DISSERTATION

Search for displaced dimuons in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the CMS muon system

Ausgeführt zum Zwecke der Erlangung des akademischen Grades
Doktor der technischen Wissenschaften

unter der Leitung von
Univ. Doz. Dipl.-Ing. Dr. techn. Claudia-Elisabeth Wulz

am Institut für Hochenergiephysik der Österreichischen Akademie der Wissenschaften
im Rahmen des Doktoratskollegs "Particles and Interactions" (DK-PI)
finanziert durch den österreichischen Wissenschaftsfonds (FWF)

Eingereicht an der Technischen Universität Wien
Fakultät für Physik

von

Dipl.-Ing. Sebastian Templ BSc
Matrikelnummer 01125396

Wien, 30. September 2021

(Unterschrift Verfasser) (Unterschrift Betreuerin)

Zusammenfassung

Long-lived particles (LLPs, "langlebige Teilchen") erscheinen in zahlreichen, gut begründeten Theorien jenseits des Standardmodells der Teilchenphysik (SM) auf natürliche Weise. Solche LLPs könnten in hochenergetischen Teilchenkollisionen erzeugt werden und anschließend in bekannte Teilchen zerfallen, welche wiederum eine messbare Distanz entfernt vom LLP-Erzeugungsort detektiert werden könnten. Diese Arbeit präsentiert eine allumfassende ("inklusive") Suche nach neutralen, exotischen LLPs in Kollisionsereignissen mit Paaren von gegenteilig geladenen Myonen, welche einem gemeinsamen, dislozierten Vertex entstammen. Das zugrundeliegende Datenset wurde vom CMS-Detektor während der LHC "Run 2"-Periode aufgenommen und umfasst Proton-Proton-Kollisionen mit einer integrierten Luminosität von $97.6\,\mathrm{fb}^{-1}$ bei einer Schwerpunktsenergie von $\sqrt{s} = 13\,\mathrm{TeV}$. Die Suche ist suszeptibel für LLP-Zerfälle außerhalb des CMS-Trackingsystems, weil sie auf Myonen basiert, die lediglich mit dem CMS-Myonensystem rekonstruiert werden. Um die relevanten Trigger- und Rekonstruktionseffizienzen für Myonen in Abhängigkeit der Dislokation zu messen und die systematischen Unsicherheiten und Skalenfaktoren für die Analyse zu bestimmen, wurden Performance-Studien mit Myonen aus kosmischer Strahlung durchgeführt, welche im Jahr 2016 von CMS aufgezeichnet wurden. Die Ergebnisse werden im Kontext von zwei Benchmark-Modellen interpretiert: (1) einem Szenario, in welchem Myonen das finale Zerfallsprodukt eines generischen, langlebigen Skalarbosons sind, welches wiederum einem nicht-SM Higgsboson entstammt ("BSM Heavy Scalar"-Modell), und (2) dem "Hidden Abelian Higgs"-Modell, in welchem langlebige, skalare "dark photons" zu dislozierten Myonenpaaren führen. Für beide Interpretationen werden Obergrenzen für den Produktionswechselwirkungsquerschnitt für ein umfangreiches Hypothesenspektrum von LLP-Massen und -Lebenszeiten präsentiert. In allen Szenarien herrscht gute Übereinstimmung zwischen erwarteten und beobachteten oberen Limits und es wird kein statistisch signifikanter Anstieg über den prognostizierten SM-Hintergrund verzeichnet.

Abstract

Long-lived particles (LLPs) naturally appear in many well-motivated theories beyond the Standard Model of Particle Physics (SM). These LLPs could be produced in high-energy particle collisions and subsequently decay into known particles, which can be detected a measurable distance away from the LLP point of creation. This work presents an inclusive search for neutral, exotic LLPs in events with pairs of oppositely-charged muons originating from a common displaced vertex. The underlying dataset was recorded by the CMS detector during the LHC Run 2 and comprised an integrated luminosity of $97.6\,\mathrm{fb}^{-1}$ of proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13\,\mathrm{TeV}$. The search is sensitive to LLP decays beyond the CMS tracking system because it uses muons reconstructed by the CMS muon detectors only. To measure the relevant muon trigger and reconstruction efficiencies as a function of displacement and to assess systematic uncertainties and scale factors in the analysis, performance studies were carried out with cosmic-ray muons recorded by CMS in 2016. The results are interpreted in the context of two benchmark models: (1) a scenario in which the final-state muons are the decay products of a generic long-lived, scalar boson from a non-SM Higgs initial state ("BSM heavy scalar" model), and (2) the "Hidden Abelian Higgs model" which features long-lived, scalar dark photons as the intermediate LLP states giving rise to displaced dimuons in the final state. Upper limits on the production cross-sections in both interpretations are presented for a wide range of LLP mass–lifetime hypotheses. In all scenarios, there is good agreement between expected and observed upper limits and no statistically significant excess over the predicted SM background is reported.

Acknowledgments

This work would not have been possible without the support from colleagues, family, and friends.

First and foremost, I would like to extend my gratitude to Claudia-Elisabeth Wulz, who readily took me on as her Ph.D. student in 2018. Since then, her continuous supervision, availability, and readiness to help me with any issues have provided an excellent framework for carrying out my Ph.D. project.

I would also like to thank the members of the CMS displaced dimuon analysis group: Wolfgang Adam, Robert Cousins, Abhigyan Dasgupta, Alberto Escalante del Valle, Alice Florent, Jay Hauser, Muhammad Ansar Iqbal, Suchita Kulkarni, Ivan Mikulec, William Nash, Mangesh Sonawane, Slava Valuev, and Claudia-Elisabeth Wulz. The level of teamwork and commitment in this collaboration between HEPHY, UCLA, and CERN was outstanding. In particular, I would like to express my appreciation for the countless hours spared by Alberto and Ivan to regularly discuss the analysis progress and provide invaluable support. Especially in the early phase of my Ph.D. involvement, Alberto was never shy of lending me an ear, explaining the intricacies of the CMS collaboration and its detector, and nudging the analysis in the right direction. He significantly impacted my scientific training and helped shape my working style as a data analyst and researcher, and I would like to thank him for that sincerely.

In addition to my focus on data analysis, I have taken part in CMS Level-1 trigger activities from the beginning of my Ph.D. employment. I would like to thank the HEPHY Trigger Group – Claudia-Elisabeth Wulz, Bernhard Arnold, Herbert Bergauer, Priya Sajid Hussain, and Manfred Jeitler – for the pleasant collaboration over the years. Further, I am grateful to many people in the CMS L1T DPG and L1T Menu team, who not only welcomed me in my early Ph.D. days but who also provided me with the opportunity and challenge of leading the L1 Menu team between the years 2019 and 2021 – Cristina Botta, Andrew Brinkerhoff, Aaron Bundock, Olivier Davignon, Elisa Fontanesi, Jaana Heikkilä, Benjamin Radburn-Smith, Alexandre Zabi, and Mateusz Zarucki, among others. Being part of the L1 trigger management team was an instructive experience, which gave me further insights into some of my strengths and weaknesses and proved to be useful in both my professional and personal life.

I would also like to thank Jochen Schieck and Christoph Schwanda for their help and advice over the years as a student and HEPHY employee. Simone Krüger and Barbara Weber were instrumental in ensuring smooth handling of all administrative matters regarding my Ph.D. and all its contractual extra hurdles, and their always prompt support should not go unnoticed.

I am also indebted to my former high-school physics teacher, Bernhard Kinast, who sparked my interest in physics and eventually motivated me to enroll in a physics program at university. Without his early encouragement, I would not have taken up what had then seemed an (almost too) daring undertaking to me.

On a more personal note, I would like to express my gratitude to my closest Ph.D. colleagues at the institute – Janik Andrejkovic, Dominic Blöch, Priya Sajid Hussain, Lukas Lechner, and Felix Ulrich-Pur – who were invaluable friends in good Ph.D. times as in bad. Of course, these are not the only people who have made my (working) days better, and I would like to thank all fellow students, friends, and colleagues from the institute and the university with whom I could spend enjoyable times over the years.

A Ph.D. program is not an easy thing to follow through – particularly, in the environment of a large collaboration like CMS, where one typically has many more

responsibilities, meetings, presentations, and jobs than are directly relevant for one's thesis project – and finding the time required to carry out the necessary task provides a daily challenge. I could not have mastered this challenge without a network of family and friends, who always were supportive and understanding, especially in times of increased workload when I often could not give them all the attention they deserved.

I would like to give my heartfelt thanks to those who, over the past years, shared in my work most directly and were the best fellows imaginable – my beloved parents Sabine and Robert, my wonderful sister Johanna, as well as my dear friends Lukas Lechner, Claudia Klein, Nicole Zechner, Lukas Weichhart, Dominic Blöch, Caroline Gossler, Christina Reumiller, Andreas Lechner, Mariko Päßler, Felix Ulrich-Pur, and Sebastian Ratzenböck, to name but a few.

This work has been supported by the Austrian Science Fund FWF, Doctoral program No. W1252-N27.

Contents

Introduction

In the quest of searching for answers to some of the most fundamental questions, such as *What is the universe made of?*, curious and diligent minds of the last century developed a seemingly consistent and comprehensive picture of all known elementary particles and their electromagnetic, weak, and strong interactions – the mathematical framework known as the *Standard Model (SM)* of particle physics [1–3].

The SM is capable of describing with an astonishing degree of accuracy a breadth of physical phenomena from the lowest energies up to the $\mathcal{O}$(TeV) scale, which have been confirmed time and again in experiments.

One of the most recent confirmations of SM predictions came from the Large Hadron Collider (LHC) [4–7] when the Higgs boson [8, 9] – an elusive particle indicative of the then-hypothesized (and now-established) generation mechanism of particle masses – was discovered by the ATLAS [10] and CMS [11] collaborations, its couplings to gauge bosons were measured [12–14], and its coupling to third-generation fermions was directly observed [15–17].

Before long, the SM has become the much-heralded success story of modern particle physics.

Without a doubt, the development of the SM is one of the crowning achievements of modern physics. Notwithstanding all the praise, there are significant shortcomings[1] of the theory in terms of yet wholly unexplained phenomena and observations.

Arguably the most striking deficit of the theory is its disregard of gravity, which is simply not included in the framework (and any attempt to formulate a complete quantum theory of gravity has so far failed). The SM "gets away" with this and is still perfectly capable of describing particle-physics processes to the minute detail because the gravitational force is negligibly weak on the scale of particles, atoms, and molecules, compared to the other fundamental forces of nature.

However, this relative difference in interaction strength poses a pressing problem in itself. The weak interaction is about 10^{24} times stronger than the gravitational interaction, a scale difference so large that it is regarded as "unnatural" by many experts. To better understand this, one needs to keep in mind that the weak interaction is mediated by the W and Z bosons, which, in turn, obtain their masses via their couplings to the Higgs field [18]. Theoretical a-priori expectations of the size of the vacuum expectation value Higgs field would put its value either at zero or at the Planck mass [19], $m_p = (\hbar c/G)^2 \approx 1.2 \times 10^{19}\,\text{GeV}$ [20]. However, the observed Higgs field value turns out to have a size of

[1] Some of these shortcomings are not explicitly discussed in the following because they are probably not directly connected to this work. Examples include the inability of the SM to account for the observed asymmetry in matter-antimatter abundance (*Where is the antimatter in the universe?*), its contribution to Dark Energy (*What makes the universe expand at an accelerated rate?*), or the nature of neutrino masses (*What gives neutrinos their small but non-zero masses?*).

250 GeV, leading to a field configuration that, according to quantum mechanics, should be unstable and quickly change into one of the extreme values. Yet, more than 13.8 billion years after the Big Bang, we still measure a value of 250 GeV. This problem is dubbed the *hierarchy problem* [19].

Several solutions to the hierarchy problem have been proposed, including new particles and new forces capable of "correcting" the Higgs boson mass to its observed value (*e. g.*, Supersymmetry, abbreviated as SUSY [21]) or a new understanding of gravity (such as extra dimensions [22]).

Presently, however, there is no experimental evidence for any of them at the energies probed so far (up to a few TeV). The hierarchy problem of the "unnatural" and "fine-tuned" Higgs boson mass remains an open question.

However, most analyses in the first round of LHC searches targeted specific experimental signatures with *promptly*-decaying particles (such as some hypothesized SUSY particles), thus leaving a large volume of the model phase space that could be featuring yet-unknown long-lived particles (LLPs) unexplored [23, 24].

Another shortcoming of the present-day SM is that it does not provide a suitable candidate particle for Dark Matter (DM).

Even though no experiment to date has convincingly and directly identified a DM particle, the indirect evidence for the abundance of DM in the universe is compelling. For instance, studies of the anisotropies in the cosmic microwave background radiation or the rotational behavior of (clusters of) galaxies strongly suggest that about 85% of all matter in the universe is, in fact, dark matter [25]. On top of that, famous examples like the Bullet Cluster [26, 27] or the MACS J0025.4-1222 galaxy cluster [28] underline the expectation of the particle character of DM.

At the LHC, most initial DM searches hoped to find DM in the form of significant amounts of missing energy in the detectors and often targeted mono-particle signatures. But none of them has been able to report any statistically significant deviations from SM expectations.

However, as it turns out, in many DM scenarios, it can be much more promising to search for particles that mediate the interactions between the SM and the dark sector (*e. g.*, dark photons [29, 30]). The corresponding model phase space has been only marginally explored so far, with current experimental bounds being weak, non-existent, or covering only small volumes of phase space [31, 32].

The present work aims to fill this gap, at least to some degree, by covering known and unknown phase space with a generic and inclusive search for pairs of displaced muons across a broad range of LLP masses and lifetimes. This signature-driven approach allows sensitivity to a broad range of BSM physics models, such as SUSY, DM, heavy-neutrino, or hidden-sector scenarios, to name but a few.

Eventually, the results will be interpreted in the context of two common BSM benchmark models, *i. e.*, the Hidden Abelian Higgs Model (HAHM) [33, 34] and the generic BSM LLP production via a heavy scalar [35].

To this end, the document is structured as follows.

Chapter 2 introduces and outlines some of the core concepts of the modern SM formulation before discussing the opportunities and challenges of searches for long-lived particles (LLPs) beyond the Standard Model (SM). A particular focus on LLPs producing muon signatures in the detector is made.

Chapter 3 presents the experimental foundations of the search, *i. e.*, the Large Hadron Collider, the Compact Muon Solenoid Experiment (CMS) experiment, and the reconstruction of particles in the detector.

Chapter 4 describes the primary analysis. First, it introduces the used data samples and signal triggers and delineates the general analysis strategy, followed by a detailed treatment of muon object performance at all relevant stages of particle reconstruction, evaluated on a dedicated sample of actually-displaced muons (*i. e.*, muons from atmospheric cosmic rays). Subsequently, the core part of the analysis are detailed studies of the event selection, the prediction of SM backgrounds in the search regions, the involved systematic uncertainties, and the statistical interpretation of the results.

Finally, chapter 5 concludes the presented search and gives a brief outlook on the major improvements for displaced dimuon searches in CMS that can be expected in the near future.

At various places throughout the document, it is relevant to know which specific muon reconstruction algorithm (see Sec. 3.3) produced the particle tracks under discussion. The default assumption will be the "offline" reconstruction. This means that a muon track property, such as the transverse momentum p_T, will correspond to the offline-reconstructed value unless stated otherwise (*e. g.*, $p_T^{(\mathrm{L1})}$, $p_T^{(\mathrm{L2})}$, ...).

The acronyms used throughout this work are collected at the end of the document.

This work was carried out in close collaboration with analysis groups from HEPHY, UCLA, and CERN. In a team effort like this, the separation of tasks does not always strictly follow the outlines of individual theses and publications, with many people working on many different tasks in a way that benefits the team as a whole and not only a specific written document, necessarily. In fact, the fruitful collaboration within the HEPHY-UCLA-CERN team involved sharing as much of the workload as possible (e. g., by developing a common code framework for processing the data and establishing common analysis methods). Therefore, while this thesis focuses on the author's main contributions to the displaced dimuon search, some degree of overlap with existing [36] and future publications produced by this team and its members is accepted in favor of a complete and self-contained description of the analysis project.

Theoretical and practical aspects of searches for long-lived particles

The Standard Model of elementary particle physics successfully describes most of experimental reality. Yet, undoubtedly, it is not capable of accounting for and explaining *all* natural phenomena, such as Dark Matter or the scale hierarchy between the SM interactions and gravity. This section outlines some of the core ideas of the established SM and motivates searches for physics beyond the SM with a particular focus on the promise of long-lived particle signatures.

2.1 The Standard Model of elementary particles

This section gives an overview of the known elementary particles and their interactions via the fundamental forces of nature. First, the basic "vocabulary" is introduced in Sec. 2.1.1 by describing the particles and their broad classifications according to their properties. Afterward, in Sec. 2.1.2, the mathematical framework of the quantum-field theory of the known particles and their interactions – the *Standard Model (SM)* of present-day particle physics [1–3] – is discussed, outlining some of its core ideas and concepts.

This section does not attempt to provide a complete description of the SM. Instead, it motivates key notions that are needed in later chapters, such as the gauge-field structure of the theory, the SM-Higgs boson, the generation of particle masses via the mechanism of spontaneous symmetry breaking, or the kinetic mixing between bosonic gauge fields.

Much of this discussion is based on the explanations in [18], an excellent and pedagogical introduction to fundamental symmetries and how they make the SM emerge.

2.1.1 Overview of particles and interactions

All known elementary particles come in two variants: *fermions* and *bosons*. Two fermions can never assume the same particle state – a circumstance commonly known as the *Pauli exclusion principle* – while an arbitrary number of bosons can occupy the same state. This simple fact results in these two particle types behaving entirely differently: Fermions make up matter, and bosons are responsible for the forces of nature. In other words, atoms are made of fermions, but the forces acting between and within atoms are mediated by bosons.

There are four known fundamental forces in nature:

- Electromagnetic force (mediator particle: massless photons)
- Weak force (mediator: massive W^+, W^- and Z^0 bosons)

- Strong force (mediator: massless gluons)
- Gravity (hypothesized[1] mediator: massless gravitons)

The Standard Model of elementary particles (SM) incorporates all forces but the gravitational interaction. All attempts to formulate a complete quantum theory for gravity have failed so far. However, due to the relative weakness of the gravitational force compared to the other ones, the SM can describe the microscopic world of fundamental particles and their interactions highly accurately without accounting for gravity.

Whether mediators have mass or not has important consequences for the nature of their interactions: Each of the interactions is closely linked to a fundamental symmetry, and the experimental observation that the mediators of the weak interaction have non-zero mass indicates that the related symmetry is broken. In fact, the very mechanism for this symmetry breaking gives rise to the masses of *all* particles. This concept of "spontaneous symmetry breaking" and the connection to another fundamental mediator – the *Higgs* boson – will be discussed in more detail in Sec. 2.1.2.

Elementary particles can only interact via a given force if they carry a corresponding charge. Reversely, if a given particle has zero charge, it is not subject to the corresponding interaction.

The charges associated with each force in the SM are:

- Electromagnetic force: *electric* charge
- Weak force: *isospin*
- Strong force: *color* charge

The color charge is used to further divide fermions into two sub-categories: *quarks* (which are color-charged and therefore subject to the strong force) and *leptons* (which are color-neutral and therefore not strongly interacting). For example, color-charged particles like *up* and *down* quarks are the main building blocks of the nuclear components of atoms (*i. e.*, protons and neutrons). The lightest leptons (*i. e.*, electrons) constitute the "shell" around atomic nuclei, which defines the characteristics of all chemical reactions between atoms and molecules in nature.

Fermions come in three generations, each consisting of two quarks and two leptons. Tab. 2.1 gives an overview of the fermions in the SM and some of their most relevant properties.

In addition to their masses and charges, fundamental particles have further discriminating properties. Most importantly, each particle has a *spin*. In fact, it is the spin property that differentiates between fermions and bosons: The fundamental fermions all have spin $\frac{1}{2}$, while the fundamental bosons have spin 1. The only exception is the Higgs boson, which is the only presently known boson with spin 0.

Moreover, for each fundamental particle, there is a corresponding *anti-particle*, which has opposite physical charges such as electric charge. For example, the electron's ($q = -e$) anti-particle has charge $q = +e$, but otherwise identical properties. In the case of the electron, the anti-particle has a dedicated name – the *positron* – but, generally, anti-particles are

[1] There is presently no experiment sensitive enough to provide direct evidence for the existence of gravitons.

TABLE 2.1: *The fermions in the standard model, grouped into three generations of quarks and leptons, together with their charges and masses. [37] The mass values are rounded, and their measurement uncertainties are omitted in favor of clarity.*

	Generation 1	**Generation 2**	**Generation 3**	**Electric charge**	**Isospin**	**Color**	**Masses (Gen. 1, Gen. 2, Gen. 3)**
Quarks	Up Down	Charm Strange	Top Bottom	$+\frac{2}{3}e$ $-\frac{1}{3}e$	$+\frac{1}{2}$ $-\frac{1}{2}$	yes yes	2.2 MeV, 1.28 GeV, 173.0 GeV 4.7 MeV, 0.95 GeV, 4.18 GeV
Leptons	electron neutrino electron	muon neutrino muon	tauon neutrino tauon	0 $-e$	$+\frac{1}{2}$ $-\frac{1}{2}$	no no	$< 2 \cdot 10^{-3}$, < 0.19, < 18.2 MeV 0.511, 105.66, 1776.86 MeV

denoted by a simple "anti-" prefix. (Examples are the *anti-up* quark, *anti-neutrinos*,...) Lastly, some particles are their own anti-particles, like the photon.

After this generic, "big-picture" overview of the particles and interactions in the SM, the next section will discuss a selection of concepts in more detail which are at the core of the modern, quantum-field theoretical formulation of the SM.

2.1.2 Structure and fundamental symmetries of the quantum-field theory of elementary particles

There is no physical theory or mathematical framework that does not rely on a minimal set of assumptions, *i. e.*, a collection of axioms or external inputs that cannot be derived from more fundamental principles. The modern formulation of the SM is no exception. It has many ingredients that one needs to input "by hand" to derive a theory that accurately describes all experimental results. Some of the properties of the SM that cannot be derived from underlying principles are

- most of the coupling constants of the fundamental interactions and the masses of elementary particles,
- the observation that there are exactly *three* forces in the SM following from the symmetry groups U(1), SU(2), and SU(3) but no additional interactions (*e. g.*, there is no known force corresponding to SU(4)),
- the realization that there are *three* generations of fermions and no further ones (which is strongly suggested by experiments),
- the circumstance that particles have *three* possible spin configurations (0, $\frac{1}{2}$ or 1), but no elementary particles with other configurations seem to exist (*e. g.*, there are no elementary spin-$\frac{3}{2}$ particles).

However, a remarkable feature of the present-day understanding of the SM is that despite these "free parameters," a surprisingly small set of assumptions is needed to derive the full theory. It is sufficient to impose a minimal set of basic symmetries to deduce an accurate description of the fundamental quantum fields that eventually give rise to the particles and interactions constituting our experimental reality.

This section briefly outlines how one can derive our current understanding of particle physics from a few basic symmetry properties. The terminology and concepts introduced here will help to motivate theories beyond the SM, which will be the topic of Sec. 2.2 and subsequent sections. The following discussion is not a rigorous and complete derivation of the theory but merely quotes and connects some key results. For a more complete treatment, please see [18] or other standard particle physics resources.

Throughout this section, natural units will be used (*i. e.*, assuming $c = 1$ and $\hbar = 1$).

The mathematical principle underlying all modern particle physics theories is the *Lagrangian formalism*. Its basic idea is that something like Fermat's principle for light[2] also exists for massive objects. One defines an *action* functional S,

$$S[q(t)] = \int \mathcal{L} dt \ , \tag{2.1.1}$$

where $\mathcal{L}$ is a non-constant parameter commonly referred to as the *Lagrangian*. Generically, $\mathcal{L}$ is a function of the position $q(t)$ and the velocity of the object, $\mathcal{L} = \mathcal{L}\left(q(t), \partial_t q(t)\right)$.

Modern particle physics theories, such as the SM, have the Lagrangian describe fields rather than particles. In this case, the Lagrangian density[3] is

$$\mathscr{L} = \mathscr{L}\left(\Phi(\vec{x}, t), \partial_\mu \Phi(\vec{x}, t), \vec{x}, t\right) \tag{2.1.2}$$

and the corresponding action S is defined as

$$S = \int d^4x \, \mathscr{L}\left(\Phi^i, \partial_\mu \Phi^i\right) \tag{2.1.3}$$

for different field components i of the field Φ. Space and time are treated on equal footing in this context and provide the foundation on which the fields $\Phi(\vec{x}, t)$ act.

Finding the $\mathscr{L}$ that minimizes the action S will reveal the equations of motion (EOM) that describe the physical system in question. These are the so-called *Euler-Lagrange equations*.

In summary, the modern formulation of the SM sets out to find the $\mathscr{L}$ that correctly describes the configuration and interactions of the field components, which give rise to all known particles and forces (except gravity).

The starting point for the SM Lagrangian is the most fundamental assumptions of special relativity:

- The speed of light, c, is the same in all reference frames.
- The physical laws are identical in all inertial reference frames.

The set of all transformations complying with these symmetry requirements is called the Poincaré group. This "full spacetime symmetry group of nature" encompasses rotations and boosts (*i. e.*, the Lorentz group) as well as translations, and it constitutes the group-theoretical basis of the SM. Hence, in the context of the Lagrangian formalism, imposing

[2] For the example of light, Fermat's principle postulates that light always follows the path $q(t)$ between two points that minimizes the travel time between those two points. Mathematically, one defines an action functional of a given path $q(t)$ as $S[\boldsymbol{q}(t)] = \int dt$ and asks for the particular path $q(t)$ that minimizes this action.

[3] The Lagrangian density $\mathscr{L}$ is related to the Lagrangian introduced earlier by $\mathcal{L} = \int \mathscr{L} \, d^3x$.

the invariance of physical laws across inertial reference frames is equivalent to requiring the Lagrangian of the SM to be invariant under any transformations of the Poincaré group.

To establish the complete framework of the SM, one further needs to identify and exploit other symmetries of the Lagrangians describing spin-$\frac{1}{2}$ fields (which are associated with the Dirac equation), spin-1 fields (associated with the Proca equation), and spin-0 fields (associated with the Klein-Gordon equation) and also introduce the concept of *spontaneous symmetry breaking*. Moreover, each symmetry in the final SM Lagrangian will be connected to a conserved quantity – a conserved *charge*[4] – by virtue of Noether's theorem [38].

The remainder of this section is dedicated to briefly discussing these ideas and describing how they connect to form the full quantum-field description of elementary particles and their interactions.

U(1) transformations

The Lorentz-invariant Lagrangian describing free spin-$\frac{1}{2}$ fields is

$$\mathcal{L}_{\text{Dirac}} = \bar{\Psi}\left(i\gamma_\mu\partial^\mu - m\right)\Psi \tag{2.1.4}$$

and its corresponding EOM is the famous Dirac equation, $(i\gamma_\mu\partial^\mu - m)\Psi = 0$, describing spin-$\frac{1}{2}$ particles and fields.

$\mathcal{L}_{\text{Dirac}}$ is also invariant under global U(1) transformations of the form

$$\Psi \to \Psi' = e^{ia}\Psi \ , \tag{2.1.5}$$

where a is an arbitrary constant.

However, it is not symmetric under *local* U(1) transformations of the form

$$\Psi \to \Psi' = e^{ia(x)}\Psi \ , \tag{2.1.6}$$

where $a = a(x)$ now depends on the position x in space-time, as such transformations produce an extra term

$$-\left(\partial_\mu a(x)\bar{\Psi}\gamma^\mu\Psi\right) \ . \tag{2.1.7}$$

Global transformations, though, are a special case of a given set of *local* transformations. A theory inhibiting only global, but not local, symmetries, therefore, seems implausible (and is certainly in conflict with special relativity when the choice of the factor a would immediately fix its value across the whole universe).

A solution to this problem is offered when investigating the symmetries of the Lagrangian describing free and massless spin-1 fields:

$$\mathcal{L}_{\text{Maxwell}} = \partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu \tag{2.1.8}$$

This Lagrangian is symmetric under global transformations of the form

$$A_\mu \to A'_\mu = A_\mu + a_\mu \ , \tag{2.1.9}$$

[4]The expression of a *charge* might not be the ideal nomenclature for some conserved quantities, like momentum or energy (which follow from translational symmetry in space and time, respectively). However, following the spirit of [18], the concept of such a conserved *charge* should be understood in a generic, abstract manner rather than be taken too literally. The linguistic choice of calling *any* conserved quantity arising from a continuous symmetry a *charge* should serve the purpose of underlining the universality of Noether's theorem.

but, again, $\mathcal{L}_{\text{Maxwell}}$ is not *locally* invariant under similar transformations when $a_\mu = a_\mu(x)$. However, this Lagrangian is found to be locally symmetric under transformations

$$A_\mu \to A'_\mu = A_\mu + \partial_\mu a(x) \ . \tag{2.1.10}$$

Now, the pieces can be put together: An extra term $A_\mu \bar{\Psi}\gamma^\mu\Psi$ in $\mathcal{L}_{\text{Maxwell}}$, which couples the fields A_μ, $\bar{\Psi}$, and Ψ together, would provide a term under local transformations of Eq. 2.1.10 that exactly cancels out the extraneous term 2.1.7 that broke the local U(1) symmetry of $\mathcal{L}_{\text{Dirac}}$. In other words, adding a term $A_\mu \bar{\Psi}\gamma^\mu\Psi$ makes the combined Lagrangian locally U(1) invariant, and a complete description of a system with massive spin-$\frac{1}{2}$ fields and massless spin-1 fields is consequently given by

$$\mathcal{L} = \underbrace{-m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi}_{\text{free spin-}\frac{1}{2}\text{ fields}} + \underbrace{A_\mu\bar{\Psi}\gamma^\mu\Psi}_{\text{extra coupling term}} \underbrace{+\partial^\mu A^\nu\partial_\mu A_\nu - \partial^\mu A^\nu\partial_\nu A_\mu}_{\text{free spin-1 fields}} \ . \tag{2.1.11}$$

With the coupling constant g (describing the strength of the electromagnetic interaction) and the *covariate derivative* $D_\mu \equiv \partial_\mu + igA_\mu$, this Lagrangian becomes

$$\mathcal{L}_{\text{QED}} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu D^\mu\Psi - \frac{1}{2}(\partial^\mu A^\nu\partial_\mu A_\nu - \partial^\mu A^\nu\partial_\nu A_\mu) \ . \tag{2.1.12}$$

This is the Lagrangian for the quantum field theory of electrodynamics ("quantum electrodynamics"). It accurately describes fermions (spin-$\frac{1}{2}$ particles) and their electromagnetic interaction via photons (spin-1 mediators).

To recap, $\mathcal{L}_{\text{QED}}$ was assembled by starting from free spin-$\frac{1}{2}$ and spin-1 theories and subsequently introducing interaction terms that couple these fields together in a way that makes the resulting Lagrangian invariant under local U(1) transformations. The conserved quantity originating from U(1) symmetry by virtue of Noether's theorem is the electric charge.

SU(2) transformations

It turns out that U(1) is not the only internal symmetry of the discussed Lagrangians.

For two massless[5] spin-$\frac{1}{2}$ fields, ψ_1 and ψ_2, by adding two copies of $\mathcal{L}_{\text{Dirac}}$ according to

$$\mathcal{L}_{D_1+D_2} = i\bar{\Psi}\gamma_\mu\partial^\mu\Psi \ , \tag{2.1.13}$$

with the doublets $\Psi := \begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix}$ and $\bar{\Psi} := \begin{pmatrix}\psi_1 & \psi_2\end{pmatrix}$, one obtains a Lagrangian that is invariant under global SU(2) transformations,

$$\Psi \to \Psi' = e^{ia_i\frac{\sigma_i}{2}}\Psi \ , \tag{2.1.14}$$

where a_i are arbitrary constants and $\sigma_i/2$ are the three generators of SU(2) with the Pauli matrices σ_i (the sum over the index $i = 1, 2, 3$ is implicit).

