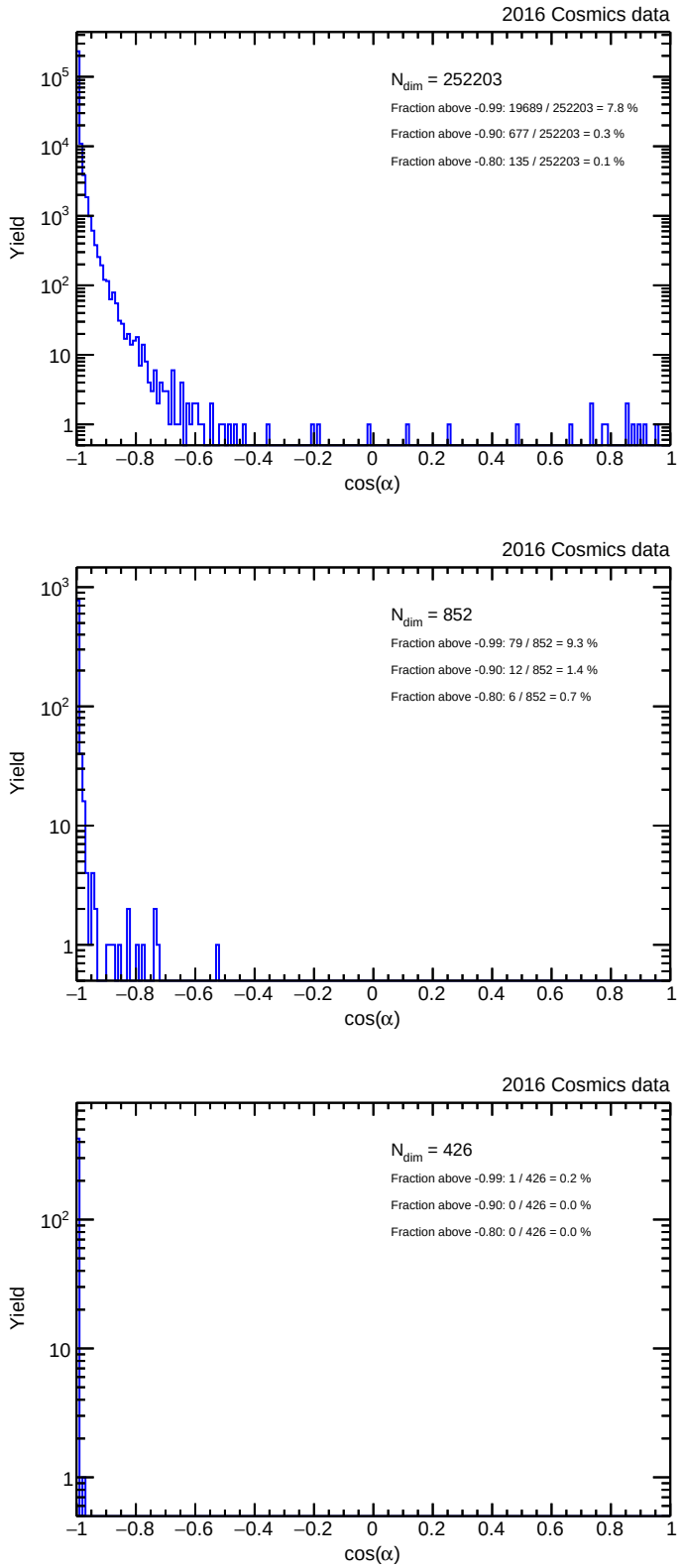


Figure 4.30: Distributions of $\cos\alpha$ for STA-STA (top), STA-TMS (middle), and TMS-TMS (bottom) dimuons in a dedicate sample of cosmic ray muons. Additional details are provided in Reference [Val19a].

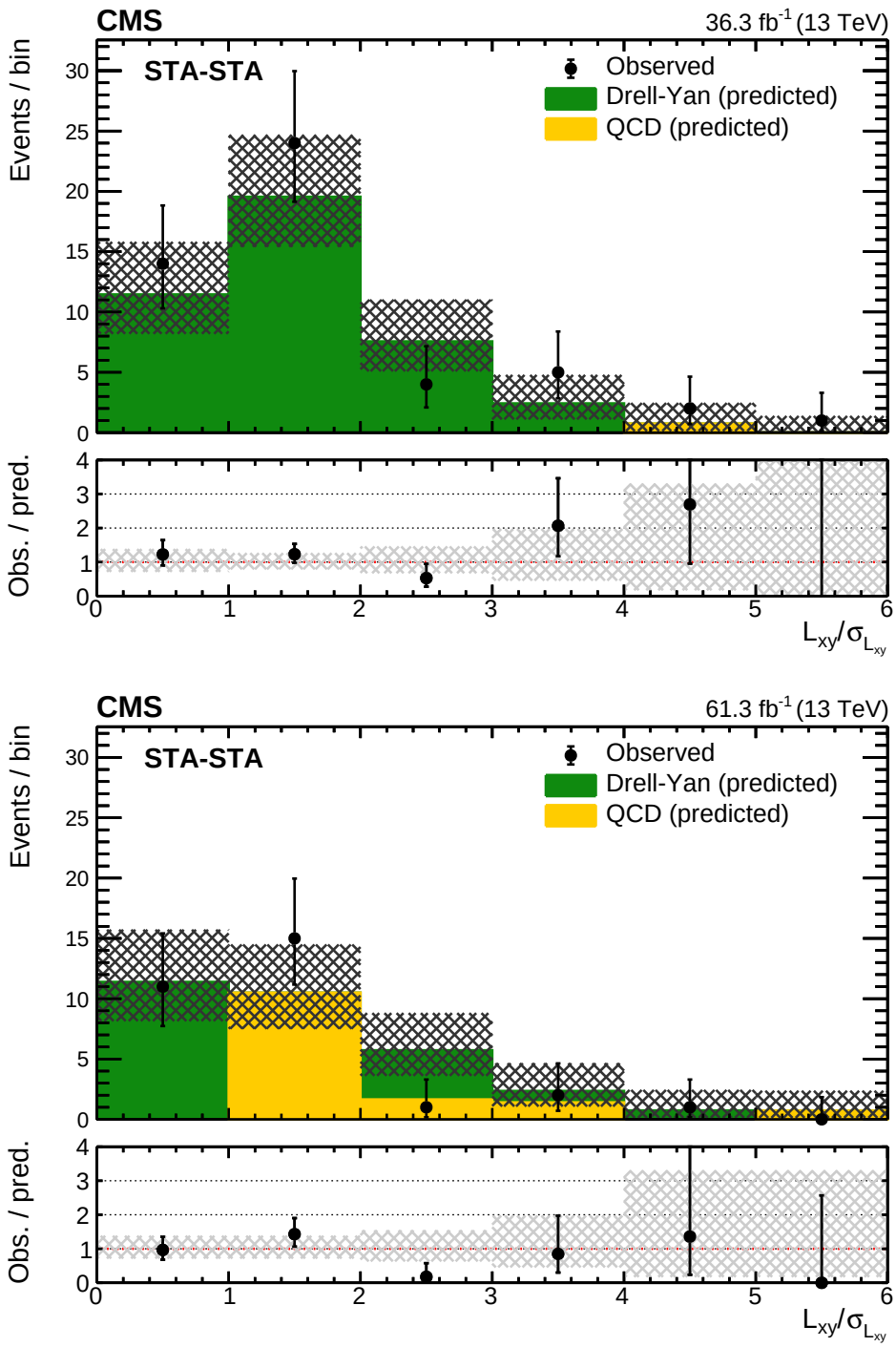


Figure 4.31: Validation of the background estimation procedure as a function of $L_{xy}/\sigma_{L_{xy}}$ for STA-STA dimuons with $L_{xy}/\sigma_{L_{xy}} < 6$. The observed distribution (black circles) are overlaid with stack histograms containing the expected number of DY (green) and QCD (yellow) background events in 2016 (top) and 2018 (bottom). The lower panels show the ratio of the observed to predicted numbers of events. Shaded area shows the statistical uncertainty in the background prediction.

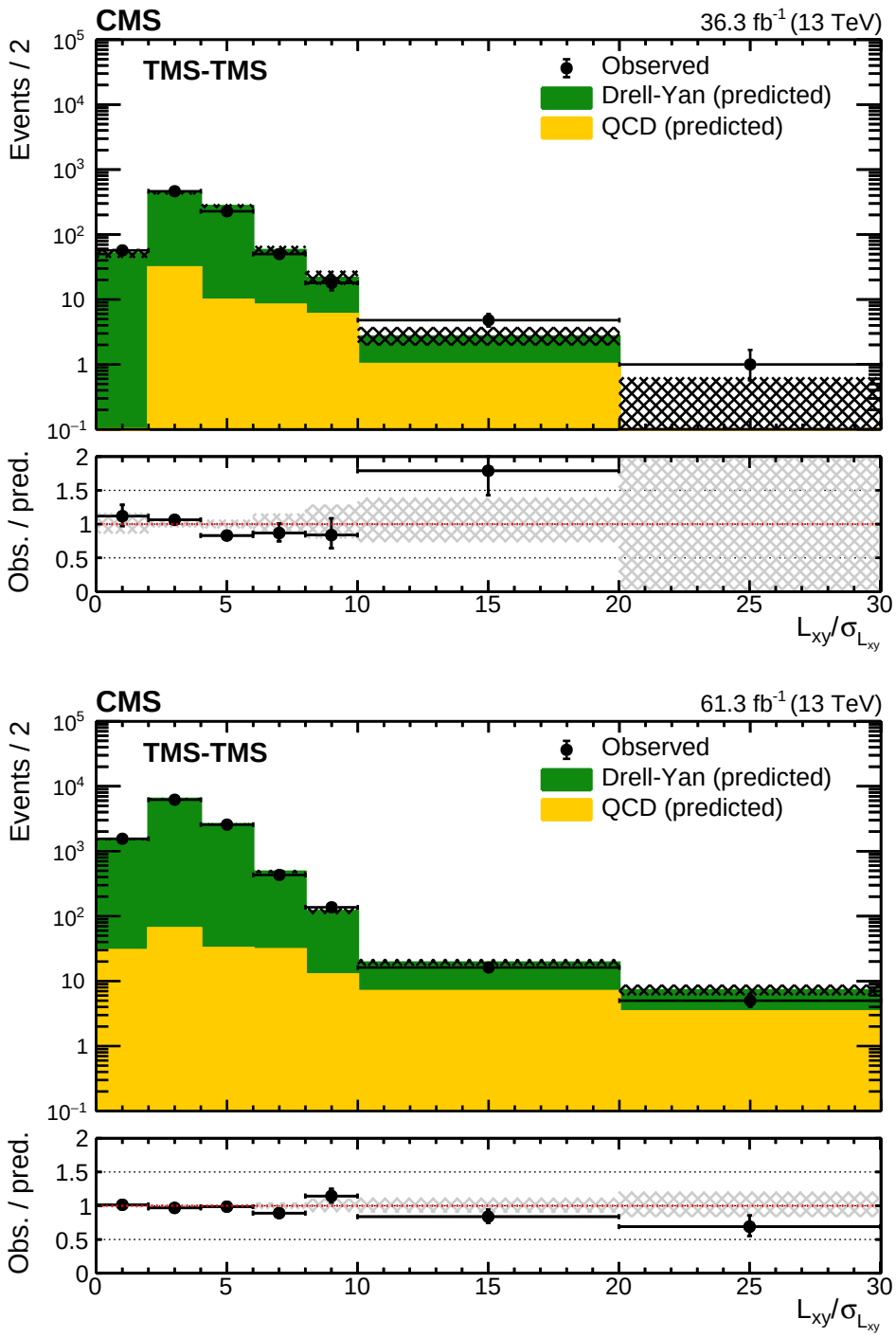


Figure 4.32: Validation of the background estimation procedure as a function of $L_{xy}/\sigma_{L_{xy}}$ for TMS-TMS dimuons with $2 < \min(d_0/\sigma_{d_0}) < 6$. The observed distribution (black circles) are overlaid with stack histograms containing the expected number of DY (green) and QCD (yellow) background events in 2016 (top) and 2018 (bottom). The lower panels show the ratio of the observed to predicted numbers of events. Shaded area shows the statistical uncertainty in the background prediction.

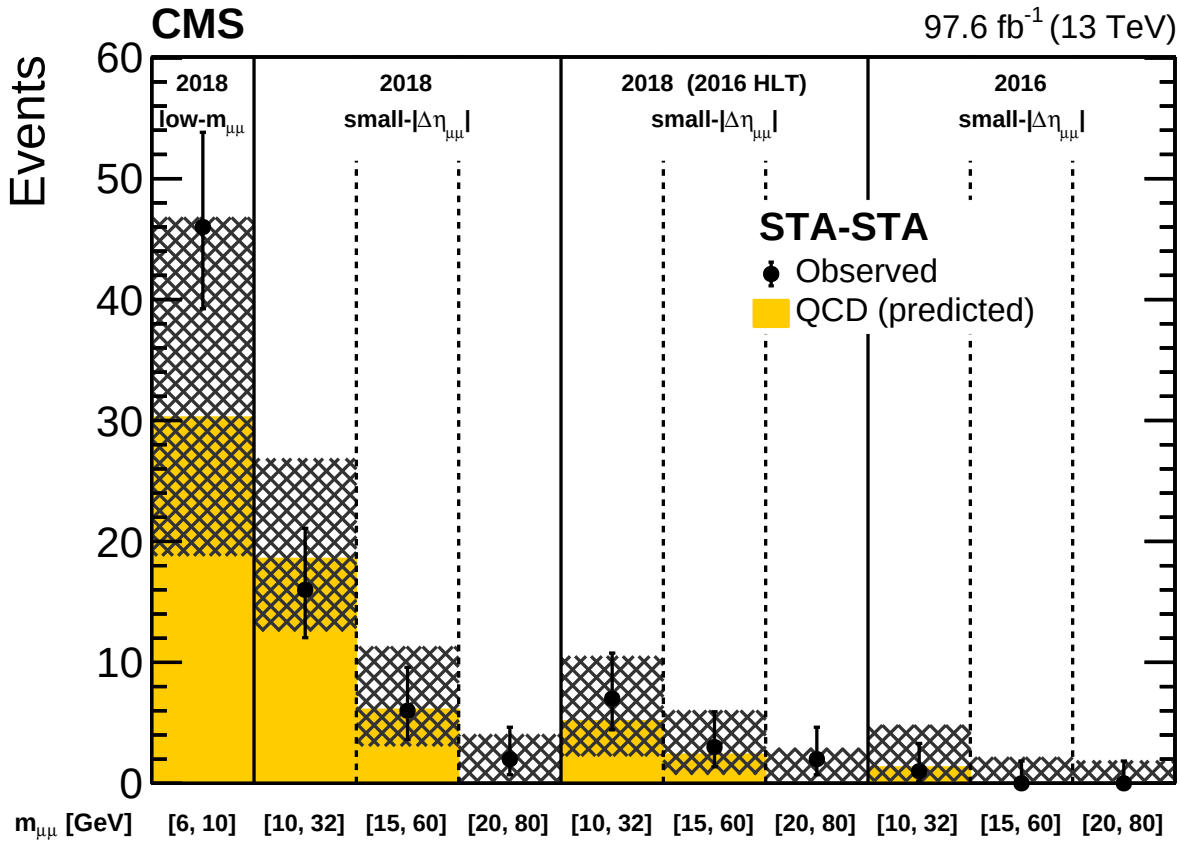


Figure 4.33: Validation of the background estimation procedure for STA-STA dimuons as a function of $m_{\mu\mu}$ in various regions. The observed number of events (black circles) is compared with the result of the background prediction (yellow) for low mass events and those with small-$|\Delta\eta_{\mu\mu}|$. The first bin includes low-mass events measured in 2018, unavailable in 2016 due to the trigger. The remaining bins include both the observed and predicted number of events with small-$|\Delta\eta_{\mu\mu}|$ for both 2018 and 2016 data. Bins five through seven include events measured in 2018 which would have passed the tightened trigger requirement present during 2016 data-taking. Shaded area shows the statistical uncertainty in the background prediction.

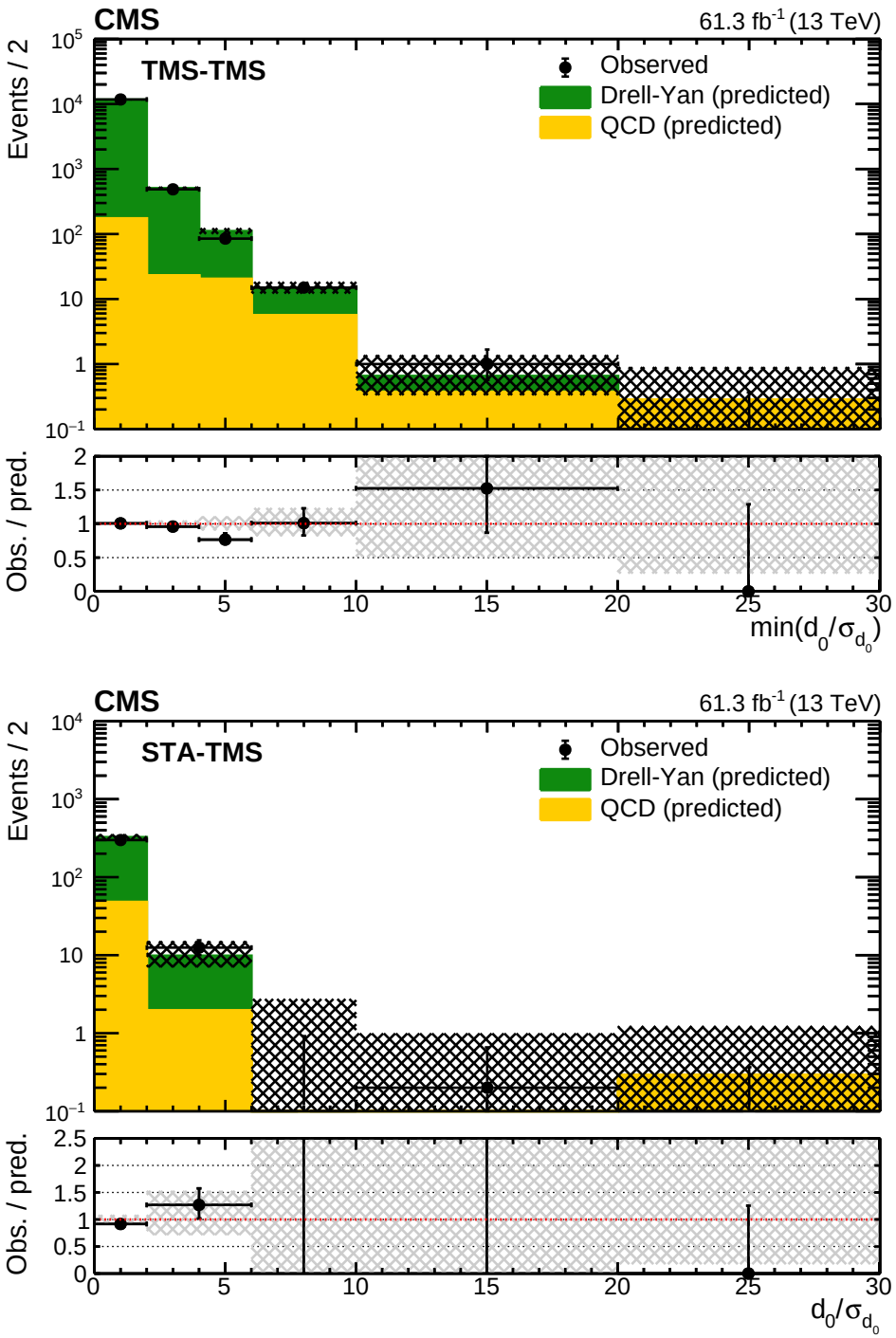


Figure 4.34: Validation of the background estimation procedure for TMS-TMS (top) and STA-TMS (bottom) dimuons with $\pi/4 < |\Delta\Phi| < \pi/2$ as a function of $\min(d_0/\sigma_{d_0})$ and d_0/σ_{d_0} respectively. The observed distributions (black circles) are compared to the results of the background prediction method applied to events with $\pi/2 < |\Delta\Phi| < 3\pi/4$. The stacked histograms show the expected numbers of DY (green) and QCD (yellow) background events. The last bin includes events in the overflow. The lower panels show the ratios of the observed to predicted numbers of events. Shaded area shows the statistical uncertainty in the background prediction.

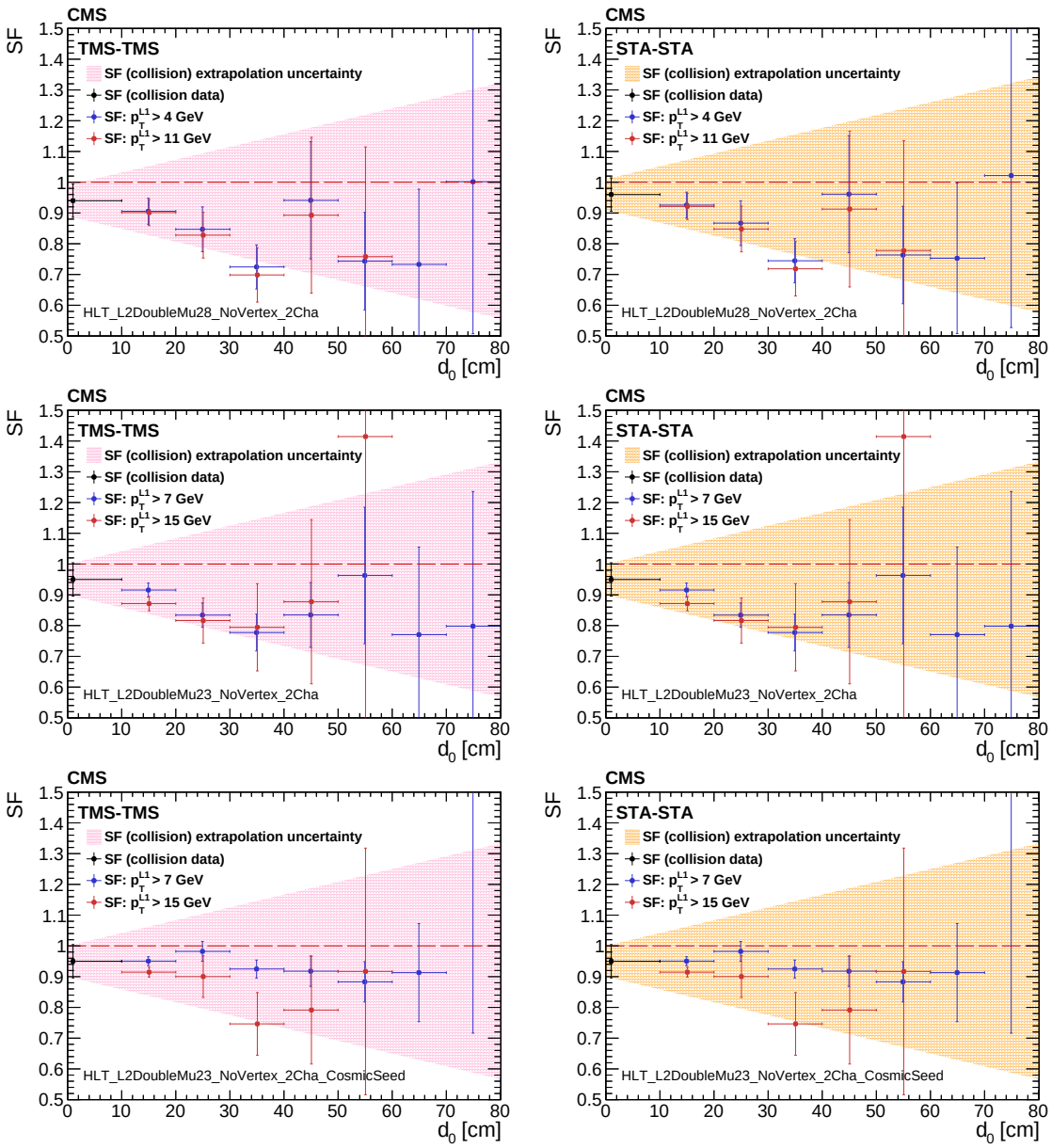


Figure 4.35: Single muon trigger efficiency scale factors for TMS (left) and STA (right) muons. Scale factors are calculated for the HLT configurations used in both the 2016 (top) and 2018 (middle and bottom). Each muon is assigned a scale factor measure from collision data (black dots), which is verified as a function of d_0 using cosmic data at a given L1 p_{T} threshold (red and blue). A d_0-dependent systematic uncertainty is then assigned equal to the size of the corresponding bands.