However, similar to before, the Lagrangian is not *locally* SU(2)-symmetric. Again, an extra term in the Lagrangian helps to restore this symmetry. This time, the extra term

[5] The mass terms in Eq. 2.1.13 are omitted intentionally at this point because they would spoil the SU(2) symmetry. The solution to this apparent shortcoming will be the additional coupling to a spin-0 field and the mechanism of spontaneous symmetry breaking, which will be discussed later.

needs to be $i\bar{\Psi}\gamma_\mu W_i^\mu \sigma_i \Psi$ and describes interactions between two spin-$\frac{1}{2}$ fields $\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix}$ and three spin-1 fields W_i^μ, which transform according to

$$(W_\mu)_i \to (W'_\mu)_i = (W_\mu)_i + \partial_\mu a_i(x) + \varepsilon_{ijk} a_j(x)(W_\mu)_k \ . \tag{2.1.15}$$

Analogously to the assembly of the locally U(1) invariant Lagrangian discussed earlier, one obtains the locally SU(2) invariant $\mathcal{L}$ by adding the Lagrangians of the free spin-$\frac{1}{2}$ fields, the free spin-1 fields, and an extra coupling term:

$$\mathcal{L} = \underbrace{i\bar{\Psi}\gamma_\mu\partial^\mu\Psi}_{\text{free spin-}\frac{1}{2}\text{ fields}} + \underbrace{\bar{\Psi}\gamma_\mu\frac{\sigma_i}{2}W_i^\mu\Psi}_{\text{extra coupling term}} \underbrace{-\frac{1}{4}(W_{\mu\nu})_i(W^{\mu\nu})_i}_{\text{free spin-1 fields}} , \tag{2.1.16}$$

with the field strengh tensors $(W_{\mu\nu})_i = \partial_\mu(W_\nu)_i - \partial_\nu(W_\mu)_i + \varepsilon_{ijk}(W_\mu)_j(W_\nu)_k$.

The conserved charge originating from SU(2) symmetry is called isospin.

Mass terms and unification of SU(2) and U(1)

Combining the Lagrangians established so far (Eq. 2.1.12 and 2.1.16) yields the locally U(1) and SU(2) invariant Lagrangian

$$\mathcal{L} = \bar{\Psi}\gamma_\mu\left(i\partial^\mu + gB^\mu + g'\sigma_i W_i^\mu\right)\Psi - \frac{1}{4}(W_{\mu\nu})_i(W^{\mu\nu})_i - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \ , \tag{2.1.17}$$

where g and g' are the coupling constants, and B^μ and $(W^\mu)_i$ the U(1) and SU(2) *gauge fields*, respectively.

As indicated before, adding mass terms like $m_1\bar{\Psi}\Psi$ and $m_2(W^\mu)_i(W_\mu)_i$ to Eq. 2.1.17 would spoil the SU(2) symmetry. But we know from experiments that the particles which these Lagrangians are supposed to describe[6] have non-zero masses (see Tab. 2.1). This indicates that the underlying symmetry is "broken", *i. e.*, the symmetry exists at high energy scales but spontaneously breaks at lower energies.

As it turns out, the Lagrangian 2.1.17 can obtain mass terms while preserving SU(2) symmetry if $\mathcal{L}$ includes interactions with a spin-0 field [8, 9], as will be illustrated in what follows.

Starting from the globally U(1)-invariant Lagrangian for a complex spin-0 field, ϕ,

$$\mathcal{L}_{\text{spin-0}} = \frac{1}{2}\left(\partial_\mu\phi^\dagger\partial^\mu\phi - m^2\phi^\dagger\phi\right) \ , \tag{2.1.18}$$

one adds the next-higher power of ϕ to this Lagrangian. This reads, after renaming some variables according to conventions,

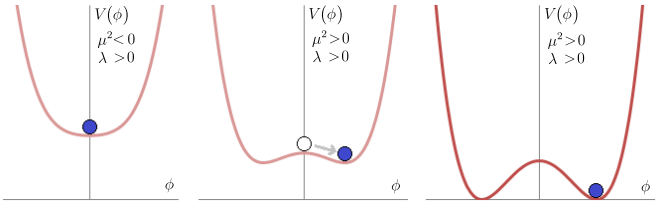
$$\mathcal{L}_{\text{spin 0 + higher power}} = \partial_\mu\phi^\dagger\partial^\mu\phi + \rho^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2 \ . \tag{2.1.19}$$

The extra term $-\lambda(\phi^\dagger\phi)^2$ preserves the symmetries and only adds self-interactions of the field ϕ.

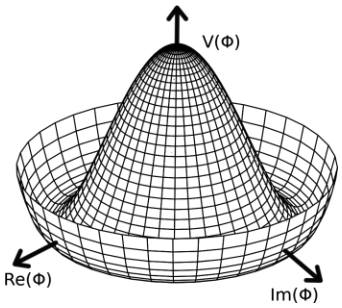
Now, one can define a locally U(1) and SU(2) invariant Lagrangian for doublets of spin-0 fields,

$$\Phi := \begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix} \ , \tag{2.1.20}$$

[6] For example, $\begin{pmatrix}\nu_e\\ e^-\end{pmatrix}$ is described by Ψ, or the three bosons of weak interaction are described by $(W_\mu)_i$.



(A) *The Higgs potential $V(\Phi)$ in two dimensions for $\lambda > 0$ and different values of ρ, representing the evolution of $V(\Phi)$ in a cooling universe. Figure adapted from [39].*



(B) *The Higgs potential $V(\Phi)$ in three dimensions.*

FIGURE 2.1: *Illustration of the shape and evolution of the Higgs potential $V(\Phi)$.*

as

$$\mathcal{L}_{\mathrm{U(1)+SU(2)}} = \left[(\partial_\mu - ig'\sigma_i(W_\mu)_i - igB_\mu)\,\Phi^\dagger\right] \cdot \left[(\partial^\mu + ig'\sigma_i(W^\mu)_i + igB^\mu)\,\Phi\right] - V(\Phi)\ , \tag{2.1.21}$$

with the so-called *Higgs potential* $V(\Phi)$,

$$V(\Phi) = -\rho^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2\ . \tag{2.1.22}$$

It is assumed that the parameters ρ and λ are temperature-dependent. This means that their values have changed over the universe's evolution, and so has the shape of the Higgs potential. In the early universe, which exhibited high energies, $|\rho^2| < |\lambda|$ and therefore the minimum of the Higgs potential – its so-called *vacuum expectation value* – was unambiguously at $\Phi = 0$, as Fig. 2.1 (left) indicates. With passing time, the universe cooled and transitioned into a period with $|\rho^2| > |\lambda|$. In this configuration, the Higgs potential has *non-zero* minima, $\Phi > 0$. More concretely, there are infinitely many minima (for every value of φ),

$$\Phi_{\mathrm{min}} = \sqrt{\frac{\rho^2}{2\lambda}}e^{i\varphi}\ . \tag{2.1.23}$$

In general, both components of the doublet 2.1.20 have to "choose" their respective minimum, $\Phi_{\mathrm{min}} = \begin{pmatrix}\phi_{1,\mathrm{min}}\\ \phi_{2,\mathrm{min}}\end{pmatrix}$. But an economical choice (without loss of generality) is

$$\Phi_{\mathrm{min}} = \begin{pmatrix}0\\ \sqrt{\frac{\rho^2}{2\lambda}}\end{pmatrix} \equiv \begin{pmatrix}0\\ \frac{v}{\sqrt{2}}\end{pmatrix}\ , \tag{2.1.24}$$

with $v \equiv \sqrt{\rho^2/\lambda}$, for brevity.

This very process of "choosing" one minimum out of infinitely many possibilities is called *spontaneous symmetry breaking* (and in this context also referred to as *electroweak symmetry breaking*).

To investigate the consequences of this symmetry breaking, one needs to shift the field Φ to its new, non-zero minimum and perform a series expansion around that minimum. The result of this operation (in *unitary gauge*) is

$$\Phi_{\mathrm{min,un}} = \begin{pmatrix}0\\ \frac{v+h}{\sqrt{2}}\end{pmatrix}\ , \tag{2.1.25}$$

where h is a physical field called the *Higgs field*.

Substituting the shifted field $\Phi_{\text{min,un}}$ in Eq. 2.1.21 and ignoring self-interactions of the Higgs field as well as interactions between the Higgs and other fields (*i. e.*, ignoring h) results in

$$\mathcal{L} = \underbrace{\frac{1}{4}v^2 g'^2}_{=M_W^2}(W^+)_\mu(W^-)^\mu + \underbrace{\frac{1}{4}v^2(g^2+g'^2)}_{=M_Z^2} Z_\mu^2 + \underbrace{\frac{1}{4}v^2 \cdot 0}_{=(\text{photon mass})^2} \cdot A_\mu^2 \ , \tag{2.1.26}$$

where the spin-1 fields $W_\pm^\mu$ are combinations of the original fields W_1^μ and W_2^μ according to

$$W_\pm^\mu \equiv \frac{1}{\sqrt{2}}\left(W_1^\mu \mp i W_2^\mu\right) \tag{2.1.27}$$

and A^μ and Z^μ are orthogonal linear combinations of the original fields W_3^μ and B^μ after a basis rotation with the *weak mixing angle* θ_W (also known as the "Weinberg angle")

$$\begin{pmatrix} A^\mu \\ Z^\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B^\mu \\ W_3^\mu \end{pmatrix} \ . \tag{2.1.28}$$

To recap, starting from a Lagrangian without mass terms for the spin-1 fields W_i^μ and B^μ (see Eq. 2.1.17), after including interactions with a doublet of spin-0 fields and after spontaneous symmetry breaking, one is left with a Lagrangian that now includes new terms. Remarkably, these new terms can be interpreted as mass terms proportional to M_W^2 and M_Z^2 for the mediators of the weak force, *i. e.*, the $W^\pm$ and Z bosons. One of the spin-1 fields, A_μ, remains massless even after symmetry breaking. This is the photon field of electromagnetism, reflecting the experimentally established fact that photons are indeed massless.

Another essential experimental observation [40] is that the weak interaction exclusively affects left-chiral particles (and right-chiral anti-particles). This means that, in nature, the discrete symmetry known as *parity* is maximally violated in weak interactions. Mathematically, one can implement this circumstance rather straightforwardly, simply by introducing (a) SU(2) *doublets* of left-chiral fields, $\Psi_L = \begin{pmatrix} (\psi_1)_L \\ (\psi_2)_L \end{pmatrix}$, which can transform into each other via weak interactions, and (b) SU(2) *singlets* of right-chiral fields, $(\psi_1)_R$ and $(\psi_2)_R$, which do not interact via the weak force.[7] All other fundamental interactions are parity-conserving.

Fermion masses cannot be incorporated into the Lagrangians directly without spoiling the SU(2) symmetry. The fact that the weak interaction violates parity underlines this notion, as Lorentz-invariant mass terms always combine left- with right-chiral fields. Again, the solution is to introduce appropriate coupling terms with spin-0 fields and subsequently break the electroweak symmetry.

A coupling term that is invariant under SU(2), U(1), and Lorentz transformations is the so-called *Yukawa term*

$$\mathcal{L} = -\lambda_f \left(\bar{\Psi}_L \Phi \psi_{2R} + \bar{\psi}_{2R} \Phi \Psi_L\right) \ , \tag{2.1.29}$$

[7] For a detailed discussion, the reader is asked to refer to [18] or any other standard modern particle physics textbook.

with the *Yukawa coupling* λ_f. Such a term describes the interaction between the fermions and the Higgs field.[8]

To illustrate this further, consider charged leptons. Using the more suggestive notation $\Psi = \begin{pmatrix} \nu_l \\ l \end{pmatrix}$, with $l = e, \mu, \tau$ (for "electron", "muon" and "tauon"), as well as $l = \begin{pmatrix} l_L \\ l_R \end{pmatrix}$, the coupling between the electrically charged lepton fields and the spin-0 field, $\bar{\Psi}_L \gamma_\mu \frac{\sigma_i}{2} W_i^\mu \Psi_L$ (compare Eq. 2.1.16), after symmetry breaking becomes

$$\mathcal{L}_{\text{mass}} = -\underbrace{\frac{\lambda_l v}{\sqrt{2}}}_{m_l}(\bar{l}l) - \underbrace{\frac{\lambda_l h}{\sqrt{2}}}_{c_l}(\bar{l}l) \ , \tag{2.1.30}$$

where the mass of a given lepton l is

$$m_l = \frac{\lambda_l v}{\sqrt{2}} \tag{2.1.31}$$

and the coupling strength between this lepton and the Higgs field is

$$c_l = \frac{\lambda_l h}{\sqrt{2}} = \frac{m_l h}{v} \ . \tag{2.1.32}$$

The last equation poses a profound insight: The mass of a charged lepton is proportional to its coupling strength to the Higgs field. The stronger this coupling, the heavier the lepton. This is true not only for charged leptons but for all other fermions[9] (*i. e.*, neutral leptons and quarks). Various precision measurements of the Higgs couplings were carried out at the LHC and confirmed this theoretical prediction [8, 9].

SU(3) transformations

Similarly to the previous discussion of SU(2) symmetry for two spin-1 fields, as discussed previously, one can identify another internal symmetry and construct a Lagrangian for *three* fermion fields q_i that is locally invariant under SU(3) transformations of the form

$$Q \to Q' = e^{iT_A \theta_A(x)} Q \ , \tag{2.1.33}$$

with the triplet of spin-$\frac{1}{2}$ fields $Q \equiv \begin{pmatrix} q_i \\ q_2 \\ q_3 \end{pmatrix}$ and the eight SU(3) generators $T_A = \frac{\lambda_A}{2}$ (where λ_A are conventionally chosen to be the 3×3 Gell-Mann matrices [41]).

To make the appropriate Lagrangian, $\mathcal{L} = i\bar{Q}\partial_\mu \gamma^\mu Q - \bar{Q} m Q$, locally SU(3) invariant, one again adds coupling terms between the spin-$\frac{1}{2}$ fields and new spin-1 fields to obtain the final Lagrangian, $\mathcal{L}_{\text{QCD}}$, which accurately describes strongly interacting particles in the framework of Quantum Chromodynamics (QCD) [42–44]:

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4}\mathcal{G}^A_{\alpha\beta}\mathcal{G}_A^{\alpha\beta} + \bar{Q}(iD_\mu \gamma^\mu - m)Q \ . \tag{2.1.34}$$

[8] The ψ_1 field is treated analogously, but with the charge-conjugated Higgs field $\widetilde{\Phi} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \Phi^*$ in Eq. 2.1.29 instead of Φ.

[9] The derivations are analogous to the case of charged leptons, but not further discussed here.

In $\mathcal{L}_{\text{QCD}}$, the sum over Roman capital letters (A, B,...) is implicit. $\mathcal{G}^A_{\alpha\beta}$ is the field strength tensor for the spin-1 gluon fields, $\mathcal{G}_\alpha \equiv T^C G^C_\alpha$,

$$\mathcal{G}_{\alpha\beta} = \partial_\alpha \mathcal{G}_\beta - \partial_\beta \mathcal{G}_\alpha - g[\mathcal{G}_\alpha, \mathcal{G}_\beta] \ , \tag{2.1.35}$$

with T^C being the generators of SU(3).

The covariate derivative, D_μ, in $\mathcal{L}_{\text{QCD}}$ is defined via

$$D_\alpha = \partial_\alpha + igT^C G^C_\alpha = \partial_\alpha - ig\mathcal{G}_\alpha \tag{2.1.36}$$

The above equations are entirely analogous to the SU(2) case, except for the fact that they describe *three* spin-1 fields and have different generators with different commutator properties. Specifically, the generators T_A follow

$$[T_A, T_B] = if^{ABC}T^C \ , \tag{2.1.37}$$

with f^{ABC} being the *structure constants* of SU(3), which are antisymmetric under permutation of any two indices.

Again, new conserved quantities appear from global SU(3) symmetry by virtue of Noether's theorem. Particularly, one can extract three new charges, called *color* charges (with the artificial labels "red", "green", "blue"), which are assigned to each strongly interacting elementary particle. This means that each quark is an SU(3) triplet, Q, and inside each triplet, the same quark appears in three different color charges. Particles that do not interact via the strong force are SU(3) *singlets* with zero color charge.

Moreover, the non-commutative nature of the SU(3) generators, Eq. 2.1.37, has two important consequences for the phenomenology of the strong interaction [45, 46]:

- **Color confinement:** To form color-neutral states, quarks come in various combinations but are never observed individually. For example, mesons are quark-antiquark pairs ($\bar{q}q$) with color and anti-color forming a color-neutral compound, while baryons (qqq) are combinations of different colors that form an overall color-neutral state. Other possibilities like tetraquarks ($q\bar{q}q\bar{q}$) or pentaquarks ($qqq\bar{q}q$), although being relatively rare in nature, have been experimentally produced recently. [47, 48] When *individual* quarks are produced in high-energy particle collisions, a $q\bar{q}$ pair is quickly created from the vacuum to re-establish a color-neutral state. This process is called *hadronization*.

- **Asymptotic freedom:** Due to the mechanism of hadronization, there are no color-charged currents over a macroscopic distance. Only at short distances (corresponding to high energies) can quarks be seen as *free* particles, as the strong coupling constant decreases with growing momentum transfer. This circumstance is known as the *asymptotic freedom* of the strong interaction. It allows for perturbative calculations at *high* energies (in contrast to all other interactions, for which perturbative approaches only work in the *low*-energy domain).

The full picture: Fermions and bosons and their interplay

The previous sections gave a condensed overview of relevant core concepts of the field-theoretical SM formalism. The following paragraphs will put these pieces together, repeat some of the main ideas, and thus outline the complete picture of the SM.

Instead of addressing the particles that we see in experiments directly, the modern SM formulation describes *fields*. Mathematically, it is convenient to work in the established framework of Lagrangian densities, which are the starting point for the derivation of the equations of motions of a physical system (by virtue of their Euler-Lagrange equations).

Even though the SM framework uses *fields* as the fundamental objects, there is a direct connection between these fields and the known particles: Each particle can be produced or destroyed by its corresponding quantum field. The fundamental fields of the SM – and, consequently, the elementary particles – come in three spin configurations: spin $\frac{1}{2}$ (fermions), spin 1, and spin 0 (both bosons). Matter is made of fermions; bosons mediate the forces between matter.

The Lagrangians describing one, two, or three spin-$\frac{1}{2}$ fields exhibit internal symmetries. These symmetries give rise to various conserved quantities through Noether's theorem, which are interpreted as charges. However, these symmetries are only *global* ones, which seems implausible given the principles of special relativity.

It turns out that the Lagrangians can be made *locally* invariant by including additional coupling terms between the spin-$\frac{1}{2}$ fields and new spin-1 fields.

Local U(1) invariance requires one gauge field, A_μ, and the corresponding Lagrangian correctly describes electromagnetic interactions. The mediator created by the field A_μ is the photon. The conserved quantity from U(1) symmetry is the electric charge.

Local SU(2) invariance requires three gauge fields, W_i^μ $(i = 1, 2, 3)$, and the corresponding Lagrangian describes weak interactions. The mediators created by the fields W_i^μ are the W^+, W^- and Z bosons. The conserved quantity from SU(2) symmetry is isospin.

Local SU(3) invariance requires eight gauge fields, G_A^μ $(A = 1, 2, \ldots, 8)$, and the corresponding Lagrangian describes strong interactions. The mediators created by the fields G_A^μ are eight different gluons. The conserved quantity from SU(3) symmetry is color.

From experiments, we know that all particles (except photons and gluons) have non-zero mass. However, explicit mass terms in the Lagrangian would spoil SU(2) symmetry. The Higgs mechanism presents a solution. By introducing additional terms that couple the spin-$\frac{1}{2}$ and spin-1 fields to a new spin-0 field (*i. e.*, the Higgs field), subsequently breaking the SU(2) symmetry spontaneously and expanding the Higgs field around the newly obtained and asymmetric minimum, one dynamically generates the appropriate mass terms in the Lagrangian.

Finally, the full gauge group describing the fermions and bosons of the SM and their interactions is the unitary product group

$$G_{\mathrm{SM}} = SU(3)_C \times SU(2)_L \times U(1)_Y \ . \tag{2.1.38}$$

2.2 Long-lived particles beyond the Standard Model

In this section, the general notion of a long-lived particle (LLP) and the various mechanisms that can generate macroscopic particle lifetimes will be discussed. Searches for LLPs from BSM processes will be motivated, and a brief overview of the opportunities and challenges of such endeavors will be given. Finally, the discovery potential of displaced *muons* will be discussed, and the two common benchmark BSM signal models that are used in the context of this work will be described.

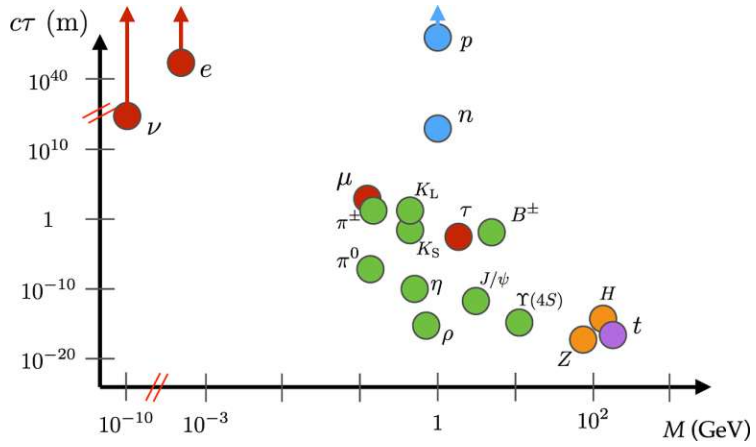


FIGURE 2.2: *Lifetimes and masses of a variety of SM particles. Figure taken from [49].*

2.2.1 Exploring new phase space with long-lived particle searches

The notion of a long-lived particle (LLP) is not a new one. The known SM particles have lifetimes ranging over many orders of magnitude – from the short-lived Z boson (lifetime $\tau \approx 2 \times 10^{-25}\,\mathrm{s}$) to the neutron ($\tau \approx 900\,\mathrm{s}$) to the proton ($\tau \gtrsim 10^{34}$ years, or more than 10^{24} times the age of the universe) to the electron (stable).

Fig. 2.2 illustrates the lifetimes and masses of known particles (in lifetime units[10] of $c\tau$). It shows that many of the known particles decay within micrometers or much less, especially the heaviest ones, like the Z, the Higgs, or the top quark. In the context of collider experiments (such as the CMS experiment), such very short-lived particles or their decay products are called *promptly* decaying particles.

On the other hand, particles that traverse the entirety of the detector volume without decaying are treated as *stable* final-state particles (like electrons or protons).

In between those two corner cases is the realm of *long-lived* particles, *i. e.*, particles that decay a (potentially) sizeable – measurable – distance away from their point of creation yet have lifetimes short enough to decay within the detector volume. Of course, this definition is not an exact one but rather motivated by the presently existing experiments and their measurement capabilities. In practice, one could call a particle a *long-lived* particle if it has a lifetime $c\tau$ somewhere between a few millimeters and a few kilometers when traveling with near light speed.

> “***Long-lived particles (LLPs)*** *naturally appear both in the Standard Model (SM) and in many theories beyond the SM (BSM), with their masses and lifetimes varying over a broad range of values. Present-day discussions of LLPs usually address BSM LLPs with lifetimes ranging from millimeters to kilometers (depending on the experimental context).*”

An important realization is that LLPs do not only appear in the SM. In fact, many BSM models naturally predict new and exotic particles with macroscopic lifetimes. While there exists a breadth of BSM models and a flurry of BSM LLP production scenarios, long

[10] Strictly speaking, the quantity $c\tau$ (*i. e.*, the proper particle lifetime τ multiplied by the speed of light c) has the dimension of length and, therefore, is not a direct measure for (life)time. However, since most of the particles under discussion move with near light speed, their lifetime is often conveniently expressed in length units of $c\tau$.

lifetimes are typically generated by one of three mechanisms [24]:

- A combination of scale hierarchies in interactions that mediate decays (*e. g.*, decays via heavy off-shell mediators),
- small couplings, or
- phase space considerations (*e. g.*, small mass splitting between the particles in a decay chain or large multiplicities of particles in the final state).

To make these LLP generation mechanisms more concrete, one can consider the following scenario: Assumption of a heavy mediator X with mass m_X and two BSM particles, Y and Z, with respective weak-scale masses m_Y and m_Z. The couplings between the BSM and the SM particles are defined as g_X, g_Y, and g_Z.

With this, the following configurations will lead to BSM models that contain particles with macroscopic lifetimes. [50] For an extensive collection of references for each of the models quoted below, please see section 2.2 of [24].

Decay via a heavy, off-shell mediator:

Considering the decay $Y \to X^* \to \text{SM}$, if the mediator X is heavy (with a mass up to or around the TeV scale), then $m_X/m_Y \gg 1$. This will lead to the lifetime of Y being determined by

$$c\tau_Y \approx \frac{1.2\,\text{fm}}{g_X^4}\left(\frac{m_X}{m_Y}\right)^4\left(\frac{1\,\text{TeV}}{m_Y}\right) \tag{2.2.1}$$

Example scenarios:

- $m_X \approx 100\,\text{GeV}$, $m_Y \approx 10\,\text{GeV}$, $g_X^4 \approx 10^{-7} \to c\tau_Y \approx 1\,\text{cm}$ (*e. g.*, heavy neutral lepton decays,...)
- $m_X \approx 1\,\text{TeV}$, $m_Y \approx 10\,\text{GeV}$, $g_X^4 \approx 10^{-3} \to c\tau_Y \approx 1\,\text{cm}$ (*e. g.*, Hidden Valley scenarios, Split SUSY,...)

Small mass splitting between initial and final states:

If the masses of particles Y and Z are nearly degenerate in the process $Y \to X^* \to Z + \text{SM}$, then $m_X/\Delta m \gg 1$ (with $\Delta m = m_Y - m_Z$). The lifetime of Y is then determined by

$$c\tau_Y \approx \frac{1.2\,\text{fm}}{g_X^4}\left(\frac{m_X}{\Delta m}\right)^4\left(\frac{1\,\text{TeV}}{\Delta m}\right) \tag{2.2.2}$$

Example scenario:

- $m_X \approx 100\,\text{GeV}$, $\Delta m \approx 1\,\text{GeV}$, $g_X^4 \approx 10^{-2} \to c\tau_Y \approx 1\,\text{cm}$ (*e. g.*, Compressed SUSY,...)

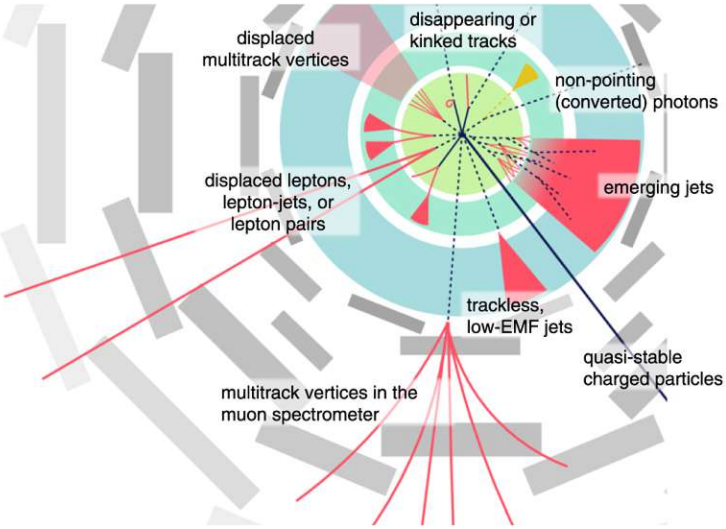


FIGURE 2.3: *An illustration of a variety of plausible detector signatures involving BSM LLPs in a collider experiment like CMS (in the ϕ–r plane). Not shown are delayed signatures such as stopped particles or delayed jets. Figure taken from [51].*

Small couplings:

If the coupling between BSM and SM particles in the process $Y \to \text{SM}$ is small, $g_Y \ll 1$, then the lifetime of Y is determined by

$$c\tau_Y \approx \frac{0.02\,\text{fm}}{g_Y^2}\left(\frac{1\,\text{TeV}}{m_Y}\right) \tag{2.2.3}$$

Example scenario:

- $m_Y \approx 1\,\text{GeV}$, $g_Y^2 \approx 10^{-12} \to c\tau_Y \approx 1\,\text{cm}$ (*e. g.*, dark photons,...)

In all of the above scenarios, plausible sets of mass values and couplings can lead to a wide range of Y lifetimes. However, the exact configuration of each of these parameters is clearly model-dependent, and, consequently, the lifetime of a new particle is commonly treated as a free parameter in experimental LLP searches.

Not only does the realization that LLPs appear naturally in many BSM models motivate to search for such processes, but also does an entirely different circumstance: SM particles with macroscopic lifetimes have masses roughly below 5 GeV and well-understood signatures in experiments – quite unlike BSM LLPs, which can have unusual and exotic signatures. Fig. 2.3 illustrates a selection of different plausible decay signatures that involve BSM LLPs. In principle, this makes any BSM processes involving LLPs potentially stand out from SM decays – some searches even being expected to be virtually background-free for a wide range of search parameter ranges.

However, many of the circumstances that make searches for unknown LLPs appealing are also responsible for these searches being experimentally challenging in many cases. The very fact that BSM LLP signatures are usually exotic in their nature often makes it difficult or even impossible to use established experimental methods, *e. g.*, strategies for event triggering, particle reconstruction, and physics simulation. Especially in experiments

that have originally been designed for the study of prompt processes (like CMS or ATLAS), new or refined methods have to be developed at most stages of the data-processing pipeline to increase the sensitivity for LLP detection. Given the vastly different characteristics of LLP decays (see again Fig. 2.3), each signature usually requires a dedicated analysis with dedicated methods.

There are several generic sources of background that LLP searches at the LHC have to deal with. They come in the form of real and misidentified particles as well as from sources inside and outside a given detector.

The following will give a brief and generic overview of them and present a distilled version of the discussion in [24]. Those background sources that are directly relevant to the work at hand (as well as the methods used to control them) are described in great detail in Sec. 4.6.

Long-lived particles from known SM decays are a relevant source of background, especially in searches for low-mass LLPs (few tens of GeV or below) in the proximity of the interaction point (IP) (few millimeters off the beam spot). In this domain, SM processes like the decays of b hadrons or b-mesons can look similar to the targeted signal process. This type of background is usually suppressed by studying both collision and MC data and identifying effective mitigation methods (*e. g.*, by applying uni-, bi- or multivariate cuts on analysis variables).

> “ ***Searches for LLPs often face the challenge*** *of atypical backgrounds, unknown particle reconstruction performance, and unreliable MC simulations. Experimental methods established in prompt particle reconstruction and analysis cannot always be applied to LLP searches directly.* ”

Real particles created in material interactions (*i. e.*, promptly-produced particles that interact with nuclei of the detector material) can fake LLP signatures that involve displaced vertices. Since vertices reconstructed from such interactions will cluster in detector regions with high material densities, the most direct approach to reducing this background is to use detailed detector material maps and veto such events based on the location of the reconstructed displaced vertices.

Cosmic-ray muons are a natural source of muons that generally appear displaced in the reference frame of a given detector. These atmospheric particles can be reconstructed as displaced muons or as displaced jets in the muon system or the calorimeters of an experiment, respectively. Exploiting known detector properties of cosmic-ray muons (*e. g.*, their preferred direction and the fact that they are usually reconstructed as two back-to-back muon objects with respect to the center of the detector), as well as the analysis of dedicated cosmic muon datasets, can often provide enough handles to control this type of background.[11]

Beam halo in collider experiments is another type of background that, like cosmic-ray muons, has its origin outside the detector. It is the result of imperfections of the colliding beams when beam particles scatter off the collimators and subsequently produce debris that can fake a displaced-particle signal in the detector. Beam halo can be controlled,

[11] Cosmic-ray muons are of particular importance for the displaced dimuon search presented in this work – not only do they constitute a major source of background, but they also serve as a crucial ingredient for the direct measurement of reconstruction performance and other analysis ingredients, as will be discussed in Sec. 4.4 and other places.

e. g., by requiring correlations between particle hits in the inner and outer detector regions, by requiring reconstructed objects to have a certain momentum/energy minimum, or by generally exploiting the fact that beam halo is dominating the very forward regions of a detector.

Cavern radiation is another general source of potential background, but particularly relevant in searches for stopped particles (*i. e.*, BSM particle decays that happen outside the typical time window of a beam bunch-crossing). Mitigation methods for cavern radiation often involve the study of events collected by randomly firing triggers when there are no collisions.

Signatures of fake particles can arise from signatures unrelated to actual detector hits but are nevertheless interpreted as evidence of measured particles. For example, such signatures can be produced by spurious detector noise. It is generally difficult to model this background in simulation accurately. However, the fact that such detector noise usually appears as a single and localized energy deposit that does not correlate in space and time with any other energy deposits in the detector generally provides good handles to control this type of background.

Algorithmically induced misreconstructions can be an issue when particle reconstruction encounters hit patterns from multiple particles that are challenging to decode, given the measurement uncertainties. Even in a background-free event, it can be algorithmically difficult to unambiguously and correctly reconstruct, *e. g.*, very close-by vertices or randomly crossing tracks. Mistakes in the reconstruction algorithms can lead to mismeasured properties of a (system of) particles and therefore mimick exotic signatures. In contrast to most of the previously discussed background types, algorithmically induced "fake particles" can be more easily studied and mitigated using MC simulation (*e. g.*, by artificially inducing and studying problematic event topologies).

2.2.2 Displaced muons as a window to unknown physics

In this section, the focus will shift from the generic picture of various LLP decay scenarios to the signature of displaced muons. The promising potential of displaced muons to discover unknown LLPs in the context of a variety of BSM models will be discussed. Finally, two generic benchmark models involving LLP decays to displaced dimuons will be presented. These models will be at the core of the search for displaced dimuons beyond the CMS tracker, which will be the subject of the remainder of this work.

> "*Searching for detector signatures of* ***displaced muons*** *is a promising route to probing unexplored phase space for a variety of BSM models. The two interpretations used in this work are the benchmark-model LLP decays $H \to XX$, $X \to \mu\mu$ ("BSM Heavy Scalar model") and $h \to Z_D Z_D \to 2\mu 2X$ (dark photons in the "Hidden Abelian Higgs Model").*"

The discussions leading up to this point have suggested that there is a large collection of plausible BSM scenarios that produce unknown LLPs, which might decay in a variety of ways into measurable final-states (*e. g.*, particles of the SM). Therefore, the experimental approach presently adopted by most LLP searches is a *signature-driven* one rather than a model-specific one. This allows individual searches to cover much more LLP phase space when the experimental findings are eventually re-interpreted in the context of a specific scenario. Moreover, relatively simple models are often sensitive to a

TABLE 2.2: *A selection of neutral LLP (X) channels leading to displaced leptons (adapted from [24]).*

Production	Topology and decay ($X \to l^+l^-$)	UV complete models
Direct pair production	sneutrino pair neutralino pair	SUSY
Heavy parent	squark pair: $\tilde{q} \to jX$ gluino pair: $\tilde{g} \to jjX$ slepton pair: $\tilde{\ell} \to \ell X$ chargino pair: $\tilde{\chi} \to WX$	SUSY
Higgs boson production	$h \to XX$(+invisible)	SUSY, DM, RHν
Resonance	$Z(Z') \to XX$(+invisible) $Z(Z') \to X$(+invisible)	Z', DM
Charged current	$W(W') \to \ell X$	RHν

large portion of the model phase space. Therefore, the following discussion focuses on *single* decay modes, more concretely, on BSM decays that lead to (at least) a pair of displaced muons ($X \to \mu^+\mu^-$) for a wide range of X lifetimes. Tab. 2.2 gives a broad overview of a selection of common BSM models that feature the decay of a neutral LLP to displaced leptons and illustrates that such generic signatures can eventually be connected to a variety of concrete BSM theories.

The search presented in this work is performed in the context of two commonly used LLP benchmark models, both of which feature displaced dimuons in the final states resulting from BSM decays with a minimal set of assumptions. These two models will be discussed in the remainder of this section.

Minimal benchmark model featuring rare Higgs boson decay and dark photons

A widely used benchmark model with a minimal set of BSM ingredients is the so-called *Hidden Abelian Higgs Model (HAHM)* [33, 34]. It describes a prototypical hidden sector featuring a spontaneously broken "dark" $\mathrm{U}(1)_D$ gauge symmetry, which is mediated by a new vector boson, the so-called *dark photon* Z_D. The SM fields are singlets under the new gauge symmetry, so the only coupling between the Z_D and the SM hypercharge gauge bosons (photon and Z boson) happens through kinetic mixing.

If the spontaneous symmetry breaking of the $\mathrm{U}(1)_D$ is a "dark Higgs mechanism" (analogous to the electroweak-symmetry-breaking Higgs mechanism in the SM), the associated dark Higgs boson, s, will generally mix with the 125-GeV SM-like Higgs, h.

Hidden sectors near the weak scale, such as the one featured in this model, are motivated by naturalness [52–56], thermal dark matter [57–59], and electroweak baryogenesis [60] and therefore represent a general expectation of BSM physics [33, 61].

Using the notation from [33], the gauge terms in the Lagrangian that describe the kinetic mixing between a broken dark Abelian gauge symmetry, $\mathrm{U}(1)_D$, and the SM hypercharge, $\mathrm{U}(1)_Y$, are collected in Eq. 2.2.4. The hatted fields in this Lagrangian denote the original fields with non-canonical kinetic terms (*i. e.*, before the field redefinitions).

$$\mathcal{L} \supset -\frac{1}{4}\hat{B}_{\mu\nu}\hat{B}^{\mu\nu} - \frac{1}{4}(\hat{Z}_D)_{\mu\nu}(\hat{Z}_D)^{\mu\nu} + \frac{1}{2}\frac{\varepsilon}{\cos\theta}(\hat{Z}_D)_{\mu\nu}\hat{B}^{\mu\nu} + \frac{1}{2}m^2_{D,0}(\hat{Z}_D)^{\mu}(\hat{Z}_D)_{\mu} \qquad (2.2.4)$$

with $\hat{B}_{\mu\nu} \equiv \partial_\mu \hat{B}_\nu - \partial_\nu \hat{B}_\mu$ and $(\hat{Z}_D)_{\mu\nu} \equiv \partial_\mu (\hat{Z}_D)_\nu - \partial_\nu (\hat{Z}_D)_\mu$, the Weinberg mixing angle θ, the kinetic mixing parameter ε between the Z_D and the SM hypercharge bosons and the dark photon mass before mixing with the SM fields $m_{D,0}$.

The combined potential for the SM and the dark Higgs fields is [33]

$$V_0(H,S) = -\mu^2|H|^2 + \lambda|H|^4 - \mu_S^2|S|^2 + \lambda_S|S|^4 + \kappa|S|^2|H|^2 \tag{2.2.5}$$

with the SM Higgs doublet H, the SM-singlet dark Higgs S and the parameter κ, which describes the mixing between the dark and the SM Higgs fields.

The existence of the kinetic mixing couplings ε and κ opens up two "interaction portals" through which the leading couplings between the hidden sector and the SM can be explored. In both descriptions, the phenomenology is determined by m_{Z_D}, m_s, ε, and κ, which are all treated as free parameters. The event kinematics and topology, however, are almost exclusively characterized by m_{Z_D} and ε, which will be demonstrated in the following.

- **The "hypercharge portal"** [33] kinetically mixes Z_D with the SM Z boson. If m_{Z_D} represents the lowest mass state in the dark sector, Z_D will exclusively decay to SM particles, with a sizable branching fraction to leptons. Fig. 2.4 (left) illustrates such a hypercharge portal decay. The branching fraction of the Z_D into leptons, $B(Z_D \to l^+l^-)$, generally depends on m_{Z_D}, as is shown in Fig. 2.5 (left). It is larger than 10% for $m_h < m_{Z_D} < m_h/2$.

  The Z_D lifetime is proportional to ε^{-2}. This means that the Z_D lifetime can become macroscopic if ε is sufficiently small. Fig. 2.5 (right) shows the Z_D lifetime in units of c as a function of m_{Z_D} for different ε values and illustrates the promise of Z_D searches over volumes covered by present-day collider experiments (whose radii typically range from millimeters to meters).[12]

  Existing searches have put multiple constraints on dark photons that are directly produced in Drell-Yan interactions [32, 62]. However, they cover only a phase space of relatively large ε and m_{Z_D} (roughly $\varepsilon \gtrsim 10^{-3}$ and $m_{Z_D} \gtrsim 10\,\text{GeV}$). The phase space at smaller ε remains widely unexplored. However, it is unfeasible to access this phase space via hypercharge-portal interactions since the direct Z_D production cross-section scales with ε^{-2}. The so-called Higgs portal, on the other hand, offers new possibilities to explore this region of phase space.

- **The "Higgs portal"** [33] kinetically mixes the SM Higgs with the dark Higgs, with the latter decaying to dark photons and eventually to final-state leptons when $m_{Z_D} < m_h/2$. The process $h \to Z_D Z_D$, which is illustrated in Fig. 2.4 (right), is possible because the h–s mixing generates non-zero coupling hZ_DZ_D.

  The following discussion assumes the dark Higgs to be decoupled ($m_s \gg m_h$), thus kinematically forbidding the process $h \to ss$. This common assumption avoids the strong constraints from direct searches.[13]

  The partial width to the lowest order in κ is

$$\Gamma(h \to Z_D Z_D) = (\kappa')^2 \frac{1}{32\pi} \frac{v^2}{m_h} \sqrt{1 - \frac{4m_{Z_D}^2}{m_h^2}} \frac{(m_h^2 + 2m_{Z_D}^2)^2 - 8m_{Z_D}^2(m_h^2 - m_{Z_D}^2)}{m_h^4} \tag{2.2.6}$$

[12] For example, a model configuration with $m_{Z_D} = 20\,\text{GeV}$ and $\varepsilon = (10^{-6}, 10^{-7}, 10^{-8})$ produces dark photons with lifetimes $c\tau_{Z_D} = (5.4 \times 10^{-2}\,\text{cm}, 5.4\,\text{cm}, 5.4 \times 10^2\,\text{cm})$.

[13] For a discussion of the several possibilities, see *e. g.* [63].

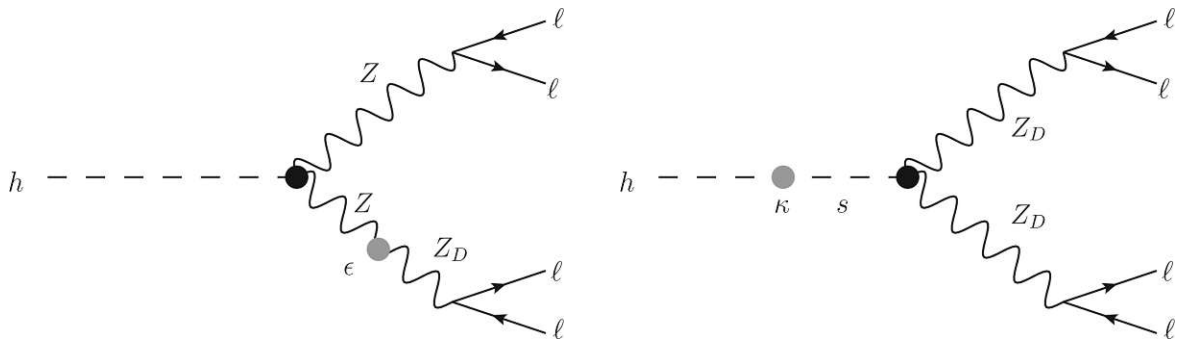


FIGURE 2.4: *Feynman diagrams of the decay of the exotic Higgs boson s to four final-state leptons via intermediate dark photons Z_D in the HAHM model.* Left: $h \to Z_D Z^{(*)} \to 4\ell$ *in the hypercharge portal.* Right: $h \to Z_D Z_D \to 4\ell$ *in the Higgs portal. Figures taken from [33].*

where $v \approx 246\,\text{GeV}$ and κ' is defined as

$$\kappa' = \kappa \frac{m_h^2}{|m_h^2 - m_s^2|} \tag{2.2.7}$$

The introduction of the dimensionless parameter κ' is convenient because it encapsulates the dependence on m_s and κ in a single parameter that controls the rate of the exotic decay of the SM Higgs boson. As a consequence, the branching fraction $B(h \to Z_D Z_D)$ is proportional to $(\kappa')^2$ (and thus also to κ^2) for a given m_{Z_D}. Plausible values for κ' are of the order of $\kappa' \approx 10^{-2}$, which corresponds to $B(h \to Z_D Z_D) = 1\%$, assuming $m_{Z_D} = 10\,\text{GeV}$, and is compatible with the existing measurements of the Higgs boson properties.

An important consequence of Eq. 2.2.6 is that $B(h \to Z_D Z_D \to 4\ell)$ does not depend on ε, as that parameter only determines the *decay length* of Z_D. The Higgs-portal production of dark photons is, therefore, the most sensitive search for the $\text{U}(1)_D$ hidden sector at $m_{Z_D} > \mathcal{O}(1\,\text{GeV})$ in the long-lived regime, $\varepsilon < \mathcal{O}(10^{-6})$.

Presently, LHC experiments are the only probe for dark photons that can explore ε values not disfavored by current precision electroweak tests [33]. However, such dark photon searches are challenging with these detectors. So far, there is only one other experimental constraint on the existence of dark photons in this domain set by the ATLAS search for displaced dimuons [64].

In conclusion, the dark photon model adds a single $\text{U}(1)_D$ gauge field and a scalar field, together with spontaneous symmetry breaking in the hidden sector. This relatively minimal extension of the SM can give rise to distinctive collider experiment signatures involving muon pairs originating from vertices that are significantly displaced from the primary interaction point.

Benchmark model featuring generic $H \to XX$ production

The collision energies provided by the LHC not only allow for the Higgs-portal dark photon production in decays of the 125-GeV Higgs boson but also heavier mediators decaying to LLPs. A wider spectrum of signal kinematic properties and topologies than described in the HAHM setting can be explored in the context of an additional benchmark model.

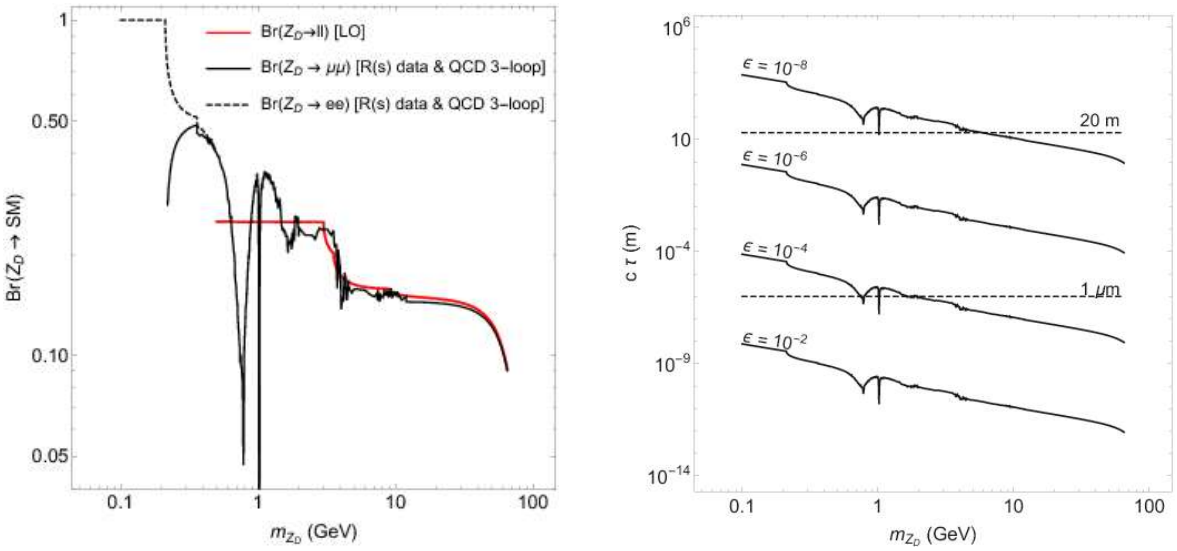


FIGURE 2.5: Left: *Branching ratio of Z_D to lepton final-states ($\mu^+\mu^-$, e^+e^-, l^+l^-) as a function of m_{Z_D}.* Right: *Decay length of Z_D as a function of m_{Z_D} for different values of ε. Figures taken from [33].*

This generic benchmark model [35] features a new and heavy scalar boson H that is produced in gluon-gluon fusion, and which decays into exotic spin-0 bosons X. These scalars X, in turn, decay into dileptons with a non-zero branching fraction. The full process, $gg \to H \to XX$, $X \to \ell^+\ell^-$, is a generalization of the decay depicted in Fig. 2.4 (right).

An example of a previous study interpreting its results in the context of this BSM Higgs benchmark model is the CMS search for displaced leptons at $\sqrt{s} = 8\,\text{TeV}$ [65, 66].

The displaced dimuon search presented in this work will use the Higgs-portal production of dark photons in the HAHM setting, $h \to Z_D Z_D$, as well as the generic $H \to XX$ production benchmark model for the interpretation of the results. The signal samples simulated for the search are designed to cover a wide spectrum of LLP masses and lifetimes to maximize discovery potential across the full available phase space. All the necessary details about the simulated signal samples will be given in later chapters when they are needed.

The CMS experiment at the CERN–LHC

The experimental foundation of the present work is the Compact Muon Solenoid Experiment (CMS) [67, 68]. It is one of the four major experiments at the CERN Large Hadron Collider [4–7], alongside the ATLAS [69], ALICE [70], and LHCb [71] experiments.

3.1 The Large Hadron Collider and its pre-accelerator chain

The Large Hadron Collider (LHC), located at the France-Switzerland border near Geneva and operated by the European Organization for Nuclear Research (CERN), is the largest and most powerful particle accelerator, storage ring, and collider to date. It was commissioned in 2008 and reported the first particle collisions in 2009, after more than a decade of planning and construction. The design principles, machine specifications, and operating parameters of the LHC are discussed in [4–7]. This section gives a birds-eye overview of the CERN accelerator complex, focusing on the LHC, and presents some of the highlights of these comprehensive references.

The LHC is housed by a circular tunnel of 26.7 km circumference, placed underground at depths ranging from 45 m to 170 m. The collider ring consists of two parallel and evacuated beam pipes in which counter-orbiting beams of charged particles (protons or heavy ions such as lead atoms) are accelerated, stored, and eventually made to collide in several interaction points.

The LHC is designed to reach energies of up to 7 TeV for protons and 2.76 TeV per nucleon for Pb ions. To first order, this connects with the *center-of-mass energy*, $\sqrt{s}$, of the collision of two identical particles via

$$\sqrt{s} = 2 \cdot E_{\text{beam}}, \tag{3.1.1}$$

which results in $\sqrt{s} = 14\,\text{TeV}$ for proton-proton collisions in the LHC [72].

However, the LHC cannot accelerate particles over the full energy range ($0 \rightarrow 7\,\text{TeV}$) but requires them to have a minimum energy of 450 GeV. Therefore, it relies on a chain of pre-accelerators. [6] The full injector sequence is shown in Fig. 3.1. It consists of the linear accelerator LINAC2[1] , the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS). The peak energies that each of these pre-

[1] The LINAC2 was eventually replaced by the more powerful LINAC4 during the LHC "Long Shutdown 2" (2019–2021).

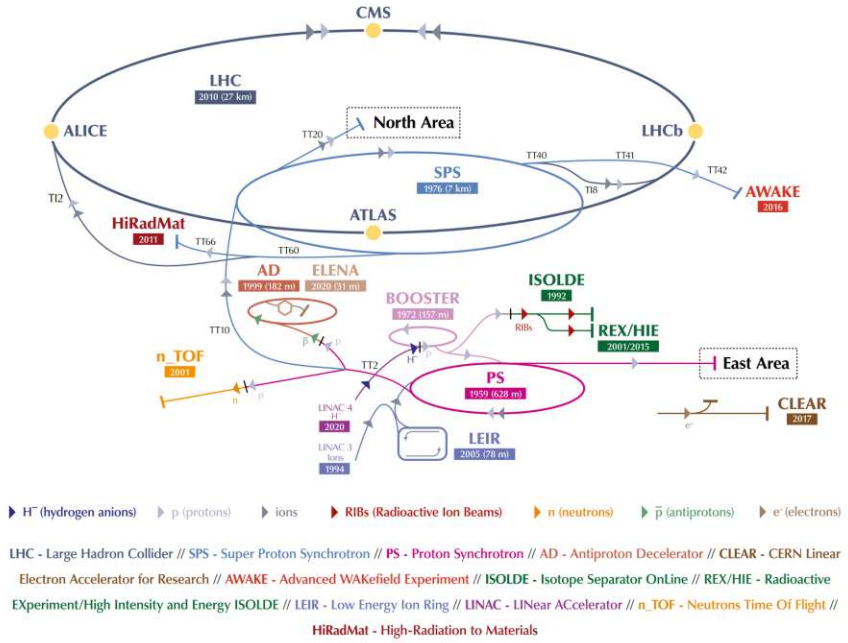


FIGURE 3.1: *The CERN accelerator complex and a selection of key facilities and experiments. Figure adapted from [73].*

accelerators reaches are 50 MeV, 1.4 GeV, 25 GeV and 450 GeV, respectively. In the LHC, particles are further accelerated to the maximum LHC energy.

The acceleration happens in eight radio-frequency (RF) cavities which operate at 2 MV per cavity at a frequency of $f_{\mathrm{RF}} = 400.789\,\mathrm{MHz}$ [7], resulting in a maximum energy gain of about 0.5 MeV per proton and per revolution.

Due to the constant gating frequency of the accelerating RF fields, defined by f_{RF}, particles in a beam are organized into longitudinally-spaced intervals, called *buckets*. The maximum number of buckets is determined by the harmonic number $h = f_{\mathrm{RF}}/f_{\mathrm{rev}} \approx 35640$, with f_{rev} being the revolution frequency [7]. A bucket containing a fill of particles is referred to as a particle *bunch*, but buckets can also be empty. The exact bunch configurations are realized in accordance with the injection cycles of the LHC pre-accelerators and optimized to both control beam interaction rates and comply with the technical limitations imposed, among others, by the heat-absorbing capacity of the cryogenic system.

The cryogenic system is responsible for cooling the superconducting magnets along the LHC beamline, which keep the charged beams in their circular orbits. 1232 dipole magnets, operated at temperatures below 2 K using superfluid Helium as coolant, provide a flux density of more than 8 T to bend the particle trajectories. Additional arrays of multipole magnets are responsible for focusing the beam, thus counter-acting the electromagnetic repulsion within the beam and preventing interactions of the beam particles with the confines of the vacuum chamber. Once the beams have reached the desired collision energy, dedicated arrays of focusing magnets at the interaction points are used to compress them to a diameter of about 10 µm, where a fraction of all beam particles collides in each revolution.

The most important figure of merit for colliders like the LHC is the *instantaneous luminosity*,

$$L = \frac{k_b N_b^2 f_{\mathrm{rev}} \gamma}{4\pi \varepsilon_n \beta^*}, \tag{3.1.2}$$

with k_b... number of bunches per ring, N_b... number of protons per bunch, γ... relativistic

gamma factor ($\gamma = E/m$), f_{rev}... revolution frequency, ε_n... normalized RMS transverse beam emittance, β^*... beta function at the interaction point. [6]

Maximizing L will maximize the rate of collision events, $\mathrm{d}N/\,\mathrm{d}t$, for a process of given cross-section σ as these quantities are directly proportional:

$$\frac{\mathrm{d}N}{\mathrm{d}t} = L \cdot \sigma \ . \tag{3.1.3}$$

A common measure for the total amount of recorded data is the *integrated luminosity*, $\mathcal{L} = \int L \,\mathrm{d}t$, which can be used to express the total expected number of events, N, for a given physical process with cross-section σ:

$$N = \int \frac{\mathrm{d}N}{\mathrm{d}t}\,\mathrm{d}t = \sigma \int L \,\mathrm{d}t = \sigma \cdot \mathcal{L} \tag{3.1.4}$$

In practice, due to a multitude of technical limitations, the instantaneous luminosity (and, consequently, also the integrated luminosity) cannot be pushed to arbitrarily large values. Examples of relevant effects acting against luminosity gain are highlighted in the following non-exhaustive list (further details can be found in [6, 7]):

- The *beam emittance*, a generic measure for the average spread of beam particle coordinates in the phase space of position and momentum, has to fit into the small aperture of the LHC magnets.

- As the beams cross each other in the collision region, *beam-beam interactions* cause non-linear perturbations of the charge distributions in the respective beams and are ultimately responsible for a spread in betatron tunes, an effect which has to be kept below a certain limit. In other words, beam-beam interactions limit the maximally attainable particle density per bunch.

- The maximum dipole field is constrained by the magnet quench limits: While the nominal dipole field is 8.33 T, the actually achievable field depends on the heat load and temperature margins inside the superconducting magnets and the cryogenic system in general and, therefore, ultimately on the beam losses during operation.

- The total intensity, $k_b \cdot N_b$, is limited by the capacity of the cryogenic system to absorb the thermal energy produced by the synchrotron radiation.

- The space-charge limit in the injectors scales with N_b/ε_n.

- Both during nominal running operations and in the case of an emergency, the beam dumping system must be able to safely and quickly absorb the particle beams – its technical limitations are another factor influencing the maximally attainable beam energies.

Throughout the LHC *Run 2* (2016–2018), the accelerator provided collisions at center-of-mass energies of $\sqrt{s} = 13\,\text{TeV}$. The typical filling scheme was 2040 bunches of $1.15 \cdot 10^{11}$ particles each and an inter-bunch spacing of 25 ns. This resulted in a total integrated luminosity of $160\,\text{fb}^{-1}$ that was delivered to the LHC experiments such as CMS (see Fig. 3.2).

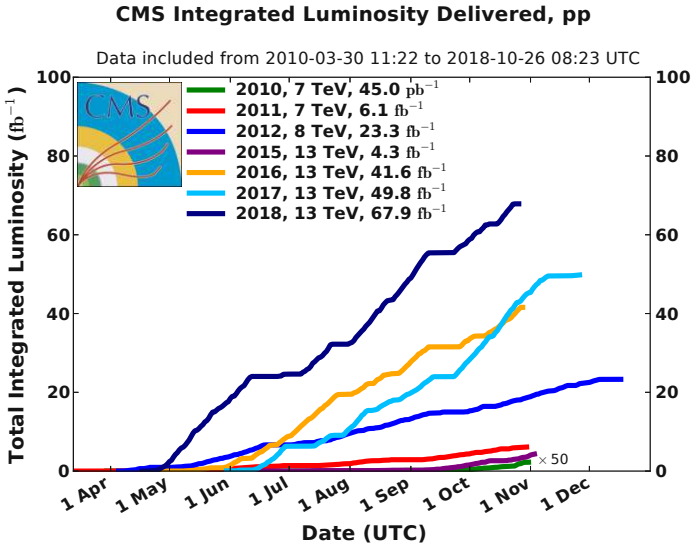


FIGURE 3.2: *Total integrated luminosity provided by the LHC to the CMS experiment during stable operations since the beginning of Run 1 in 2016. The cumulative distributions corresponding to the Run-2 data-taking period are depicted in orange, light blue, and dark blue colors for 2016, 2017, and 2018, respectively. Figure taken from [74].*

3.2 The Compact Muon Solenoid experiment (CMS)

The Compact Muon Solenoid Experiment (CMS) [67, 68] is a multi-purpose detector dedicated to mainly recording proton-proton collisions in the LHC at high instantaneous luminosities and record energies of up to $\sqrt{s} = 14\,\text{TeV}$. It is located at the LHC Interaction Point 5 (IP5), near the French city Cessy and between Lake Geneva and the Jura mountains, about 100 m underground.

The CMS detector is designed to record and measure all particles that are stable enough to traverse the detector volume without decaying.[2] However, thanks to the precise measurement of track properties, like energy and momentum, and the accurate spatial resolution, intermediate and unstable particles can be reconstructed as well, thus enabling the study of a wealth of physics processes.

The local CMS coordinate system (x, y, z) has its origin at the very center of the detector, where the nominal collision point of the two LHC beams lies. The x-axis is defined to point towards the LHC center and the y-axis to point upwards. The plane spanned by x and y axes is called the *transverse plane*. In this plane, the azimuthal angle ϕ is zero in the x-direction. The z-axis is perpendicular to the transverse plane and points along the beam in the direction defined by the *right-hand rule* (*i. e.*, it follows the beam that is counter-rotating when looking at the LHC from an aerial perspective). Instead of the polar angle θ, the pseudo-rapidity η is usually used. Differences in η are Lorentz-invariant and, therefore, this quantity is not dependent on longitudinal boosts of particles – a circumstance advantageous for describing hadron processes, where the longitudinal momenta of the colliding partons

[2] Neutrinos are an exception as they cannot be measured effectively due to their small interaction cross-section.

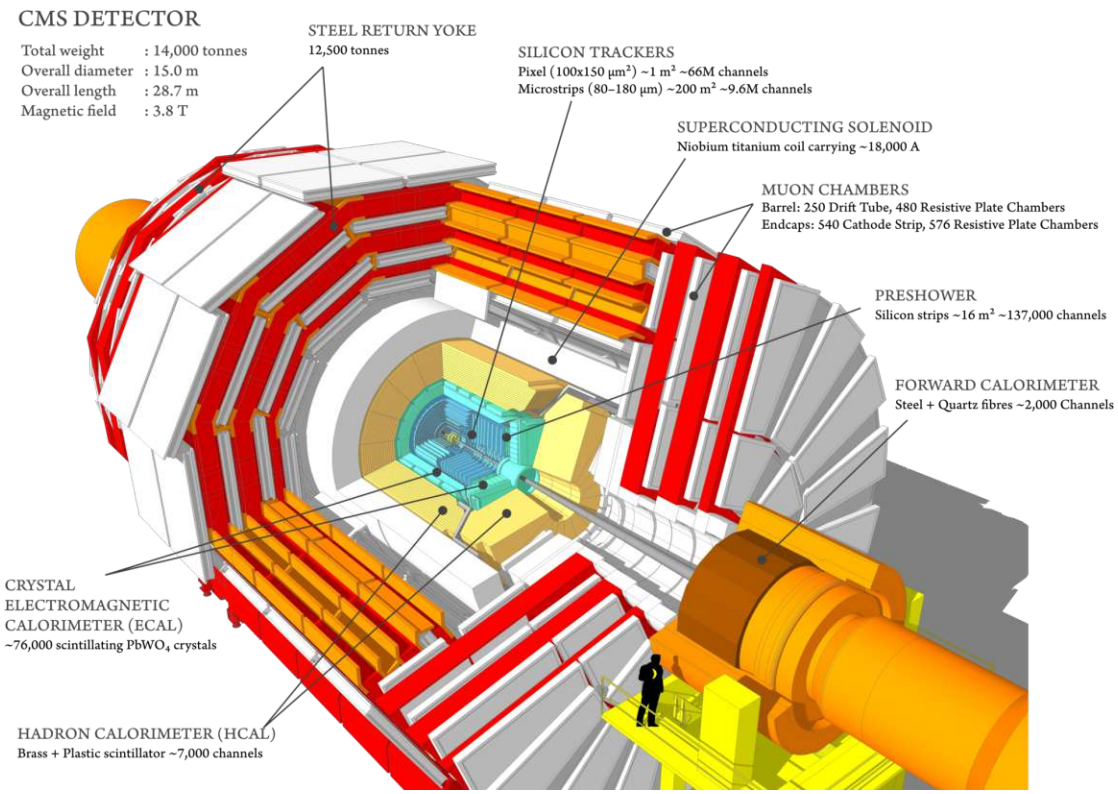


FIGURE 3.3: *Illustrative rendering of the CMS detector and its sub-detectors (banana for scale). Figure adapted from [75].*

generally cannot be known precisely enough. The pseudorapidity is defined as

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] = \frac{1}{2}\cdot\ln\left(\frac{p+p_z}{p-p_z}\right) = \operatorname{arctanh}\left(\frac{p_z}{p}\right), \tag{3.2.1}$$

with p... absolute momentum, p_z... momentum along the z direction.