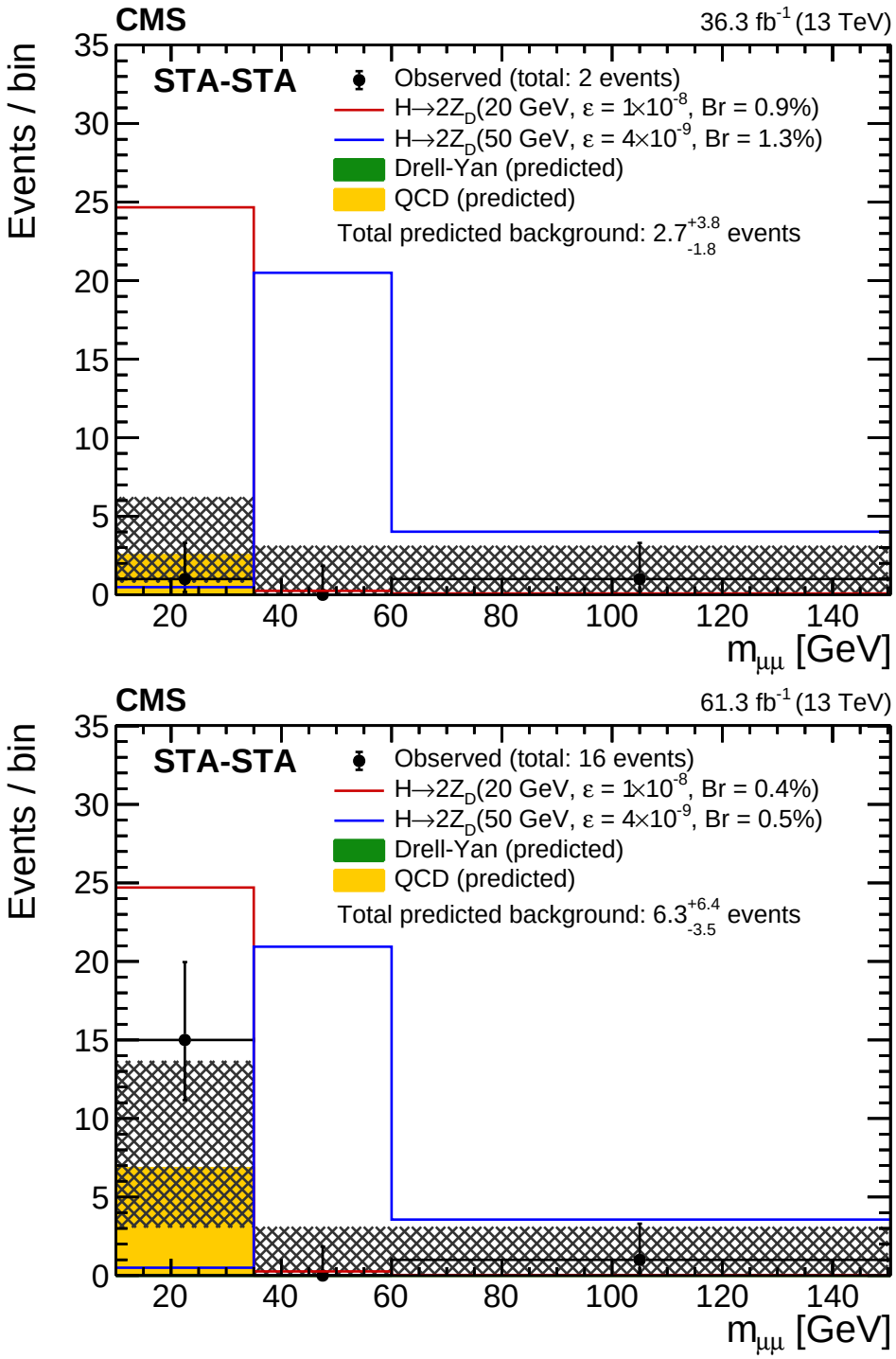


Figure 4.36: Observed number of events (black dots) in the signal region of 2016 (top) and 2018 (bottom) in the STA-STA dimuon category together with the number of expected background events (green and yellow), in representative $m_{\mu\mu}$ intervals. The last bin includes events in the overflow. The uncertainties in the total expected background (shaded area) are statistical only. The legends include the total number of observed events as well as the $m_{\mu\mu}$-integrated number of expected background events, which is obtained by applying the background evaluation method to the events in all mass intervals combined.

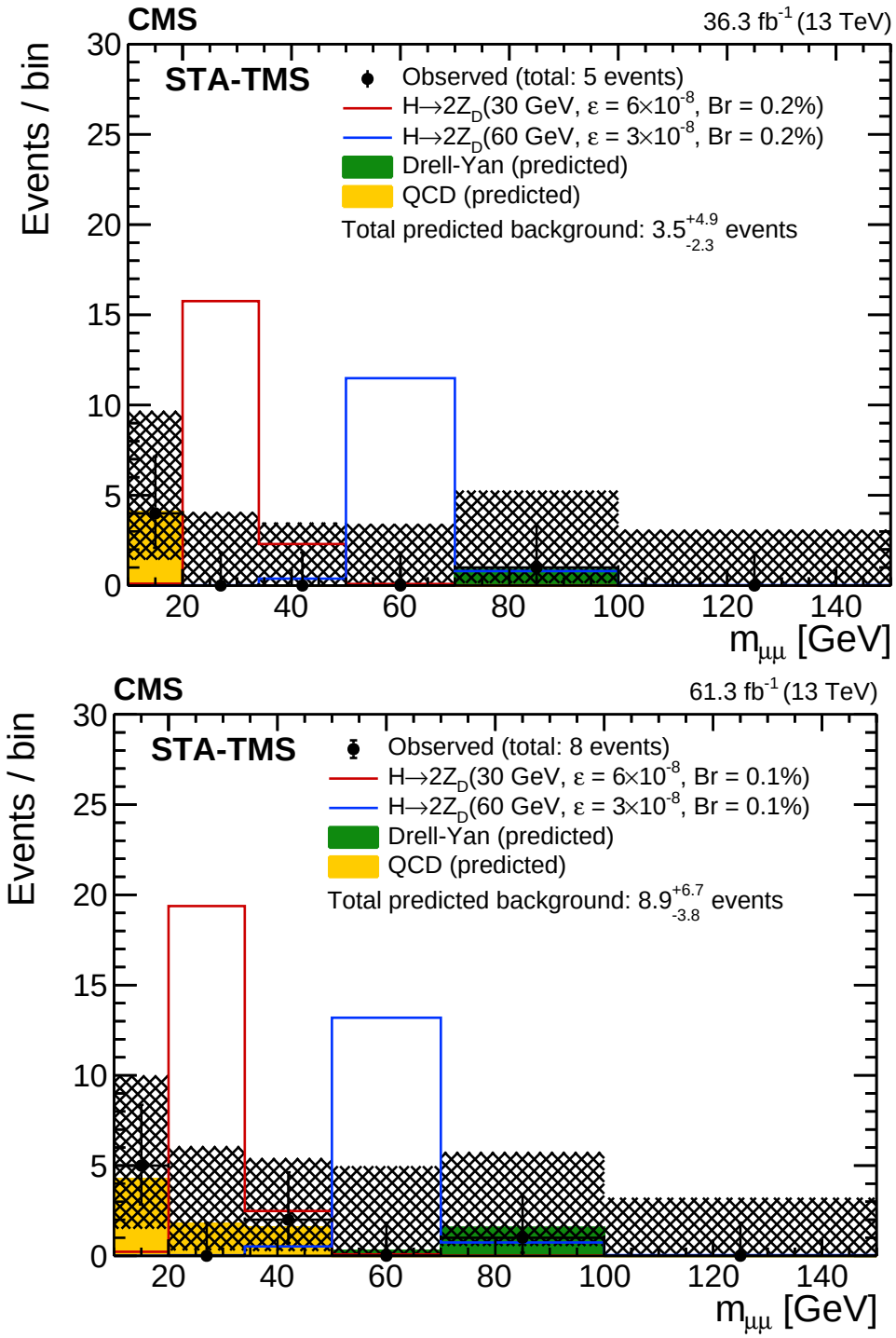


Figure 4.37: Observed number of events (black dots) in the signal region of 2016 (top) and 2018 (bottom) in the STA-TMS dimuon category together with the number of expected background events (green and yellow), in representative $m_{\mu\mu}$ intervals. The last bin includes events in the overflow. The uncertainties in the total expected background (shaded area) are statistical only. The legends include the total number of observed events as well as the $m_{\mu\mu}$-integrated number of expected background events, which is obtained by applying the background evaluation method to the events in all mass intervals combined.

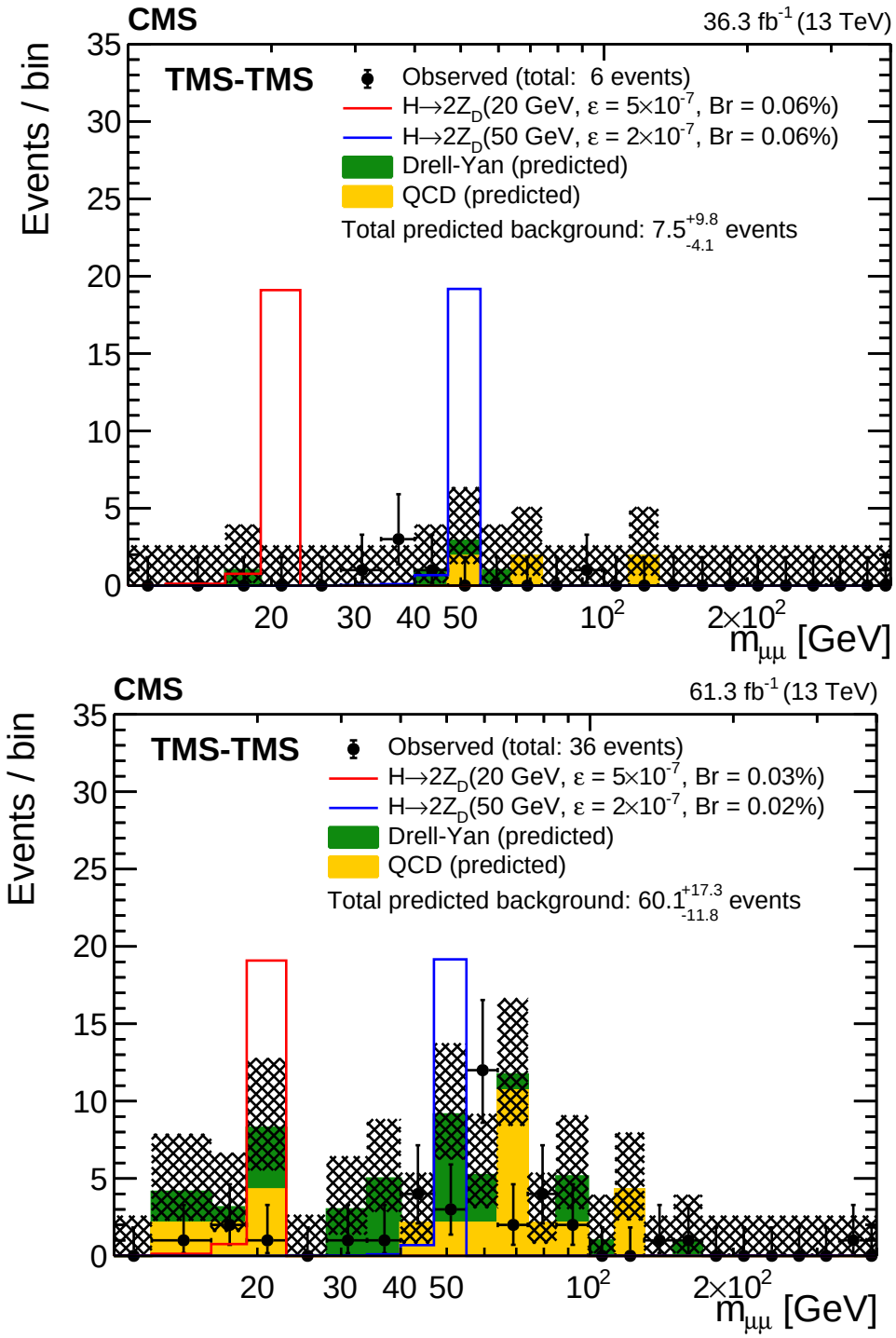


Figure 4.38: Observed number of events (black dots) in the signal region of 2016 (top) and 2018 (bottom) in the TMS-TMS dimuon category together with the number of expected background events (green and yellow), in representative $m_{\mu\mu}$ intervals. The last bin includes events in the overflow. The uncertainties in the total expected background (shaded area) are statistical only. The legends include the total number of observed events as well as the $m_{\mu\mu}$-integrated number of expected background events, which is obtained by applying the background evaluation method to the events in all mass intervals combined.

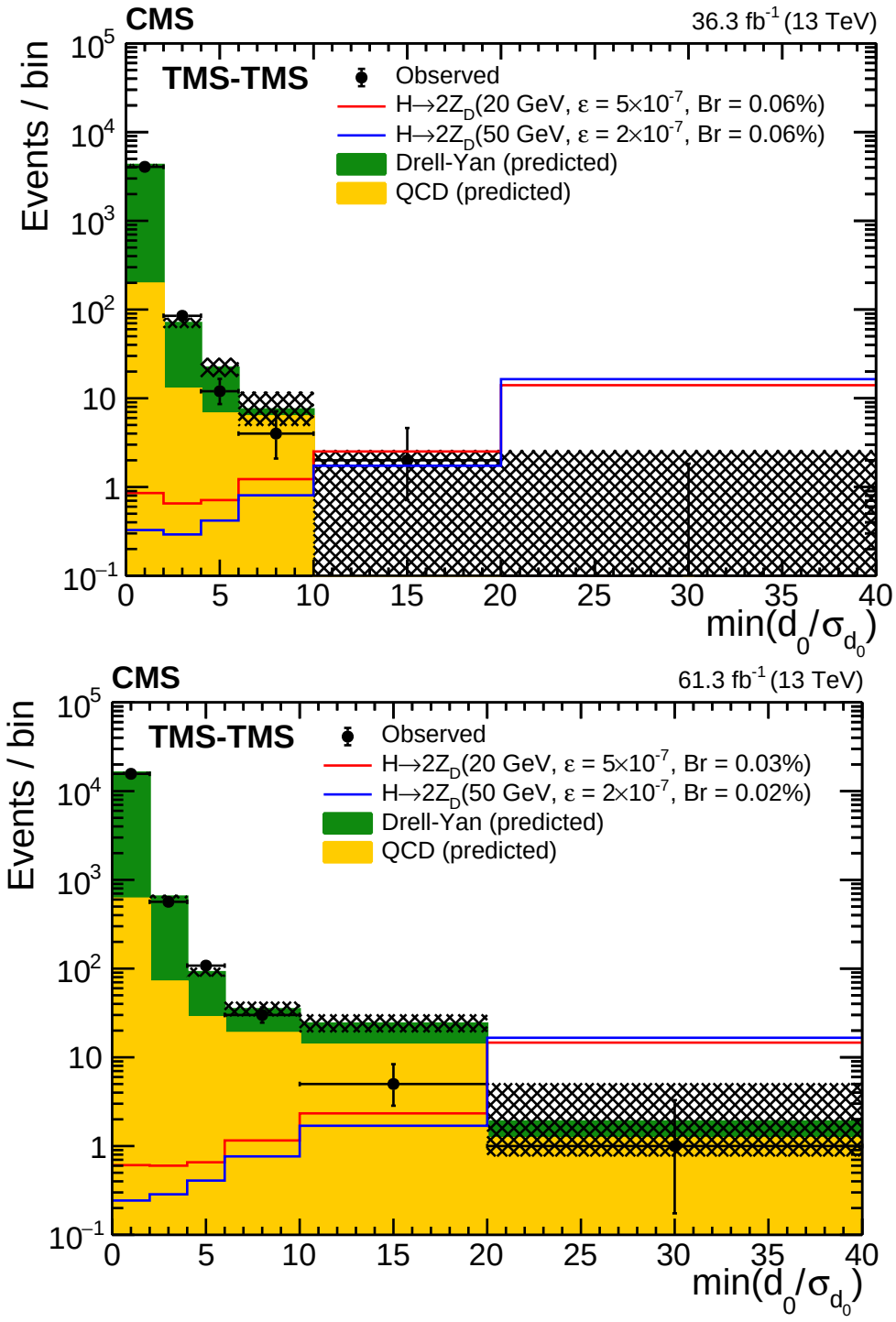


Figure 4.39: Observed number of events (black dots) in the signal region of 2016 (top) and 2018 (bottom) in the TMS-TMS dimuon category together with the number of expected background events (green and yellow), as a function of the smaller of the two d_0/σ_{d_0} values of the dimuon. The last bin includes events in the overflow. The uncertainties in the total expected background (shaded area) are statistical only.

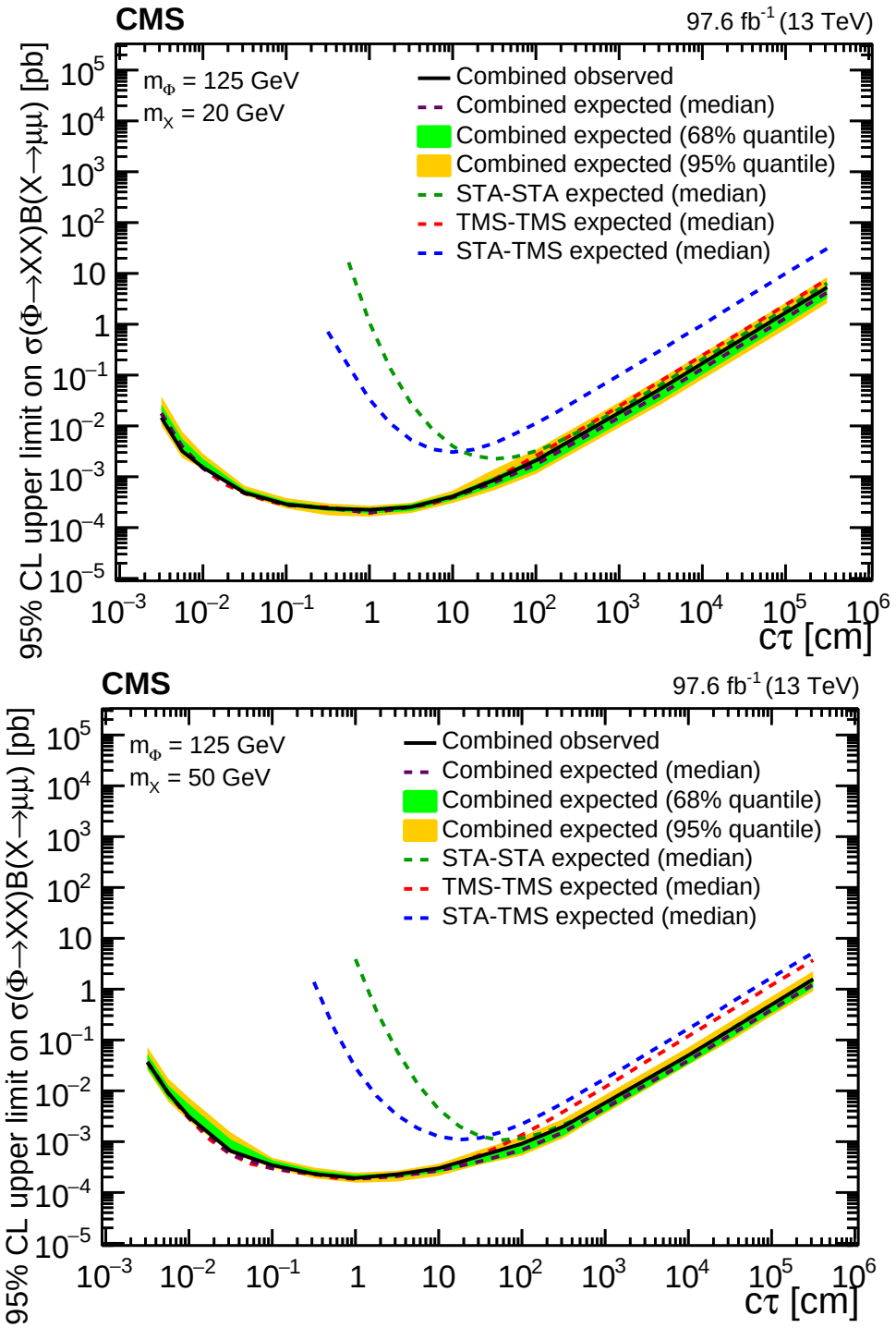


Figure 4.40: The 95% CL upper limits on $\sigma(\Phi \to \mathrm{XX})\mathcal{B}(\mathrm{X} \to \mu\mu)$ as a function of $c\tau(\mathrm{X})$ in the heavy scalar model, for $m_\Phi = 125$ GeV and (top) $m_\mathrm{X} = 20$ GeV and (bottom) $m_\mathrm{X} = 50$ GeV . The median expected limits obtained from the STA-STA, STA-TMS, and TMS-TMS dimuon categories are shown as dashed green, blue, and red curves, respectively; the combined median expected limits are shown as dashed black curves; the combined observed limits are showed as solid black curves. The green and the yellow bands correspond, respectively, to the 68% and 95% central quantiles for the combined expected limits.