The overall detector design resembles cylindrical and concentric layers around the primary interaction point (IP) of the colliding beams, oriented along the LHC beamline. CMS has a length of 21.6 m, a diameter of 14.6 m, and weighs about 14 kt. Its inner parts are suffused by a strong solenoidal magnetic field.

Fig. 3.3 shows a schematic rendering of the CMS detector and its several sub-detectors, each dedicated to the measurement of specific types of particles. The main components are described in the following sections and loosely follow [67, 68].

3.2.1 Solenoid magnet

One of the central components of the CMS detector is the solenoidal magnet [76–78] providing a magnetic field of 3.8 T in a 220-t cold bore of 6.3 m diameter and 12.5 m length. At the nominal current of 19.14 kA, the total energy stored in this volume is 2.6 GJ. The magnet uses a 4-layered winding of stabilized reinforced niobium-titanium (Nb-Ti) wires and is operated at a superconducting temperature of 4.6 K. Among the specifications that make the CMS solenoid unique and its design and construction technically challenging are its large geometry while the radial extent of the coil is relatively small ($\Delta R/R \approx 0.1$). The resulting ratio between stored energy and cold mass ($\approx 11.6\,\mathrm{kJ/kg}$) is large enough to result

in considerable mechanical deformations of the material (0.15%) during energizing, which the system had to be designed to withstand.

The magnetic flux is returned through a yoke that features five wheels and two endcaps (composed of three disks each). This return yoke has an outer diameter of the iron flats of 14 m and a barrel length of 13 m. The thicknesses of the respective iron layers in the barrel are 30 cm, 63 cm, and 63 cm and amount to a total mass of 6 kt in the barrel. In the endcaps, the yoke iron disks have thicknesses of 25 cm, 60 cm, 60 cm, respectively. The mass of iron in the endcap is 2 kt. In total, the mass of iron in the entire return yoke amounts to 10 kt.

The solenoid also serves as a hadron absorber that filters the particle flux traversing the inner tracking system (see Sec. 3.2.2) and the calorimeters (see Sec. 3.2.3) before it reaches the outer muon system (see Sec. 3.2.4).

3.2.2 Tracking system

The CMS Tracker [79, 80] is a detector system dedicated to the precise measurement of the transit of charged particles originating from collisions at and near the LHC beam spot. The innermost part of the tracking system consists of silicon pixel detectors and is referred to as the *pixel detector*. At the beginning of the LHC Run 2 in 2016, these sensors were arranged in three layers in the barrel region (at radii $r = 4.4\,\text{cm}$, $7.3\,\text{cm}$, $10.2\,\text{cm}$) and two layers in each endcap (at $z = \pm 34.5\,\text{cm}$ and $z = \pm 46.5\,\text{cm}$). After the 2016 data-taking period, the pixel detector was upgraded to feature four barrel layers and three layers in each endcap for improved tracking performance. [81] The innermost layer is located as close to the LHC beam pipe as possible to maximize the detector potential to resolve large numbers of tracks near the primary IP. The individual pixel sensors have a size of $100 \times 150\,\mu\text{m}^2$ and provide a granularity fine enough to cope with the high particle-flux densities expected during typical LHC running conditions.

In the outer parts of the tracking system, particle-flux densities are smaller, and silicon microstrip sensors provide adequate hit resolution in these detector regions. These modules are organized in four distinct zones. The four concentric detector layers in the Tracker Inner Barrel (TIB) are located at $r = 255.0\,\text{mm}$, $339.0\,\text{mm}$, $418.5\,\text{mm}$, and $498.0\,\text{mm}$ from beam axis and span from $z = -700\,\text{mm}$ to $z = +700\,\text{mm}$. The two inner TIB layers host double-sided detector modules with $80\,\mu\text{m}$ strip pitch, whereas the outer two layers feature single-sided modules with $120\,\mu\text{m}$ strip pitch. The resulting single-point resolution is 23–34 µm in transverse direction and 23 µm longitudinally.

The Tracker Inner Disk (TID) constitutes three disks of detector modules placed between $z = \pm 800\,\text{mm}$ and $z = \pm 900\,\text{mm}$ on either side. Each disk consists of three rings spanning radii of $200\,\text{mm} \lesssim r \lesssim 500\,\text{mm}$. The inner two rings hold back-to-back modules, while the outer ring holds single-sided modules. Their strip pitches vary between 80 µm and 140 µm.

The Tracker Outer Barrel (TOB) and the Tracker Endcaps (TEC) define the outer parts of the tracking system. Their silicon strip detectors are respectively arranged in six concentric and nine transverse layers. The cylindrical TOB layers are located within $555\,\text{mm} < r < 1160\,\text{mm}$ and $|z| < +1180\,\text{mm}$, while the TEC layers on each side are located between $\pm 1240\,\text{mm} < z < \pm 2800\,\text{mm}$ and have radial extents from 220 mm to 1135 mm.

Fig. 3.4 visualizes the tracker system geometry and the four tracker regions.

The total material budget of the tracker should be kept at a minimum to keep energy losses of particles small when they traverse the tracker volume. But it is ultimately determined by the silicon detectors themselves as well as the readout electronics and the support

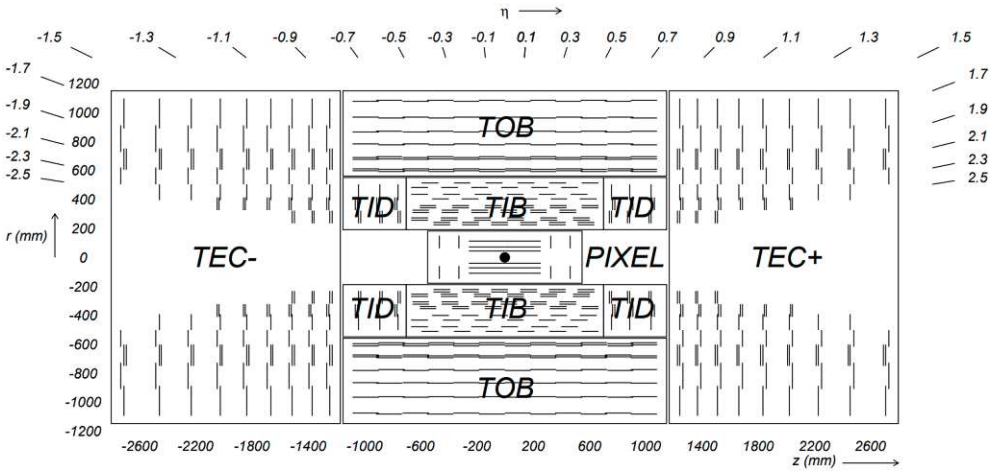


FIGURE 3.4: *Schematic cross-section of the CMS tracker. Detector modules are indicated by lines. (Double lines represent back-to-back modules.) The names of the various sub-detectors are overlaid on the respective tracker regions. Figure taken from [67].*

and cooling structures. The tracker material budget is not isotropically distributed around the IP but varies in different pseudo-rapidity intervals from 0.4 to 0.8 radiation-length equivalents.

Charged particles traversing tracking system sensors induce electron-hole pairs in the semiconductors, which can subsequently be recorded as a particle *hit* in a given sensor. Multiple hits across the tracker layers serve as input for the reconstruction of the full particle trajectory – called a particle *track* – employing dedicated pattern recognition and fitting algorithms. Thanks to the presence of the magnetic field, which bends the trajectories of charged particles, the energy and momentum as well as the electric charges of particle tracks can be reconstructed highly accurately.

Most of the technical challenges for the design of the CMS tracker stem from the harsh collision environment imposed by the LHC run conditions. The detector and readout electronics have to operate reliably at frequencies of 40 MHz (corresponding to a 25-ns bunch-crossing interval). This puts stringent time constraints on the track reconstruction, and thus limits the computational complexity of the involved algorithms. Moreover, all tracker components have to endure radiation-heavy environments for the full time span of operation.

Efficiency and precision of the track reconstruction were extensively studied, *e. g.*, with muons from Z boson decays. Tracking efficiencies of more than 99% were measured both in data and simulation with only a small dependence on collision pile-up. [82] Measurements with charged pions from $D^{\pm}$ decays provided an alternative performance measurement and showed a good data-simulation agreement of the tracking efficiency [83]. The primary vertex resolution was measured to be less than 15 µm and 20 µm in the transverse and the longitudinal directions, respectively. [84]

3.2.3 Calorimeters

Outside the tracking system, mostly within the confines of the solenoid, lies the CMS calorimeter system. This detector array is designed to stop all particles except muons and neutrinos to measure their total energy. The accurate measurement of particle energy

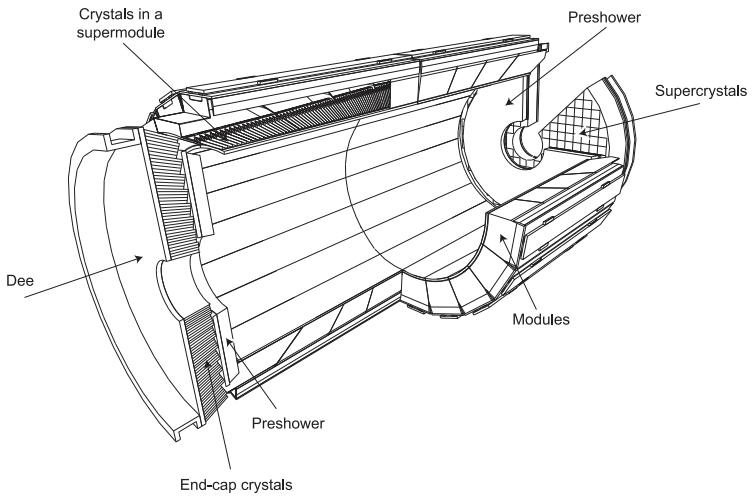


FIGURE 3.5: *Schematic drawing of the CMS Electromagnetic Calorimeter (ECAL). Figure taken from [67].*

requires high granularity of its components. The calorimeter system is divided into two subdetectors, each tailored to the measurement of specific types of particles.

Electromagnetic calorimeter

The CMS Electromagnetic Calorimeter (ECAL) [85] measures charged particles (mostly electrons and positrons) and photons.

The ECAL is subdivided into a cylindrical barrel part (EB) and two endcap parts (EE). The EB has inner and outer radii of $r_{\text{in}} = 124\,\text{cm}$ and $r_{\text{out}} = 175\,\text{cm}$, respectively, and extends up to $|\eta| < 1.479$. The EE is located at $z = \pm 314\,\text{cm}$ and covers $1.479 < |\eta| < 3.0$. Precision measurements are possible up to $|\eta| < 2.6$, whereas the remaining coverage helps to augment the measurement of energy flow in the forward direction. Fig. 3.5 shows an illustration of the ECAL.

The used detector technology is scintillating lead-tungstate ($PbWO_4$) crystals. [85] These crystals have a high density, a short radiation length, and a small Molière radius. This allows the crystals to be relatively small, resulting in an overall compact and radiation-resistant calorimeter. In the EB, each of them has a front-facing area of $2.2\,\text{cm} \times 2.2\,\text{cm}$ and a length of 23 cm. The crystal length is determined by the requirement of a total thickness of about 26 radiation lengths at $|\eta| = 0$, which limits the longitudinal shower leakage of highly-energetic electromagnetic showers. Crystals in the EE are shaped slightly differently, with front cross-sections of $2.86\,\text{cm} \times 2.86\,\text{cm}$ and lengths of 22 cm. To compensate for the coarser granularity in the EE and improve its spatial resolution of the energy measurement, an additional ECAL Preshower Detector (ES) is installed in front of the endcap calorimeters, consisting of two silicon microstrip modules and two lead absorber planes.

The ECAL features 61200 $PbWO_4$ crystals (360-fold in ϕ and $(2 \cdot 85)$-fold in η) in the barrel part and 7324 crystals in each of the endcap calorimeters. The total EB crystal volume is $8.14\,\text{m}^3$ and amounts to a weight of 67.4 t. The EE crystals are $2.90\,\text{m}^3$ and weigh 24.0 t. [67]

The optical signal induced by particles hitting the scintillators is converted into an electrical signal by photodetectors, which are attached to each crystal, and in further consequence digitized for signal processing purposes. The photodetector technologies used are avalanche photodiodes (APD) in the EB and vacuum phototriodes (VPT) in the endcaps.

Hadron calorimeter

The CMS Hadron Calorimeter (HCAL) [86] measures strongly-interacting particles (such as protons, neutrons, pions, and kaons) that pass through the ECAL. The HCAL practically covers the full solid angle and can be considered *hermetic* in that sense. This makes the HCAL particularly important for the measurement of missing transverse energy (MET), *e. g.*, as a result of decays involving neutrinos or exotic particles.

The HCAL is a sampling detector featuring alternating layers of plastic scintillators and non-magnetic brass absorbers. Traversing particles can produce hadronic showers in the brass layers, which lead to measurable scintillation light in the subsequent scintillator layers.

The HCAL is largely located between the ECAL (outer edge at $r = 1.77\,\mathrm{m}$) and the solenoid magnet (inner edge at $r = 2.95\,\mathrm{m}$). Fig. 3.6 shows a schematic cross-section of the sub-detector. It is divided into several geometric regions.

The HCAL Barrel (HB) constitutes the innermost part of the HCAL system with a radial extent from $r \simeq 1.2\,\mathrm{m}$ to $r \simeq 3\,\mathrm{m}$, a total length of about 9 m, providing coverage of $|\eta| < 1.3$ [67] It contains 2304 towers[3] with a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$.

The HCAL Outer (HO) wraps around the HB. It matches the ϕ segmentation of the muon system (which is the outermost CMS subdetector) and provides coverage up to $|\eta| < 1.26$. The HO is placed outside the solenoid. It complements the HB and measures the tails of energetic hadron showers that pass through both the HB and the magnet system.

The HCAL Endcap (HE) subdetector covers $1.3 < |\eta| < 3.0$ and contains 2304 towers with sizes depending on the $|\eta|$ region (about 0.087 in $\Delta\eta$ and 5°–10° in ϕ).

Finally, the HCAL Forward (HF) is placed outside the muon system endcaps on either side of the detector (at $z = \pm 11.2\,\mathrm{m}$), where it provides coverage for $3.0 < |\eta| < 5.0$. Its length on either side is 165 cm. One particular technical challenge for the HF is radiation hardness. Each pp collision deposits an average energy of 760 GeV in the HF, compared to 100 GeV in the rest of the calorimeter system. The active medium chosen to withstand this radiation dose is quartz fibers (polymer cladding with a fused-silica core), which are expected to endure in this harsh environment for at least a decade.

3.2.4 Muon system

Muons are favorable final-state particles for precision measurements of certain quantities since – compared to other particles like electrons – they retain most of their energy when traversing the inner subdetectors. Due to the general discovery potential of decays featuring muons, the CMS detector was designed to identify and measure muons efficiently.

The CMS muon system [67, 87] comprises the outermost layers of the detector. Fig. 3.3 and 3.6 show its placement and orientation with respect to the other detector components.

Three main detector technologies are deployed in the muon system: drift tube (DT) chambers, cathode strip chambers (CSCs), and resistive plate chambers (RPCs). Similar to the other subsystems, the muon system is divided into a central, cylindrical barrel section and two planar endcap regions. The combined coverage of the muon system is $|\eta| < 2.4$ without any gaps in acceptance. This allows for high reconstruction efficiency of promptly produced muons for the full η range.[4]

[3] In the context of the calorimeters, a "tower" refers to the deposited energy in a given detector region summed over many layers of tiles in radial depth.

[4] The muon reconstruction efficiency drops at $|\eta| = 0.25$ and $|\eta| = 0.8$ (regions between two DT wheels) and at $|\eta| = 1.2$ (transition region between barrel and endcap detectors), both of which are known ef-

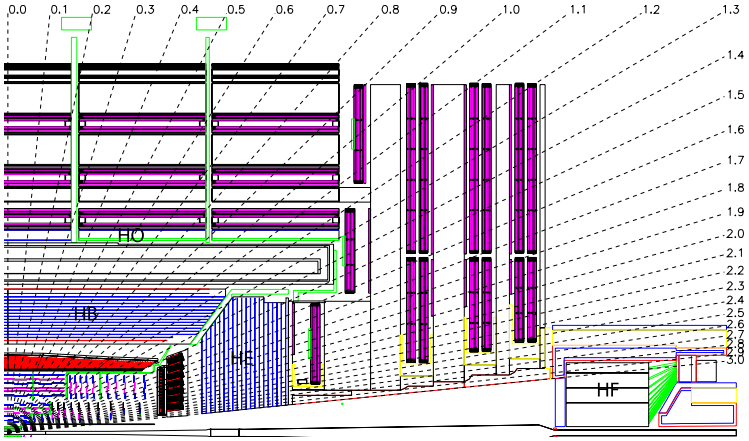


FIGURE 3.6: *Schematic cross-section of the CMS Hadron Calorimeter (HCAL) and the muon system (purple color) in the r–z plane. The HCAL sub-modules (HB, HE, HO, and HF) are indicated by their overlaid acronyms. Figure taken from [67].*

In addition to the identification and measurement of muons, the muon system provides the capability of triggering on the p_T of muons. (See Sec. 3.2.5 for a description of the CMS trigger.)

The following sections describe the muon barrel and muon endcap detectors in more detail and contextualize some of their relevant design parameters.

Muon barrel

The muon barrel region spans $|\eta| < 1.2$ and is suffused with a uniform magnetic field of 3.8 T. In this region, the rate of collision muons is relatively low (due to small neutron-induced background). In this environment, it is feasible to use drift chambers with standard rectangular drift cells for muon tracking purposes.

The four muon barrel layers ("stations") are interspersed among the flux return yoke layers and house 60 drift tube (DT) chambers in each of the three inner stations and 70 DT chambers in the outermost station.

The smallest independent design unit of a DT chamber is a so-called superlayer (SL), which consists of four layers of rectangular drift cells. In the outer two SLs, the charge-collecting wires run parallel to the beamline and thus measure tracks in the magnetic bending plane (r, ϕ). The inner SL wires are perpendicular to the beamline and provide measurements of the muon track z position along the beamline.

A drift chamber consists of three[5] SLs. Each chamber has its drift cells offset by a half-cell width relative to neighboring cells to remove any dead spots. During muon track reconstruction, hit information from multiple stations is combined (if possible) for improved measurement and minimized background noise.

The tube cross-section of DT chambers is 13 mm × 42 mm and was optimized under constraints like mechanical stability, limited space within the muon system, and redundancy requirements.

Muons with $p_T > 40$ GeV are often accompanied by electromagnetic cascades. The DT chambers were designed to cope with such additional cascades as well as with uncorrelated

fects. [87]

[5]The fourth station misses the z-measuring layer and thus consists of only *two* SLs.

background hits from neutrons or photons. This is achieved by a combination of good tracking efficiency and measurement redundancy due to several layers of separated drift cells per station.

The DT chambers are capable of providing a muon time measurement with a resolution of a few nanoseconds, which, *i. a.*, can be used for triggering and beam-crossing-time identification. However, when the LHC reaches its full instantaneous luminosity, they are not quite sufficient to accurately measure the eventual background rates and the correct bunch-crossing time after all. Hence, an additional and complementary trigger system is installed in the muon barrel and in the endcap regions consisting of resistive plate chambers (RPCs). RPCs have fast response times (albeit with coarser position resolution than the DTs) and thus provide excellent bunch-crossing association, even in the presence of the high collision and background rates expected at the LHC. Additionally, RPCs help to resolve track reconstruction ambiguities when there are multiple hits in a DT chamber. In total, the barrel muon system embeds six RPC layers.

Finally, to optimize muon momentum resolution, the positions of the detectors in the muon system are measured with respect to each other and the inner tracker by a sophisticated alignment system.

Muon endcaps

The muon system endcaps are located on either side of the muon barrel and provide coverage for $0,9 < |\eta| < 2.4$. There are four stations in each endcap, positioned perpendicular to the beamline and embedded between the flux return plates. Between $0.9 < |\eta| < 1.2$, the muon system layers are aligned such that a muon from the IP traverses both barrel and endcap detectors. This η region, therefore, is referred to as *overlap region*, while the muon system at $|\eta| > 1.2$ is called the *endcap region*.

The muon endcap layers experience higher rates and background levels than the barrel layers and are suffused by a large and non-uniform magnetic field. Cathode strip chambers (CSCs) are the chosen detector technology for this collision environment, as these detectors provide fast response times, fine segmentation, and radiation resistance and can be operated in large and inhomogeneous magnetic fields.

CSCs are multiwire proportional chambers featuring six anode wire panes interleaved among seven cathode panels. Accurate measurement of the r coordinates of tracks is possible due to the azimuthal alignment of the wires. The track ϕ coordinates are obtained by interpolating the charges induced in strips, which run lengthwise at constant $\Delta\Phi$ widths.

CSCs are used both for precision measurements and high-efficiency muon triggering. Like in the muon barrel, the (inner three) endcap stations are supplemented with RPC layers to improve their performance further.

3.2.5 Trigger and data acquisition

During typical data-taking conditions, the LHC can provide more than a billion pp collisions per second in the CMS detector [67, 68]. Assuming a plausible event size of $\mathcal{O}(1\,\mathrm{MB})$ in CMS, this would result in a total data output rate of $\mathcal{O}(1\,\mathrm{PB/s})$. This is entirely infeasible for the CMS data acquisition (DAQ) system, which has a maximum rate of about 100 MB/s. Considering that most of the beam particle collisions in the interaction point are *not* the result of pp hard scattering, this makes most of the collision events unlikely to contain "interesting" physics (in the context of the CMS physics program).

The CMS trigger system [88, 89] constitutes an "event filter" which aims to select relevant events at a rate that is manageable for the DAQ pipeline. This two-tiered trigger consists of a hardware and a software trigger, called Level-1 Trigger (L1) and High-Level Trigger (HLT), respectively. At each level, the trigger implements a "trigger menu", which is a collection of algorithms for event triggering. Trigger algorithms at the L1 trigger (so-called *seeds*) are pre-filters for algorithms at HLT (so-called *paths*). At either trigger stage, an event is accepted if at least one of the seeds/paths fires (and otherwise discarded).

The L1 trigger is an array of field-programmable gate arrays (FPGAs) that implement up to 512 different seeds. Data recorded by the detector are buffered for about 3.2 µs before they get overwritten by newly-incoming data. Every L1 seed must reach a trigger decision within this short time frame – and, in fact, most seeds are designed to process trigger data in about 1 µs.

However, the stringent time constraints at L1 do not allow for arbitrarily complex trigger algorithms. Moreover, the L1 trigger uses only information from the muon system and the ECAL, with reduced granularity to construct *trigger primitives* for common particles, such as muons, electrons, or photons, as well as other detector objects, such as jets or local sums of transverse (missing) energy. The L1 trigger reduces the data-taking rate from 40 MHz (LHC rate) to about 100 kHz.

The HLT is a computing cluster that implements trigger algorithms in software. Thanks to the reduced data input rate – the HLT receives only those events selected by the L1 trigger – there is more time per event to process event data (about 100 ms). This allows for more sophisticated track reconstruction algorithms and the use of the full detector readout data at HLT, bringing the HLT reconstruction performance close to the offline reconstruction performance (described in Sec. 3.3.3). The HLT further reduces the data-taking rate to less than 1 kHz and about 1 GB/s, which are manageable levels for the DAQ system.

Several factors drive the interaction rate and thus the total rate of a given L1/HLT menu, such as instantaneous luminosity and pile-up. Both these LHC conditions vary within a given data-taking run or LHC fill as fewer and fewer particles are available over time due to preceding collisions and beam loss effects. To keep the interaction rate in the detector roughly constant over time, physics triggers are usually available in several versions (*main* and *backup* seeds), each tailored to specific intervals of instantaneous luminosity. Additionally, seeds/paths can be *prescaled* with a certain numerical factor depending on the current instantaneous luminosity. (For example, a prescale (PS) value of 2 means that every second event is rejected a priori.) Standard, multi-purpose CMS physics triggers, however, are usually unprescaled, *i. e.*, they will accept every triggered event.

3.3 Reconstruction of muons and other particles in CMS

Hit information and energy deposits are collected by all subdetectors in CMS. Different particles, given their respective charges and interactions, leave different signatures in the detector. Fig. 3.7 schematically shows examples of such distinct particle signatures.

All particles produced in beam collisions first traverse the tracker [79, 80]. If they are charged, they produce *hits* in the tracker layers, which are used to reconstruct their trajectories (*tracks*) and their points of origin (*vertices*). Charged particle tracks are bent

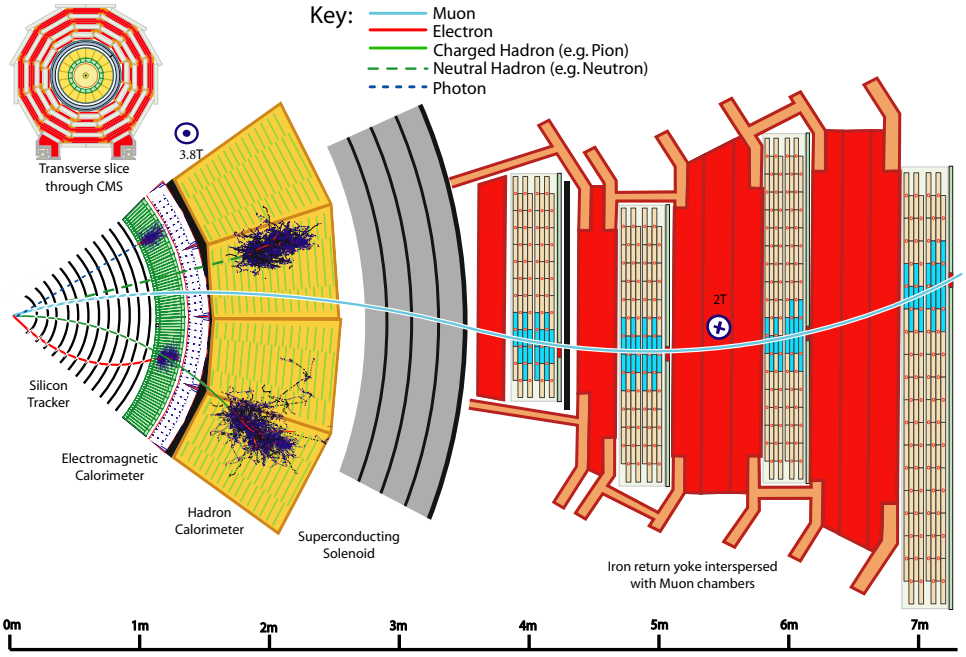


FIGURE 3.7: *Interactions of different particle types with the subdetectors of CMS (schematic). The distinct hit/shower patterns across the detector subsystems are combined in the Particle Flow (PF) algorithm to form hypotheses of stable particles. Figure taken from [90].*

by the magnetic field, which allows for a measurement of particle charge and momentum. Neutral particles traverse the tracker without producing any hits.

In the Electromagnetic Calorimeter (ECAL) [85], electrons and photons produce electromagnetic showers, which are detected as clusters of energy. These particles are fully absorbed in the ECAL volume, allowing for a measurement of their energy (including information about the direction of the energy flow).

Hadrons, both charged and neutral, traverse the tracker and the ECAL. While they might deposit some energy in these inner detectors, most of their energy is absorbed in the outer calorimeter, the Hadron Calorimeter (HCAL) [86]. The HCAL stops all particles (except muons and neutrinos) and measures their energies and directions through clustering of the induced hadronic showers. It is even possible, at least to some extent, to infer the type of particle that leads to a shower in the HCAL.

Muons deposit small amounts of energy in most of the subdetectors, but are the only particles that produce hits in the detector layers beyond the HCAL, *i. e.*, in the muon system. [67, 87]

Neutrinos have a very small interaction cross-section with matter. Therefore they traverse all detector layers and escape direct detection entirely.

The process of forming particle tracks from hit information – thus generating hypotheses of stable particles with estimated energies, momenta, and type – is called reconstruction of *physics objects*. All CMS subdetectors are finely segmented and granular, and the magnetic field is strong enough to separate energy deposits in the calorimeters. These properties allow for a kind of physics object reconstruction and global event description that is known as *particle flow* reconstruction.

3.3.1 Overview of Particle Flow reconstruction

The Particle Flow (PF) algorithm [90] processes information from all CMS subdetector systems and generates a list of final-state particles. Thus, PF provides a global event description that includes the following objects:

- Jets from either hadrons or photons: The jet energy measurement is inclusive in the calorimeters, whereby the identification of singular particles within a jet is not attempted. The reconstruction of jets and missing momentum ($p_{T\,\mathrm{miss}}$) only requires the calorimeters.
- Further, jets can be tagged as originating from hadronic τ decays and b-quark hadronization. This tagging utilizes information from the tracker to measure the properties of the charged particle track.
- Isolated electrons and photons, mainly measured by the ECAL.
- Muon identification based on information from the muon system.

Moreover, PF is capable of identifying particles from pile-up (PU) interactions and allows for PU mitigation methods.

In general, the PF reconstruction has three stages:

1. Generation of PF objects: This preprocessing step collects data from all subdetector systems and builds tracks of charged particles (from tracker information), clusters of energy deposits (from calorimeter information), and tracks from hit information in the muon system
2. Correlation of PF objects from these three detector regions.
3. Inference of particle hypotheses and computation of derived objects, such as jets, MET, primary and secondary vertices etc.

The following sections detail the reconstruction of muons at the various stages of data taking. The technicalities of the reconstruction of other particle types are not critical for the work at hand, and detailed descriptions can be found in [90].

3.3.2 Online muon reconstruction

The online muon reconstruction refers to the measurement of muon candidate objects at the different CMS trigger stages, *i. e.*, Level-1 Trigger (L1) and High-Level Trigger (HLT) [88, 89, 91, 92].

L1 muon objects

At the L1, muon candidates are reconstructed using information from the muon system only. Different track finding algorithms are deployed in different geometrical detector regions: The barrel muon track finder (BMTF) uses hits or trigger primitives (TPs) (*i. e.*, short track segments) in the barrel muon stations (DT and RPC detectors) and provides muon candidates in $0 < |\eta| < 0.83$. The overlap muon track finder (OMTF) produces muon candidates in $0.83 < |\eta| < 1.24$, where DT, RPC, and CSC detectors provide the hit

information. Finally, the endcap muon track finder (EMTF) obtains information only from CSC detectors to form muon candidates in $1.24 < |\eta| < 2.4$.

The TPs carry rudimentary position coordinates (θ and ϕ) as well information about muon direction and timing (with respect to a collision bunch crossing). Due to the magnetic field, the position coordinates – mainly ϕ – differ between consecutive stations. From this, an angular deflection $\Delta\phi$ and thus a track $p_T^{(\mathrm{L1})}$ value can be deduced.

Ideally, the reconstruction at L1 would use algorithms that reproduce the track parameters as closely as possible to their true values. Such algorithms are generally used at the latest reconstruction stage (the so-called *offline* reconstruction, see Sec. 3.3.3). However, since the L1 trigger has to operate under stringent time requirements, the computational complexity of the L1 reconstruction algorithms is limited and, generally, less accurate than the offline algorithmic counterparts. Still, the performance of the L1 muon reconstruction for prompt muons in Run 2 leads to L1 efficiencies of about 93–97% (depending on the applied muon quality criteria) and is independent of muon p_T for a wide range of p_T. [91]

A crucial property of the L1 muon reconstruction algorithms utilized throughout Run 2 is the fact that they assume each muon track to originate in the very center of the CMS detector, *i. e.*, in the point where the LHC beams collide. This *beam-spot-constraint* clearly improves the resolution of prompt muons, *i. e.*, muons being produced at or close to the PV, and helps reducing the L1 trigger rates.

However, if a muon is produced at a considerable distance off the beam spot, the implicit beam-spot requirement results in additional bending of the muon track that is purely artificial. The overly bent L1 track of such a *displaced muon* is assigned an artificially small p_T. If, then, cuts on the $p_T^{(\mathrm{L1})}$ of a muon candidate are applied, a displaced muon can fall below this $p_T^{(\mathrm{L1})}$ threshold, be rejected by the L1 trigger, and therefore irrecoverably lost for later offline analysis. Ultimately, this leads to trigger inefficiencies for non-prompt muons as a function of the muon displacement. This inefficiency mechanism will be dubbed the *L1 bias* for displaced muons and the topic of Sec. 4.4.3.

> “ *The so-called **L1 bias for displaced muons** is a consequence of a particular property of the online muon reconstruction at the L1 trigger in Run 2, i. e., the implicit assumption that each muon track originates at the beam–beam interaction point (**“beam-spot constraint”**). For muons with non-zero transverse impact parameter ($d_0 > 0\,\mathrm{cm}$), this can lead to considerable trigger inefficiencies when $p_T^{(L1)} > 0\,\mathrm{GeV}$ L1 trigger requirements are applied.* ”

L2 and L3 muon objects

About 20–25% (depending on the data-taking year) of all muon candidates from L1 do not correspond to offline reconstructed muons. [91] Reconstruction algorithms at the HLT stage have enough time available to employ more sophisticated computations compared to the L1 reconstruction. They are therefore able to further filter and identify muons and, consequently, further reduce the data-taking rate.

At the HLT stage, muons are reconstructed in two steps and the resulting muon candidates are called L2 and L3 muons, respectively, depending on whether they do or do not exclude tracker information in the reconstruction process.

L2 muons are built using information from the muon system only. At the beginning of the L2 reconstruction, initial track states (so-called *L2 seeds*) are generated from segment

patterns in and across the DT and CSC segments. For these L2 seeds, geometric matches ($\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.3$) to an L1 muon are required. They are further assigned a direction as well as a $p_T^{(\mathrm{L2})}$ value, the latter being obtained from a $\Delta\phi$ parametrization reflecting the bending angle of the segment with respect to a neighboring segment or the origin of the CMS coordinate system.

Standalone muon tracks are then built, starting from the L2 seed state, using the Kalman filter technique [93] – a recursive algorithm that progressively constructs tracks, detector layer by detector layer, employing pattern recognition while iteratively updating the track parameters.

In one iteration, track reconstruction is performed "inside-out," where the tracks are extrapolated outwards to the muon system and correlated with measurements in the muon chambers. In another iteration, the track is built along the opposite direction, "outside-in". The combination of these two iterations allows the reconstruction to remove potential biases resulting from the initial seed.

As an (optional[6]) last step, a beam-spot constraint is applied on the muon track, forcing it to have its origin in the point of pp interaction. This improves track parameter resolution for *prompt* muons, yielding an overall L2 reconstruction efficiency (with respect to L1 muons) of more than 99.5%, independent of p_T. [91]

However, similarly to the L1 case, the artificial beam-spot-constraint generally introduces trigger inefficiencies for *displaced* muons when $p_T^{(\mathrm{L2})}$ cuts are applied. While the beam-spot constraint always applies to *L1* muons (in the trigger configuration used throughout Run-2 data-taking), the *L2* muon reconstruction comes in two configurations, *i. e.*, with and without the beam-spot constraint. The latter is suited for the reconstruction of muons that are produced off the pp collision point, *e. g.*, cosmic-ray muons (in dedicated data-taking runs) or displaced muons resulting from LLP decays in hypothesized BSM processes. L2 efficiencies for displaced muons will be discussed in Sec. 4.4.4.

The beam-spot-constrained L2 muon reconstruction corresponds to the offline "standalone muon reconstruction," while the version without the beam-spot constraint corresponds to the offline "displaced standalone muon reconstruction." Both these types of offline reconstruction will be described in more detail in Sec. 3.3.3.

The ultimate muon reconstruction algorithm at the HLT produces so-called L3 muons. L3 muon reconstruction is not restricted to measurements from the muon system alone but utilizes all available information, including the inner tracking detector. An L3 muon is reconstructed following one of two approaches: Either, matching a track in the inner tracker with an L2 muon and subsequently performing a combined fit that uses information from both the tracker and the muon spectrometer, or matching a track to an L1 muon and thus identifying it without performing a combined fit [91].

Conceptually, these two L3 approaches correspond to the "global muon" and the "tracker muon" algorithms that are used in the offline reconstruction [94].

Because this work searches for displaced muons beyond the CMS tracker, it is not reliant on L3 muons but only on L2 muons (which do not require tracker hits for muon tracks). A more detailed discussion of the various L3 muon reconstruction algorithms is therefore omitted here but can be found in [91] and references therein.

[6] The beam-spot constraint was applied consistently in the 2016 reconstruction. Alternative trigger algorithms without this constraint were added only in 2018. Please see Sec. 4.1 for more details.

3.3.3 Offline muon reconstruction

In the standard CMS muon reconstruction for pp collisions [91, 95, 96], muon tracks are first built independently in the inner tracker and the muon system. Then, similar to the HLT muons described in the previous section, the final reconstructed muon objects come in three major types:

- *Standalone muons (SA muons)* are reconstructed using information from the muon system only. This is done by gathering all available information from the DT, RPC, and CSC detectors along the muon trajectory using a Kalman-filter technique [93]. Groups of DT or CSC segments are the seeds for the SA muon reconstruction.

  There are different flavors of SA muon reconstruction, which are defined via their respective handling of the beam-spot constraint.
  The original SA reconstruction developed before the start of the LHC operations is available both with and without the beam-spot constraint in the track fit. The latter version is more suited for the reconstruction of displaced muons and is the basis for two subsequent modifications.
  The first modification, developed for the Run-1 search for displaced dimuons [97], is called the *refitted standalone muon (RSA muon)* reconstruction. It performs an additional track refit to reduce biases caused by the beam-spot constraint in the SA seed generator, which results in RSA muons featuring better spatial and momentum resolution than SA muons.
  The second modification is a more recent attempt to further reduce the beam-spot-related biases of the SA reconstruction and is called the *displaced standalone muon (DSA muon)* reconstruction [98, 99]. It replaces the seed generator of the SA reconstruction with that used in the dedicated reconstruction of cosmic-ray muons.

- *Tracker muons* are reconstructed starting from tracks in the tracker, which are propagated outwards to the muon system if they have $p_T > 0.5\,\text{GeV}$ and $p > 2.5\,\text{GeV}$. For the extrapolated tracks, loose matching to DT or CSC segments is attempted, and the tracker track is called a tracker muon if such a match can be found. Hence, the tracker muon reconstruction relies on the tracker measurement and uses the muon system only for the purpose of tagging the particle type.

- *Global muons* are constructed by propagating SA muon tracks from the muon system inwards to search for matching tracker tracks. The Kalman-filter technique is again used in a combined fit, where it uses information from both the tracker and the SA muon tracks to build a final global muon track.
  Similar to the SA reconstruction, there is a variant of the global muon reconstruction that lifts the implicit beam-spot requirement and is seeded by DSA muons in an attempt to reduce related biases. Muons reconstructed with this modified algorithm are called “displaced global muons” [98].

Because the work at hand focuses on displaced muons outside of the tracker volume, the most relevant algorithm is the DSA muon reconstruction. However, as will be described in Sec. 4.3, the full CMS Run-2 search for displaced dimuons takes on a more holistic approach to muon reconstruction and, in fact, uses an optimal mix of the different types of reconstruction depending on the targeted detector region by partitioning the dimuon event

content into several distinct sets. Therefore, due to this partitioning, even the present muon-system-only search depends on tracker information, albeit only indirectly. For promptly-produced muons within the geometrical acceptance of the muon system, about 99% are reconstructed either as global muons or as tracker muons (and very often as both) [91].

However, similarly high efficiencies cannot be assumed for *displaced* muons in general, despite the existence of reconstruction algorithms targeting such signatures.

> “*The type of offline muon most directly relevant for this work is the* ***displaced standalone muon (DSA muon)****. It is reconstructed without any beam-spot constraint or tracker information and is thus capable of providing muon candidates outside the tracker.*”

The careful measurement of the reconstruction performance of displaced muons at all stages of the data-taking pipeline (L1, HLT, and offline) and the parametrization of efficiencies by the relevant variables will be the topic of the next chapter.

Search for displaced dimuons

The core analysis of this work – the search for pairs of displaced muons within the CMS muon system as the result of long-lived particle (LLP) decays – is the subject of this chapter.

The following will describe the used signal triggers and data samples, the general analysis strategy, the performance studies of displaced muon objects, the optimized event selection, the background estimation strategy and predictions, the relevant systematic uncertainties, and finally, the results of the search in terms of limits on the production cross-section in different signal benchmark interpretations.

In doing so, the reader's knowledge of the preceding chapters is assumed, in particular the different types of muon reconstructions described in Sec. 3.3 and the benchmark signal models in Sec. 2.2.1.

4.1 Signal triggers

Muon triggers in CMS have traditionally been developed for muons that originate at or near the beam spot. However, these triggers are not necessarily optimized for non-prompt muons. In fact, their performance deteriorates dramatically for muons produced off the beam spot and even drops to zero for particles produced in the outer half of the tracker and beyond.

To retain trigger sensitivity to muons produced over a wide range of displacement, the search utilizes triggers that only rely on hit information in the muon system. These implement an online track reconstruction that is conceptually equivalent to the offline standalone muon (SA muon) reconstruction (see Sec. 3.3.3).

An "L2 version" of the displaced standalone muon (DSA muon) reconstruction (see Sec. 3.3.3) was added during the 2018 data-taking period, which further enhanced the trigger capabilities for displaced muons.

2016 triggers

The trigger filtering the 2016 search dataset used in the search is called `HLT_L2DoubleMu28_NoVertex_2Cha_Angle2p5_Mass10`. This HLT path, inherited from the Run-1 search for displaced muons [65, 66], requires two beam-spot-unconstrained muons at the L2 online reconstruction stage. Each of those muons must have $p_T^{(\mathrm{L2})} > 28\,\mathrm{GeV}$ and $|\eta| < 2.0$. Additional trigger requirements are:

- `2Cha`: Hits in at least two muon system segments are required. This helps to reduce poorly reconstructed muons and hadronic punch-through that reaches the inner layers of the muon system.

- `Angle2p5`: The 3D angle between the two L2 muons is required to be smaller than 2.5 rad. This cut targets the trigger-level rejection of cosmic-ray events, as those typically feature two back-to-back reconstructed muons (see Sec. 4.4).

- `Mass10`: Finally, a minimum threshold of 10 GeV on the invariant mass of the two L2 muon candidates is imposed.

The HLT path uses an L2 muon reconstruction whose initial stage of the muon track reconstruction (the track "seeding") utilizes a *seed generator* that is optimized for promptly produced particles. This type of algorithm is referred to as "pp-seeded" or *hltL2* reconstruction in this work.

The following two L1 triggers are seeds for the HLT path, which will fire if at least one of them gets activated:

- `L1_DoubleMu_11_4` (active throughout the 2016 data-taking period)

- `L1_DoubleMu_12_5` (disabled in about 4% of the data-taking runs in 2016)

Each of these L1 seeds requires two L1 muon candidates passing the $p_T^{(\mathrm{L1})}$ thresholds respectively indicated in these seed names (in units of GeV).

Of particular importance is the circumstance that the L1 reconstruction in Run 2 includes the *beam-spot requirement*, *i. e.*, it assumes the muon tracks to originate at the location of the beam spot. Consequently, the $p_T^{(\mathrm{L1})}$ is underestimated for displaced muons that, in reality, do not point back to the beam spot. If, for such muons, cuts on $p_T^{(\mathrm{L1})}$ are applied, displaced muons can be mistakenly rejected by the L1 trigger and irrecoverably lost for later analysis. This is indeed the case for the above L1 seeds.

2017 triggers

The 2017 trigger configuration did not include the double-muon triggers described above or any other triggers suitable for the displaced muon search. Therefore, the 2017 dataset is not used in the presented analysis.

2018 triggers

In 2018, improved versions of the 2016 triggers were deployed. Not only does the HLT path `HLT_DoubleL2MuNoVtx_2Cha` have lower $p_T^{(\mathrm{L2})}$ thresholds ($p_T^{(\mathrm{L2})} > 23$ GeV) for the two muons, but it also drops the requirements on the 3D angle and invariant mass. These modifications are motivated by increased signal efficiencies – particularly, higher sensitivity to the "Higgs portal" channel of the Hidden Abelian Higgs Model (HAHM) – with an acceptable increase in trigger rate (about 1 Hz). Similar to the 2016 HLT path, `HLT_DoubleL2MuNoVtx_2Cha` exhibits "pp-seeded" (or *hltL2*) muon reconstruction at L2.

Additionally, the path `HLT_DoubleL2Mu23NoVtx_2Cha_CosmicSeed` was introduced in 2018. It is equivalent to the `HLT_DoubleL2MuNoVtx_2Cha` path except for the crucial difference of a modified SA muon reconstruction. Specifically, the `CosmicSeed` version of the signal trigger uses a track seed generator that is optimized for cosmic-muon reconstruction. This type of L2 muon reconstruction is referred to as "cosmic-seeded" or *hltL2Cosmic* reconstruction. It improves the L2 reconstruction efficiencies of displaced, non-pointing muons over the previously-used, beam-spot-constrained reconstruction algorithms.

TABLE 4.1: *Re-emulated triggers for the displaced muon object studies.*

Re-emulated triggers	L1 seeding	L2 requirements	p_T requirements	Domain of application
`HLT_L1SingleMuOpen`	"Open"	–	$p_T^{(L1)} > 0\,\text{GeV}$	Reference for (L1) trigger studies
`HLT_L2Mu10_NoVertex_pp`	"Open"	"pp track seeding"	$p_T^{(L1)} > 0\,\text{GeV}$, $p_T^{(L2)} > 10\,\text{GeV}$	Reference for triggers in 2016/2018
`HLT_L2Mu10_NoVertex_CosmicSeed`		"cosmic-muon track seeding"	$p_T^{(L1)} > 0\,\text{GeV}$, $p_T^{(L2)} > 10\,\text{GeV}$	Reference for triggers in 2018

However, the cosmic-seeded L2 reconstruction performs slightly less well for prompt and prompt-like muons. The analysis, therefore, uses a logical OR of the two HLT paths to filter the used data sets.

A combination of four L1 triggers seeds either of the 2018 signal HLT paths:

- `L1_DoubleMu_15_5_SQ` (The "SQ" stands for "single-muon quality", which is a set of more stringent track requirements than the implicit "double-muon quality" requirements imposed on the other seeds.)
- `L1_DoubleMu_15_7`
- `L1_TripleMu_5_3_3` (active in about 90% of the 2018 runs)
- `L1_TripleMu_4_4_4` (active in about 10% of the 2018 runs)

Re-emulated trigger components

For the purpose of displaced muon object studies (particularly for the studies of the corresponding trigger efficiencies), dedicated samples with re-emulated trigger content were produced. Most importantly, the re-emulated samples were filtered by a `SingleMuOpen` L1 seed, a type of algorithm that has no requirements on the $p_T^{(L1)}$ and thus does not suffer from the L1 bias of non-pointing muons. The customized trigger menu additionally features trigger algorithms for single-muon objects with low online-p_T thresholds, which are suited as reference paths for the study of trigger efficiencies. To utilize the results for the analysis, the performance of each "muon leg" of the double-muon signal triggers is assumed to be independent of the other leg. This means that the efficiencies obtained for each muon leg are eventually multiplied to get the full double-muon efficiency. Tab. 4.1 lists the re-emulated trigger components. The re-emulated datasets are described in more detail in Sec. 4.4.

4.2 Data and simulated samples

In addition to the pp-collision data sets of Run 2, the search relies on various other data, both recorded in dedicated data-taking runs and simulated via MC methods. This section describes all samples that are relevant to the analysis.

4.2.1 Run-2 collision data

The main dataset was produced by proton–proton collisions at a center-of-mass energy $\sqrt{s} = 13\,\text{TeV}$ and collected in the years 2016 and 2018 during the LHC Run 2. In 2017, no

suitable trigger was included in the trigger configuration, and therefore data of this year are excluded from the analysis. Events of interest are filtered by dedicated "double muon" triggers described in Sec. 4.1.

Because the analysis does not rely on calorimeter information, one can maximize the number of valid data-taking runs by including even those runs that had problematic calorimeter states. Thus, a total integrated luminosity of $(36.3 \pm 0.9)\,\mathrm{fb}^{-1}$ and $(61.3 \pm 1.5)\,\mathrm{fb}^{-1}$ can be achieved in the years 2016 and 2018, respectively. The full Run-2 dataset for the search comprises $97.6(17)\,\mathrm{fb}^{-1}$ of recorded pp-collision data.

About half of the 2016 data (the first $19.8\,\mathrm{fb}^{-1}$) was affected by saturated APV readout chips of the tracker front-end electronics. This circumstance resulted in a lower tracking efficiency in this early data-taking period. Data recorded afterward (late 2016 and the full 2018 dataset) are not affected. Even though the presented analysis is based on DSA muons, *i. e.*, muons that are reconstructed using the muon system only (see Sec. 3.3.3), these tracker inefficiencies in 2016 have an indirect effect on the number of available DSA muons. This is a consequence of the analysis design, as explained in Sec. 4.3.

4.2.2 Signal models and simulated signal samples

The search is intended to be as generic and inclusive (model-independent) as possible. To that extent, two signal benchmark models with a wide range of signal kinematic properties and a minimal set of BSM ingredients are used: The Hidden Abelian Higgs Model (HAHM) and the BSM heavy scalar model ($H \to XX$) both predict displaced muons from long-lived particle (LLP) decays with typical transverse momenta near and above the thresholds of the used triggers. A general description of these models can be found in Sec. 2.2.2; specific details on the simulations are given in what follows.

Properties of the simulated HAHM samples

HAHM signal samples ($h \to Z_D Z_D$, $Z_D \to \mu\mu$) were produced on a discrete grid of dark photon mass, m_{Z_D}, and kinetic mixing parameter, ε. Tab. 4.2 details the m_{Z_D}–ε grid and lists the used branching fractions $\mathcal{B}(Z_D \to \mu\mu)$ as well as the mean displacements of the formed dimuon vertices.

On a technical level, the dark photon production was modeled at leading order by MADGRAPH5_aMC@NLO [100] (version 2.4.2). The 2016 and 2018 samples use the NNPDF2.3 LO and NNPDF3.1 NNLO PDF sets [101], respectively. PYTHIA 8.2 [102] was then used to simulate the decays of the dark photons. For each signal point in Tab. 4.2, 750 000 events were generated, respectively using the CUETP8M1 [103] and CP5 [104] tunes for 2016 and 2018 to model the underlying event. Finally, the propagation through and the interaction with the detector material was simulated with Geant4 [105].

To save processing time and storage, only events containing at least one muon in the final state were kept. The corresponding filter efficiency ranges between 20% and 28% and generally depends on $\mathcal{B}(Z_D \to \mu\mu)$.

Signal lifetimes in between the generated grid points can be constructed at a later analysis stage by appropriately reweighting the signal distributions from generated-lifetime samples.

TABLE 4.2: *Specifications of the simulated HAHM samples. The values of* $\mathcal{B}(Z_D \to \mu\mu)\mathcal{B}(Z_D \to \mu\mu)$ *are based on Table 2 of [33].*

m_{Z_D} [GeV]	$\mathcal{B}(Z_D \to \mu\mu)$	ε	$c\tau_{Z_D}$ [mm]	$\langle L_{xy} \rangle$ [cm]
10	0.1538	1×10^{-6}	1.14	0.77
		5×10^{-7}	4.55	2.89
		1×10^{-7}	114	72
		3×10^{-8}	1264	800
20	0.1476	5×10^{-7}	2.17	0.72
		2×10^{-7}	13.6	4.14
		5×10^{-8}	217	66.4
		1×10^{-8}	5425	1665
30	0.1437	3×10^{-7}	3.90	0.80
		1×10^{-7}	35.1	6.73
		3×10^{-8}	390	74.7
		7×10^{-9}	7165	1374
40	0.1462	2×10^{-7}	6.21	0.86
		8×10^{-8}	38.8	5.10
		2×10^{-8}	621	81.4
		5×10^{-9}	9937	1301
50	0.1257	2×10^{-7}	4.42	0.47
		6×10^{-8}	49.1	4.37
		1×10^{-8}	1768	158
		4×10^{-9}	11049	982
60	0.1069	1×10^{-7}	11.8	0.70
		4×10^{-8}	73.7	3.93
		7×10^{-9}	2405	129
		2×10^{-9}	29464	1591

TABLE 4.3: *Specifications of the simulated heavy scalar benchmark samples.*

m_H [GeV]	m_X [GeV]	$c\tau_X$ [mm]	Generated number of events per $c\tau_X$
125	20	13, 130, 1300	200 000 (2μ 2jets), 100 000 (4μ)
	50	50, 500, 5000	
200	20	7, 70, 700	100 000 (2μ 2jets), 50 000 (4μ)
	50	200, 200, 2000	
400	20	4, 40, 400	50 000 (2μ 2jets), 25 000 (4μ)
	50	8, 80, 800	
	150	40, 400, 4000	
1000	20	2, 20, 200	30 000 (2μ 2jets), 20 000 (4μ)
	50	4, 40, 400	
	150	10, 100, 1000	
	350	35, 350, 3500	

Properties of the BSM Heavy Scalar model

As already mentioned in Sec. 2.2.2, the SM-like Higgs is not necessarily the only particle initiating an LLP decay featuring displaced muons in the final state. In principle, the LHC energies allow for the production of even heavier mediators.

To explore a larger phase-space of signal kinematic properties and topologies than the HAHM, the "BSM heavy scalar" benchmark model ($H \to XX$, $X \to \mu\mu$) features decays of generic scalars H to spin-0 mediators X, with a broad spectrum of masses m_H and m_X.

Two sets of these benchmark signals were produced, depending on whether one or both of the LLPs were forced to decay to muons: In the scenario $H \to XX \to 4\mu$, both X's decay to final-state muon pairs. For $H \to XX \to 2\mu\,2\text{jets}$, only one X decays to a muon pair while the other produces pairs of light quarks ($u\bar{u}$, $d\bar{d}$, $s\bar{s}$) with equal probabilities. Because the quarks quickly hadronize, the resulting final states are collectively referred to as "jets" and not further considered in the analysis. Tab. 4.3 lists the simulated grid of m_H, m_X, and lifetime $c\tau_X$. Each set of (m_H, m_X) was produced in three lifetime settings, corresponding to mean transverse decay lengths of about 3 cm, 30 cm, and 250 cm. The dimuon acceptance and selection efficiencies depend on the mass of the signal. Therefore, the number of generated events varies with m_H.

The heavy scalar signal samples were generated with PYTHIA 8 [102] (v8.212 in 2016 and v8.230 in 2018) using the CUETP8M1 [103] and CP5 [104] tunes as well as the NNPDF2.3 LO and NNPDF3.1 NNLO PDF sets [101] in 2016 and 2018, respectively. Again, Geant4 [105] simulated the particle propagation through the detector.

4.2.3 Background Monte-Carlo samples

There are no SM processes that produce LLPs with masses larger than about 5 GeV, as already mentioned in Sec. 2.2.1. The present work targets LLPs with masses larger than 10 GeV. Therefore, all potentially observed "SM background" above the 5 GeV mass must be due to muon track misreconstruction. Such instrumental effects cannot be faithfully simulated with Monte Carlo (MC) methods, and so the measurements of the various background components are obtained from recorded data directly.

Still, simulated background samples from SM processes are useful, at least to some extent. In the presented search, they are used to design and optimize some of the basic

TABLE 4.4: *Simulated SM backgrounds used in the analysis. For each process, the table shows kinematic cuts in the event generator (if any), the total production cross-section, σ, and the equivalent luminosity, $\mathcal{L}_{int}^{equiv} = N_{events}/\sigma$, per year.*

Process	Kinematic cuts	σ [pb]	$\mathcal{L}_{\text{int}}^{\text{equiv}}$ [fb^{-1}] 2016	2018
$Z/\gamma^* \to ll$	$10 < m_{ll} < 50\,\text{GeV}$	18445	1.22	2.14
	$m_{ll} > 50\,\text{GeV}$	6077	13.5	16.5
$t\bar{t} \to b\bar{b}ll\nu\nu$		87.3	906	737
tW, $\bar{t}W$		35.8	388	484
$WW \to ll\nu\nu$		12.2	164	646
WZ		47.1	85	82
ZZ		16.5	120	120
W + jets		61527	0.48	1.15
QCD μ-enriched	$\hat{p}_T > 20\,\text{GeV}$ $\lvert\eta(\mu)\rvert < 2.5$ $p_T(\mu) > 15\,\text{GeV}$	302672	0.07	0.07

signal selection criteria and as a guide to understanding generic features of the various background components.

The generated MC background sources and some of their relevant properties are listed in Tab. 4.4. In 2016, they were simulated with the CMS software framework [106] CMSSW_7_1 and reconstructed with CMSSW_10, while in 2018, they were produced using CMSSW_10_2. Simulations of parton showering and hadronization were performed by PYTHIA 8.2 [102] using the CUETP8M1 tune [103] in all 2016 samples and the CP5 tune [104] in all 2018 samples. (The only exception is the 2016 WW sample, whose parton showering was simulated using HERWIG++ 2.7.1 [107]. In all cases, particle propagation and material interactions were simulated using Geant4 [105].

4.2.4 Cosmic-ray data and signals with re-emulated trigger

In the search for displaced muons, it is crucial to understand the detector and reconstruction performance over the full range of relevant muon displacement. Previous muon object studies in CMS [94, 95, 108] focused on measuring prompt and prompt-like muons, such as muons produced in pp collisions. However, these methods cannot necessarily be employed for muons that originate beyond the inner tracker layers. Moreover, as alluded to earlier, MC simulations of muons in the displaced regime are not guaranteed to depict reality sufficiently accurately.

One way – and arguably the *only* way – to assess the performance of *displaced* muon reconstruction at its various reconstruction stages is to use muons from cosmic rays. This work makes an effort to use cosmic-ray muons recorded in dedicated runs in 2016 to measure the performance of displaced L1, L2, and offline muons and derive scale factors and systematic uncertainties relevant to the displaced dimuon search.

A critical property of the used cosmic-ray dataset is that it was filtered by an L1 seed that does not pose any requirements on the $p_T^{(\text{L1})}$ (a so-called *SingleMuOpen* seed). This means that no artificial L1 inefficiencies due to the L1 bias[1] of displaced muons are present

[1] This *L1 bias* was introduced in Sec. 3.3.2 and will be the subject of Sec. 4.4.3.

in this dataset, as even tracks with significantly mismeasured $p_T^{(\mathrm{L1})} \geq 0\,\mathrm{GeV}$ will trigger an event. Unlike the collision datasets, which are already filtered by seeds with non-zero $p_T^{(\mathrm{L1})} > 0\,\mathrm{GeV}$, this opens up the possibility of measuring the relevant L1 trigger efficiencies directly from recorded data.

Other auxiliary triggers included in the re-emuation are `HLT_L2Mu10_NoVertex_pp` and `HLT_L2Mu10_NoVertex_CosmicSeed`, described in Sec. 4.1 and listed in Tab. 4.1. Due to their reduced $p_T^{(\mathrm{L2})}$ thresholds ($p_T^{(\mathrm{L2})} > 10\,\mathrm{GeV}$), they serve as single-muon reference triggers for the nominal double-muon signal triggers of the search.

The following introduces the data samples used for the muon object performance studies. A full discussion of these studies will be given in Sec. 4.4.

Cosmic-ray data

The cosmic-ray data samples were recorded in dedicated runs in 2016 when there was no beam but still a magnetic field of $B = 3.8\,\mathrm{T}$. This configuration gives access to a clean sample of cosmic-ray muons. Such muons traverse the CMS detector primarily from (mostly) top to bottom and typically produce hits in both the upper and the lower detector hemispheres.

The used "cosmics" runs were recorded in two different settings of the CMS global muon trigger (uGMT): The *bottomOnly* mode only triggers events with activity in the lower detector half – a configuration that is typically used in CMS during dedicated cosmics data-taking. The *base* mode enables triggering in both detector halves and represents the usual setting for pp data-taking runs. As a baseline for the displaced muon objects studies, datasets recorded in the *bottomOnly* mode are used.

The total amount of recorded cosmic-ray events used for the muon object studies amounts to 1436 lumi-sections[2] in the *bottomOnly* mode.(They correspond to two data-taking eras in mid-2016, *i. e.*, eras D and E.)

Signal samples with re-emulated triggers

Later sections will often compare results obtained from recorded cosmic-ray muons to the ones from simulated signal processes. For these purposes, the targeted signal process is the generic LLP production via a heavy scalar, $H \to XX \to 2\mu\,2\mathrm{jets}$, as described in Sec. 2.2.2 and 4.2.2.

At the time of the muon object studies (mid-2018), official signal samples were not yet available in CMS for all relevant data-taking years. Hence, to faithfully study the performance for both 2016 and 2018 settings, which have different configurations of the signal triggers, a set of signal samples was privately produced, re-emulating the L1 and HLT trigger appropriately. The technical specifications used in the sample production are identical to the one for simulated signals in Sec. 4.2.2. These samples were later validated with equivalent official samples.

Like the cosmic-ray data, these re-emulated signals are filtered by a *SingleMuOpen* L1 seed without $p_T^{(\mathrm{L1})}$ requirements. Although technically simulated in a 2016 configuration, the obtained signal samples are suited to study the online and offline reconstruction performance of cosmic-ray muons for both 2016 and 2018.

[2] In CMS, a "lumi-section" refers to the amount of data recorded over 2×10^{18} bunch orbits (corresponding to a time of about 23.3 s) at constant instantaneous luminosity.

TABLE 4.5: *Specifications of the simulated signal events with re-emulated trigger.*

m_H [GeV]	m_X [GeV]	$c\tau_X$ [mm]	Generated number of events per $c\tau_X$
125	20	13, 130, 1300	20 000
	50	50, 500, 5000	
200	20	7, 70, 700	20 000
	50	20, 200, 200	

Further, it is important to keep in mind that the kinematic properties of typical cosmic-ray events generally differ from typical signals, as will be further discussed later.

The mass–lifetime grid and the number of generated events of re-emulated signals is summarized in Tab. 4.5.

Simulated atmospheric cosmic rays

An ingredient for some of the displaced muon performance studies are *simulated* cosmic-ray events. Such a "cosmics simulation" was privately produced with PYTHIA 8 [102] in a technical configuration identical to the signal samples described in Sec. 4.2.2.

Cosmic-ray muon events are filtered using a cylindrical volume defined via $r < 300\,\text{cm}$ and $|z| < 600\,\text{cm}$. Two cosmic-simulated samples have been produced with different time parameters, MinT0 and MaxT0. These parameters are proxies for possible segment time measurements of incident cosmic-ray particles in the upper and lower hemispheres, respectively:

- $(-12, 12)$ ns, where the average segment time in the upper leg is around $-12\,\text{ns}$.
- $(-40, 0)$ ns, where the average segment time in the upper leg is around $-40\,\text{ns}$.

The sample with $(-40, 0)$ reproduces the typical time distribution expected in events that trigger in the lower hemisphere. Because of this, it is used as a reference sample for simulated cosmic rays.

One limitation of the cosmics simulation is that it contains the trigger emulation used in the displaced dimuon search. This configuration does not reproduce the uGMT *bottomOnly* setting, and therefore it is not used in the muon object studies of Sec. 4.4 in explicit measurements of the trigger performance.

4.3 Analysis strategy

Particles created in the collisions of beam particles have their origin in a point called the primary vertex (PV), which is close to the beam spot (BS). All known particles are either stable (over the extent of the detector volume) or decay immediately. In either case, they form tracks with PVs at or near the BS and are referred to as *prompt* particles.