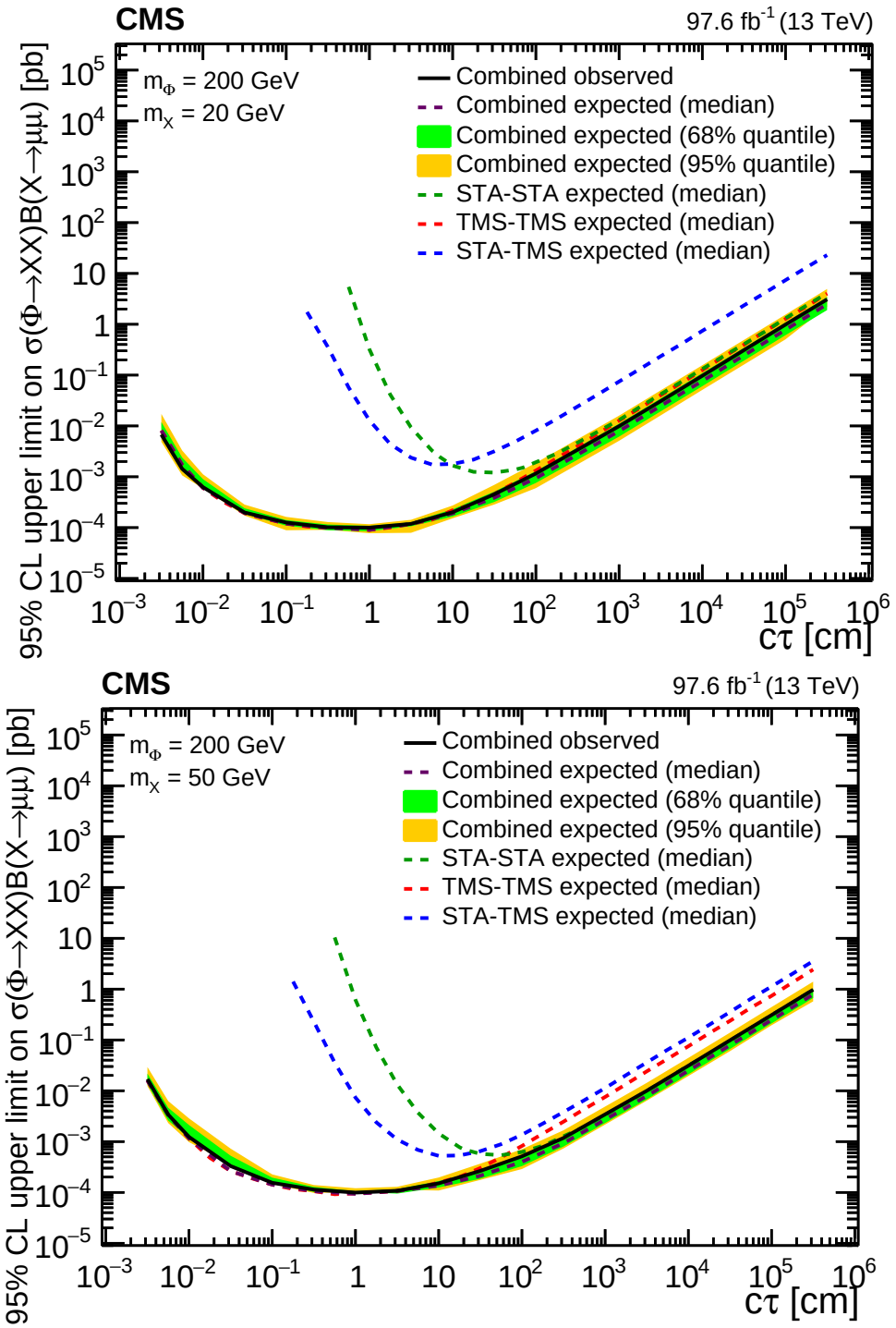


Figure 4.41: The 95% CL upper limits on $\sigma(\Phi \to \mathrm{XX})\mathcal{B}(\mathrm{X} \to \mu\mu)$ as a function of $c\tau(\mathrm{X})$ in the heavy scalar model, for $m_\Phi = 200$ GeV and (top) $m_\mathrm{X} = 20$ GeV and (bottom) $m_\mathrm{X} = 50$ GeV . The median expected limits obtained from the STA-STA, STA-TMS, and TMS-TMS dimuon categories are shown as dashed green, blue, and red curves, respectively; the combined median expected limits are shown as dashed black curves; the combined observed limits are showed as solid black curves. The green and the yellow bands correspond, respectively, to the 68% and 95% central quantiles for the combined expected limits.

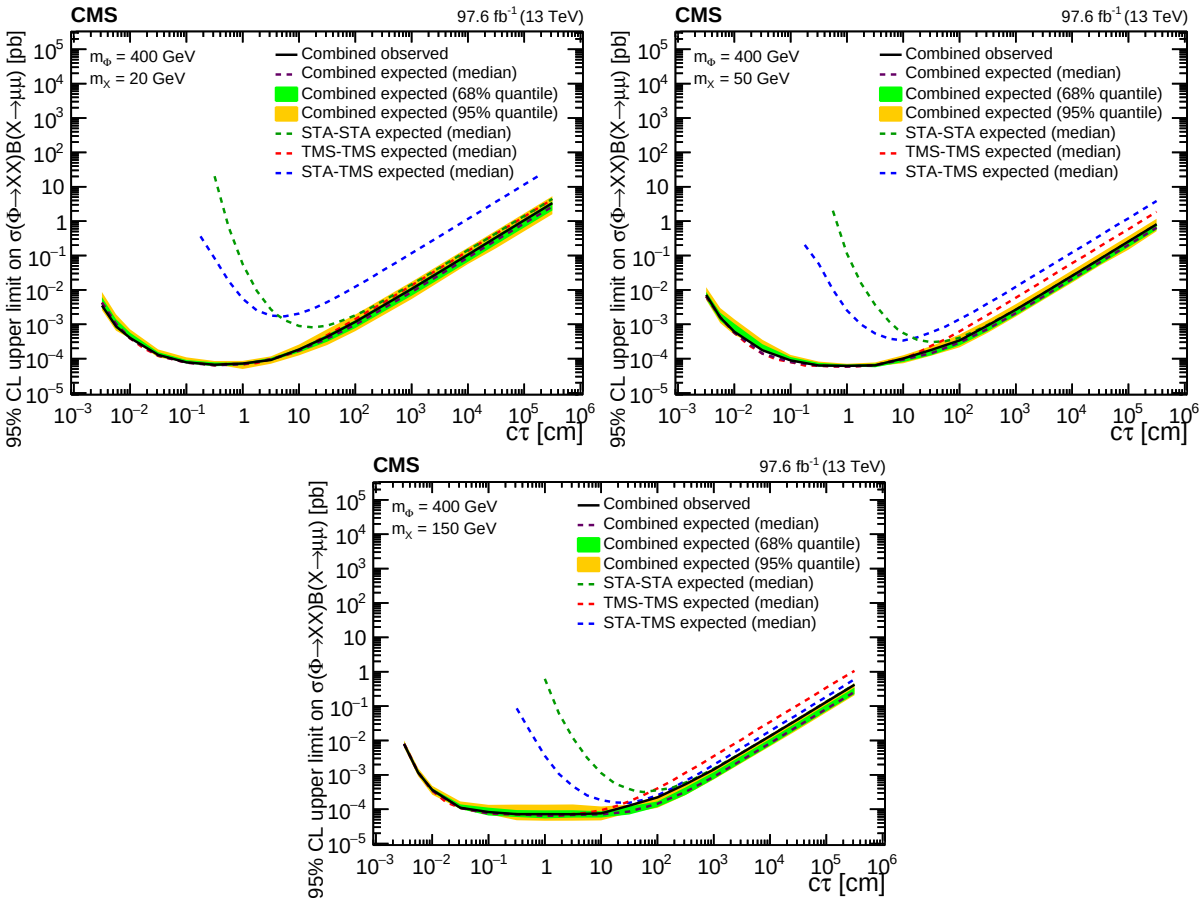


Figure 4.42: The 95% CL upper limits on $\sigma(\Phi \to \mathrm{XX})\mathcal{B}(\mathrm{X} \to \mu\mu)$ as a function of $c\tau(\mathrm{X})$ in the heavy scalar model, for $m_\Phi = 400$ GeV and (upper left) $m_\mathrm{X} = 20$ GeV , (upper right) $m_\mathrm{X} = 50$ GeV , and (lower) $m_\mathrm{X} = 150$ GeV . The median expected limits obtained from the STA-STA, STA-TMS, and TMS-TMS dimuon categories are shown as dashed green, blue, and red curves, respectively; the combined median expected limits are shown as dashed black curves; the combined observed limits are showed as solid black curves. The green and the yellow bands correspond, respectively, to the 68% and 95% quantiles for the combined expected limits.

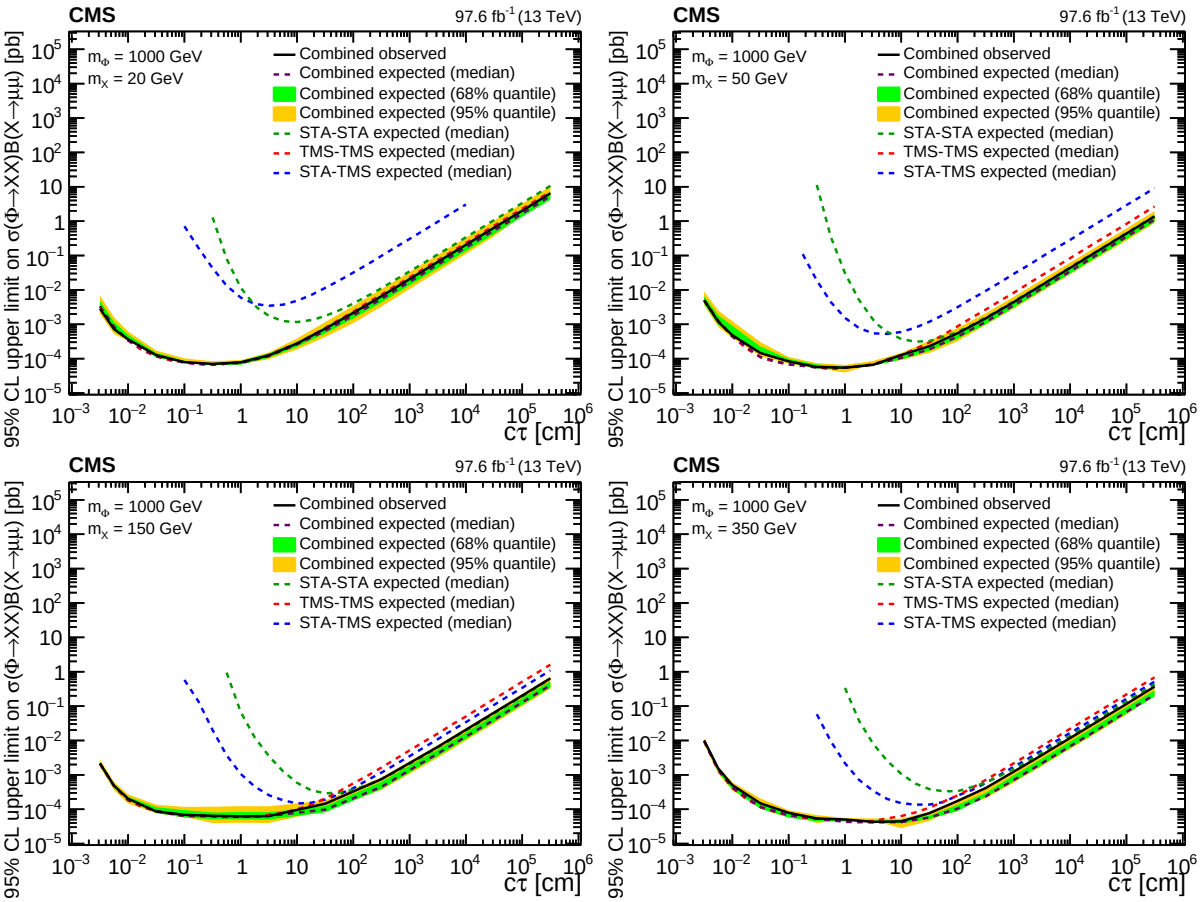


Figure 4.43: The 95% CL upper limits on $\sigma(\Phi \to \mathrm{XX})\mathcal{B}(\mathrm{X} \to \mu\mu)$ as a function of $c\tau(\mathrm{X})$ in the heavy scalar model, for $m_\Phi = 1$ TeV and (upper left) $m_\mathrm{X} = 20$ GeV , (upper right) $m_\mathrm{X} = 50$ GeV , (lower left) $m_\mathrm{X} = 150$ GeV , and (lower right) $m_\mathrm{X} = 350$ GeV . The median expected limits obtained from the STA-STA, STA-TMS, and TMS-TMS dimuon categories are shown as dashed green, blue, and red curves, respectively; the combined median expected limits are shown as dashed black curves; the combined observed limits are showed as solid black curves. The green and the yellow bands correspond, respectively, to the 68% and 95% quantiles for the combined expected limits.

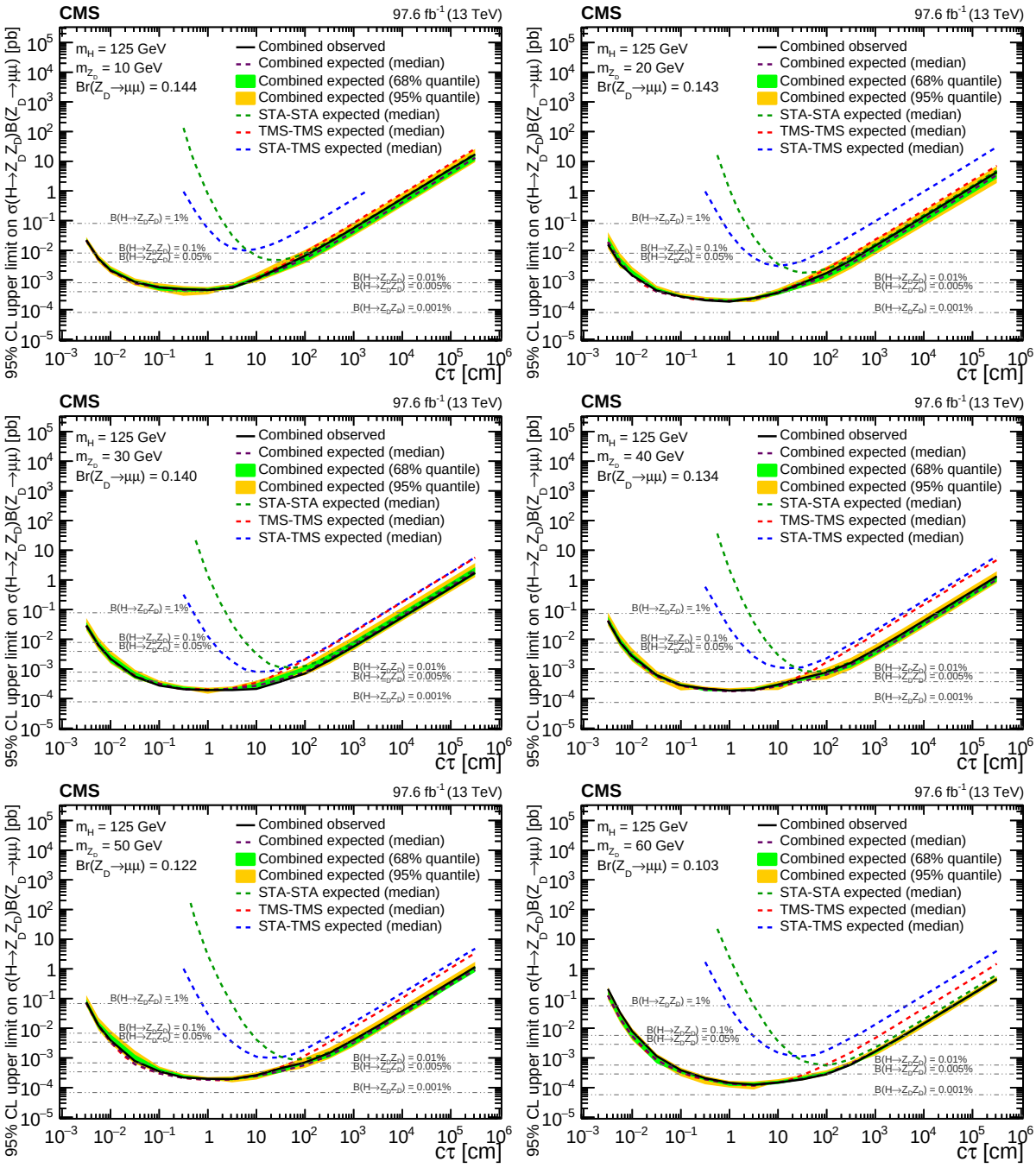


Figure 4.44: The 95% CL upper limits on $\sigma(\mathrm{H} \to \mathrm{Z_D Z_D})\mathcal{B}(\mathrm{Z_D} \to \mu\mu)$ as a function of $c\tau(\mathrm{Z_D})$ in the HAHM model, for $m(Z_\mathrm{D})$ ranging from 10 GeV (upper left) to 60 GeV (lower right). The median expected limits obtained from the STA-STA, STA-TMS, and TMS-TMS dimuon categories are shown as dashed green, blue, and red curves, respectively; the combined median expected limits are shown as dashed black curves; the combined observed limits are showed as solid black curves. The green and the yellow bands correspond, respectively, to the 68% and 95% central quantiles for the combined expected limits. The horizontal lines in gray correspond to the theoretical predictions for values of $\mathcal{B}(\mathrm{H} \to \mathrm{Z_D Z_D})$ indicated next to the lines.

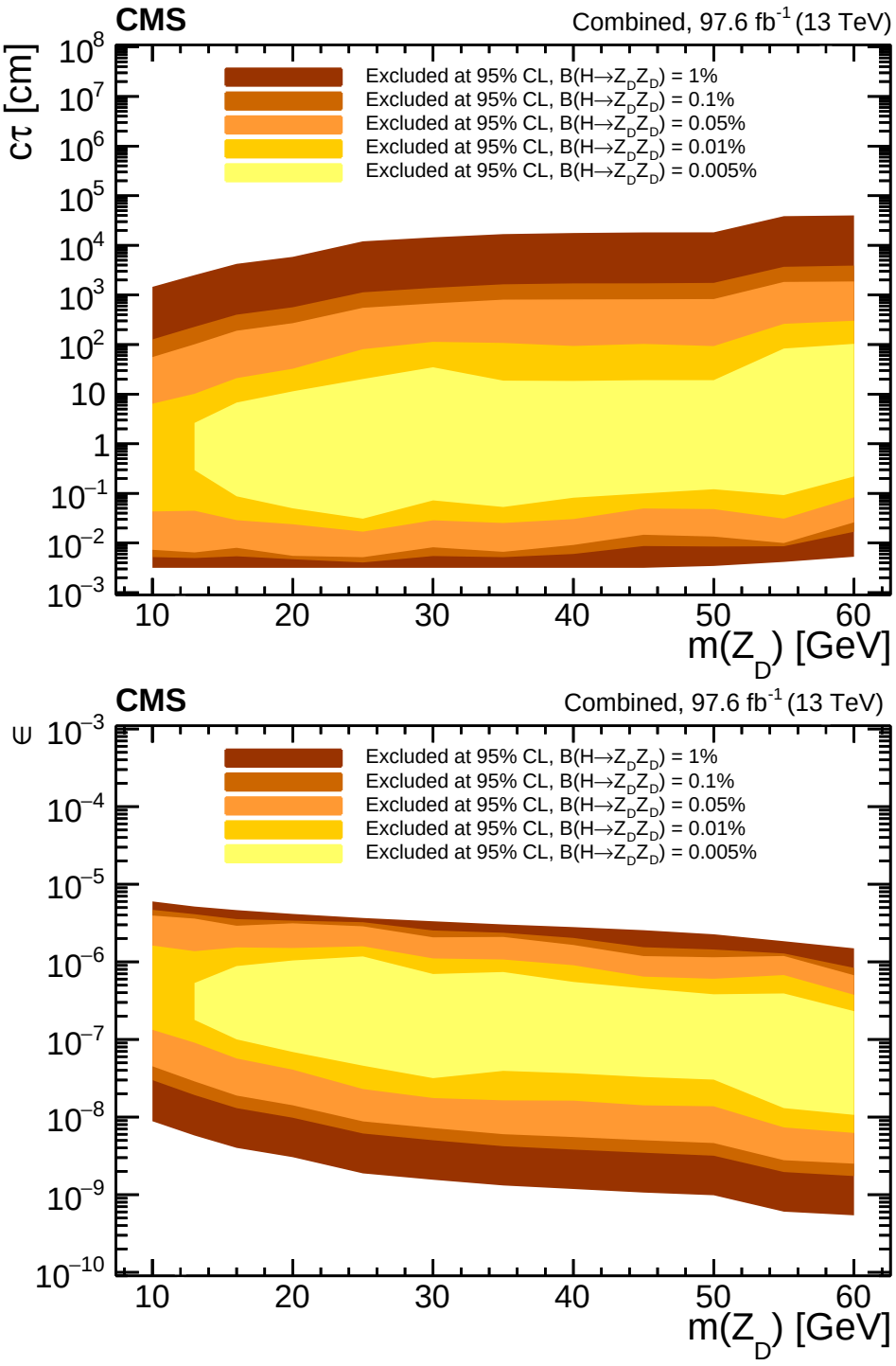


Figure 4.45: Observed 95% CL exclusion contours in the HAHM model, in the (top) ($m(Z_\mathrm{D})$, $c\tau(\mathrm{Z_D})$) and (bottom) ($m(Z_\mathrm{D})$, ϵ) planes. The contours correspond to several representative values of $\mathcal{B}(\mathrm{H} \to \mathrm{Z_D Z_D})$ ranging from 0.005% to 1%. The results of the measurements at twelve mass points (10, 13, 16, 20, 25, 30, 35, 40, 45, 50, 55, and 60 GeV) are connected by straight lines.

CHAPTER 5

Improving Low-Level Muon Resolution in the Endcap Muon Chambers

Resolve, and thou art free.