A long-lived particle (LLP) produced at the PV, on the other hand, will cover some distance before decaying to muons. The decay products will again form a vertex, called the common vertex (CV), which is displaced from the PV. Fig. 4.1 provides an illustration of a typical LLP decay to muons and the key variables involved. Muon pairs stemming from a CV are called *dimuons*, and if their common origin is off the beamspot by more than a few

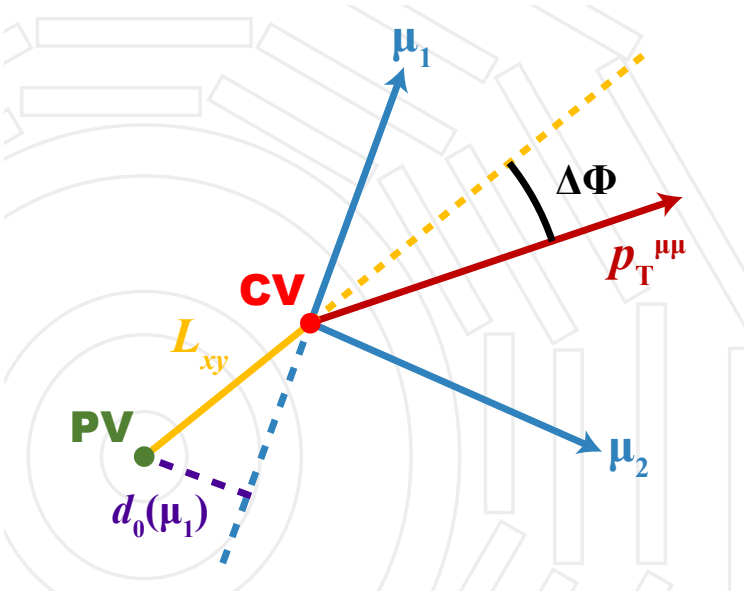


FIGURE 4.1: *Illustration of a typical LLP decay that leads to the formation of a displaced dimuon from the final-state muons μ_1 and μ_2. The common vertex (CV) of the displaced dimuon is located off the primary vertex (PV), separated by a distance L_{xy} in the transverse plane. The angle between the L_{xy} vector and the dimuon momentum $p_T{}^{\mu\mu}$ is the so-called dimuon collinearity angle $\Delta\Phi$. The impact parameter d_0 denotes the shortest transverse distance between an (extrapolated) muon track and the PV.*

hundred micrometers, they are commonly called *displaced dimuons*. The individual muons of a given dimuon are referred to as the muon *legs* of the dimuon.

The magnitude of the dimuon displacement in the transverse detector plane, *i. e.*, the absolute distance between PV and the CV, is called the transverse decay length L_{xy}. The measurement uncertainty of L_{xy}, denoted as $\sigma_{L_{xy}}$, is computed by combining the (transverse) position uncertainties of the PV and the CV. A measure for the significance of the observed displacement compared to zero is the so-called L_{xy} *significance*, defined as $L_{xy}/\sigma_{L_{xy}}$. This quantity is large for a well-measured displaced dimuon (which has small $\sigma_{L_{xy}}$ and large L_{xy}) but small for promptly-produced backgrounds, like those from SM Drell-Yan processes.

Any dimuon produced by a massive LLP has a non-zero opening angle, α, between its muon legs. Consequently, the momentum vectors of the muon legs point in different directions than the momentum of the preceding LLP, whose momentum is along the L_{xy} vector. The *dimuon* momentum, on the other hand, can still be aligned with the L_{xy} vector. Notably, this is the case for a typical two-body LLP decay as a consequence of momentum conservation. In contrast, for common background processes, the dimuon momentum is not correlated with the L_{xy} vector. This motivates the definition of the variable $\Delta\Phi$, which is the angle between the dimuon momentum and the L_{xy} vectors in the transverse plane, as a handle to differentiate between signal and background processes.

Another key variable being characteristic of dimuons of an LLP decay is the so-called transverse impact parameter d_0. It is defined as the shortest distance between the PV and the inwards-extrapolated track of the muon. While promptly-produced particles have small values of d_0, regardless of their orientation, muons created in a displaced vertex can have large d_0. Small- and large-d_0 tracks are referred to as *pointing* and *non-pointing* tracks,

respectively, throughout this work.

A generic LLP signal process of the form $H \to XX \to 2\mu$ (see Sec. 2.2.2) cannot be characterized by just *one* of the key variables introduced above but instead requires a combination of them. The masses of the involved particles, m_H and m_X, and the LLP lifetime, $c\tau_X$, are generally free parameters, and different values will produce vastly different characteristics in terms of the discussed key variables. To illustrate this point, one can consider the following extreme scenarios:

- Heavy, long-lived mediator: $m_X \approx m_H/2$, large $c\tau_X \to$ large L_{xy}, large α, large d_0 (non-pointing muons)
- Light, long-lived mediator (producing boosted particles): $m_X \ll m_H/2$, large $c\tau_X \to$ large L_{xy}, small α, small d_0 (pointing muons)
- Heavy, but short-lived mediator: $m_X \approx m_H/2$, small $c\tau_X \to$ small L_{xy}, large α, potentially small d_0.

The above examples show, depending on the parameters of the signal model decay, that one can produce pointing and non-pointing muons with small or large displacements. The performance of the CMS detector is not uniform across the full range of event kinematic properties. Therefore, different model parameter settings will lead to decays experiencing different detector limitations. The following will briefly mention some of the limitations associated with the above example scenarios. The purpose of these considerations is to illustrate again that only a *combination* of variables can meaningfully characterize a decay, but never a single variable alone. This also means that the quantification of these and other limitations requires careful and systematic studies to disentangle the various detector effects – an absolute necessity to maximize the discovery potential of the search.

As was alluded to earlier (Sec. 3.3.2) and as will be discussed in more detail later (Sec. 4.4.3), the CMS L1 trigger suffers from a bias in the reconstruction of $p_T^{(\mathrm{L1})}$ for displaced muons, which is a consequence of the implicit beam spot constraint for L1 muon tracks. As it turns out, this bias can be parametrized by d_0 in first order, where large-d_0 tracks suffer L1 trigger inefficiencies more considerably. This means that the detector performance is generally worse for displaced muons, as they typically have large values of d_0 (*e. g.*, the first example above). However, there is a range of model parameters that produce highly-displaced but pointing muons (by featuring strongly boosted particles), which, at least in principle, are not significantly affected by the L1 bias.

Boosted topologies (*e. g.*, the second of the above examples), while favored by the L1 reconstruction to some degree, often come with another kind of limiting effect. Close-by muons (called *collinear* dimuons) suffer from reconstruction inefficiencies if the hits of the two muons overlap in the muon stations such that the individual tracks cannot be adequately resolved anymore.

While these resolution effects are only present for dimuons with the smallest opening angles, there is another limitation for dimuons on the other end of the opening-angle spectrum, *i. e.*, for dimuons with (almost) back-to-back muons legs. In this regime, the CV fitting algorithm becomes inefficient, as it often struggles to find a common vertex along a "line" of antiparallel muon tracks.

In addition to these considerations, different offline muon reconstruction algorithms (producing tracks of varying degrees of track parameter resolution) can and need to be

employed in different domains of L_{xy}. If an LLP decays well within the tracker volume, the muon reconstruction can benefit from this sub-detector's vastly superior tracking performance (compared to the muon-system). If the LLP decays outside the tracker volume, the only feasible muon reconstruction algorithms are ones solely based on muon system information. The class of algorithms based on the standalone muon (SA muon) reconstruction makes it possible to measure displaced muons outside of the CMS tracker, albeit at the cost of significantly lower tracking precision. On the other hand, the displaced regime that requires SA-based algorithms is virtually free from SM backgrounds, unlike the prompt-like regime, which has to deal with larger levels of background despite the superior muon track resolution.

To recap, the above discussion shows that it is possible to identify a handful of key variables good for discriminating signal process candidates from known backgrounds. However, different kinematic characteristics of the decay combined with varying detector performance make the analysis design and its optimization for discovery potential a multidimensional problem, which requires trade-offs between signal efficiency and background rejection in many cases.

The analysis searches for vertices formed by an intersection of trajectories of two oppositely-charged muons and displaced from the IP. The targeted range of displacement reaches from several micrometers to several meters.

There is no single muon reconstruction algorithm giving optimal performance over the full target range, as outlined earlier. To leverage the full CMS muon reconstruction potential, the analysis utilizes a combination of reconstruction algorithms, each tailored to a specific range of displacement.

The general idea is that the analysis selection starts with the most efficient SA muons and substitutes them with more accurately reconstructed global and tracker muons whenever such muons are available. The SA-type algorithm with the best performance for the targeted signal kinematic properties and topology was found to be the displaced standalone muon (DSA muon) reconstruction (see Sec. 3.3.3). [36] Tracker-based muons that serve as potential replacements for the DSA muons are represented by an ensemble of global and arbitrated tracker muons.[3] Such muons that are reconstructed as both global and tracker muons are called PAT muons, their name reflecting their implementation as a signal object in the underlying software framework of the CMS Physics Analysis Tools (PAT).

4.4 Displaced muon performance

Previous CMS studies have measured the performance of online and offline muon reconstruction algorithms in events with *prompt* muons, *i. e.*, muons that were produced within the first tracker layer. [94, 95, 108] While generally high reconstruction performance at the various stages of the CMS muon reconstruction has been shown for this class of muons, one cannot expect similarly high efficiencies and sharp resolutions for muons produced farther off the interaction point (IP).

The results presented in this section aim to extend previous studies by measuring the CMS muon reconstruction capabilities as a function of the offline-reconstructed muon transverse momentum p_T, the transverse impact parameter d_0, and the longitudinal impact pa-

[3] In this context, "arbitrated" denotes that no sharing of muon segments between tracker muons is allowed.

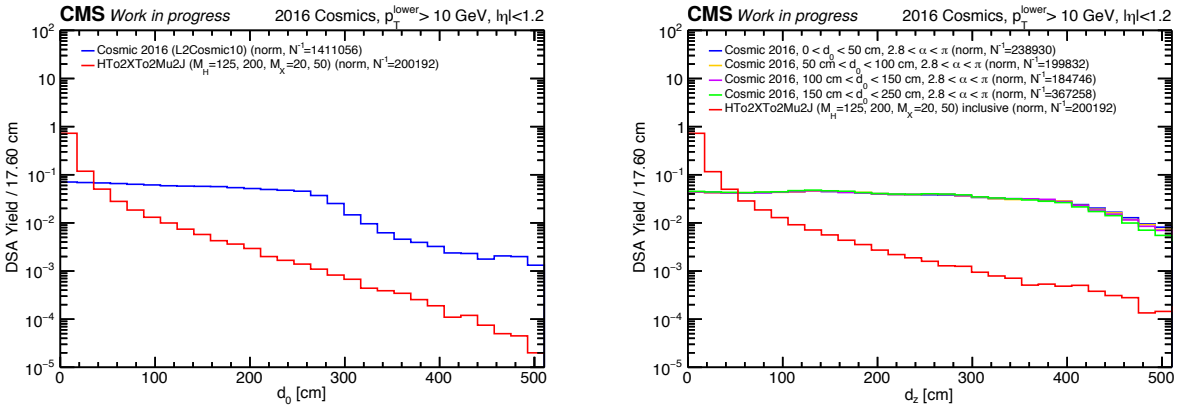


FIGURE 4.2: *Distributions of* d_0 (left) *and* d_z (right) *for cosmic-ray data (blue) and simulated signal (red). The* d_z *distribution of cosmic-ray data are shown in several bins of* d_0.

rameter $|d_z|$. Since no SM process would lead to muons with displacements of up to a few meters as the result of pp collisions, a different source of muons is required to perform these measurements, *i. e.*, muons from atmospheric cosmic rays.

Cosmic radiation is produced in the interactions of highly-energetic particles of cosmic origin with nuclei in the upper Earth atmosphere. A certain fraction of these interactions leads to the formation of mesons (primarily pions and kaons), whose decays subsequently produce muons in a broad energy spectrum that traverse the atmosphere and reach the Earth's surface. Because the CMS experiment is located about 100 m underground, cosmic radiation arriving at the detector mainly consists of cosmic-ray *muons* due to the natural filtering properties of the overburdened rock. These muons cover energies up to and beyond the LHC reach.

Moreover, particle trajectories from cosmic-ray muons reconstructed in the CMS detector cover a wide range of displacement. As Fig. 4.2 shows, they span the full range of d_0 that is expected from a displaced muon BSM signal. Unlike the simulated signal, where the d_0 distribution reflects the underlying model parameters, the d_0 distribution in cosmic-ray data only depends on the trajectory of the incoming particle and is relatively flat up to $d_0 \leq 3\,\text{m}$. The drop around $d_0 \approx 3\,\text{m}$ in cosmic-ray data roughly coincides with the inner and outer boundaries of the solenoid magnet (at radii of 3.16 m and 3.47 m, respectively). Further, d_0 and d_z are uncorrelated in cosmic-ray data, as they merely depend on the (random) position of the incoming cosmic rays and not on underlying model parameters. This allows d_0 and d_z to be studied separately. The d_z of cosmic-ray muons is practically flat up to about $d_z \approx 4\,\text{m}$, where the drift tube (DT) coverage ends. (Cosmic-ray reconstruction becomes inefficient outside of the barrel because the vertical endcap detector layers are aligned roughly parallel to the incident cosmic rays.)

In summary, their broad kinematic range makes cosmic-ray muons an ideal source to measure the detector performance of displaced muons [96, 109].

This section will present detailed measurements of the displaced muon performance in CMS (reconstruction, identification, dimuon vertexing) at all stages of the muon reconstruction (online and offline algorithms). Moreover, the scale factors and systematic uncertainties relevant to the displaced dimuon search (see Chap. 4) will be derived. Many of the discussed concepts will be based on the descriptions in earlier sections, *e. g.*, the CMS apparatus in Sec. 3.2, or the online and offline muon reconstruction algorithms in Sec. 3.3.2

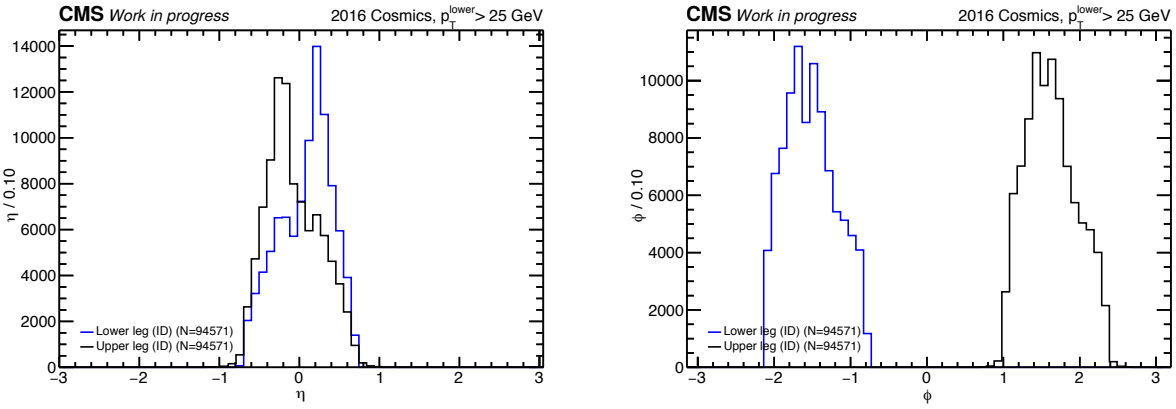


FIGURE 4.3: *Distributions of* η (left) *and* ϕ (right) *for the upper (black) and lower (blue) legs of cosmic-ray muons reconstructed as pairs of back-to-back DSA muons.*

and 3.3.3, respectively.

The datasets used for the studies in this section are the cosmic-ray data samples, simulated atmospheric cosmic-rays, and the signal samples with re-emulated triggers (described in Sec. 4.2), which are all filtered by the single-muon HLT paths summarized in Tab. 4.1. The direct use of double-muon triggers, like the signal triggers of the analysis, is infeasible since the typical time differences between upper and lower cosmic legs ($\Delta t \approx 40\,\text{ns}$) are large enough to exceed the nominal bunch-crossing time of $\Delta t_{\text{BX}} = 25\,\text{ns}$ and are therefore discarded by the CMS trigger rules.

4.4.1 Reconstruction of cosmic-ray muons

A cosmic-ray muon traversing the CMS detector is typically reconstructed as two separate DSA muons, one in the upper detector hemisphere (at $\phi > 0$) and one in the lower hemisphere (at $\phi < 0$). These two distinct reconstructed muons are respectively referred to as the *upper* and *lower legs* of the cosmic-ray muon.

Fig. 4.3 shows the η and ϕ distributions for each the upper and the lower legs in cosmic-ray events with a back-to-back DSA muon pair. The observed η asymmetry is characteristic for cosmic rays and is a consequence of the cavern shaft (filled with air), which is located off the detector center and which has different "muon filter" properties than the surrounding rock.

From the detector-center point of view, the lower leg of the cosmic-ray muon travels in the same direction ("inside-out") as a muon produced in a collision event. In contrast, the upper leg's direction of travel is opposed to a promptly produced muon. This has important consequences for the charge assignment of the muon legs. While the lower-leg charge matches the actual charge of the cosmic-ray muon, the upper leg is labeled with the opposite charge (assuming no track reconstruction mistakes). In other words, a perfectly well reconstructed cosmic-ray muon produces a pair of oppositely charged muons in the detector, much like the targeted signature from an LLP decay. Cosmic-ray muons, therefore, provide a clean control sample of displaced muons that can be used to study the detector performance at all relevant stages of track reconstruction.

Most of the methods presented in this section rely on tagging one leg and measuring the efficiency of the other leg, comparing different reconstruction algorithms for an individual

leg, or comparing track parameters between the upper and lower legs.

4.4.2 Muon identification and dimuon vertexing

To study the performance of displaced muons with cosmic-ray data in a configuration relevant for the primary analysis, cosmic-ray muon legs are selected in a similar manner as in the analysis. This reduces the rate of misreconstructed tracks and makes it possible to eventually derive the relevant scale factors and systematic uncertainties for the search.

Specifically, DSA-reconstructed muons are used that fulfill the following set of track quality criteria referred to as the *displaced muon ID*:

- DSA muons reconstructed using segments in at least two different muons stations.
- The reconstructed muon is required to have at least 18 or 12 valid hits in the DT or CSC chambers, respectively.
- The fitted track must fulfill $\chi^2_{\text{track}}/\text{ndof} < 2.5$ and $\sigma_{p_T}/p_T < 1.0$ to ensure proper measurement of p_T.

The displaced muon ID was optimized using simulated signals and a collision data sample enriched in signal-like background events (see Sec. 4.6 for the corresponding control regions in the primary analysis). Details on the ID optimization procedure can be found in Sec. 4.5.

The two oppositely-charged legs of a reconstructed cosmic-ray muon can, at least in principle, undergo a common vertex fit to form a displaced dimuon. The common vertex (CV) is an essential object as it gives access to the dimuon mass $m_{\mu\mu}$, the transverse decay length L_{xy}, and its uncertainty $\sigma_{L_{xy}}$. Moreover, the resolution of the individual track parameters (p_T, η, ϕ, d_0, d_z) can be improved by refitting these quantities to the CV.
The ID criteria for common vertices formed from DSA-DSA muons used in the object studies reflect the dimuon selection criteria used in the search and are a subset of the entire analysis selection described in Sec. 4.5.

The dimuon vertex ID for cosmic-ray studies is represented by the following set of criteria:

- Each DSA muon leg must satisfy the displaced muon ID and $p_T^{(\text{DSA})} > 10\,\text{GeV}$
- $\chi^2_{\text{vertex}}/\text{ndof} < 20$
- $m_{\mu\mu} > 10\,\text{GeV}$
- Opposite charges of the dimuon legs
- DCA $< 50\,\text{cm}$, *i. e.*, distance of closest approach (DCA) between the tracks
- $N(\text{segments}) > 4$ (5 if $\Delta\eta_{\mu\mu} < 0.1$), *i. e.*, the sum of the segments of each DSA muon leg.

The variable $\cos\alpha$ is used both in the primary analysis and in the cosmic-ray object studies, albeit in opposite ways. In the former case, cosmic-ray muons are an unwanted background and are effectively suppressed by requiring $\cos\alpha > \alpha_{\text{min}}$ (where α_{min} is -0.8 or -0.9, depending on the year of data taking). In the latter case, the $\cos\alpha$ handle is used to obtain a clean sample of back-to-back cosmic-ray muons with $\cos\alpha < -0.8$ for the study of signal-like vertices.

However, the very fact that both studies use mutually exclusive $\cos\alpha$ ranges poses a limitation for the applicability of cosmic-ray dimuon results to the primary analysis. Moreover, the back-to-back-leg characteristic of well-reconstructed cosmic-ray muons introduces another difficulty: The position of a common vertex along two (anti-)parallel tracks is not well defined. Therefore, the vertexing performance for cosmic-ray dimuons (which typically approach the limit $\cos\alpha \approx -1$) is expected to be insufficient.

Cosmic-ray dimuons are still the only physical source to probe the dimuon vertex fitter over a broad range of displacement, so they are not wholly discarded despite their limited potential. However, the displaced vertexing with cosmic-ray dimuons is only considered a "sanity check" of the vertex fit parameters across the full span of displacement. This means that the performance of the fitter (*i. e.*, the absolute dimuon vertex efficiency) will not be directly compared to signal simulation in any of the following studies.

4.4.3 L1 muon performance

Most collision events happening in the detector are discarded at the L1 trigger stage after rudimentary physics object reconstruction and never stored for subsequent processing. The L1 trigger was optimized to reduce the data-taking rate from the LHC bunch-crossing rate to a manageable level by filtering events that are most likely to contain "interesting" physics processes.

However, like many other components in CMS, the L1 trigger was tailored to promptly-decaying particles, and it is not immediately clear how the trigger performance of prompt-like objects translates to the displaced regime.

This section attempts to answer this question for muons by measuring the L1 performance as a function of the transverse impact parameter and other key variables.

Efficiency of the `L1SingleMuOpen` trigger

All data samples used in the object studies with cosmic-ray muons are filtered by the `L1SingleMuOpen` trigger, which is an algorithm that does not impose any minimum thresholds on the $p_T^{(\text{L1})}$ of muon candidates. This means that the efficiency of the `L1SingleMuOpen` trigger itself cannot be measured in cosmic-ray data directly but only in simulation using MC information. However, the absence of a direct efficiency measurement in cosmic-ray data is not critical for the efficiency estimation over the full range of muon displacement because the major source of L1 inefficiency originates in the $p_T^{(\text{L1})}$ scale and resolution (as Sec. 4.4.3 will uncover).

Throughout this work, a "generator-" or "gen-level" particle refers to a particle created in the simulation by the event generator. Such a GEN particle is subsequently propagated through the detector volume while its material interactions are simulated using MC methods. In an ideal detector, a GEN particle will eventually result in a reconstructed track.

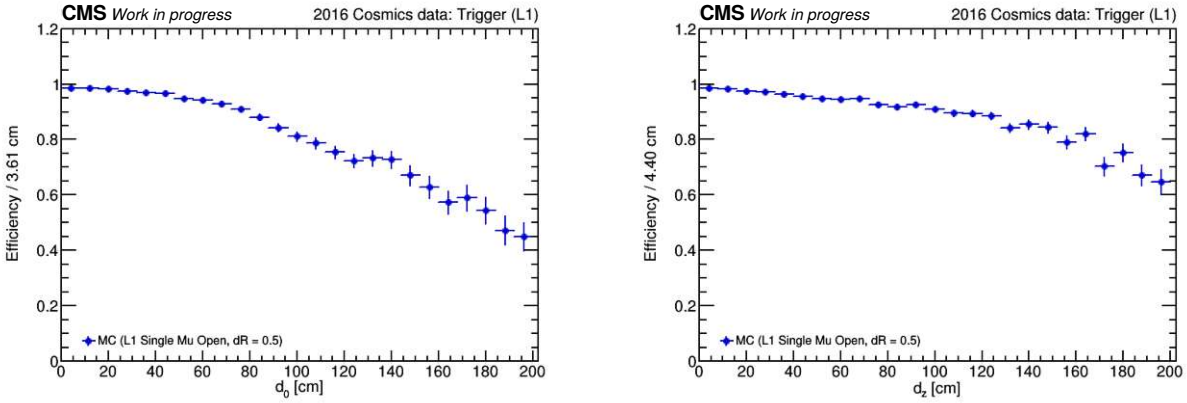


FIGURE 4.4: *Efficiency of the* `SingleMuOpen` *trigger as a function of* d_0 (left) *and* d_z (right) *as measured in signal simulation.*

In the following, the `L1SingleMuOpen` reconstruction efficiency is computed in simulated signal events that contain a generator-level muon with the following kinematic properties:

- $p_T^{(\mathrm{GEN})} > 10\,\mathrm{GeV}$
- $|\eta^{(\mathrm{GEN})}| < 1.2$
- $|{d_z}^{(\mathrm{GEN})}| < 200\,\mathrm{cm}$
- ${L_{xy}}^{(\mathrm{GEN})} < 330\,\mathrm{cm}$

These criteria ensure that the generator-level muons are produced within the detector volume where they can – at least in principle – be triggered at the L1.

The `L1SingleMuOpen` efficiency is subsequently computed as the fraction of events with geometric matches between a GEN muon and an L1 muon if the `L1SingleMuOpen` trigger has accepted the event. A geometric match[4] is found if $\Delta R = \sqrt{(\Delta\eta_{\text{gen-L1}})^2 + (\Delta\phi_{\text{gen-L1}})^2} < 0.5$, where the subscripts "gen-L1" indicate that the corresponding quantities are measured as differences between GEN and L1 muons.

The resulting `L1SingleMuOpen` efficiency is shown in Fig. 4.4 as a function of the generator-level d_0 and d_z. It represents the reconstruction efficiency of L1 muons, which is found to be 90% and 80% at $d_0 = 60\,\mathrm{cm}$ and $d_0 = 100\,\mathrm{cm}$, respectively. For $d_z < 100\,\mathrm{cm}$, this L1 efficiency is above 90%.

Scale and resolution of $p_T^{(\mathrm{L1})}$

This section aims to measure the L1 $p_T^{(\mathrm{L1})}$ scale and resolution by comparing the p_T values of online- (L1) and offline-reconstructed (DSA) muons.

The used datasets were recorded by the `L1SingleMuOpen` trigger (in the *bottomOnly* uGMT mode). Only events containing at least two DSA muons are selected.

In each of these events, the lower DSA leg must fulfill the following requirements:

- It must satisfy the displaced muon ID and also $10\,\mathrm{GeV} < p_T^{(\mathrm{DSA})} < 60\,\mathrm{GeV}$, $|\eta| < 1.2$, and $|d_z| < 50\,\mathrm{cm}$.

[4]The ΔR requirement is stricter if the matching does not involve an L1 muon, because the "higher-level" reconstruction algorithms have generally better resolution.

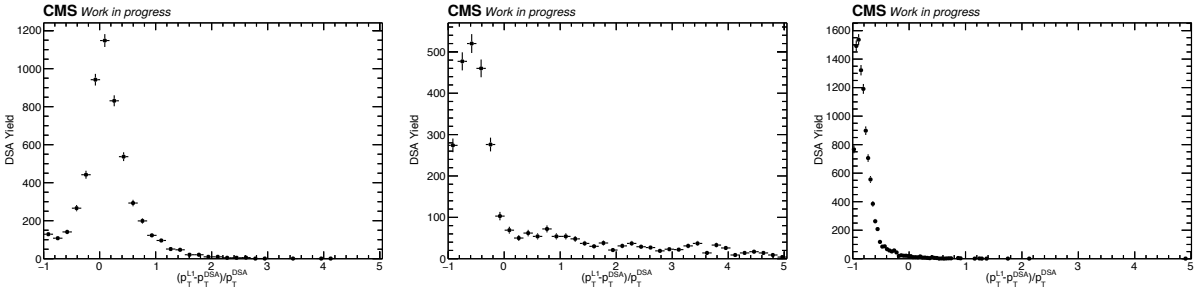


FIGURE 4.5: *Distributions of $R^{(L1)}$ measured with cosmic-ray muons in three bins of d_0.* Left to right: $0\,\text{cm} < d_0 < 5\,\text{cm}$, $15\,\text{cm} < d_0 < 20\,\text{cm}$, *and* $90\,\text{cm} < d_0 < 100\,\text{cm}$.

- A geometric match with an L2 muon having $p_T^{(\text{L2})} > 10\,\text{GeV}$.
- In the case of cosmic-ray data (*i. e.*, not in signal MC), the 3D angle between the muon legs must be large, $\alpha > 2.9$. The upper leg is merely used for tagging well-reconstructed cosmic-ray muons but does not play any further role in the p_T measurement.

Finally, the L1 muon is obtained from the matched L2-muon seed.

The $p_T^{(\text{L1})}$ resolution and scale are defined as

$$R^{(\text{L1})}(d_0) = \frac{p_T^{(\text{L1})} - p_T^{(\text{DSA})}}{p_T^{(\text{DSA})}} \tag{4.4.1}$$

This definition leverages the fact that the resolution of DSA muons (see Sec. 4.4.6) is generally significantly better than the one of L1 muons. (Conceptually, the offline $p_T^{(\text{DSA})}$ is therefore treated as the "true" p_T, which the $p_T^{(\text{L1})}$ can be compared to.)

Fig. 4.5 shows the results of the $R^{(\text{L1})}$ measurement in three distinct d_0 bins. The maxima of these distributions are indicative of the $p_T^{(\text{L1})}$ scale and the widths of the $p_T^{(\text{L1})}$ resolution.

For each d_0 bin, the L1 performance is characterized by the maximum and the full width at half maximum (FWHM) of the $R^{(\text{L1})}$ peak, referred to as the "resolution peak parameters"). The bin widths of the $R^{(\text{L1})}$ histograms are dynamically adjusted in each d_0 bin to ensure a robust measurement of the peak parameters.

While Fig. 4.5 showcases only a few d_0 bins, the full measurement is summarized in Fig. 4.6. Here, the resolution peak parameters are shown as a function of d_0. The results obtained from cosmic-ray data and signal simulation are overlaid.

Remarkably, the $R^{(\text{L1})}$ maxima shift towards negative values as d_0 increases, reflecting an underlying $p_T^{(\text{L1})}$ mismeasurement ($p_T^{(\text{L1})} < p_T^{(\text{DSA})}$) in the displaced regime. The $p_T^{(\text{L1})}$ misreconstruction is maximal already at $d_0 \geq 35\,\text{cm}$, where the $R^{(\text{L1})}$ peak approaches the value -1, indicating $p_T^{(\text{L1})} \ll p_T^{(\text{DSA})}$. The $p_T^{(\text{L1})}$ scale dependence is a direct consequence of the implicit beam-spot constraint in the L1 muon reconstruction, which forces muon tracks to originate in the BS, thus artificially lowering the bending radius and reconstructed $p_T^{(\text{L1})}$ of L1-reconstructed displaced muons.

The $R^{(\text{L1})}$ peak widths (FWHM) increase with d_0 (starting from the prompt case), indicating deteriorating resolution performance for non-pointing muons. The FWHM values

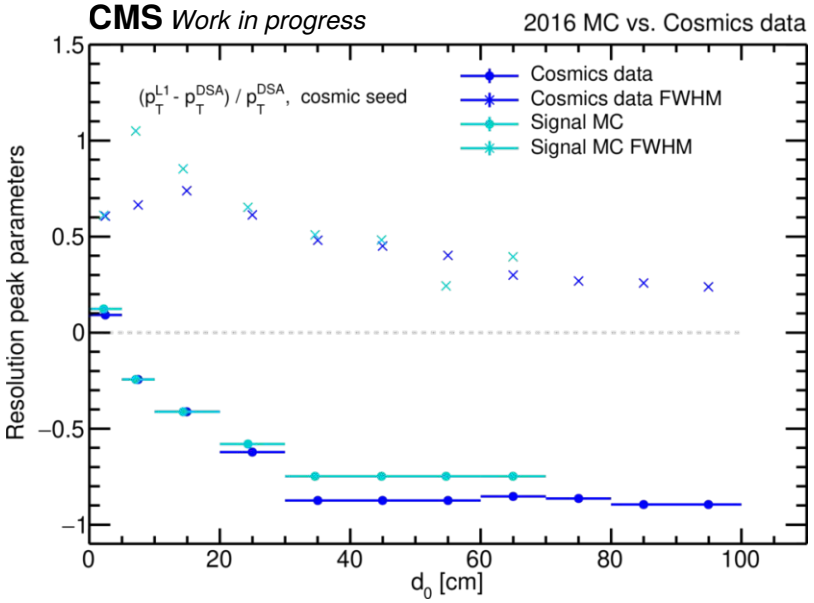


FIGURE 4.6: *$p_T^{(L1)}$ scale and resolution as a function of the offline d_0 as measured in cosmic-ray data (blue) and signal simulation (teal). The underlying $R^{(L1)}$ distributions per d_0 bin are characterized by the $R^{(L1)}$ peak maximum ($p_T^{(L1)}$ scale, depicted with solid circles) and the FWHM ($p_T^{(L1)}$ resolution, depicted with x-shaped markers).*

are largest around $d_0 \approx 15\,\text{cm}$ and start to decrease again for even larger d_0 values. However, while this might naively suggest better resolution performance at $d_0 > 15\,\text{cm}$, the real cause for the FWHM decrease is the underlying $p_T^{(\text{L1})}$ scale bias.

L1 trigger efficiency of Run-2 triggers

As the previous section showed, the $p_T^{(\text{L1})}$ scale is intrinsically biased towards lower $p_T^{(\text{L1})}$ for non-pointing muons – a direct consequence of the beam-spot constraint of L1 muons. A dataset that is filtered by the `L1SingleMuOpen` trigger, like the one used for the object studies, is not affected by this $p_T^{(\text{L1})}$ bias. However, the data recorded in typical pp collision runs is filtered by L1 seeds with $p_T^{(\text{L1})} > 0\,\text{GeV}$ to keep the trigger rate at a manageable level.

The signal triggers used in the Run-2 search for displaced dimuons are primarily seeded by `L1_DoubleMu_*` algorithms with various $p_T^{(\text{L1})}$ thresholds, depending on the year of data-taking. Their most relevant $p_T^{(\text{L1})}$ requirements are (see Sec. 4.1):

- 2016: $p_T^{(\text{L1})} > 4\,\text{GeV}$ and $p_T^{(\text{L1})} > 11\,\text{GeV}$
- 2018: $p_T^{(\text{L1})} > 7\,\text{GeV}$ and $p_T^{(\text{L1})} > 15\,\text{GeV}$

To measure the efficiencies of the L1 triggers used in the analysis, samples of lower-leg DSA muons and their corresponding L1 muons are obtained similarly to the previous Sec. 4.4.3, except now requiring $p_T^{(\text{DSA})} > 33\,\text{GeV}$ in 2016 and $p_T^{(\text{DSA})} > 28\,\text{GeV}$ in 2018. These updated $p_T^{(\text{DSA})}$ requirements select muons well above the respective $p_T^{(\text{L2})}$ thresholds of the HLT paths and thus ensure that any effects of the L2 muon reconstruction are suppressed.

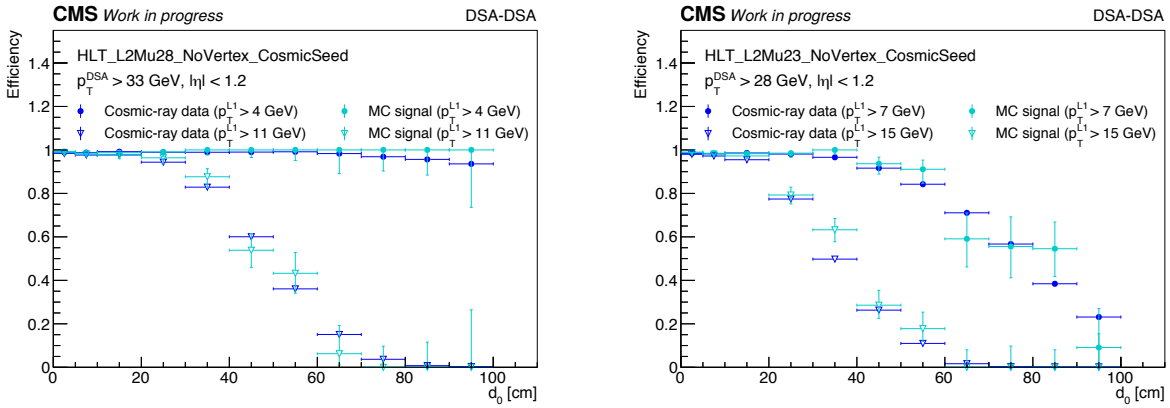


FIGURE 4.7: *L1 trigger efficiencies as a function of d_0, measured in cosmic-ray data (blue) and signal simulation (teal). The applied $p_T^{(L1)}$ thresholds reflect the L1 seeding of the Run-2 signal triggers in 2016* (left) *and 2018* (right).

Fig. 4.7 shows the resulting L1 efficiencies for the relevant $p_T^{(\mathrm{L1})}$ thresholds as measured in cosmic-ray data and signal simulation as a function of d_0. For each year, the efficiency associated with the highest $p_T^{(\mathrm{L1})}$ threshold dominates the total L1 efficiencies of non-pointing muons. The L1 efficiencies are above 20% for $d_0 < 60\,\mathrm{cm}$ and $p_T^{(\mathrm{L1})} > 11\,\mathrm{GeV}$ in 2016, and for $d_0 < 50\,\mathrm{cm}$ and $p_T^{(\mathrm{L1})} > 15\,\mathrm{cm}$ in 2018.

The general L1 efficiency trends in cosmic-ray data are sufficiently reproduced in signal MC, as the differences of the underlying η distributions between these sample types are a sub-dominant effect at the large-d_0 regime that is relevant for the search.

The total trigger efficiencies of the Run-2 signal triggers are a combination of the L1 and L2 efficiencies. This section presented the bare L1 trigger efficiencies by removing L2 reconstruction effects. The following section will provide results for the bare L2 efficiencies, which generally come with a richer set of available muon reconstruction algorithms. (The total Run-2 displaced muon trigger efficiencies will be presented afterward.)

4.4.4 L2 muon performance

The reconstruction of L2 muons happens at the HLT stage, where more sophisticated algorithms can be deployed than at L1.

Two different versions of L2 muon reconstruction are available in Run-2 data (see Sec. 4.1):

- "pp-seeded L2 reconstruction" (denoted as *hltL2*): The algorithms used in the HLT paths `HLT_L2DoubleMu28_NoVertex_Angle2p5_Mass10` (2016 data) and `HLT_DoubleL2MuNoVtx_2Cha` (2018 data) implement L2 muon reconstruction with a track seeding optimized for promptly-produced muons
- "Cosmic-seeded L2 reconstruction" (denoted as *hltL2Cosmic*): In 2018 data, the HLT path `HLT_DoubleL2MuNoVtx_2Cha_CosmicSeed` is available, whose track seeding is optimized for cosmic-ray data-taking. A similar path does not exist for 2016 data.

This section presents the measurements of hltL2 and hltL2Cosmic reconstruction efficiencies, their $p_T^{(\mathrm{L2})}$ scales and resolutions (with respect to their offline $p_T^{(\mathrm{DSA})}$), and the

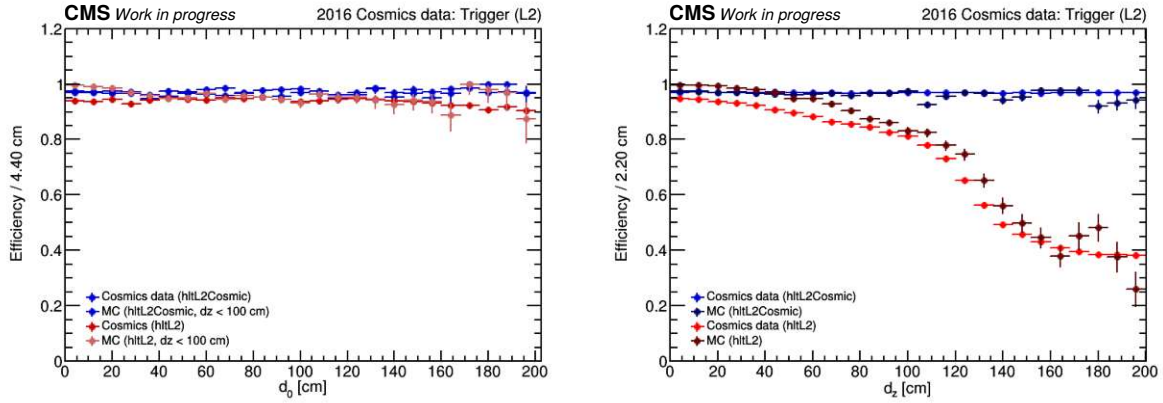


FIGURE 4.8: *L2 reconstruction efficiencies for hltL2 and hltL2Cosmic muons, measured with cosmic-ray data and signal simulation in events containing a L1 muon and a DSA muons.* Left: *Efficiencies as a function of* d_0. Right: *Efficiencies as a function of* $|d_z|$.

corresponding $p_T^{(\mathrm{L2})}$ trigger turn-ons in cosmic-ray data and signal simulation.

L2 muon reconstruction efficiency

The general idea of the L2 muon reconstruction efficiency measurement is to select events with a well-reconstructed offline muon in data filtered by a $p_T^{(\mathrm{L1})}$-inclusive L1 seed and compare to the subset of events that also fired an `HLT_L2Mu10_NoVertex*` trigger.

Specifically, the used dataset was recorded by the `HLT_L1SingleMuOpen` path. Events are selected by requiring a lower-leg DSA muon with the following properties:

- Matching a muon reconstructed by the `HLT_L1SingleMuOpen` trigger
- Fulfilling the displaced muon ID (see Sec. 4.4.2)
- $p_T^{(\mathrm{DSA})} > 20\,\mathrm{GeV}$, $|\eta| < 1.2$, $-2.1 < \phi < -0.8$
- $|d_z| < 200\,\mathrm{cm}$

The efficiency is then defined as the fraction of events that also fired an `HLT_L2Mu10_NoVertex*` path, where the asterisk encodes for the two flavors of L2 reconstruction, *i. e.*, hltL2 and hltL2Cosmic muons reconstructed with $p_T^{(\mathrm{L2})} > 10\,\mathrm{GeV}$. The measurements are done separately for hltL2 and hltL2Cosmic muons.

The resulting efficiencies are shown in Fig. 4.8. The pp-seeded reconstruction (red coloring in the figure) produces efficiencies that vary as a function of longitudinal impact parameter d_z. Particularly, the hltL2 efficiencies rapidly drop around $d_z \gtrsim 100\,\mathrm{cm}$. This general behavior is reproduced in signal simulation. In contrast, the hltL2Cosmic efficiencies are stable with d_z, both in cosmic-ray data and signal MC.

To study the efficiency behavior as a function of the transverse impact parameter d_0, any d_z-dependent effects are removed by (a) requiring $d_z < 100\,\mathrm{cm}$ in simulation and (b) reweighting the d_z in cosmic-ray data to match the signal MC. The resulting efficiencies show stable behavior with d_0 for both hltL2 and hltL2Cosmic muons in cosmic-ray data and simulation.

The considerable d_z dependence of the pp-seeded L2 reconstruction already indicates the importance of the cosmic-seeded HLT path for the search for displaced muons, as the

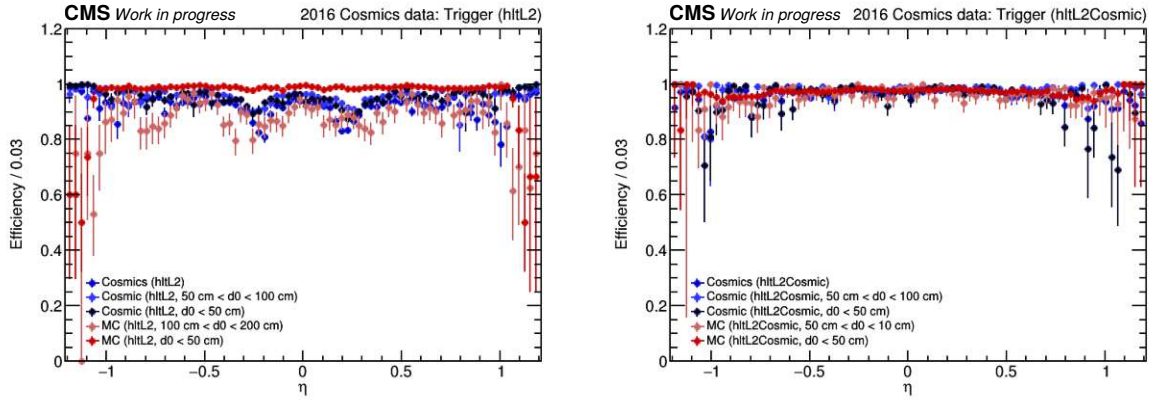


FIGURE 4.9: *L2 muon reconstruction efficiencies as a function of η, measured with cosmic-ray data and signal simulation in events containing a L1 muon and a DSA muons. The results represent the rates of hltL2* (left) *and hltL2Cosmic* (right) *muons with* $p_T^{(L2)} > 10\,\text{GeV}$.

latter performs stably over the full range of d_0 and d_z to which the underlying L1 trigger is sensitive.

Fig. 4.9 shows that the L2 efficiencies exhibit a mild $|\eta|$ dependence, which generally reflects the geometry of the CMS detector. These geometrical effects are more pronounced in htlL2 muons and practically negligible in hltL2Cosmic muons. The measurements were additionally performed in bins of d_0, and the $|\eta|$-dependent effects are reproduced in all cases. The only exception are hltL2 muons in the signal simulation, which do not match the cosmic-ray efficiencies in the overlap regions, regardless of the probed d_0 bin.

$p_T^{(L2)}$ scale and resolution

This section presents the measurement of the $p_T^{(L2)}$ scale and resolution. The general idea is similar to the equivalent L1 measurement in that the $p_T^{(L2)}$ of a well-reconstructed cosmic-ray muon leg is compared to its $p_T^{(DSA)}$ value.

The following measurements use the dataset described in the previous Sec. 4.4.4. Events are selected by requiring at least two DSA muons with the following criteria for the lower-leg DSA muon:

- Matching a muon reconstructed by the `HLT_L1SingleMuOpen` trigger
- It must satisfy the displaced muon ID and also $10\,\text{GeV} < p_T^{(DSA)} < 60\,\text{GeV}$, $|\eta| < 1.2$.
- $|d_z| < 50\,\text{cm}$
- A geometric match ($\Delta R < 0.2$) with an L2 muon having $p_T^{(L2)} > 10\,\text{GeV}$ (whereby hltL2 and hltL2Cosmic muons are measured separately).
- In the case of cosmic-ray data (*i. e.*, not in signal MC), the 3D angle between the muon legs must be large, $\alpha > 2.9$.

The $p_T^{(L2)}$ scale and resolution are defined with reference to the offline $p_T^{(DSA)}$, according to

$$R^{(L2)}(d_0) = \frac{p_T^{(L2)} - p_T^{(DSA)}}{p_T^{(DSA)}} \tag{4.4.2}$$

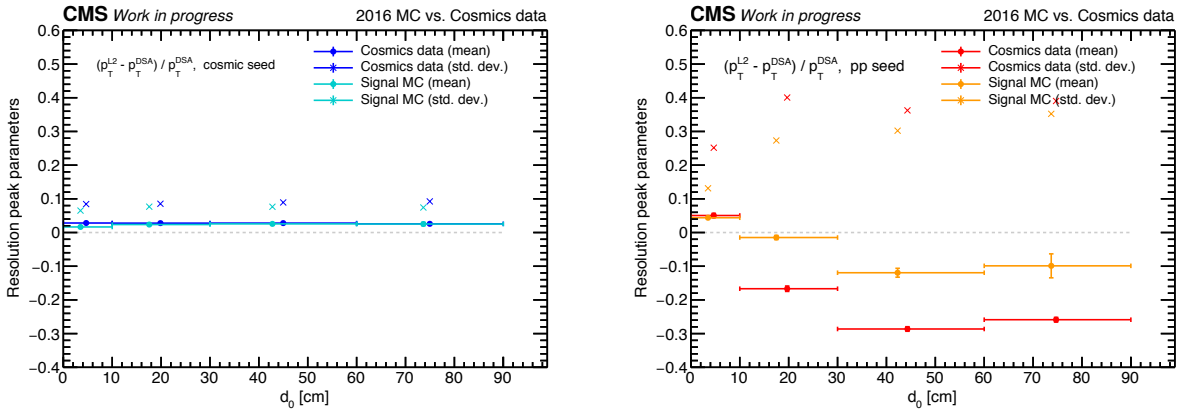


FIGURE 4.10: *$p_T^{(L2)}$ scale and resolution as a function of the offline d_0 as measured in cosmic-ray data (blue and red) and signal simulation (teal and orange). The underlying $R^{(L2)}$ distributions per d_0 bin are fitted with Gaussian functions whose mean and standard deviation characterize the $p_T^{(L2)}$ scale (depicted with solid circles) and $p_T^{(L2)}$ resolution (depicted with x-shaped markers).*

The appropriate $R^{(\mathrm{L2})}$ distributions are obtained separately for hltL2 and hltL2Cosmic muons, each in cosmic-ray data and signal simulation and several bins of d_0. Subsequently, the cores of these distributions (*i. e.*, the $R^{(\mathrm{L2})}$ peaks) are fitted with Gaussian functions. Special care was taken that every fit converges. The mean of the fitted Gaussian in a given d_0 bin serves as a proxy for the $p_T^{(\mathrm{L2})}$ scale; the fitted standard deviation is a measure for the $p_T^{(\mathrm{L2})}$ resolution.

These $R^{(\mathrm{L2})}$ characteristics are shown for hltL2 (red coloring) and hltL2Cosmic muons (blue coloring) in Fig. 4.10. The $p_T^{(\mathrm{L2})}$ reconstruction performance is stable with d_0 for hltL2Cosmic muons but shows a clear degradation for hltL2 muons.

$p_T^{(\mathrm{L2})}$ trigger turn-on profiles

As a consequence of the measured $p_T^{(\mathrm{L2})}$ scale and resolution, the hltL2Cosmic version of the signal triggers is expected to outperform the hltL2 version for non-pointing muons. Specifically, the $p_T^{(\mathrm{L2})}$ turn-on profile of the hltL2Cosmic path will not depend on d_0, while the hltL2 muons will inhibit a general broadening of their HLT turn-on profile resulting in deteriorating HLT efficiencies as a consequence of their degraded $p_T^{(\mathrm{L2})}$ resolution.

The measurements in this section are performed with events filtered by a `HLT_L2Mu10_NoVertex*` trigger, which requires a reconstructed L2 muon with $p_T^{(\mathrm{L2})} > 10\,\mathrm{GeV}$. Again, hltL2 and hltL2Cosmic configurations are measured separately. Events are selected analogously to the previous Sec. 4.4.4.

The $p_T^{(\mathrm{L2})}$ turn-on efficiencies are defined as the fraction of events whose probed muons also pass an additional cut $p_T^{(\mathrm{L2})} > {p_T}^{\mathrm{HLT}}$, where ${p_T}^{\mathrm{HLT}}$ is a threshold corresponding to one of the two p_T requirements in the signal HLT paths of the primary search (see Sec. 4.1):

- ${p_T}^{\mathrm{HLT}} > 28\,\mathrm{GeV}$, corresponding to the 2016 signal HLT path of the displaced dimuon search. This path only exists in the hltL2 muon reconstruction version and any results presented for hltL2Cosmic muons with this ${p_T}^{\mathrm{HLT}}$ threshold are purely illustrative.

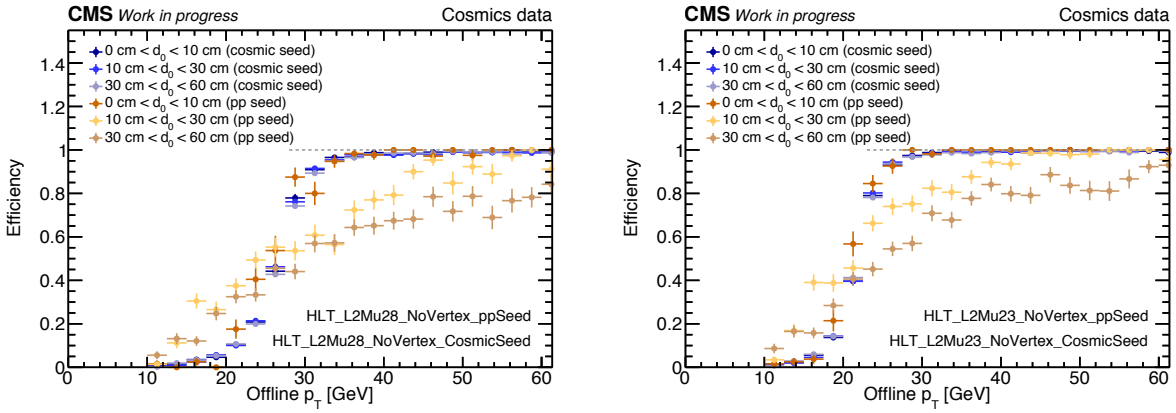


FIGURE 4.11: *Trigger turn-on profiles as a function of the offline* $p_T^{(DSA)}$ *in different* d_0 *bins for hltL2 (brown coloring) and hltL2Cosmic muons (blue coloring), as measured in cosmic-ray data filtered by* `HLT_L2Mu10_NoVertex*`. *The results represent the rate of L2 muons with the 2016 HLT threshold* $p_T^{(L2)} > 28\,\text{GeV}$ (left) *and the 2018 HLT threshold* $p_T^{(L2)} > 23\,\text{GeV}$ (right).

- $p_T{}^{\text{HLT}} > 23\,\text{GeV}$, corresponding to the 2018 signal HLT paths, which exist in both hltL2 and hltL2Cosmic configurations.

The resulting turn-on profiles as a function of the offline $p_T^{(\text{DSA})}$ are shown in Fig. 4.11. Distributions in blue colors represent measurements with the cosmic-seeded hltL2Cosmic muons, and brown colors show the efficiencies of pp-seeded hltL2 muons. One can clearly see the broadened profiles of hltL2-type triggers for non-prompt d_0 bins, which only reach their respective efficiency plateaus at $p_T^{(\text{DSA})} \gg p_T{}^{\text{HLT}}$. Triggers of the hltL2Cosmic type show a turn-on behavior on par with the hltL2 result in the prompt-like regime ($d_0 < 10\,\text{cm}$ in the figure). However, the sharpness of the hltL2Cosmic turn-on prevails also for larger values of d_0 without any noticeable performance drops. These conclusions are valid for both $p_T{}^{\text{HLT}}$ configurations, *i. e.*, for the HLT paths of both relevant data-taking years.

While trigger turn-on profiles are generally preferred to be as similar to a step-function as possible (which would correspond to ideal $p_T^{(\text{L2})}$ resolution performance), the imperfections in the presented measurements open up the possibility to probe muons with $p_T^{(\text{DSA})} < p_T{}^{\text{HLT}}$, at least to some degree of sensitivity. For example, relatively "soft" muons (*i. e.*, muons with small values of true p_T) near $p_T \approx 20\,\text{GeV}$ still have a non-zero probability of passing the $p_T^{(\text{L2})}$ trigger requirement (about 30–40% in some cases). The displaced dimuon search benefits from the limited $p_T^{(\text{L2})}$ resolution and selects muons having offline p_T values as low as $p_T^{(\text{DSA})} > 10\,\text{GeV}$, as this was found to improve signal efficiency for LLP decays in the context of the used benchmark models (see Sec. 2.2.2).

4.4.5 Combined trigger efficiencies and trigger scale factors

The previous sections quantified the trigger efficiencies and turn-on profiles as well as the p_T resolutions and scales for L1- and L2-reconstructed muons separately. In general, the full L1+L2 trigger efficiencies are the product of the individual L1 and L2 efficiencies.

This section will present the combined trigger efficiencies in configurations directly connected to the signal triggers used in the primary analysis. Following all previous studies, the

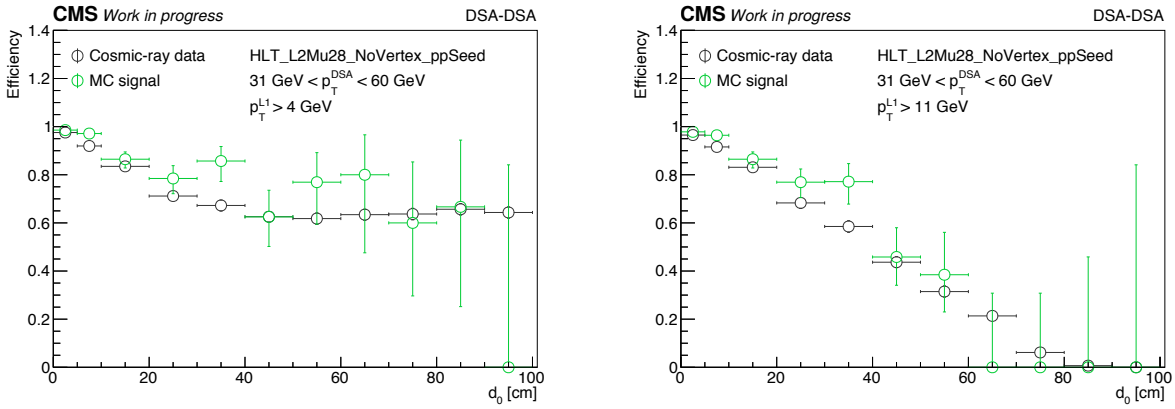


FIGURE 4.12: *Efficiencies of the trigger* `HLT_L2Mu28_NoVertex_ppSeed` *as functions of* d_0 *in cosmic-ray data (black) and signal simulation (green). The efficiency calculation is based on events filtered by* `HLT_L2Mu10_NoVertex`*. The additional cuts* $p_T^{(L1)} > 4\,\text{GeV}$ (left) *and* $p_T^{(L1)} > 11\,\text{GeV}$ (right) *correspond to the* $p_T^{(L1)}$ *thresholds used in the 2016 signal triggers.*

results will capture the key characteristics of the online reconstruction by being presented as a function of d_0.

The displaced dimuon search is based on *double*-muon triggers (as opposed to the single-muon triggers used in the muon object studies in this and the previous sections). The ultimate efficiencies for double-muon triggers can be obtained by multiplying the individual leg efficiencies. This factorization is valid as long as the muons are not overlapping (such that they are uncorrelated).

Additionally, trigger scale factors and uncertainties for the principal analysis will be derived in this section.

The following results are obtained with the same data samples and event selection as discussed in Sec. 4.4.4. Depending on the studied L2 reconstruction (hltL2 or hltL2Cosmic), the baseline events forming the efficiency denominator are filtered by the trigger `HLT_L2Mu10_NoVertex` or `HLT_L2Mu10_NoVertex_CosmicSeed`. Both these reference samples are subsets of the original `HLT_L1SingleMuOpen`-filtered samples.

The 2016 trigger configuration is shown in Fig. 4.12 in two efficiency distributions, each corresponding to a different $p_T^{(\text{L1})}$ threshold used in the seeding of the signal path `HLT_L2DoubleMu28_NoVertex_2Cha_Angle2p5_Mass10`.

The 2018 trigger configuration, `HLT_DoubleL2Mu23_NoVertex*`, is presented in Fig. 4.13, again for each of the two $p_T^{(\text{L1})}$ thresholds, as well as separately for hltL2 and hltL2Cosmic muons.

Most remarkably, all combined trigger efficiencies are significantly shaped by the underlying L1 reconstruction efficiencies, as the overall efficiencies decrease considerably with growing d_0 due to the intrinsic L1 bias for non-pointing muons. The $p_T^{(\text{L1})}$ thresholds are different between the years ($p_T^{(\text{L1})} > 4\,\text{GeV}$ and $p_T^{(\text{L1})} > 11\,\text{GeV}$ in 2016, $p_T^{(\text{L1})} > 7\,\text{GeV}$ and $p_T^{(\text{L1})} > 15\,\text{GeV}$ in 2018), so there are small differences in the L1 components of the measured efficiencies.