— Henry Wadsworth Longfellow (1807 - 1882)

The Level-1 (L1) Trigger of the CMS experiment uses coarse-grain primitives to filter proton-proton final states produced at 40 MHz down to a maximum trigger bandwidth of 100 kHz. The trigger was designed to have high efficiency for hard scattering processes which is satisfied in part by using minimum p_{T} thresholds for muon reconstruction [AT13]. This chapter describes how a muon measured in the endcap muon system results in the firing of the L1 trigger and an adjustment to this process which improves low-level muon resolution.

5.1 Cathode Strip Chambers

The endcap muon system measures muons using cathode strip chambers (CSCs) as described in Section 3.2.5.1. Each chamber contains six active layers of anode-cathode pairings that together provide a snapshot of the trajectory of the muon. Cathode charge information provides a measurement of the change in ϕ throughout the muons trajectory and is ultimately used in the determination of a muon's p_{T}.

Muon endcap (ME) chambers are identified according to the convention ME$\pm S/R/C$, where $\pm$ splits the chambers according to the sign of z of the endcap they sit, S is the station number $(1-4)$, which increases with increasing $|z|$, R is the ring number which increases radially $(1-3)$, and C is the chamber number that increases with ϕ $(1-36)$.

5.1.1 Cathode Front-End Board

Within each chamber, charge measured on all six layers of cathode strips is fed to the cathode front-end board (CFEB) which amplifies, digitizes, and stores the signals. Each chamber typically has five CFEB boards, each of which can handle 16 strips worth of cathode data. In the CFEB, charge is first amplified and then split along two separate data paths; the first processes the charge information at 40 MHz to provide muon trajectory information to the Level-1 (L1) Trigger, while the second holds the charge in a switched capacitor array to be processed with precision reconstruction if the L1 Trigger fires.

The data path which feeds the L1 Trigger uses application-specific integrated circuits (ASICs) to localize the charge on the detector. Comparator ASICs first look if the charge found on a strip exceeds a threshold, then identify which neighboring strip has the larger charge. The location of the ionization is then digitized as the half of the middle strip with a larger neighboring charge. This circuit is shown graphically in Figure 5.1. These digitized half-strip locations are referred to as comparators.

5.1.2 Trigger Mother Board

Each chamber contains either a trigger motherboard (TMB) or its chronological successor the optical trigger motherboard (OTMB) that collects and analyzes comparators from each of the chamber's CFEBs [Sev18a]. The (O)TMB uses low-level comparator information to construct track primitives known as cathode local charged tracks (CLCTs). CLCTs can then be combined with corresponding track primitives from the anode, known as anode local charge tracks (ALCTs), to construct a 2D local charged track (LCT). The (O)TMB sends up to two LCTs per bunch crossing to the muon port card (MPC). The MPC then funnels them to the endcap muon track finder (EMTF), where they are combined among chambers to reconstruct the full muon trajectory. This information is then used in determination in the firing of the L1 Trigger as explained in Section 3.2.6.1.

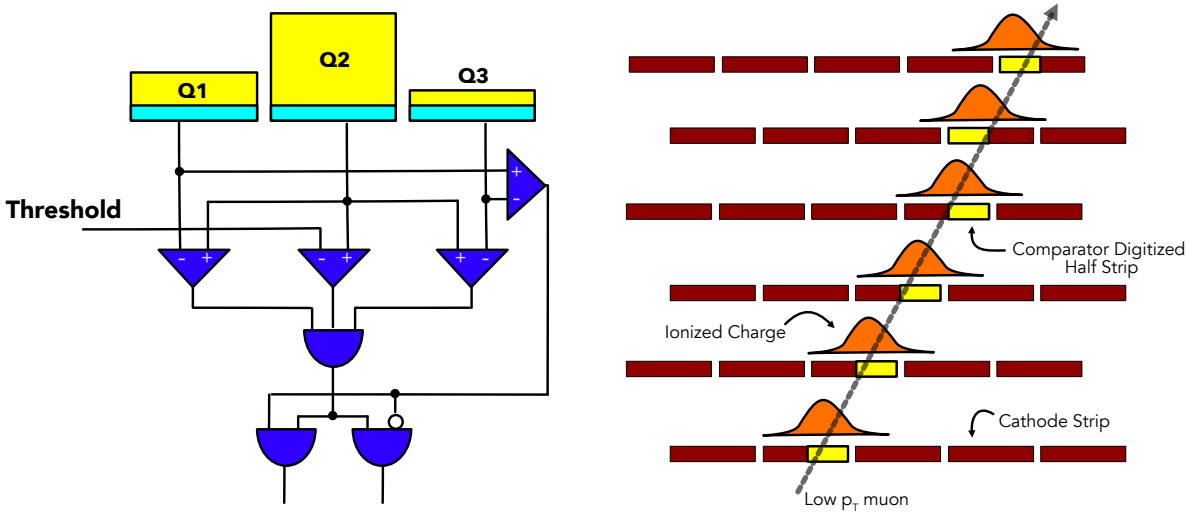


Figure 5.1: Simplified block diagram of the comparator circuit used to generate digital half-strip signals (left) together with a diagram showing the circuit in operation on all six layers when a low p_{T} muon crosses the chamber (right).

5.1.3 Cathode Local Charged Tracks

Muon p_{T} is extracted via the curvature of its trajectory in ϕ. Each CLCT measures a small snapshot of this curvature. The CLCTs themselves are composed of up to six layers of cathode hits which must fit inside a predefined pattern. They represent short, nearly straight segments of the muon trajectory.

The CLCT patterns used in Runs 1 and 2 are shown in Figure 5.2. For each event, the (O)TMB looks for the two most optimal CLCTs by sliding each pattern window along the half-strip dimension while looking for overlapping comparator firing. Comparator hits that are associated with the same muon are not always found within the same bunch crossing, therefore a 100 ns margin of error is given while running the pattern matching algorithm. Comparators are found in one of 16 readout time bins each corresponding to 25 ns. The distribution of comparator timings is centered at the middle bin and a CLCT is formed only when comparators fall within four consecutive bins.

Before the CLCT itself is formed, less stringent objects known as pre-CLCTs are created, which require only three layers and are used to identify which CFEB to read out in the event of an L1 Accept. If a pre-CLCT is found, a CLCT may then be formed if the number of matched layers exceeds an additional layer threshold [Hau04]. Nominally during Runs 1 and 2, CLCTs were formed only if at least four layers of comparator hits were found to reduce spurious background caused by detector noise, such as single layer neutron-induced hits.

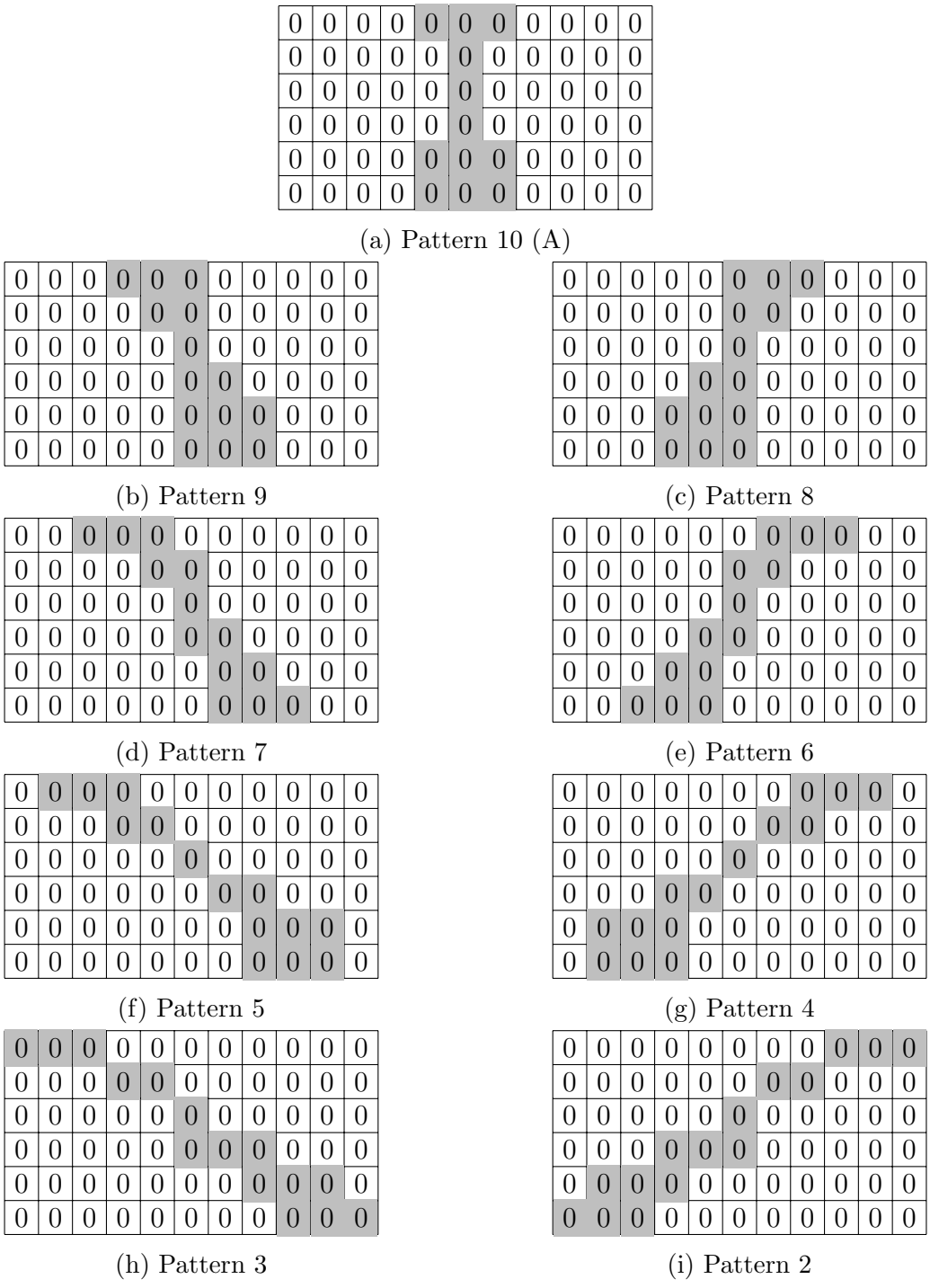
The optimal pattern is selected as the CLCT found with the largest amount of matched layers. Ties are broken by looking for the straightest pattern (i.e. most like Pattern 10), and then by the CLCT with the lowest half-strip position. Patterns used in Runs 1 and 2 contain a key half-strip (KHS) on their third layer used to identify the position of the CLCT. An identical algorithm is run for the second CLCT, although a configurable half-strip window is put in place around the first CLCT to avoid finding the same track twice.

Position resolution of the patterns used in Runs 1 and 2 are roughly a half-strip [Hau99] (~ 5 mm), but their acceptance width can lead to a range of muons that produce the same CLCT, thereby leading to a poorer p_{T} resolution. This is shown for two example tracks in Figure 5.3, where CLCTs relayed to EMTF lack available information on the muon track due to the acceptance width of the pattern.

5.1.4 Offline Reconstruction

The second data path within the CFEBs described in Section 5.1.1 is used as a buffer that is read from upon an L1 accept signal. In the buffer, cathode data is continuously overwritten to block RAM in 50 ns time bins [Cox16]. An offline analog to comparator hits, known as reconstructed hits (or `recHits`), are created by fitting to this data; the charge distribution along the cathode strips is fit to within the time bin with the largest pulse height. In contrast with the comparator hit, the `recHit` itself is a two-dimensional object, composed of both a precise cathode position derived from the procedure above and a precise anode position produced using anode data.

Offline algorithms then form muon "segments" by effectively fitting a line to the `recHits`



(a) Pattern 10 (A)

(b) Pattern 9

(c) Pattern 8

(d) Pattern 7

(e) Pattern 6

(f) Pattern 5

(g) Pattern 4

(h) Pattern 3

(i) Pattern 2

Figure 5.2: CLCT patterns used during LHC Run 1 and 2. Each horizontal digit represents a cathode half-strip, while the vertical axis represents the six layers of the CSC chamber. Grayed boxes denote locations accepted by the pattern when looking for a match. Layers are enumerated from one to six from top to bottom. Half-strips are ordered left to right from lowest to highest.

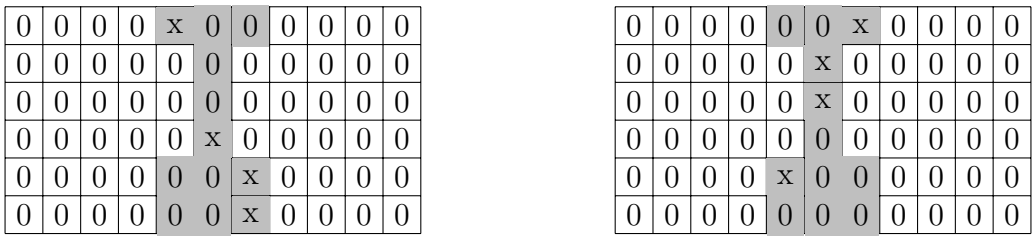


Figure 5.3: Example of two CLCTs matching Pattern 10 of the Run 1 and 2 pattern set which would be indistinguishable by EMTF. Only information pertaining to the amount of layers, pattern ID, and KHS position are encoded in CLCTs used in Run 1 and 2. Each horizontal digit represents a cathode half-strip, while the vertical axis represents the six layers of the CSC chamber. The gray outline represents the CLCT pattern, while x denotes the half-strip location of a comparator firing. Layers are enumerated from one to six from top to bottom. Half-strips are ordered left to right from lowest to highest.

in the chamber. Each segment is formed from a minimum of at least three layers of `recHit`s. Segments are then used in the track fitting procedure described in Section 3.2.7. The software used to produce both `recHit`s and segments is a standard package included in `CMSSW` [cms].

5.2 Updating the CLCT Algorithm

By using a lookup table (LUT) and an alternate scheme for encoding the CLCTs, one can break the CLCT degeneracy for underlying tracks with different signatures as seen in Figure 5.3.

We first note that the underlying comparator logic precludes no more than one comparator from firing in any consecutive window of three half-strips. This is because each comparator must have at least one full strip with a lesser charge between it and another potential firing as shown in Figure 5.1 (left). Therefore we may encode the location of a comparator firing within any three half-strips using just two bits to describe the four possible cases. The convention chosen for the encoding is shown in Table 5.1.

Hit Overlap	Binary Representation
000	00
X00	01
0X0	10
00X	11

Table 5.1: Binary representation of hit overlap on a given layer within a three half-strip window. Comparator hits are shown from lowest to highest half-strip number.

Layer Count	Number of Codes
0	1
1	18
2	135
3	540
4	1215
5	1458
6	729

Table 5.2: Comparator Codes by layer count for a given three half-strip wide pattern. Calculated using Equation 5.1.

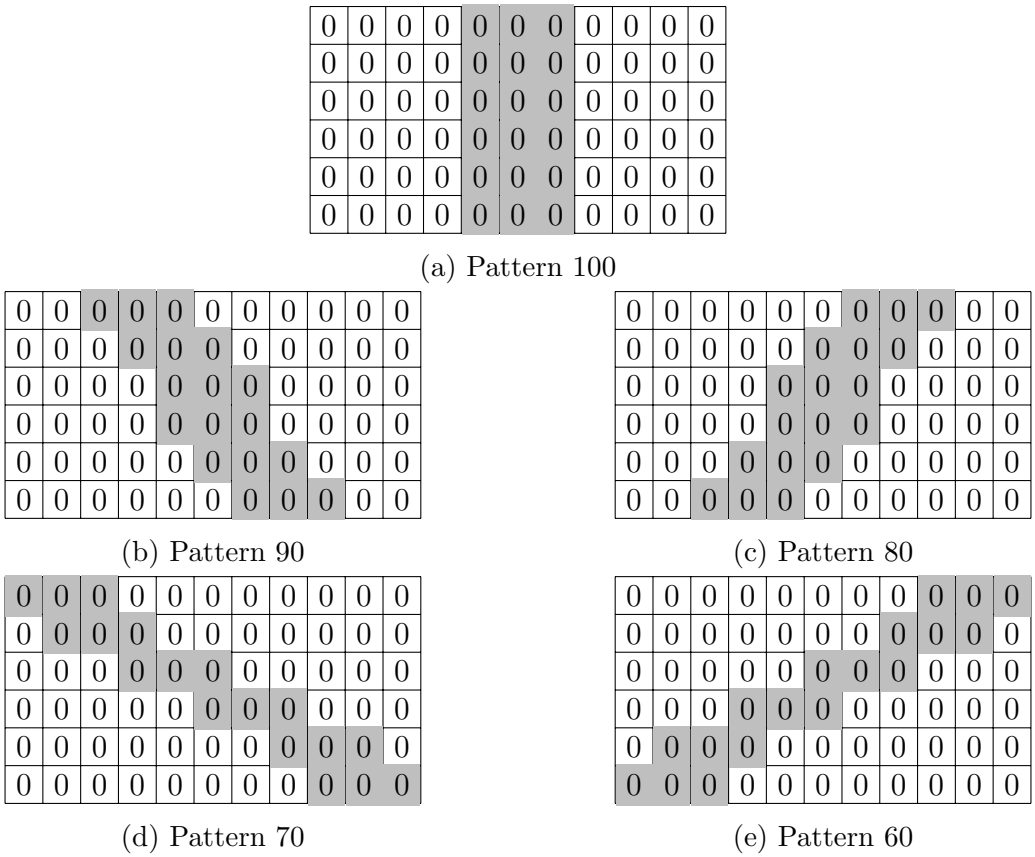
5.2.1 Comparator Codes

A set of patterns built out of three half-strip windows is shown in Figure 5.4. The exact pattern of comparator hits falling within these patterns is encoded via a concatenation of the bits as shown in Table 5.1. One defines the *comparator code* (CC) as this 12-bit concatenation; with the convention that the lowest layer represents the least significant bits. For each pattern, the 12-bit encoding allows for 4096 unique codes. An example calculation is shown in Figure 5.5. The breakdown of the number of codes c which contain hits in n layers is given by

$$c(n) = 3^n \frac{6!}{(6-n)!n!}, \tag{5.1}$$

and enumerated for each layer count in Table 5.2.

The OTMB provides enough memory to store a LUT with an entry for each comparator code. Moreover, the memory lookups can be added with no additional latency by inserting them where there are delays of one bunch crossing (25 ns) needed to maintain synchronization



(a) Pattern 100

(b) Pattern 90

(c) Pattern 80

(d) Pattern 70

(e) Pattern 60

Figure 5.4: CLCT patterns to be used in updated algorithm for Run 3. Each horizontal digit represents a cathode half-strip, while the vertical axis represents the six layers of the CSC chamber. Grayed boxes denote locations accepted by the pattern when looking for a match. Layers are enumerated from one to six from top to bottom. Half-strips are ordered left to right from lowest to highest.

0	0	0	0	x	0	0	0	0	0	0
0	0	0	0	0	x	0	0	0	0	0
0	0	0	0	0	x	0	0	0	0	0
0	0	0	0	0	0	x	0	0	0	0
0	0	0	0	0	0	x	0	0	0	0
0	0	0	0	0	x	0	0	0	0	0

(a) Pattern 100

Layer	6	5	4	3	2	1
Binary	10	11	11	10	10	01

(b) Half-strip to binary encoding

$\rightarrow$

Comparator Code
3049

(c) Decimal representation

Figure 5.5: An example encoding of a distribution of comparators found in a pattern specified in Section 5.2.1. In Figure 5.5a, each horizontal digit represents a cathode half-strip, while the vertical axis represents the six layers of the CSC chamber. The gray outline represents the CLCT pattern, while x denotes the half-strip location of a comparator firing. Layers are enumerated from one to six from top to bottom. Half-strips are ordered left to right from lowest to highest. These hits are then translated according to Table 5.1.

as shown in Figure 5.6. The figure shows the sorting sequence resulting in up to two CLCTs per bunch crossing; each CFEB is sorted to produce one or zero CLCT "candidates" (Best 1 of 32), then CLCT "candidates" are sorted to choose the best CLCT from the whole chamber (Best 1 of 7), shown for ME11 for which there are 7 CFEBs [Pec18]. The updated algorithm removes the delay flip-flop needed to stay synchronized with CMS and leverages this time to access the comparator code entry as described above. The LUT which uses comparator codes as an index is defined as the comparator code lookup table, or CCLUT.

5.3 Study Methodology

To quantify the gain in L1 Trigger resolution brought by the improved algorithm, studies of CLCT resolution were performed using a data sample of J/Ψ decays to dimuons. These low mass resonances provide muon tracks with a range of curvatures that span the space of possible CLCTs. Comparisons are made between both the CLCT patterns used in Run 1 and 2, and those described in Section 5.2.1.

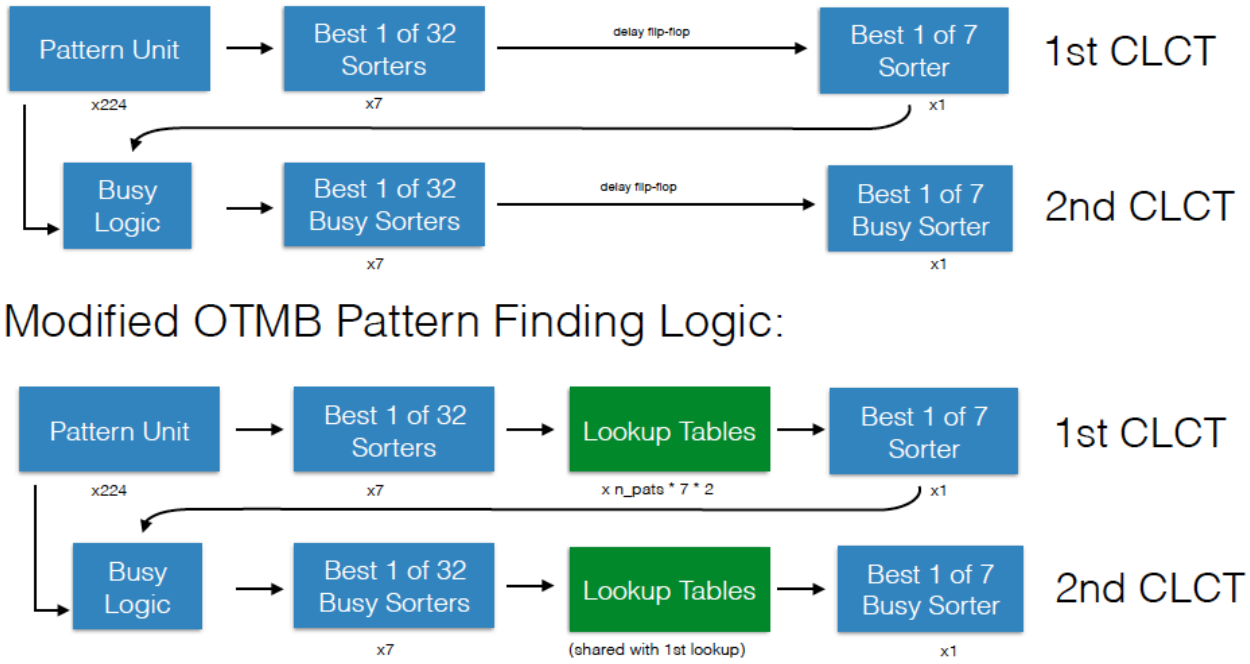


Figure 5.6: The electronic scheme for introducing look-up tables into the existing FPGAs of the CSC trigger motherboards without additional latency. The existing scheme for selecting up to two muon stubs per chamber is shown on top, while the new scheme shown on the bottom inserts look-up tables where currently there are delays of one bunch crossing (25 ns). The graphic is shown for ME11, for which there are $7 \times 32 = 224$ half-strips per layer. Reproduced from Reference [Pec18].

5.3.1 Data Samples

Datasets used in the study are listed in Table 5.3. As described in Section 4.3.2, these datasets are similarly chosen according to the recommendations of the PdmV group [Mey21]. Since the study requires only well-measured muons, the `MuonPhys` JSON files are used, which do not require proper working conditions in either the ECAL or HCAL and therefore increase the overall luminosity. The study primarily uses `Charmonium` datasets targeted at low-mass meson resonances with a charm quark such as the J/Ψ, and cross-checks its results using Z decays found in the `SingleMuon` dataset recorded in 2018. The `RAW` data format is used to access the low-level trigger primitives which are not saved in later stages of reconstruction (e.g. `AOD`). The `2017D Charmonium` dataset is shown as a representative sample in Figure 5.7.

Dataset	Run Range	$\mathcal{L}_{\text{int}}$ [fb^{-1}]
`/Charmonium/Run2017C-v1/RAW`	$299368 - 302029$	9.84
`/Charmonium/Run2017D-v1/RAW`	$302031 - 302663$	4.28
`/SingleMuon/Run2018D-ZMu-PromptReco-v2/RAW-RECO`	$320500 - 321012$	32.51
Total		46.63

Table 5.3: CMS data sets used in the analysis: data set names, run ranges certified for muon physics analysis by CMS, and corresponding integrated luminosities [Mey21]. Total non-overlapping integrated luminosities are given by 14.12 fb^{-1} and 32.51 fb^{-1} for 2017 and 2018 respectively.

5.3.2 Muon Selection

To study the performance of low-level algorithms on comparator data, we first make a selection of well-measured muons and associate them to the low-level information which resulted in their reconstruction.

In each event, muons are preselected by requiring they are reconstructed by both the standalone and global reconstruction algorithms as described in Section 3.2.5. Muons are then paired, and those pairs with opposite charges and an invariant mass within 0.1 GeV of the J/Ψ are selected. The subset of these muons with $p_{\text{T}} > 2$ GeV and segments

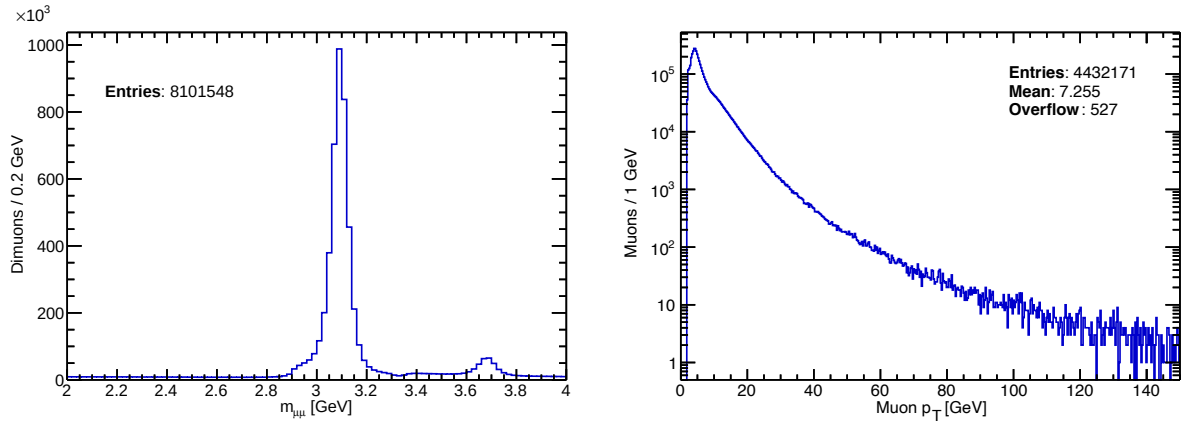


Figure 5.7: Distribution of the dimuon invariant mass (left) and muon p_{T} (right) for the `2017D Charmonium` dataset. The mass distribution is shown before the 0.1 GeV window is selected, and the p_{T} distribution shows only muons passing the full selection described in Section 5.3.2. A large resonance is seen in the mass distribution suggesting a high purity of J/Ψ events in the final selection.

measured in the endcap muon system are those used in this study, following the study in Reference [Liu16]. When the study is repeated selecting for Z decays, we instead require an invariant mass within 10 GeV of the Z and select muons with $p_{\mathrm{T}} > 10$ GeV .

5.3.3 CLCT Emulation

Recorded CLCTs found in the muon system are matched to segments belonging to selected muons to quantify CLCT performance as a function of various muon quantities. A match is defined as the closest CLCT to a segment, if any. Segments within a half-strip of the chamber edge are disregarded to prevent bias in subsequent efficiency measurements.

Additionally, an emulation of the (O)TMB was written to quantify the performance of CLCTs produced according to the updated CCLUT scheme. Emulated CLCTs resulting from this procedure are labeled as pseudo-CLCTs (pCLCTs) to distinguish them from those produced during online reconstruction. pCLCTs are created using either comparator data which is saved as a result of a pre-CLCT, or by using the cathode position of offline `recHits` rounded to the nearest half-strip. The emulation follows the same algorithm as described in Section 5.1.3.

Validation of the emulator is shown in Figure 5.8 which compares performance between the recorded CLCTs and the emulated pCLCTs. In this comparison, pCLCTs are created using rounded `recHit` positions. For each muon segment as defined in Section 5.3.2, one looks for a matching CLCT and pCLCT in the same chamber. The plot shows the relative frequency of the different pattern IDs given in Figure 5.2 and a strong level of agreement between recorded data and emulation. One may also note that when the muon p_{T} reaches 12 GeV, roughly 80% of all muons yield the same CLCT pattern, and at 6 GeV nearly 100% of muons produce only one of three CLCTs.

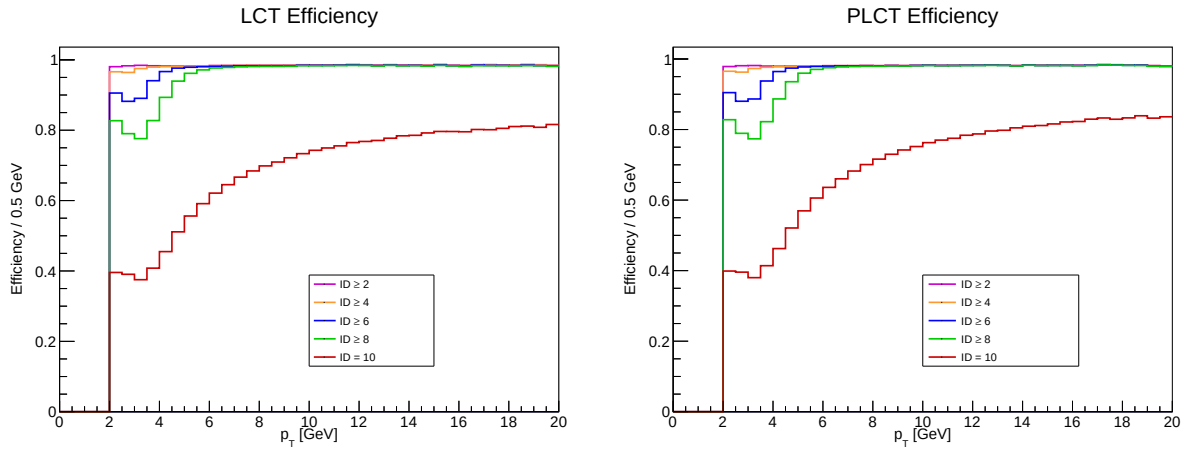


Figure 5.8: Comparison of CLCT (left) and emulated pCLCT (right) efficiency as a function of muon p_{T}. For each bin, the denominator is defined as the sum of valid segments belonging to muons described in Section 5.3.2 with a p_{T} within that bin. The numerator is the same sum including only segments for which there is a matched (p)CLCT of the specified pattern ID as shown in Figure 5.2. pCLCTs are produced using `recHit`s with a position rounded to the nearest half-strip to emulate comparators.

A similar validation of the emulator was then done using comparators recorded after a pre-CLCT. Agreement between CLCTs output by the TMB and those produced by the emulation is shown quantified in two ways: a "perfect match" between the CLCT and pCLCT if they have an identical KHS, layer count, and pattern ID, and a "match" which requires only that they are both found in the same chamber. The results are shown in Tables 5.4 and 5.5 which check the performance against the first CLCT and that inclusive of all CLCTs, respectively. The remaining discrepancies are small but not entirely understood. Likely related is the fact that for around 0.3% of segments, recorded comparators had more than

one hit within the three half-strip window, counter to the principle upon which the CCLUT scheme is based. These segments were ignored in both the numerator and denominator in the study and those that follow. Leaving these issues aside and having overall sufficient agreement, the remainder of the chapter uses emulated pCLCTs to be synonymous with real CLCTs.

	Passing / Failing	Percentage (%)
Matches	13236 / 13236	100.00
Perfect Matches	13221 / 13236	99.89

Table 5.4: Efficiency for emulation to produce only the first (of up to two) CLCTs output by the TMB. A perfect match is defined as identical KHS, layer count and pattern ID between the two. A match is a loosened criteria only requiring an unmatched emulated CLCT found in the same chamber.

	Passing / Failing	Percentage (%)
Matches	13495 / 13534	99.71
Perfect Matches	13479 / 13534	99.59

Table 5.5: Efficiency for emulation to produce all CLCTs output by the TMB. A perfect match is defined as identical KHS, layer count and pattern ID between the two. A match is a loosened criteria only requiring an unmatched emulated CLCT found in the same chamber.

5.4 Reducing the CLCT Layer Threshold

CLCTs nominally require at least four layers of comparators to be formed; however, the CCLUT scheme allows one to distinguish each of the 540 three-layer comparator codes present for each pattern (Table 5.2). Therefore, it is reasonable to quantify how much efficiency we could gain if we were to instead lower this layer threshold from four to three.

Segments are only produced using a minimum of three layers of `recHit`s, therefore we lower the layer requirement needed to form a CLCT to the same value. Comparators are not recorded unless a CLCT is formed with at least four layers, therefore we use `recHit`s in the CLCT emulation which are stored in the event of a pre-CLCT (Section 5.1.4). The expected effect of lowering the layer threshold is shown as a function of muon p_{T} in Figure 5.9 using

the same pattern set as were used in Runs 1 and 2. In total, reducing the minimum amount of layers increased segment efficiency by 1.6%.

These results were presented at the CSC Weekly Meeting on March 15, 2018 [Nas18c] and resulted in the change being implemented on the ME+1/1/11 OTMB firmware on August 23, 2018. The effect of the change was then summarized at the CSC Weekly Meeting on November 7, 2018 [Nas18a]. Using muons from Z decays as described in Section 5.3.2, lowering the layer threshold brought in between $0.4-2.7\%$ of additional CLCTs associated with Z muon segments, depending on the chamber location and firmware version. This increase is to be compared with roughly a 30% higher CLCT rate overall brought by background caused by the lowered CLCT requirements. Figure 5.10 shows the distribution of CLCTs that are output by the OTMB that are matched to Z muons (left) and all of those recorded (right). Additional three-layer CLCTs in the top figure can be interpreted as an increased efficiency for real muons brought the layer decrease of roughly 0.5%. The bottom figure can be understood as the effective rate increase of roughly 30%. Data was taken from the `SingleMuon` trigger because of the limited luminosity and relatively higher readout of muons from Z bosons.

Overall, the inclusion of CLCTs with a layer count of three was found to give non-negligible increases in efficiency and the increased rate does not meaningfully strain upstream reconstruction [Nas18a]. With the additional granularity of CLCTs afforded by the CCLUT scheme, this increased CLCT rate is expected to be further suppressed. The performance of the CCLUT scheme including CLCTs created with three layers is described in the following section.

5.5 CCLUT Performance

Positions and slopes of muon segments described in Section 5.1.4 are used as truth when measuring local CLCT resolution. We first note that each pattern has its own characteristic offset between the key half-strip used to define the CLCT position and positions of the muons

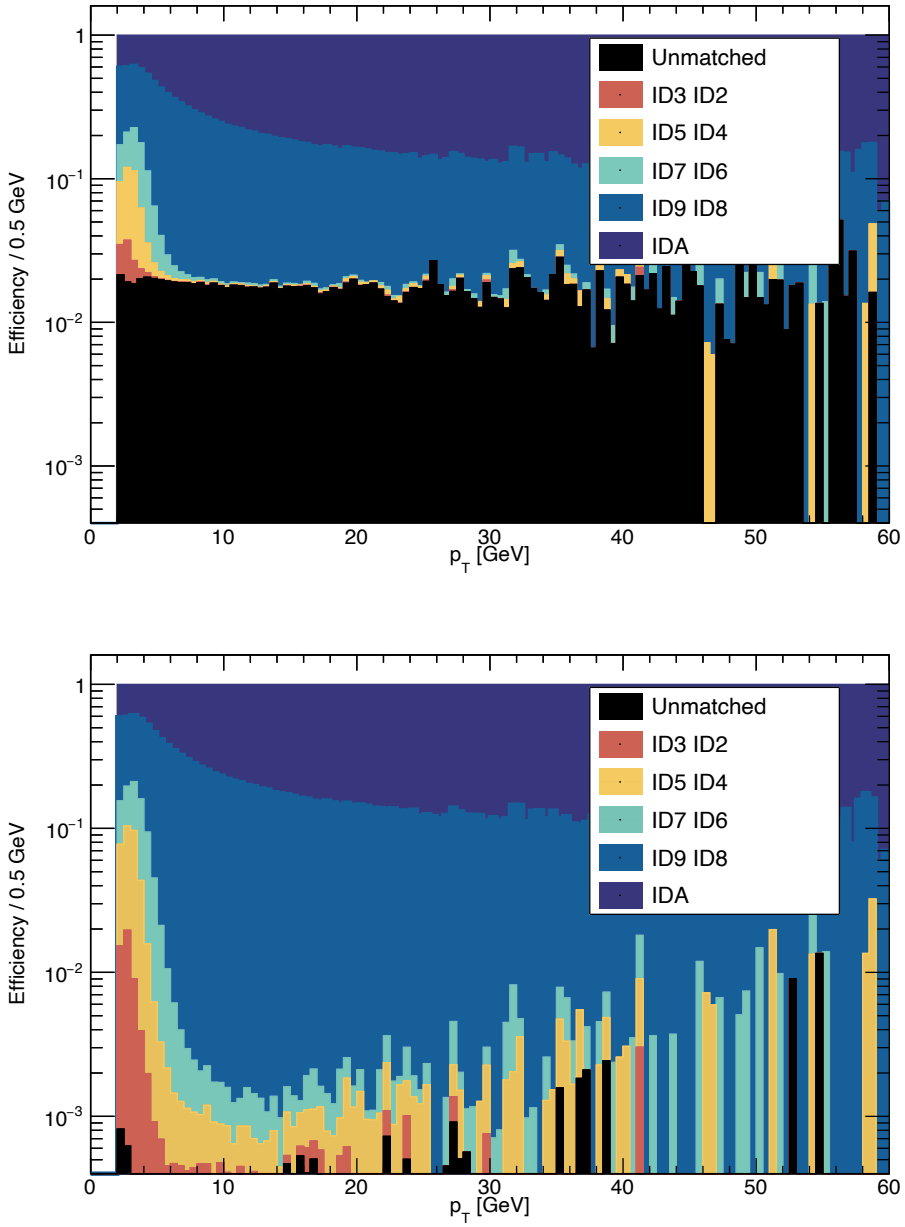


Figure 5.9: Comparison of CLCT-to-segment matching efficiencies, requiring that the CLCT have at least four layers (top) and three layers (bottom). The plot shows a stacked histogram such that all possible matching outcomes are accounted for. CLCTs are emulated in each endcap muon chamber using `recHits` information. Results suggest an efficiency gain of roughly 2% across the p_{T} spectrum if we include CLCTs with only three layers.

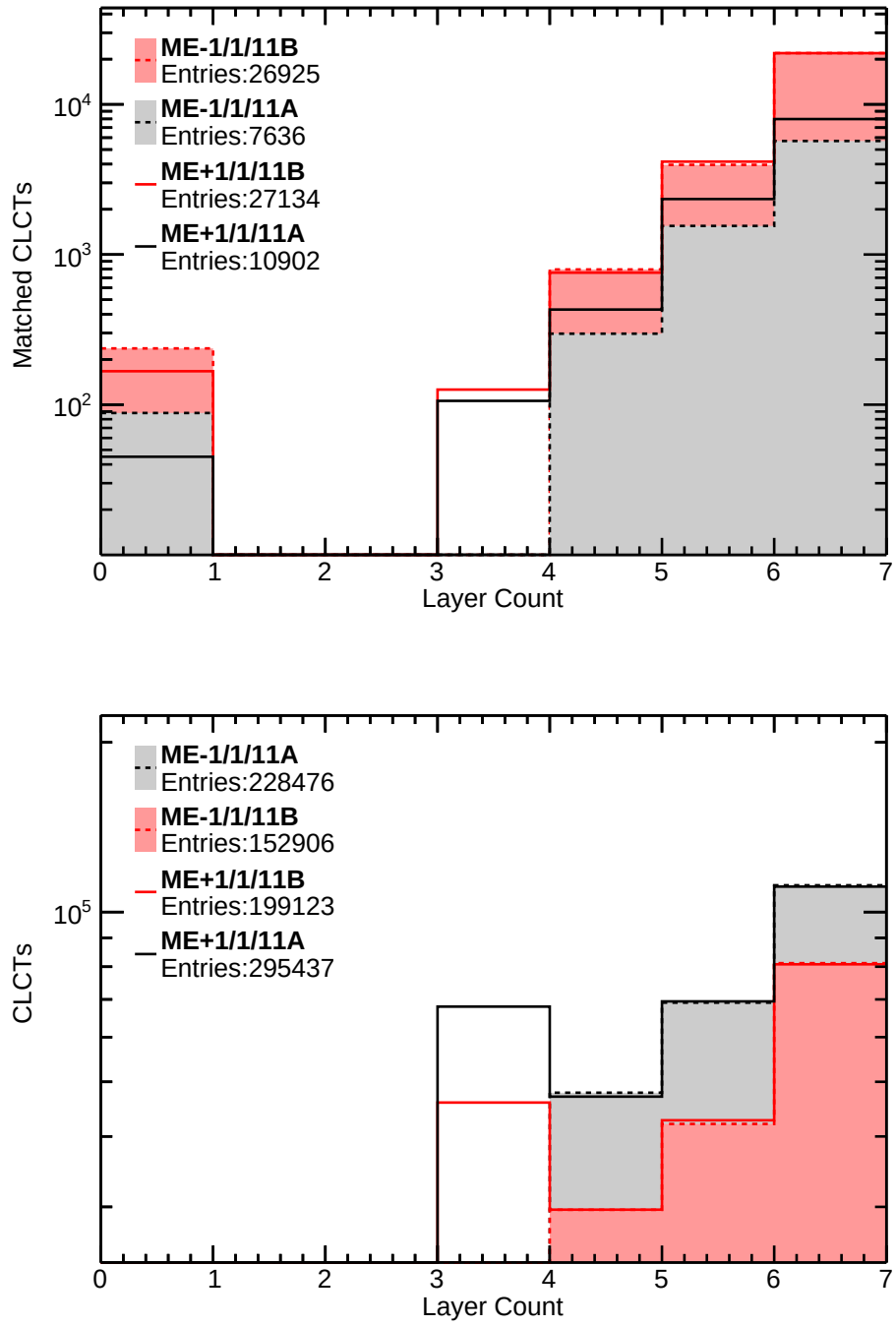


Figure 5.10: CLCT distribution following the loosening of the layer requirement from four to three in ME+1/1/11. The top plot shows the layer distribution of CLCTs which are matched to a muon originating from a Z decay (0 bin holds unmatched segments). The bottom plot shows the layer count of all CLCTs output by the OTMB. The curves are overlaid with the same results obtained from the chamber on the opposing endcap (ME-1/1/11) without the firmware change that acts as a nominal reference. Segments for which there is no match are placed in the left-most bin in the top plot.

which are reconstructed as a result of the CLCT. In the case of the CCLUT scheme, this offset can be further broken down by looking at the distribution of segment positions for each pattern-comparator-code (PCC) combination.

To measure the resolution of the patterns used in Runs 1 and 2, the distribution of segments around the KHS of each pattern has its mean centered and then each distribution is summed, yielding the effective resolution of the process as a whole. The same methodology is used for the CCLUT scheme, although distributions are centered with a PCC granularity. The distribution of segment position offsets is shown in Figure 5.11. An identical procedure is done to measure the corresponding slope resolution.

0	0	0	0	x	0	0	0	0	0	0
0	0	0	0	x	0	0	0	0	0	0
0	0	0	0	0	x	0	0	0	0	0
0	0	0	0	0	x	0	0	0	0	0
0	0	0	0	0	x	0	0	0	0	0
0	0	0	0	0	0	x	0	0	0	0

(a) Pattern 100, Comparator Code 3749

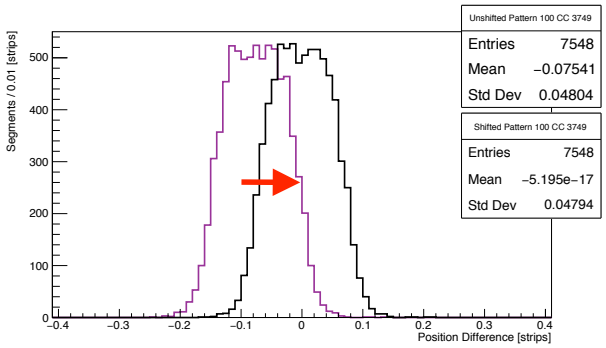


Figure 5.11: Pattern 100, comparator code (CC) 3749 (left) and the distribution of muon segment positions matched to it using emulated CLCTs defined under the CCLUT scheme (right). Position difference is defined as the muon segment position subtracted from the key half-strip of the CLCT, measured in cathode strips. This shifting is done individually for each PCC combination, and the full resolution of method is derived by its sum. The gray outline represents the CLCT pattern, while x denotes the half-strip location of a comparator firing. Layers are enumerated from one to six from top to bottom. Half-strips are ordered left to right from lowest to highest.

5.5.1 Accounting for Missing CCLUT Entries

Not all PCC combinations are found to have matching segments. Combinations without matching segments can be attributed to several factors: having a layer count less than the minimum, being degenerate with another combination which is found by the algorithm

earlier, limited statistics of the study, or if the combination is low quality such that it is never associated with a real muon.

Using five patterns yields a total of $5 \times 4,096 = 20,480$ possible PCCs, of which only a limited number are used. The total number of comparator codes with a layer count below the minimum layer count is given by Table 5.2 as 154, which reduces the total amount of PCCs to $19,710$.

Within one pattern, the same distribution of comparator hits can in some cases be described by multiple sets of comparator codes. This can be seen in Figure 5.12, where the same set of comparators is covered by two distinct comparator codes.

0	0	0	0	0	0	x	0	0	0	0
0	0	0	0	0	0	x	0	0	0	0
0	0	0	0	0	x	0	0	0	0	0
0	0	0	0	0	x	0	0	0	0	0
0	0	0	0	0	x	0	0	0	0	0
0	0	0	0	0	x	0	0	0	0	0

(a) Pattern 100, Comparator Code 2735

0	0	0	0	0	x	0	0	0	0	0
0	0	0	0	0	x	0	0	0	0	0
0	0	0	0	x	0	0	0	0	0	0
0	0	0	0	x	0	0	0	0	0	0
0	0	0	0	x	0	0	0	0	0	0
0	0	0	0	x	0	0	0	0	0	0

(b) Pattern 100, Comparator Code 1370

Figure 5.12: Example of two sets of comparator codes found for pattern 100 matching the same distribution of comparator hits. Comparator code 2735 (left) is equivalent to 1370 (right) when the pattern has its KHS shifted by one. The first found by the sorting algorithm is the one with the lowest KHS (comparator code 2735). The gray outline represents the CLCT pattern, while x denotes the half-strip location of a comparator firing. Layers are enumerated from one to six from top to bottom. Half-strips are ordered left to right from lowest to highest.

One can enumerate the number of degenerate comparator codes within each pattern by first noting that they must fit within a half-strip window smaller than that which defines the pattern. The amount of comparator codes that can fit within a width of two half-strips and one half-strip are given in Figures 5.13 and 5.14 respectively. We then note that among all $4,096$ possible comparator codes, we will see *two* copies of those that fit within two half strips, and *three* copies of those that fit within just one column of half-strips. Therefore we can enumerate the number of non-degenerate comparator codes within a pattern as the number of codes passing our layer requirement $(3,942)$ minus the number of codes that fit within two half-strips (656), minus the number that fit within one-half strip (42). The

$$c_2(n) = 2^n \frac{6!}{(6-n)!n!} \qquad (5.2)$$

Layer Count	Number of Codes
0	1
1	12
2	60
3	160
4	240
5	192
6	64

Figure 5.13: Amount of comparator codes for a given layer count that fit within a window of only two half-strips. Within the standard pattern window of three half-strips, equivalent codes matching this requirement show up twice, as can be inferred from Figure 5.12. One can arrive at the amount of non-degenerate codes that fit within three layers by subtracting those enumerated in the table (right) and those in Figure 5.14.

$$c_1(n) = \frac{6!}{(6-n)!n!} \qquad (5.3)$$

Layer Count	Number of Codes
0	1
1	6
2	15
3	20
4	15
5	6
6	1

Figure 5.14: Amount of comparator codes for a given layer count that fit within a window of only one half-strip. Within the standard pattern window of three half-strips, equivalent codes matching this requirement show up thrice. One can arrive at the amount of non-degenerate codes that fit within three layers by subtracting those enumerated in the table (right) and those in Figure 5.13.

additional set of codes fitting within one half-strip is included in those that fit within two half-strips. This leaves us with $3,244$ non-degenerate comparator codes passing our layer requirement within each pattern. Without considering degeneracy between patterns, this leaves us with only $16,220$ of the initial $20,480$ PCCs we had started with.

A smaller subset still is needed when matching to real muon segments in actual data. Figure 5.15 (left) shows that roughly only $10,000$ unique PCCs are needed to describe comparators associated with two million muon segments found in the `2017C` dataset. When creating our lookup tables, this then leaves us with many entries that would not be filled. Furthermore, there are many cases where only a limited number of segments are matched

to a specific PCC, leading to an offset calculated using severely limited statistics, as seen in Figure 5.15 (right).

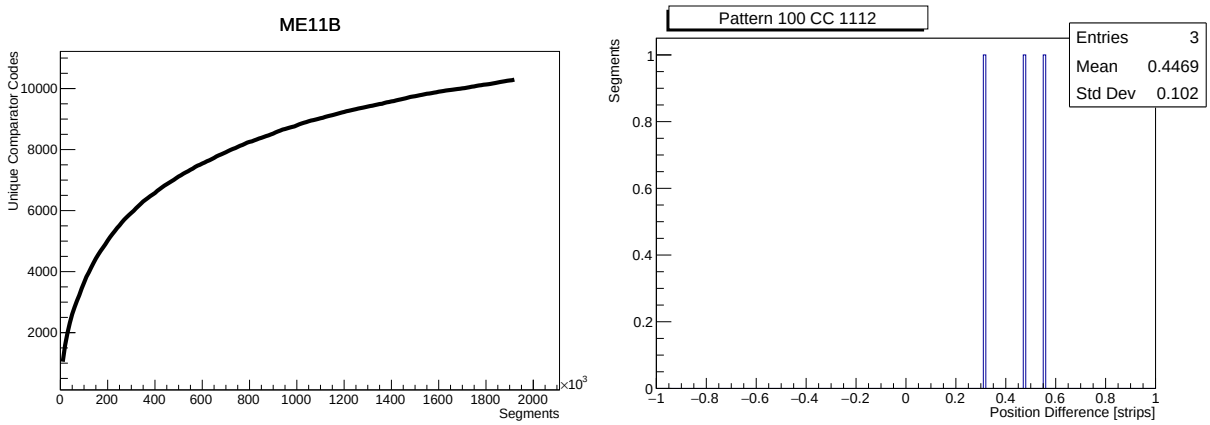


Figure 5.15: Unique PCCs needed to cover muon segments in the ME11B chamber (left). Muon segments are taken from J/Ψ decays in the `2017C` dataset. Due to limited statistics, some PCCs which are found to be associated with real muons less often have very few segments associated to them that can be used to calculate a corresponding offset with (right).

In cases where there are either no corresponding segments matched to a PCC, or where the statistics are limited, line fits are used as a proxy. Specifically, a straight line is fit to the comparator arrangement, giving us both a position and slope measurement which can be compared to the segment. Segments themselves are not simply line fits to the `recHits` [Cox16], which compels us to use line fits only in cases where they are our best choice. The minimum amount of matched segments needed to use the offsets from the segment distribution over the line fits is chosen to maximize the overall resolution. Figure 5.16 shows the position resolution as a function of N_t, the minimum number of segments needed to be matched to a PCC for the segment offset to be used over the line fit. A minimum number of 10 segments is chosen as a result of the study.

5.5.2 CCLUT Resolution

The local position and slope resolution obtained by the method is shown in Figure 5.17. Offsets are calculated using muon segments from the `2017C` dataset and resolution is measured with respect to muons segments found in the `2017D` dataset. The position resolution

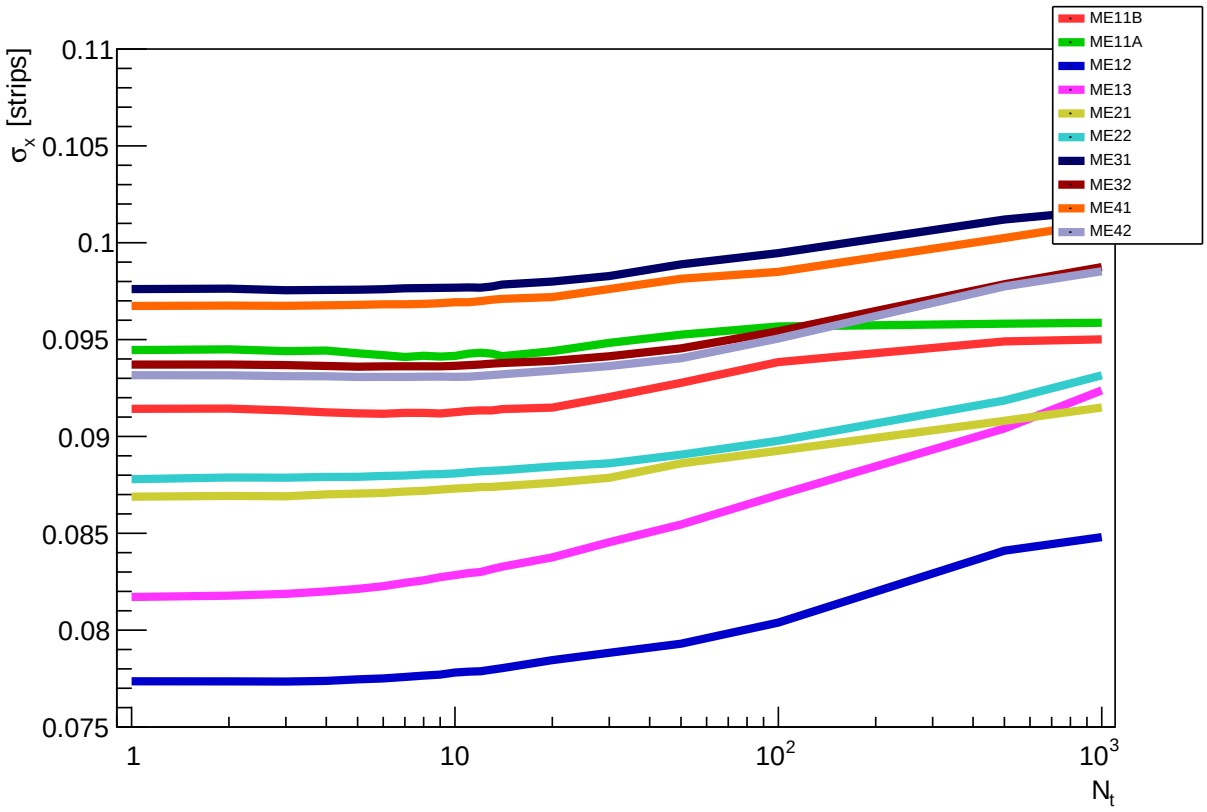


Figure 5.16: Position resolution as a function of N_t, the minimum number of segments needed to be matched to a PCC before the segment offset is used in place of the line fit. Segment offsets are found in the `2017C` dataset and the resolution shown is calculated using a complementary set of muons taken from the `2017D` dataset. A threshold of 10 segment matches is chosen before which line fits are used.

improves by nearly a factor of two with respect to those used in Run 1 and 2, from 0.17 to 0.09 strips. The improvement in the slope resolution is also substantial, going from 0.11 to 0.08 strips per layer.

5.6 Implementation and Outlook

The CCLUT scheme should substantially improve both local position and slope resolution and is actively being integrated into the L1 Trigger path, targeting completion for the start of Run 3 [Dil21]. Modifications to the LCT data format necessary to accommodate this update have been proposed to the Endcap Muon Track Finder (EMTF) group [Nas18b] and are being developed in both software and firmware.

Quantification of overall L1 muon resolution gain remains a work in progress as gain from the improved trigger primitives are propagated through trigger algorithms in track finders upstream. Resolution is expected to improve not only for prompt muons but also for those that are produced away from the primary vertex, where a beam-spot constraint cannot be used to aid the fit. As described in Chapter 4, the largest source of signal efficiency loss for the displaced dimuons search was the trigger. Therefore, long-lived analyses stand to strongly benefit from a L1 Trigger which is more capable of resolving displaced muon p_{T}.

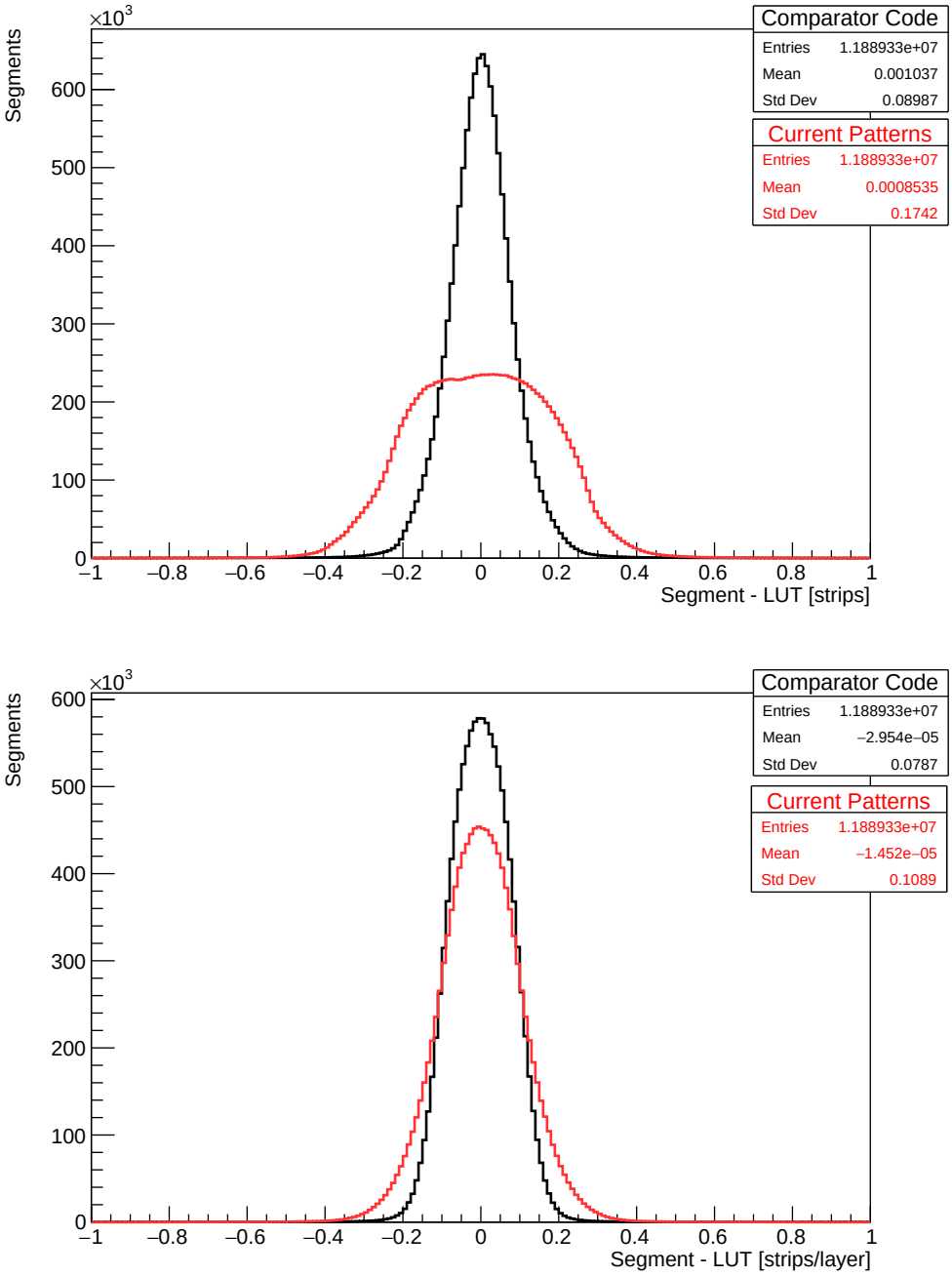


Figure 5.17: Position (top) and slope (bottom) resolution of CLCTs produced using the comparator code scheme (black) and the patterns from Runs 1 and 2 (red). The resolution is defined as the difference with respect to muon segment position described in Section 5.3.2. Offsets were calculated using the `2017C` dataset and tested against the `2017D` dataset.

APPENDIX A

Measuring Muon Momentum via Energy Deposition

There is much pleasure to be gained from useless knowledge.

— Bertrand Russell (1872 - 1970)

Within CMS, muons are typically thought of as Minimum Ionizing Particles (MIPs) and have their momentum measured according to their track curvature as explained in Section 3.2.5. However, as muon energy increases, so too does the average amount of energy they lose as they travel. By correlating the energy lost by the muon as measured in both the calorimeters and endcap muon systems, one can derive an estimate of the muon momentum without any knowledge of track fitting or the magnetic field.

A.1 Muon Energy Loss

Stopping power S is a measure of the retarding force on a particle due to interactions with matter and is equal the particle's energy loss E per unit distance x,

$$S(E) = -dE/dx. \tag{A.1}$$

A charged particle can lose energy for a variety of reasons that strongly vary with the scale of its energy. These include: ionization, e^+e^- pair production, Bremsstrahlung, and photonuclear interactions [Gro19]. All sources of energy loss excluding ionization are classified together as radiative losses and increase with the energy of the particle. Energy loss due to ionization however remains relatively flat as a function of muon energy. One may therefore define a *critical energy* E_c at which the energy loss due to radiative effects is equal to that

associated with ionization. For muons in iron, this energy is $E_{\mu c} = 347$ GeV [Gro20]. For muons approximately above this threshold, it is sufficient to parametrize their mean energy loss as [BBC52]

$$-\langle dE/dx \rangle = a(E) + b(E) \cdot E, \tag{A.2}$$

where $a(E)$ describes the ionization energy loss, and $b(E)$ includes the sum of radiative terms (e^+e^- pair production, Bremsstrahlung, and photonuclear interactions). Relative contributions of individual radiative losses are shown for muons in iron in Figure A.1.

Although the mean energy loss is parametrized by Equation A.2, this value is typically considered inappropriate to parametrize particle energy loss at high energies [Bic06, Zyl20]. This is because of non-zero probabilities for the particles to lose catastrophic amounts of energy, resulting in a significant shift of the mean away from the median value. Landau instead showed that the most probable energy loss is the appropriate parameter to characterize this energy loss [Lan44]. Figure A.2 shows the distribution of the momenta of 1 TeV muons after they pass through 3 m of iron in Monte Carlo simulation. The figure shows that the mean of the distribution is substantially different from that of either the median or the most probable value. The related Landau distribution $f(\Delta E; \mu, c)$ adequately describes this energy loss ΔE and is parametrized by both the most probable value of the energy loss μ, and a scaling parameter c.

A.2 Predicting Muon Momentum via Calorimetry

Within the LHC, only muons of more than a few GeV of momentum can escape the magnetic field and traverse the full detector. This translates to each muon measured in the muon system of CMS having energy approximately equal to that of its momentum. Furthermore, momentum loss may then be approximated as the energy loss $\Delta p \sim \Delta E$. The tracker, ECAL, HCAL, and muon system consequently sample the muon energy loss as it traverses each one. By fitting the energy loss in each one to find the total energy loss, one can therefore

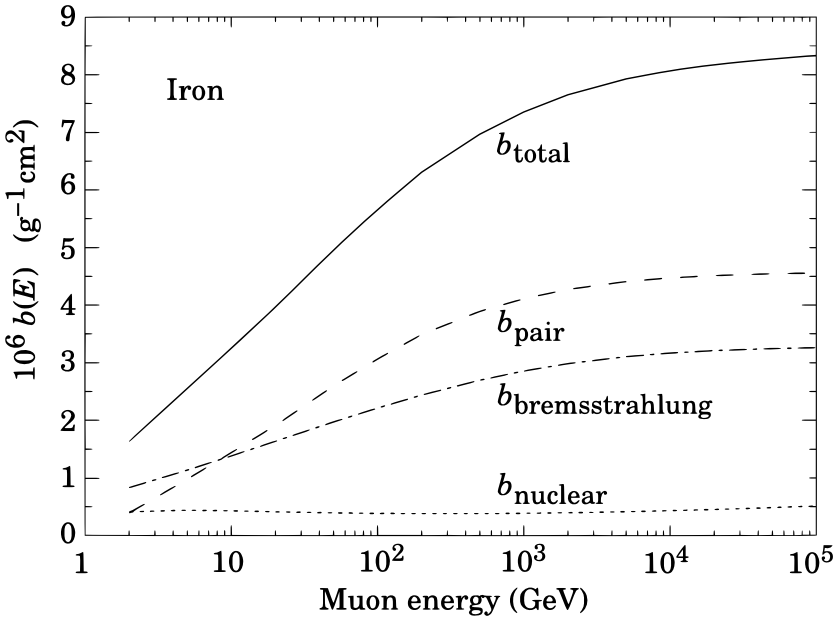


Figure A.1: Individual contributions to radiative muon energy loss and their total in iron [Gro19]. Distributions exclude post-Born corrections to the direct pair production cross section, which change the result by less than 1%.

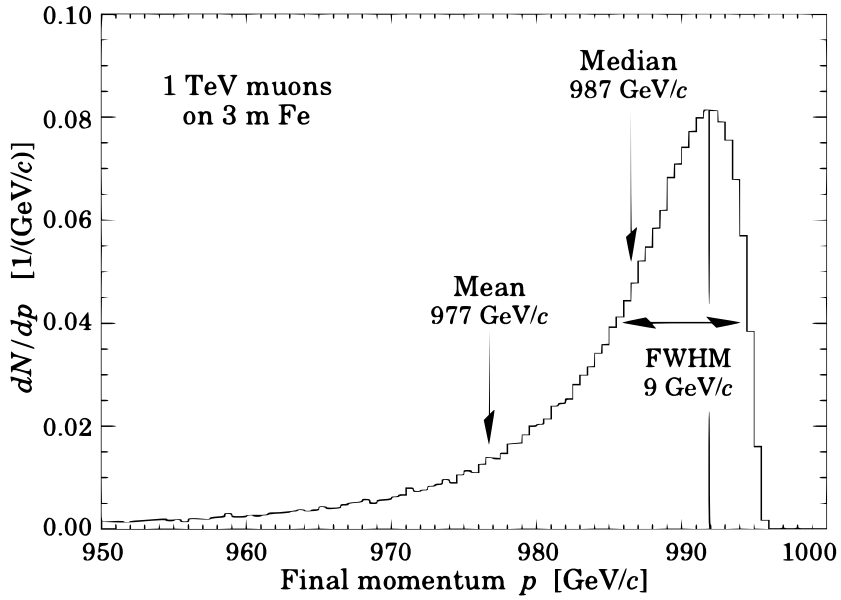


Figure A.2: Final momentum distribution after 1 TeV muons pass through 3 m of iron generated using Monte Carlo [Gro19]. The mean, median, and full-width at half maximum (FWHM) are shown together.

infer the energy of the initial particle. This section details the methodology used and results using Monte Carlo simulation.

A.2.1 Study Methodology

The algorithm was developed using Monte Carlo simulation to have truth-level knowledge of the energy of the muon in addition to the history of its energy loss as it traversed the detector. Specifically, a `MuonGun` private dataset production was made for `GEN-SIM-RAW-RECO` data using a modified `FlatRandomEGunProducer` configuration file [Eul13]. The parameters used to construct the dataset are listed in Table A.1. The range in pseudorapidity was chosen such that muons would pass through the endcap muon system, used as an element in their energy reconstruction.

Variable	Range
E_μ	$[0, 4000]$ GeV
η	$[0.9, 2.5]$
ϕ	$[-\pi, \pi]$

Table A.1: Parameter set for MuonGun Monte Carlo simulation file. Each muon is pulled from a uniform distribution among the variables listed.

A sample of roughly 1.6 million muons was produced, with their kinematic distributions shown in Figure A.3. Within each event, simulated energy loss is predicted using GEANT4[Ago03]. Notably, simulated events do *not* contain additional inelastic pp interactions (pileup), that would be found within the detector during typical working conditions. This was done to study the energy loss in the cleanest possible way while the algorithm was developed. The code developed for this study is provided in Reference [Nas22].

A.2.2 Distinguishing Muon Momenta

As the muon traverses the CMS detector, it loses its energy according to the methods described in Section A.1. Figure A.4 shows the distribution of momentum loss as a function of the generated momentum. To proceed, the distribution of ΔE for each slice of E is taken to

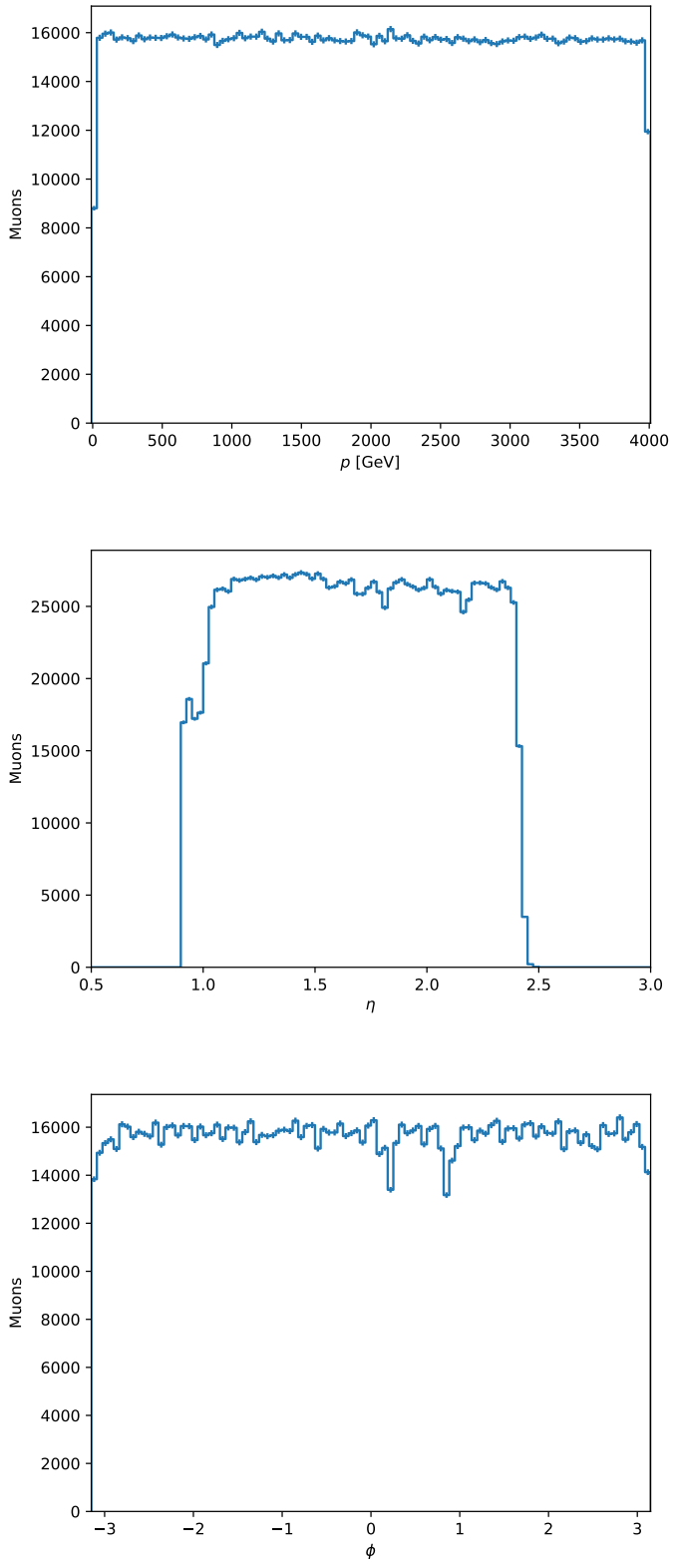


Figure A.3: Distribution of Monte Carlo generated muons in momentum (top), pseudorapidity (middle) and azimuthal angle (bottom). Each distribution is filled with 1,564,946 generated muons.

be described by the Moyal distribution, known to be a good approximation to the Landau distribution [Wal07] taking the form

$$f(x) = \exp(-(x + \exp(-x))/2)/\sqrt{2\pi} \tag{A.3}$$

with

$$x(\Delta E, \mu, c) = (\Delta E - \mu)/c. \tag{A.4}$$

Both the parameter describing the most probable value μ as well as that which describes the scale c are taken to be linearly dependent on the momentum:

$$\mu(E) = \alpha_1 + \alpha_2 \cdot E, \tag{A.5}$$

$$c(E) = \alpha_3 \cdot E. \tag{A.6}$$

A two-dimensional least-squares fit is then performed to the left distribution of Figure A.4 after it has been properly normalized. The fit achieves a $\chi^2/\text{ndf} = 2.26$, with parameters listed in Table A.2. The corresponding fitted energy loss distributions projected along both axes are shown in Figure A.5. This distribution, therefore, allows us to predict the underlying muon momentum p given a measured value of energy loss Δp.

Parameter	Value
α_1	1.067e+00 GeV
α_2	1.483e−02
α_3	1.352e+03 GeV

Table A.2: Least-squares optimized fit parameters to Equation A.3 to simulated muon deposits.

The full energy loss for each muon Δp is predicted using a neural network that treats the energy loss measured in the chambers listed in Table A.3 as learnable features. The momentum p itself is *not* trained for, given the stochastic relationship between muon energy and the amount of energy that it loses.

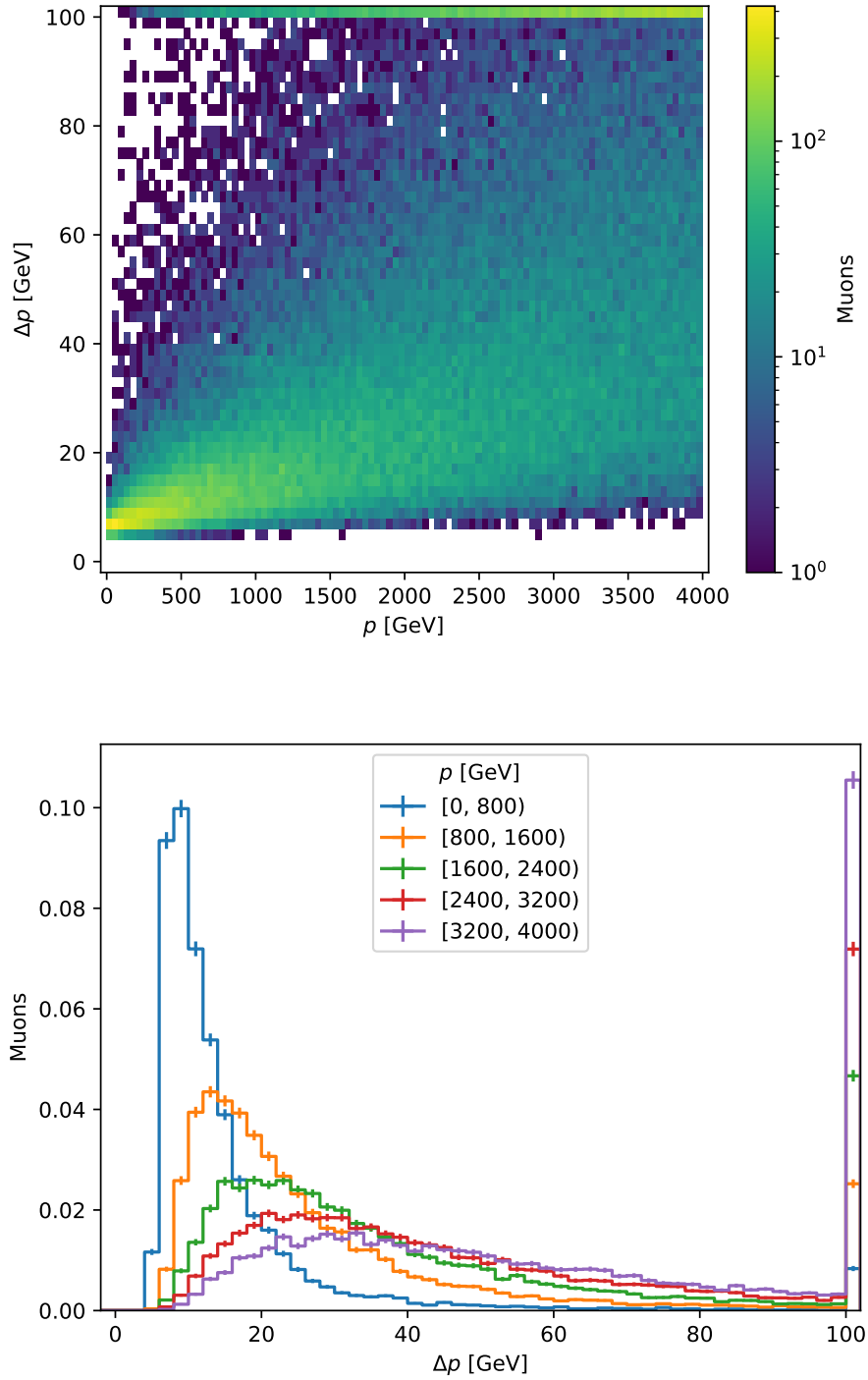
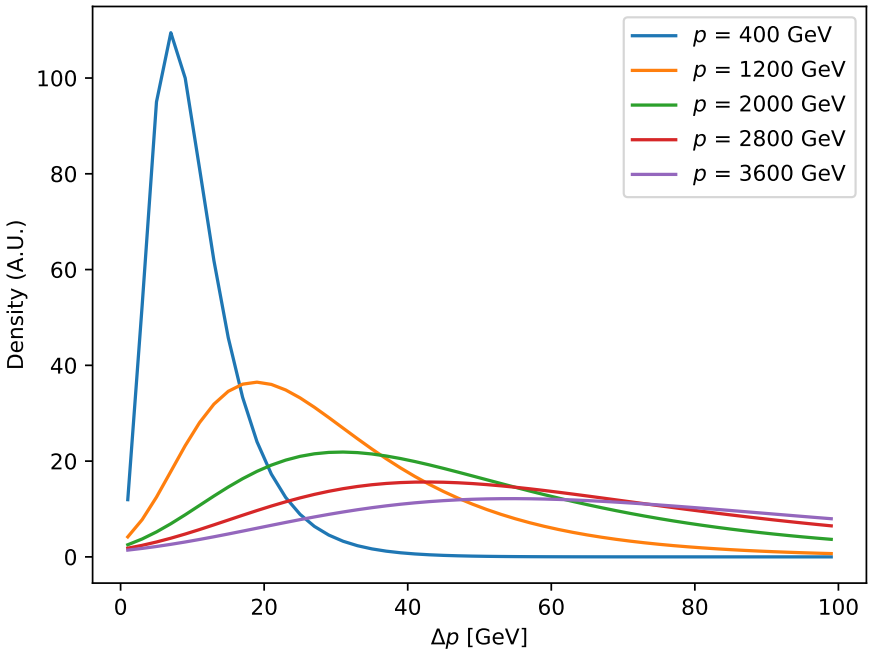


Figure A.4: Muon momentum (energy) loss Δp as a function of momentum p at the time of production. Distributions are shown as a scatter plot (top) and in normalized slices of momentum (bottom). Overflows are shown in the last bins in each plot.



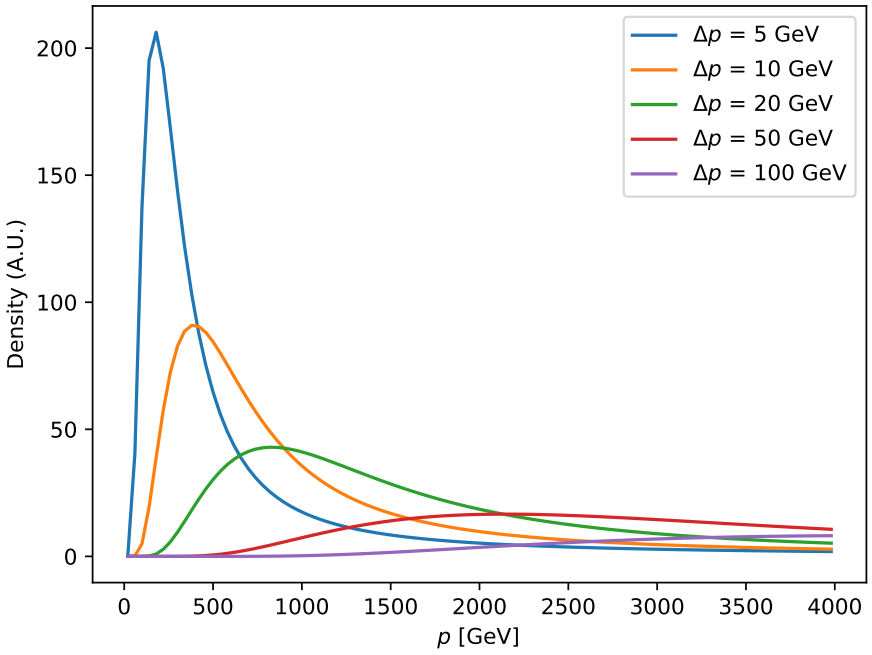


Figure A.5: Projections of the two-dimensional $\Delta p - p$ Landau fits described in Section A.2.2 along Δp (top) and p (bottom).

A.2.2.1 Model Architecture

To begin, the muon dataset is split into three components: a training set (60%) used to derive the parameters of the neural network; a validation set (20%) used to benchmark performance during training, but not explicitly used in setting parameters; and a test set (20%) used to evaluate the performance of the model after all parameters have been finalized.

Specifically, a sequential neural network is used provided by `TensorFlow` [Aba15]. The first layer is used to scale the features to have a mean of one and a standard deviation of zero. The last layer outputs the predicted momentum loss Δp. The mean absolute percentage error is used as the loss function to optimize the value typically quoted as the momentum resolution. Other relevant hyperparameters of the model, such as the number of layers, number of units per layer, the activation function, and the learning rate, are derived using the `KerasTuner` [OM19], which scans over various configurations to select those best aligned to the problem at hand. The optimal hyperparameters, as determined by the smallest validation loss (29.34%), select for four additional dense layers between those previously described, each with 32 rectified linear units [Aga18]. The optimal learning rate was selected as 0.001.

Feature
Muon η
Muon ϕ
HCAL energy deposits
ECAL energy deposits
CSC energy deposits

Table A.3: List of features used in training for the muon energy loss Δp.

Estimates on the muon energy loss Δp are then used in evaluating a confidence interval for the initial muon momentum p. The two-dimensional Landau fit is used to this end – selecting a one-dimensional slice along p corresponding to the derived Δp estimate (i.e. the right distribution in Figure A.5). This slice is treated as a likelihood $\mathcal{L}$, and both the best estimate and confidence interval are found using $-2 \ln \mathcal{L}$.

A.2.2.2 Model Performance

A comparison between the predicted and true energy loss is shown in Figure A.6 for the test data set. Modest agreement is observed with momentum loss often underestimated. Improved performance is left for future work, such as inclusion of the shape of electromagnetic shower.

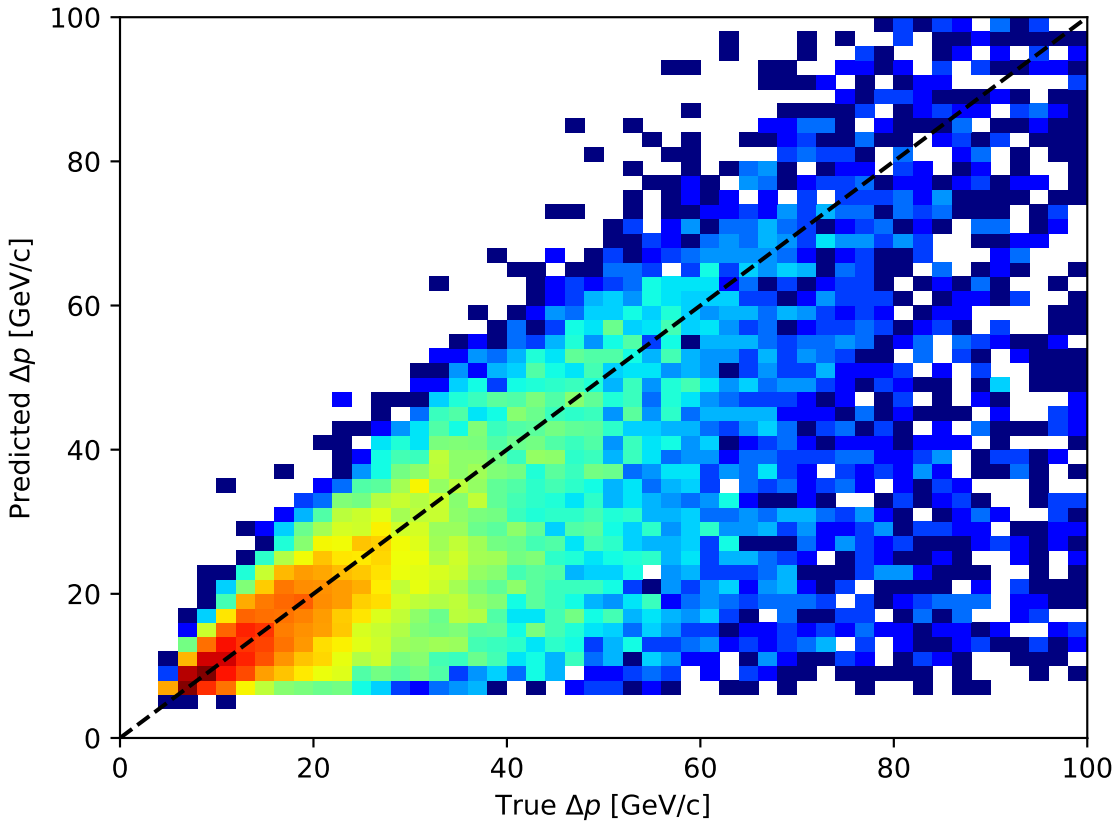


Figure A.6: Comparison between the predicted and true energy loss Δp on the test set after the model is fully trained. Higher density regions are shown in red, while less dense regions are shown in blue. The dotted line represents perfect agreement between both quantities.

The Δp output of the neural network given the energy loss in the detectors is then used to estimate the initial momentum of the muon using the two-dimensional Landau fit. A one-dimensional probability distribution is defined by the corresponding slice, and the momentum is estimated by the local minimum of the $-2 \ln \mathcal{L}$ well. The 68% confidence interval is then

calculated in the standard way [Zyl20], taking as the points where $-2\Delta \ln \mathcal{L} = 1$. Notably, there are no corrections applied to account for the Δp resolution. A sample of predictions made from this procedure is shown in Figure A.7.

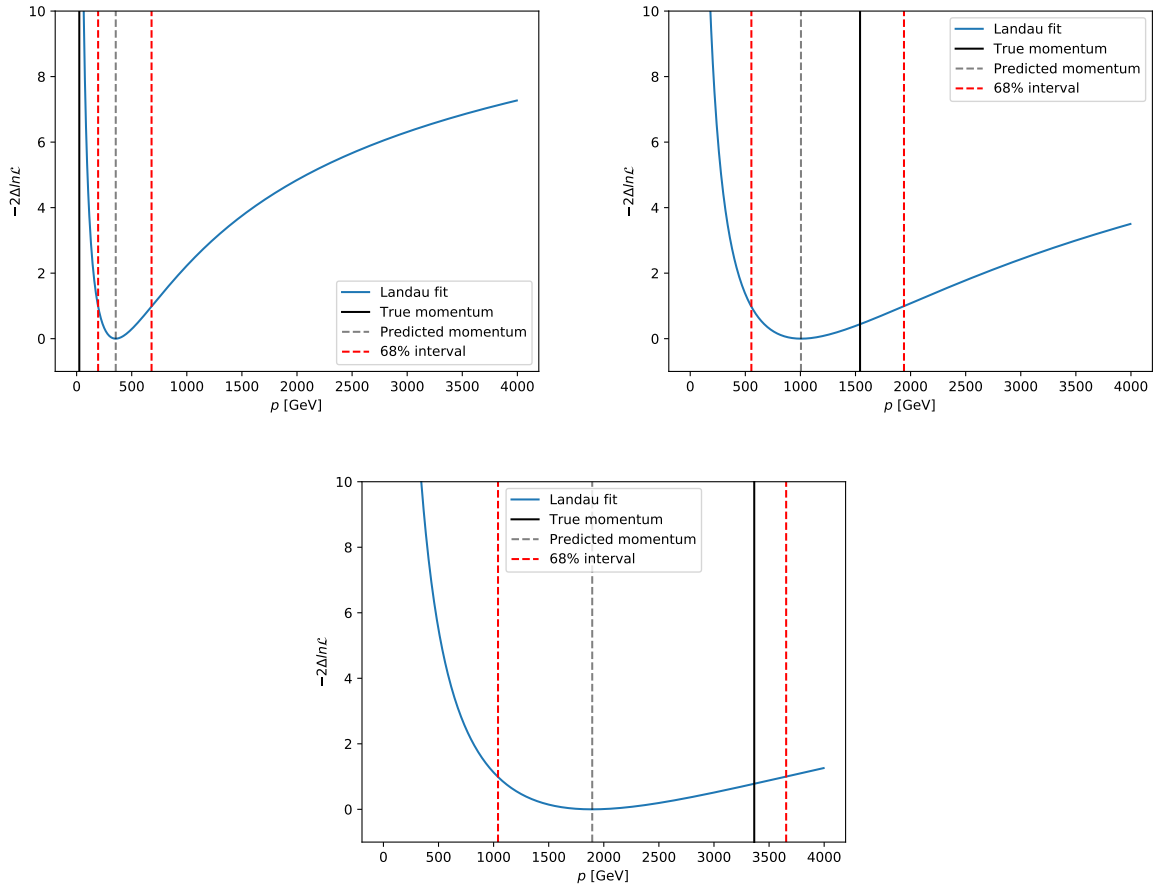


Figure A.7: Predicted momenta obtained using a log-likelihood based approach overlaid with the true generated values. Momentum loss Δp is first predicted using a neural network which treats the energy deposits as features. The output is used to slice into a two-dimensional Landau fit, which describes how a muon typically deposits its energy. This slice is treated as a likelihood function $\mathcal{L}$, and is used to estimate the underlying muon momentum.

The overall resolution obtained by the algorithm is shown in Figure A.8. A significant bias is observed. The long tail in the momentum likelihood function prevents an accurate estimate of the underlying muon momentum when selecting for the most probable value alone, despite a roughly 30% resolution on the energy loss of the muon. The use of a confidence interval is intended to account for this feature, however, the resulting resolution

is still highly asymmetric. The right sub-figure of Figure A.8, which normalizes the prediction by its confidence interval, is treated as a pull distribution and fit to a Gaussian, yielding a mean of -0.44 and width of 0.75. Marginal performance is understood as a consequence of the highly skewed relationship between the momentum and momentum loss, but additional supporting studies have not been performed.

A.2.3 Summary

This chapter presented an alternative method for measuring muon momentum. An estimate of the muon energy loss as it passes the detector is first made using a neural network. The estimated energy loss is then used to predict the underlying muon momentum from knowledge of how muons deposit their energy as energy increases. Specifically, a log-likelihood-based approach is used to this end, using a Landau distribution which evolves with increasing momentum.

The resulting resolution is orders of magnitude poorer than that achieved using tracking. Nevertheless, a similar methodology could be used to complement tracking in the next generation of colliders, where an increased center-of-mass energy will produce more multi-TeV muons, resulting in radiative losses that exceed those seen in CMS today.

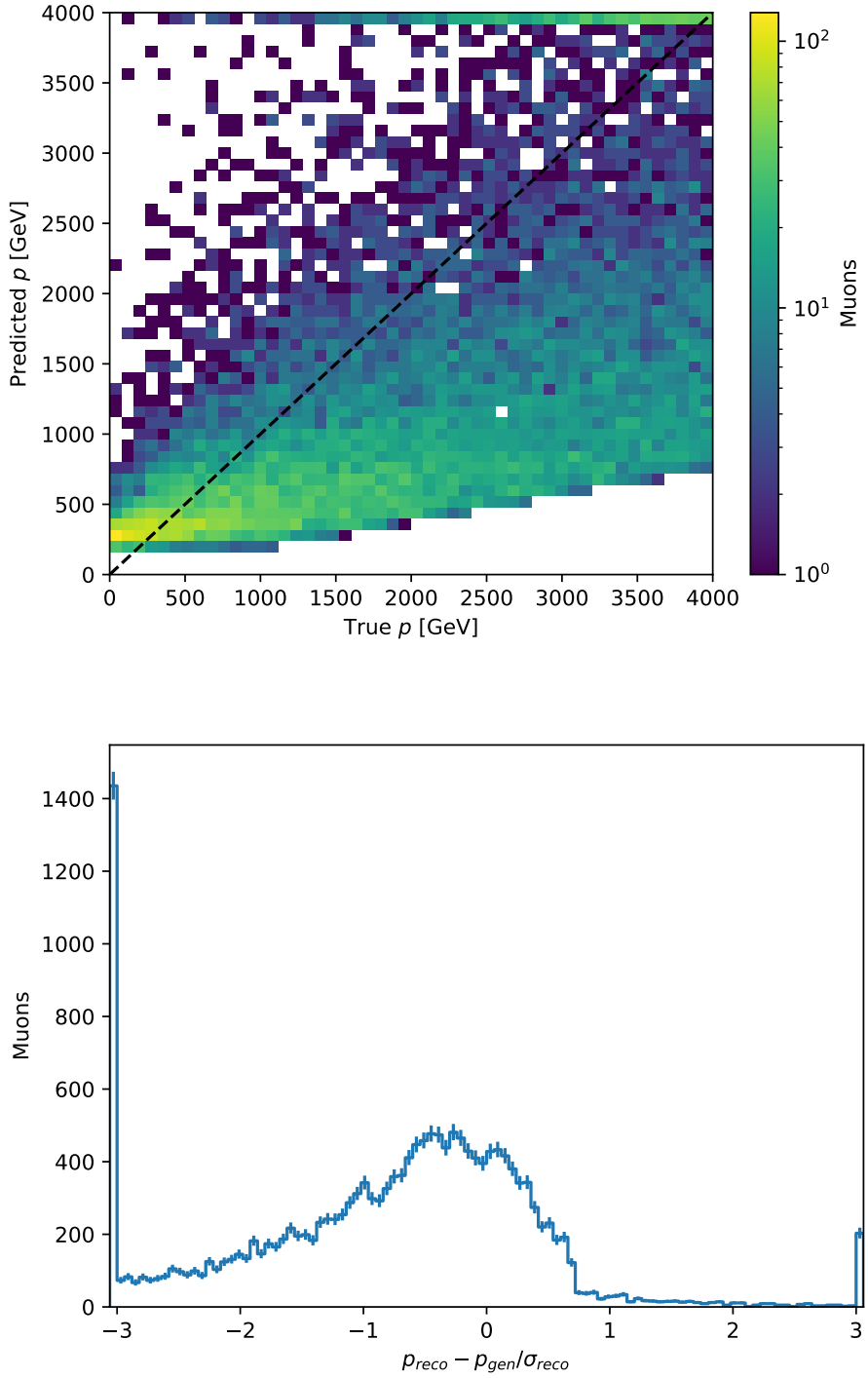


Figure A.8: Comparison of predicted and true momentum following the procedure described in the text (top). Normalized resolution (bottom) is calculated by dividing the difference in these two quantities by the size of the corresponding confidence interval. Overflows and underflows are shown in the last most bin for both plots.

REFERENCES

[Aab19] M. Aaboud et al. "Search for long-lived particles in final states with displaced dimuon vertices in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector." *Phys. Rev.*, **D99**:012001, 2019.

[Aba15] M. Abadi et al. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.", 2015. Software available from tensorflow.org.

[Abb04] M. Abbrescia et al. "Study of long-term performance of CMS RPC under irradiation at the CERN GIF." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **533**(1):102–106, 2004. Proceedings of the Seventh International Workshop on Resistive Plate Chambers and Related Detectors.

[Abb08] G. Abbiendi et al. "Muon reconstruction in the CMS detector." *CMS AN*, **2008/097**, 2008.

[AC15] J. Alimena and D. Cutts. "A study of displaced muon reconstruction." *CMS AN*, **2015/035**, 2015.

[Ada07] W. Adam. "Track and vertex reconstruction in CMS." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **582**(3):781–784, 2007. VERTEX 2006.

[AFF14] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations." *Journal of High Energy Physics*, **2014**(7):79, 2014.

[Aga18] A. Agarap. "Deep learning using rectified linear units (relu)." *arXiv preprint arXiv:1803.08375*, 2018.

[Ago03] S. Agostinelli et al. "Geant4—a simulation toolkit." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **506**(3):250–303, 2003.

[Ali17] J. Alimena et al. "Search for displaced leptons with 2016–2018 data." *CMS AN*, **2017/055**, 2017.

[AT13] D. Acosta and A. Tapper. "CMS Technical Design Report for the Level-1 Trigger Upgrade." Technical report, CMS Collaboration, 2013.

[BBC52] P. Barrett, L. Bollinger, G. Cocconi, Y. Eisenberg, and K. Greisen. "Interpretation of Cosmic-Ray Measurements Far Underground." *Rev. Mod. Phys.*, **24**:133–178, Jul 1952.

[Bic06] Hans Bichsel. "A method to improve tracking and particle identification in TPCs and silicon detectors." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **562**(1):154–197, 2006.

[Bru04] O. Bruning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004.

[bru19] *Electronics upgrade for the CMS CSC muon system at the High Luminosity LHC (September 30, 2019)*, volume 27 of *International Symposium Nuclear Electronics and Computing*, Budva, Becici, Montenegro, 2019.

[CEG15] D. Curtin, R. Essig, S. Gori, and J. Shelton. "Illuminating dark photons with high-energy colliders." *Journal of High Energy Physics*, **2015**(2):1–45, 2015. Publisher Copyright: © 2015, The Author(s).

[CER97] *The CMS hadron calorimeter project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997.

[Cer07] G. Cerminara. "The Drift Tube System of the CMS Experiment." *Nuclear Physics B - Proceedings Supplements*, **172**:71–74, 2007. Proceedings of the 10th Topical Seminar on Innovative Particle and Radiation Detectors.

[CGS08] M. Cacciari, P. Gavin, and G. Soyez. "The anti-k_t jet clustering algorithm." *JHEP*, **04**:063, 2008.

[Cli06] J. Cline. "Baryogenesis.", 2006.

[CLT08] R. Cousins, J. Linnemann, and J. Tucker. "Evaluation of three methods for calculating statistical significance when incorporating a systematic uncertainty into a test of the background-only hypothesis for a Poisson process." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **595**(2):480–501, 2008.

[cms] "CMSSW Github." `https://github.com/cms-sw/cmssw`.

[CMS06] "The CMS high level trigger." *The European Physical Journal C*, **46**(3):605–667, Apr 2006.

[CMS09] "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET." Technical report, CERN, Geneva, Apr 2009.

[CMS10] "Tracking and Primary Vertex Results in First 7 TeV Collisions." Technical report, CERN, Geneva, 2010.

[cms18] "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$ =13 TeV." *Journal of Instrumentation*, **13**(06):P06015–P06015, jun 2018.

[Col] The CMS Collaboration. *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*. Technical design report. CMS.

[Col06] The CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical design report. CMS. CERN, Geneva, 2006. There is an error on cover due to a technical problem for some items.

[Col08a] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider." *JINST*, **3**:S08003 – S08003, 2008/// 2008.

[Col08b] The CMS Collaboration. "The CMS experiment at the CERN LHC." *Journal of Instrumentation*, **3**(08):S08004–S08004, aug 2008.

[Col12a] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC." *Physics Letters B*, **716**(1):1–29, 2012.

[Col12b] The CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC." *Physics Letters B*, **716**(1):30–61, 2012.

[Col12c] The CMS Collaboration. "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s}$ = 7TeV." *Journal of Instrumentation*, **7**(10):P10002–P10002, oct 2012.

[Col14] The CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker." *Journal of Instrumentation*, **9**(10):P10009–P10009, oct 2014.

[Col15] The CMS Collaboration. "Search for long-lived particles that decay into final states containing two muons, reconstructed using only the CMS muon chambers." Technical report, CERN, Geneva, 2015. `https://cds.cern.ch/record/2005761`.

[Col21] The CMS Collaboration. "CMS Luminosity - Public Results.", 2021.

[com11] "Procedure for the LHC Higgs boson search combination in Summer 2011." Technical report, CERN, Geneva, Aug 2011.

[Cox16] T. Cox. "Introduction to Local Reconstruction and simulation.", 2016. `https://indico.cern.ch/event/522500/`.

[CPS12] M. Cacciari, Gavin P., and G. Soyez. "FastJet user manual." *Eur. Phys. J. C*, **72**:1896, 2012.

[Das19] A. Dasgupta. *Search for Long-Lived Particles Decaying to Displaced Dimuons at 13 TeV and Study of Neutron-Induced Background Hits in the Muon System of the Compact Muon Solenoid*. PhD thesis, UCLA, Los Angeles, 2019. `https://cds.cern.ch/record/2690341`.

[Dav16] S. Davis. "Interactive Slice of the CMS detector.", 2016. http://cds.cern.ch/record/2205172/.

[Dil21] S. Dildick et al. "Improving the performance of the CLCT pattern finder in CSCs using lookup-tables.", 2021. https://gitlab.cern.ch/tdr/notes/DN-19-059.

[EB64] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons." *Phys. Rev. Lett.*, **13**:321–323, Aug 1964.

[EB08] L. Evans and P. Bryant. "LHC Machine." *Journal of Instrumentation*, **3**(08):S08001–S08001, aug 2008.

[EF20] J. Erler and A. Freitas. "Electroweak Model and Constraints on New Physics.", 2020. https://pdg.lbl.gov/2020/reviews/rpp2020-rev-standard-model.pdf.

[Ein74] J. Einasto. "Dynamic evidence on massive coronas of galaxies." *Nature*, 1974.

[Eul13] G. Eulisse. "SingleElectronE1000_cfi.py.", 2013. https://github.com/cms-sw/cmssw/blob/6d2f66057131baacc2fcbdd203588c41c885b42c/Configuration/Generator/python/SingleElectronE1000_cfi.py.

[Flo] A. Florent. "Efficiency of Displaced Muon Trigger in 2018." https://indico.cern.ch/event/1032297/.

[Fra] A. Frankenthal. "Displaced muon identification and reconstruction." https://indico.cern.ch/event/984875.

[Fru87] R. Fruhwirth. "Application of Kalman filtering to track and vertex fitting." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **262**(2):444–450, 1987.

[Fru96] R. Fruhwirth et al. "Vertex reconstruction and track bundling at the LEP collider using robust algorithms." *Computer Physics Communications*, **96**(2):189–208, 1996.

[Fru07] R. Fruhwirth et al. "Adaptive Vertex Fitting." Technical report, CERN, Geneva, Mar 2007.

[Gri04] D. Griffiths. *Introduction to Elementary Particles*. 2004.

[Gro19] Particle Data Group. "Passage of Particles through Matter.", 2019. https://pdg.lbl.gov/.

[Gro20] D. Groom. "Atomic and Nuclear Properties of Materials.", 2020. https://pdg.lbl.gov/2021/AtomicNuclearProperties/.

[Hau99] J. Hauser. "Primitives for the CMS Cathode Strip Muon Trigger." p. 5 p, Oct 1999.

[Hau04] J. Hauser. "Experience with Trigger Electronics for the CSC System of CMS." 2004.

[Hig64] P. Higgs. "Broken Symmetries and the Masses of Gauge Bosons." *Phys. Rev. Lett.*, **13**:508–509, Oct 1964.

[Jun99] T. Junk. "Confidence level computation for combining searches with small statistics." *Nucl. Instrum. Meth. A*, **434**:435, 1999.

[Ken82] I. Kenyon. "The Drell-Yan process." *Reports on Progress in Physics*, **45**(11):1261–1315, nov 1982.

[Kha15] V. Khachatryan et al. "Search for long-lived particles that decay into final states containing two electrons or two muons in proton-proton collisions at $\sqrt{s} = 8$ TeV." *Phys. Rev. D*, **91**:052012, 2015.

[Lan44] L. Landau. "On the energy loss of fast particles by ionization." *J. Phys.*, **8**:201–205, 1944.

[Lee19] L. Lee. "Collider searches for long-lived particles beyond the Standard Model." *Progress in Particle and Nuclear Physics*, **106**:210–255, 2019. `https://www.sciencedirect.com/science/article/pii/S0146641019300109`.

[Liu16] S. Liu. "First Look at CLCT Patterns.", 2016. `https://indico.cern.ch/event/559368/`.

[Mat11] M. de Mattia et al. "Studies of standalone muon reconstruction for displaced muons." *CMS AN*, **2011/487**, 2011.

[Mey21] A. Meyer. "Physics Performance Dataset Run II Analysis Guideline.", 2021. `https://twiki.cern.ch/twiki/bin/viewauth/CMS/PdmVAnalysisSummaryTable`.

[Mob18] E. Mobs. "The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018." Aug 2018. General Photo.

[MR12] D. Morrissey and M. Ramsey-Musolf. "Electroweak baryogenesis." **14**(12):125003, dec 2012.

[Mul08] M. Mulders et al. "Muon identification in CMS." *CMS AN*, **2008/098**, 2008.

[Muoa] "Muon recommendations for 2016 legacy data and Monte Carlo." `https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonLegacy2016`.

[Muob] "Muon recommendations for 2018 ultra legacy data and Monte Carlo." `https://twiki.cern.ch/twiki/bin/view/CMS/MuonUL2018`.

[Nas18a] W. Nash. "Gain in Muon Efficiency from 3 Layer CLCTs.", 2018. `https://indico.cern.ch/event/771424/`.

[Nas18b] W. Nash. "Introduction to Comparator Codes and Status Update.", 2018. https://indico.cern.ch/event/778315/.

[Nas18c] W. Nash. "New LCT algorithm.", 2018. https://indico.cern.ch/event/713917.

[Nas20a] W. Nash. "Displaced dimuon: PAT-PAT and PAT-DSA categories.", 2020. https://indico.cern.ch/event/961945/.

[Nas20b] W. Nash. "PAT-PAT 2016 cut optimization with ZBi.", 2020. https://indico.cern.ch/event/943357/.

[Nas20c] W. Nash. "Progress towards calculating expected limits in the PAT-PAT category.", 2020. https://indico.cern.ch/event/979379/.

[Nas21a] W. Nash. "Displaced muon ID for small displacement.", 2021. https://indico.cern.ch/event/1018541/.

[Nas21b] W. Nash. "PAT-PAT: latest progress.", 2021. https://indico.cern.ch/event/1038911/.

[Nas22] W. Nash. "CSCCoffea github repository.", 2022. https://github.com/williamnash/CSCCoffea.

[OM19] T. O'Malley et al. "KerasTuner." https://github.com/keras-team/keras-tuner, 2019.

[OPY74] J. P. Ostriker, P. J. E. Peebles, and A. Yahil. "The Size and Mass of Galaxies, and the Mass of the Universe." *Astrophysical Journal Letters*, **193**:L1, October 1974.

[Pec18] A. Peck. "(O)TMB Pattern Logic.", 2018. https://indico.cern.ch/event/712513/contributions/2959915/attachments/1630667/2599369/20180410_otmb_pattern_logic.pdf.

[Phy15] "Search for long-lived particles that decay into final states containing two electrons or two muons in proton-proton collisions at $\sqrt{s} = 8$ TeV." *Phys. Rev. D*, **91**:052012, Mar 2015. https://link.aps.org/doi/10.1103/PhysRevD.91.052012.

[Pla20] Planck Collaboration. "Planck 2018 results - VI. Cosmological parameters." *A&A*, **641**:A6, 2020.

[Rea02] A. Read. "Presentation of search results: The CL_s technique." *J. Phys. G*, **28**:2693, 2002.

[Ros98] K. Rose. "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems." *Proceedings of the IEEE*, **86**(11):2210–2239, 1998.

[S15] Torbjörn S. et al. "An introduction to PYTHIA 8.2." *Computer Physics Communications*, **191**:159–177, 2015.

[Sal19] P. Salucci. "The distribution of dark matter in galaxies." *The Astronomy and Astrophysics Review*, **27**(1), Feb 2019.

[SC81] R. Santonico and R. Cardarelli. "Development of resistive plate counters." *Nuclear Instruments and Methods in Physics Research*, **187**(2):377–380, 1981.

[Sev18a] M. Sevilla. "CSC electronics and the LS1/LS2/LS3 upgrades.", 2018. `https://indico.cern.ch/event/741985/`.

[Sev18b] M. Sevilla. "Overview of the CSCs: muon systems in CMS, CSCs, gas, HV, and LV systems, CSC group.", 2018. `https://indico.cern.ch/event/741985/`.

[SFS06] W. Speer, T.and Adam, R. Frühwirth, A. Strandlie, T. Todorov, and M. Winkler. "Track reconstruction in the CMS tracker." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **559**(1):143–147, 2006. Proceedings of the X International Workshop on Advanced Computing and Analysis Techniques in Physics Research.

[Sir19a] A. Sirunyan et al. "Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13$ TeV ." *Eur. Phys. J. C*, **79**:421, 2019.

[Sir19b] A. Sirunyan et al. "Search for invisible decays of a Higgs boson produced through vector boson fusion in proton–proton collisions at $\sqrt{s} = 13$ TeV ." *Phys. Lett. B*, **793**:520, 2019.

[Sir19c] A. Sirunyan et al. "Search for new particles decaying to a jet and an emerging jet." *JHEP*, **02**:179, 2019.

[Smi36] S. Smith. "No. 532. The mass of the Virgo cluster." *Contributions from the Mount Wilson Observatory / Carnegie Institution of Washington*, **532**:1–8, January 1936.

[SZ08] M. Strassler and K. Zurek. "Discovering the Higgs through highly-displaced vertices." *Physics Letters B*, **661**(4):263–267, 2008.

[Tem] S. Templ. "Displaced Dimuon GitLab." `https://gitlab.cern.ch/DisplacedDimuons/DisplacedDimuons`.

[Ter] G. Terejanu. "Extended Kalman Filter Tutorial." `https://www.cse.sc.edu/~terejanu/files/tutorialEKF.pdf`.

[The] The CMS Collaboration. "Documentation for the Higgs Combine Tool.".

[Tro14] D. Trocino. "The CMS High Level Trigger." *Journal of Physics: Conference Series*, **513**(1):012036, jun 2014.

[Tul11] C.G. Tully. *Elementary Particle Physics in a Nutshell*. In a Nutshell. Princeton University Press, 2011.

[Vala] A. Escalante del Valle. "Displaced muon identification." `https://indico.cern.ch/event/1021115/`.

[Valb] A. Escalante del Valle. "Measurement of displaced muon performance using cosmic ray muons." `https://indico.cern.ch/event/938933`.

[Valc] V. Valuev. "Displaced muon identification." `https://indico.cern.ch/event/938933`.

[Val19a] A. Escalante del Valle et al. "Measurement of the displaced muon performance in Run II using cosmic ray muons." *CMS AN*, **2019/071**, 2019.

[Val19b] A. Escalante del Valle et al. "Search for long-lived particles decaying to a pair of muons in proton-proton collisions at $\sqrt{s} = 13$ TeV." *CMS AN*, **2018/206**, 2019.

[Wal07] C. Walck. "Handbook on Statistical Distributions for Experimentalists; International Report SUF-PFY/96-01.", 2007. `http://www.stat.rice.edu/~dobelman/textfiles/DistributionsHandbook.pdf`.

[Zwi33] F. Zwicky. "Die Rotverschiebung von extragalaktischen Nebeln." *Helvetica Physica Acta*, **6**:110–127, January 1933.

[Zyl20] P.A. Zyla et al. "Review of Particle Physics." *PTEP*, **2020**(8):083C01, 2020.