The hltL2Cosmic triggers can compensate to some degree for the increased $p_T^{(\text{L1})}$ thresholds in 2018, as their measured trigger efficiencies are practically independent of d_0. This

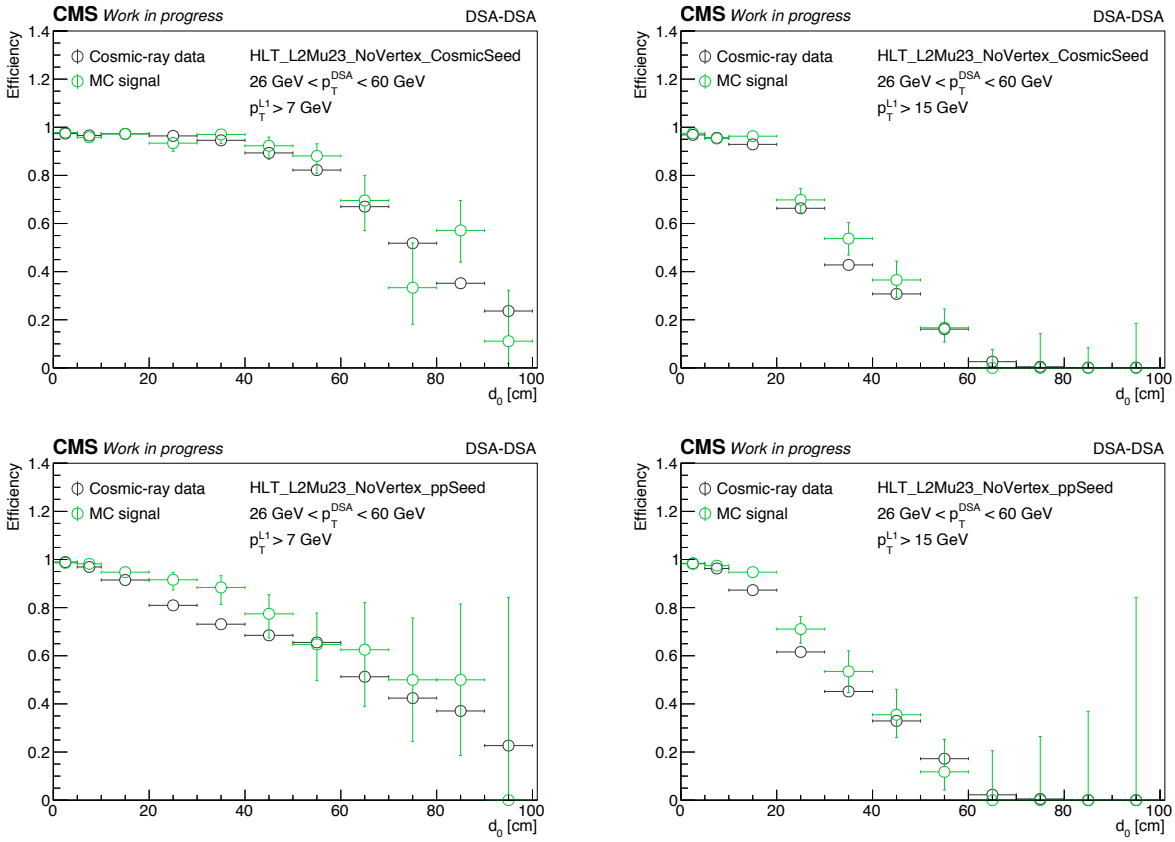


FIGURE 4.13: *Efficiencies of the trigger* `HLT_L2Mu23_NoVertex_CosmicSeed` (top) *and* `HLT_L2Mu23_NoVertex_ppSeed` (bottom) *as functions of* d_0 *in cosmic-ray data (black) and signal simulation (green). The efficiency calculation is based on events filtered by the corresponding* `HLT_L2Mu10_NoVertex*` *path. The additional cuts* $p_T^{(L1)} > 7\,\text{GeV}$ (left) *and* $p_T^{(L1)} > 15\,\text{GeV}$ (right) *correspond to the* $p_T^{(L1)}$ *thresholds used in the 2018 signal triggers.*

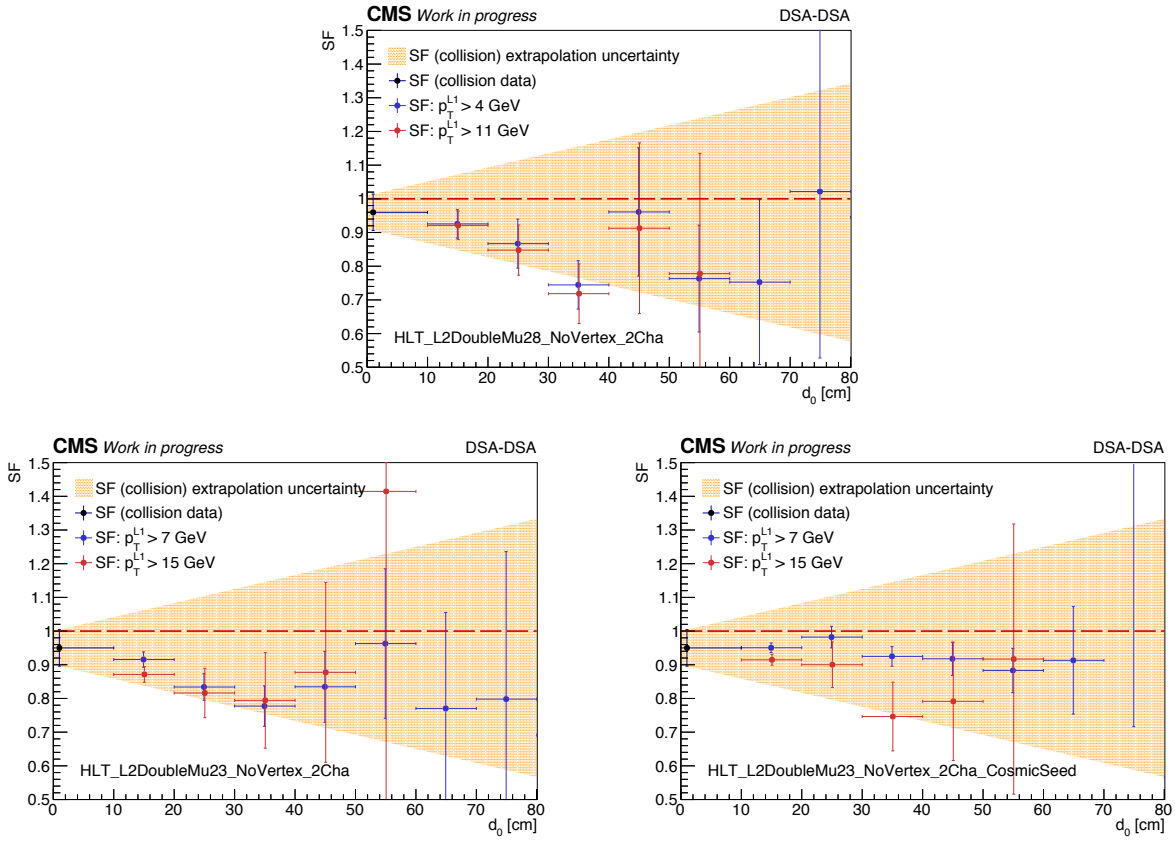


FIGURE 4.14: *Single-muon trigger efficiency scale factors for the 2016* (top) *and 2018* (bottom) *trigger configurations. The black dots indicate the prompt measurement with collision data, while the red and blue dots (each for a given $p_T^{(L1)}$ threshold) represent the measurement with cosmic-ray data. The variations of the measured scale factors are accounted for by a d_0-dependent systematic uncertainty, indicated by the orange band.*

notion is confirmed by comparing the hltL2Cosmic trigger efficiencies to the 2018 hltL2 efficiencies, which are more steeply falling with increasing d_0.

In all cases, the combined efficiencies and their dependence with d_0 are reasonably well reproduced in signal simulation. This can be further seen in the data/MC ratios of Fig. 4.14, which complement the prompt measurement ($d_0 < 10\,\text{cm}$) with the d_0-dependent cosmic-ray-muon measurement in several bins of $d_0 > 10\,\text{cm}$. These ratios (scale factors) do not exhibit any clear evidence of a d_0 dependence.

Still, the fluctuations in the data/MC ratios for $d_0 > 10\,\text{cm}$ are sizable, and there is also a non-negligible chance that parts of the variations originate from systematic effects that are intrinsic to cosmic-ray data. Therefore, the signal efficiencies in the search are corrected by the prompt trigger scale factor, and the extrapolation to the large-d_0 domain merely comes with a systematic uncertainty that increases linearly with d_0 (according to the orange bands featuring in Fig. 4.14).

4.4.6 Offline displaced standalone muon performance

The offline reconstruction is the ultimate stage of muon reconstruction, where all available detector information is combined to form the best-possible particle hypotheses. This sec-

tion presents the displaced standalone muon (DSA muon) reconstruction and identification efficiencies, the $p_T^{(\mathrm{DSA})}$ scale and resolution, and the offline dimuon vertexing performance.

DSA muon reconstruction efficiency and scale factors

The general idea behind the DSA muon reconstruction efficiency measurement – *i. e.*, the number of DSA-reconstructed muons out of all trigger-accepted muons – is to make use of the fact that an adequately reconstructed cosmic-ray muon will result in *two* back-to-back DSA tracks. Using a "tag-and-probe"-like approach in cosmic-ray data, one requires a well-reconstructed, high-quality DSA leg in the lower detector hemisphere and measures the fraction of events in which there is also a back-to-back, upper DSA leg with some loose track quality.

More specifically, the DSA leg requirements are as follows:

- Tag muon (high-quality, lower-leg DSA):
  - $|\eta| < 0.7$, $-2.1 < \phi < -0.8$, $p_T^{(\mathrm{DSA})} > 12.5\,\mathrm{GeV}$
  - $\sigma_{p_T}/p_T < 0.2$, $\chi^2_{\mathrm{track}}/\mathrm{ndof} < 2.0$, $N_{\mathrm{hits}}(\mathrm{DT}) > 30$, at least two matched stations
- Probe muon (back-to-back, upper-leg DSA):
  - Large 3D angle between the probe and the tag muons ($\alpha > 2.1$)
  - $N_{\mathrm{hits}}(DT) + N_{\mathrm{hits}}(CSC) > 0$, $\mathrm{ndof} > 0$

The comparisons shown in this section are based on (1) cosmic-ray data, with their d_z distribution reweighted to match the signal d_z distribution, (2) signal simulation, (3) cosmic-ray simulation with (MinT0, MaxT0) time parameters (-40, 0) ns and (4) cosmic-ray simulation with (-12, 12) ns. In each sample, the events are filtered by requiring an `HLT_L1SingleMuOpen` trigger to fire in the lower detector half.

The measured DSA reconstruction efficiencies are presented as functions of p_T, η, and d_0 in Fig. 4.15.

Generally, all efficiencies are flat in d_0 and p_T; the only observed dependence is in the η variable. This η dependence – specifically, the efficiency drops near $\eta = \pm 0.2$ – is a well-known effect for prompt muons passing near the CMS wheels. Coincidentally, the preferred direction of cosmic-ray muons is $\eta \approx 0.2$ in the top hemisphere (and $\eta \approx -0.2$ in the lower hemisphere), implicated by Fig. 4.3, which results from the off-center placement of the CMS cavern shaft. The cosmic-ray muon inefficiency profiles near $\eta = \pm 0.2$ are further broadened due to the underlying d_0 and d_z distributions, which are generally much flatter in non-prompt muons than in prompt-like muons (see Fig. 4.2). This argument is supported by a comparison of the d_z-reweighted and the unweighted cosmic-ray data in Fig. 4.15 (bottom).

Another observation is that the overall efficiencies are different between cosmic-ray data, cosmic-ray simulation, and signal MC. The signal simulation generally yields the highest efficiencies, and it is only marginally better than the cosmic-ray simulation with (-12, 12) ns time parameters. Ultimately, cosmic-ray data are best described by the (-40, 40) ns cosmic-ray simulation.

The lower efficiencies in all cosmic-ray samples are the result of local DT reconstruction timing effects that are specific to cosmic-ray muons (*i. e.*, local DT reconstruction parameters that are not optimized for cosmic-ray muons that have a time difference of

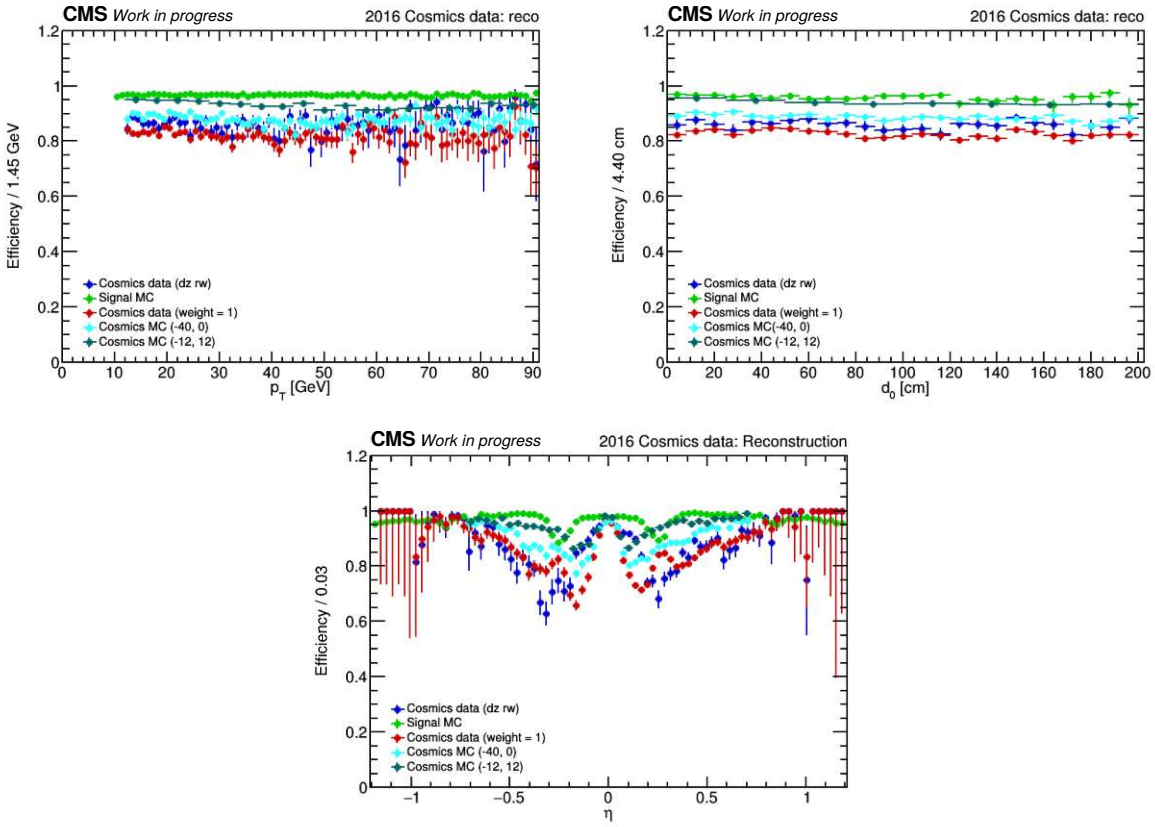


FIGURE 4.15: *DSA reconstruction efficiencies in cosmic-ray data, cosmic-ray simulations, and signal simulation, presented as functions of* p_T (top left), d_0 (top right), *and* η (bottom).

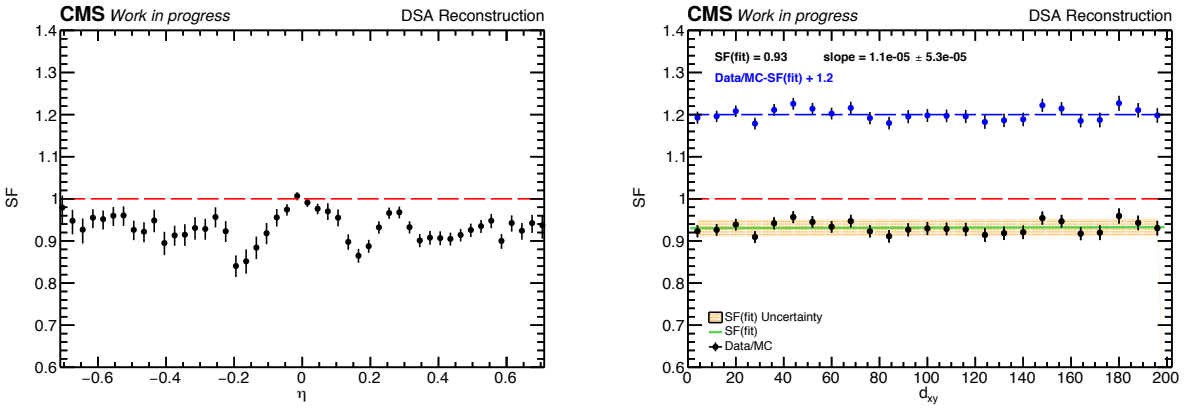


FIGURE 4.16: *DSA reconstruction scale factors obtained from the ratio of cosmic-ray data and cosmic-ray simulation, presented as functions of* η (left) *and* d_0 (right).

$\Delta t \approx 40\,\text{ns}$ between their hits in the upper and lower hemispheres (*i. e.*, the typical time it takes a relativistic particle to traverse the CMS detector). Since this is not accounted for in the efficiency measurements of Fig. 4.15, cosmic-ray samples appear with generally inferior performance. In principle, one could recover many of the muons responsible for the efficiency degradation by choosing different configurations of the DT reconstruction timing parameters. However, such fine-tuning measures are not necessary because inefficiencies resulting from the time differences between upper and lower muon legs are specific to cosmic rays and are independent of d_0 and d_z. In other words, the stability of cosmic-ray muon efficiencies as a function of the impact parameters is sufficient to characterize the DSA performance at large displacements, and the absolute efficiency scale can be fixed by an auxiliary measurement with prompt muons from collisions (*e. g.*, $Z \to \mu\mu$).

In this spirit, Fig. 4.16 compares the DSA reconstruction efficiencies in cosmic-ray data and (-40, 0) ns cosmic-ray simulation as a function of d_0. The resulting data/MC ratios define the DSA reconstruction scale factor (SF) but only characterize its d_0 behavior. The SF can be parametrized by a linear fit (green line in the figure) with an associated uncertainty of 2% (yellow band). While the DSA reconstruction SF inhibits an η dependence, it is constant across d_0 bins, as Fig. 4.17 shows, and no additional uncertainties on the η-dependence on the SF extrapolation are needed.

The absolute scale is given by the promptly-measured DSA reconstruction scale factors. These are obtained via a tag-and-probe measurement with collision-data muons from J/Ψ and Z decays and are generally dependent on muon p_T and η.[5]

Finally, the total DSA reconstruction SF is the product of the prompt SF and the extrapolated one, $\text{SF}_{\text{DSA reco}} = \text{SF}_{\text{prompt}} \times \text{SF}_{\text{extrapol}}(d_0)$.

DSA muon identification efficiency and scale factors

The basic idea behind the DSA muon identification efficiency measurement is to assess the rate of DSA legs in cosmic-ray data which pass the displaced muon ID (see Sec. 4.4.2). To obtain a suitable cosmic-ray sample of DSA muons, events filtered by the

[5]The prompt DSA reconstroction scale factors are provided centrally by the CMS "Muon Physics Objects Group" (Muon POG) following dedicated measurements of other analyzers.

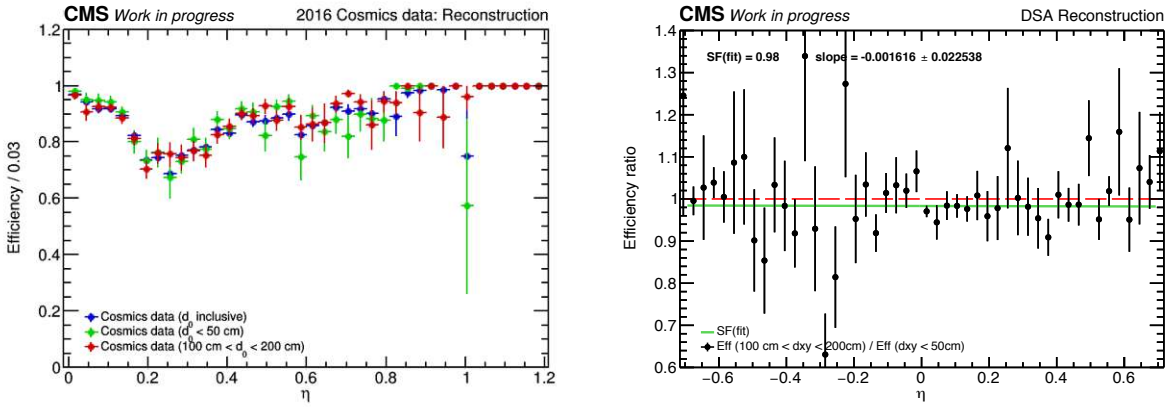


FIGURE 4.17: *d_0 stability checks of the η dependence of the DSA reconstruction.* (Left:) *DSA reconstruction efficiency inclusive in d_0 (blue), for small d_0 (green), and large d_0 (red).* (Right:) *Ratio between the DSA reconstruction efficiencies at large and small d_0, together with a linear fit (green) whose slope is consistent with 1.*

`HLT_L1SingleMuOpen` path are chosen, and the denominator of the efficiency ratio is defined as follows:

- DSA muon in the lower detector hemisphere, triggered by an `HLT_L1SingleMuOpen` path.
- Only *one* DSA muon in the lower hemisphere in total.
- Existence of a DSA leg in the upper hemisphere whose 3D opening angle is larger than $\alpha > 2.8$ with respect to the lower leg DSA muon (back-to-back requirement).
- The upper-leg DSA muon must pass the displaced muon ID.

The well-reconstructed, back-to-back DSA muons in the upper hemisphere serve the purpose of selecting events with cleanly reconstructed, individual cosmic-ray muon legs in the lower hemisphere, but they do not directly enter in the ID efficiency measurement.

This DSA-leg tagging logic via back-to-back muons is only required in cosmic-ray data but not in simulation, where one has direct access to generator-level information.

Finally, the efficiency numerator is defined as the subset of lower-leg DSA muons that fulfills the displaced muon ID.

The resulting displaced muon ID efficiencies are shown in Fig. 4.18 as functions of $p_T^{(\mathrm{DSA})}$, η, and d_0. The presented measurements are again performed in cosmic-ray data (unweighted and d_z-reweighted) and compared to signal simulation and cosmic-ray simulation (in the two time parameter settings (-40, 0) ns and (-12, 12) ns).

The efficiencies are flat in d_0 and only show a mild dependence on p_T. The η distribution again features regular efficiency drops as the result of the detector geometry, which are most strongly pronounced for the simulated signal but washed out for cosmic-ray samples (due to their different underlying d_0 and d_z profiles).

There is a clear scale difference between the signal efficiencies and the efficiencies based on cosmic-ray samples, with the signal efficiencies being larger by about 10% on average. This apparent discrepancy is due to the underlying $N_{\mathrm{hits}}(DT)$ and muon χ^2/ndof variables (see Fig. 4.19), whose modeling in signal simulation does not faithfully represent

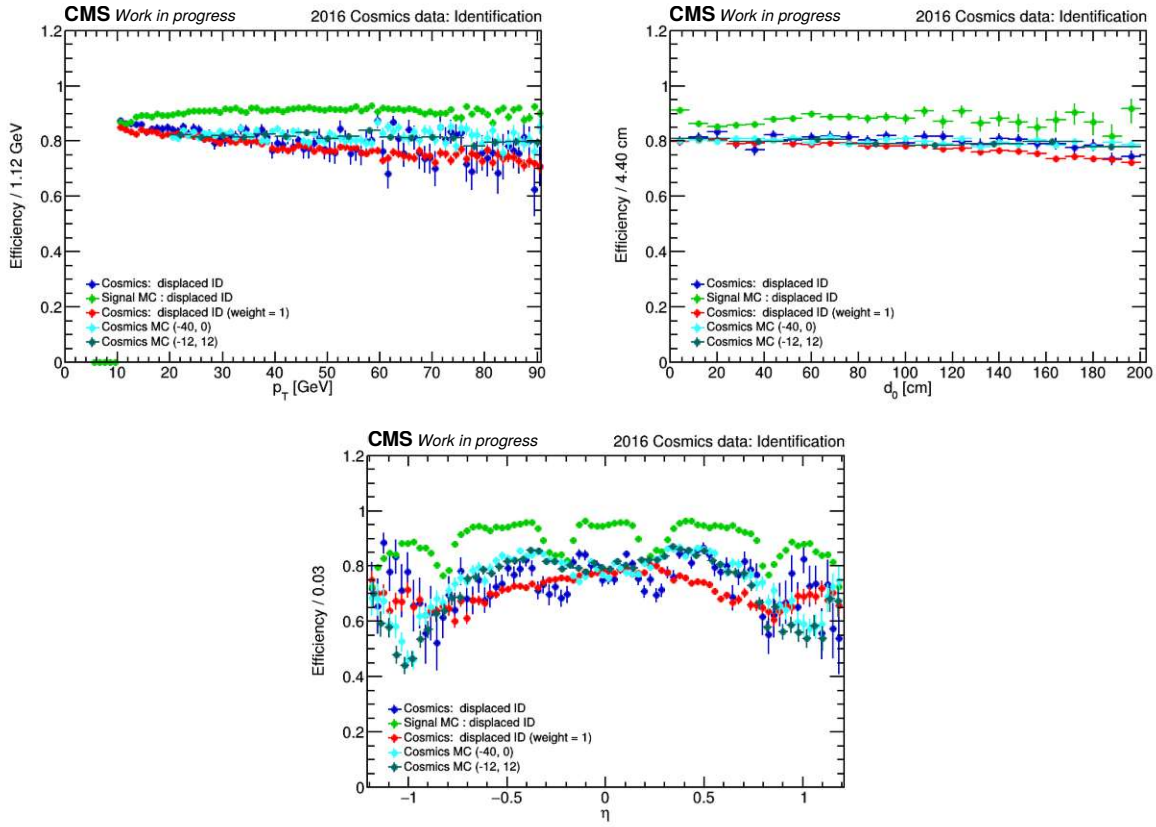


FIGURE 4.18: *DSA identification efficiencies in cosmic-ray data, cosmic-ray simulation, and signal simulation, presented as function of* p_T (top-left), d_0 (top right), *and* η (bottom).

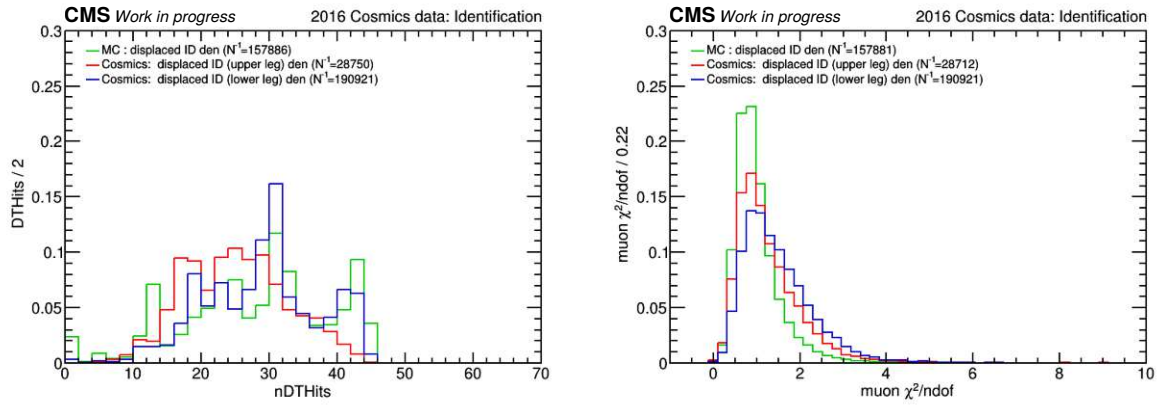


FIGURE 4.19: *Distributions of $N_{hits}(DT)$ and track-χ^2/ndof in cosmic-ray data and signal simulation.*

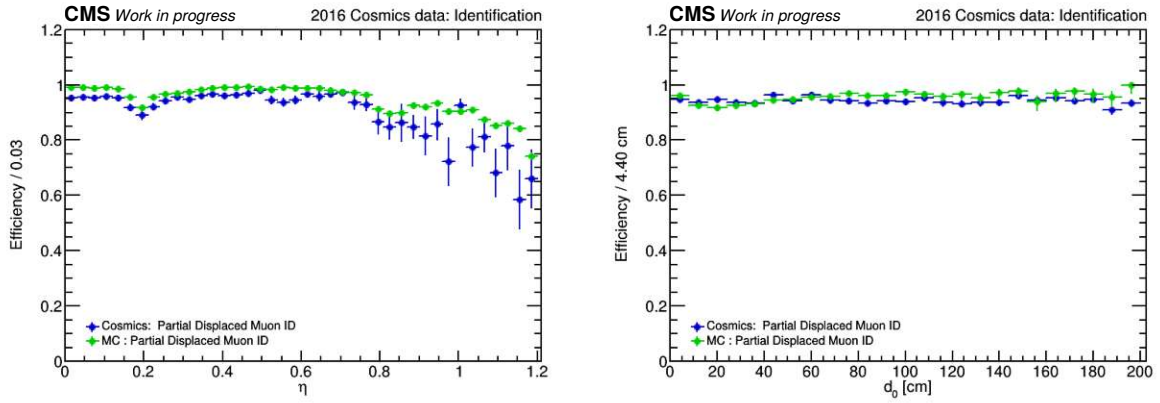


FIGURE 4.20: *Auxiliary measurement showing the DSA identification efficiencies for the "partial displaced muon ID" in cosmic-ray data (blue) and signal simulation (green) as functions of η* (left) *and d_0* (right).

the measurements with cosmic-ray data. To further support this argument, the efficiency measurement is repeated with a relaxed ID definition, in this context called the "partial displaced muon ID," which only requires $\sigma_{p_T}/p_T < 1.0$ and muon hits in at least two segments. The result of this auxiliary ID efficiency measurement, presented in Fig. 4.20, shows that the signal–data discrepancies are substantially reduced.

The differences in the $N_{\text{hits}}(DT)$ and muon χ^2/ndof variables in signal simulation compared to cosmic-ray measurements motivate a dedicated data/MC scale factor (SF) measurement performed purely in cosmic-ray samples. However, the $N_{\text{hits}}(DT)$ variable is sensitive to the modeling of the η distribution of incoming cosmic-ray muons, and simulation cannot be trusted to represent the actual cosmic-ray measurements entirely faithfully.

Again, the chosen SF strategy uses a measurement with prompt muons to fix the absolute scale of the displaced muon ID SF and uses measurements with cosmic-ray muons to extrapolate the prompt SF to the non-prompt regime of displaced muons. The prompt scale factors are measured via a "tag-and-probe" approach with muons from J/Ψ and Z decays and are generally dependent on muon p_T and η.[6] For the purpose of SF extrapolation to

[6]The prompt DSA identification scale factors are provided centrally by the CMS "Muon Physics Objects Group" (Muon POG) following dedicated measurements of other analyzers.

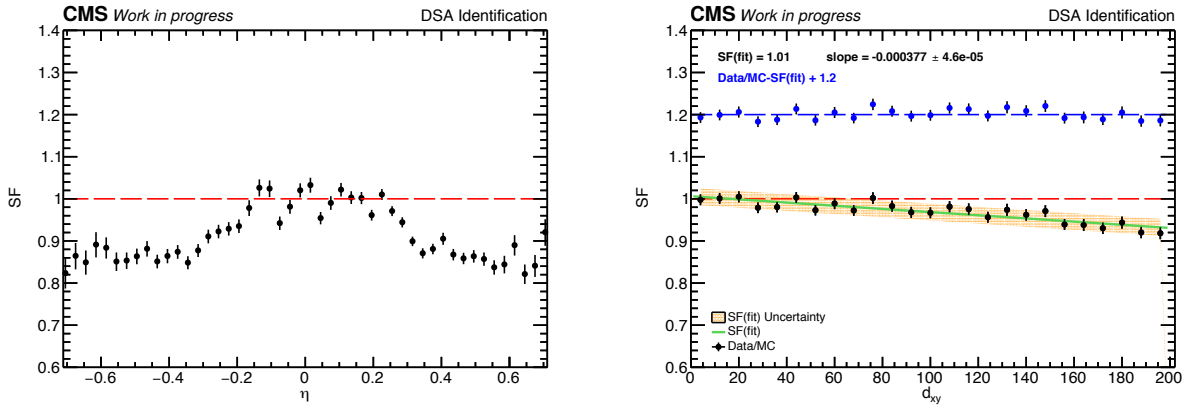


FIGURE 4.21: *DSA identification scale factors obtained from the ratio of cosmic-ray data and cosmic-ray simulation, presented as functions of* η (left) *and* d_0 (right).

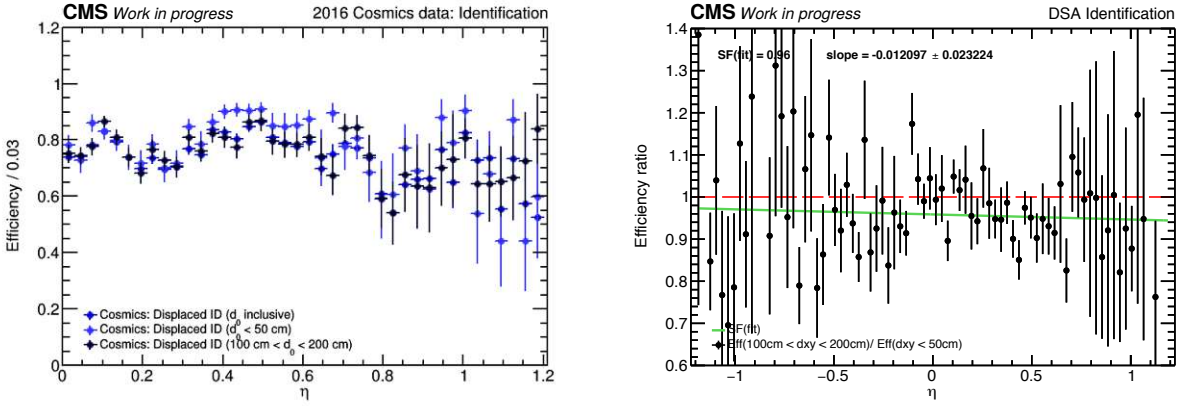


FIGURE 4.22: d_0 *stability checks of the* η *dependence of the DSA reconstruction.* (Left:) *DSA reconstruction efficiency inclusive in* d_0 *(blue), for small* d_0 *(green), and large* d_0 *(red).* (Right:) *Ratio between the DSA reconstruction efficiencies at large and small* d_0*, together with a linear fit (green) whose slope is consistent with 1.*

the displaced regime, Fig. 4.21 shows the cosmic-ray data to simulation SFs as a function of d_0. They inhibit a mild d_0 dependence, which is parametrized by a fitted linear function (green line in the figure) with an uncertainty of 2% (yellow band). The linear fit indicates a deviation from the prompt SF of about -3.5% or -7% at $d_0 = 100\,\text{cm}$ or $d_0 = 200\,\text{cm}$, respectively.

The general η dependence of the DSA identification SF is independent of d_0, as Fig. 4.22 proves. Therefore, no additional η-dependent SFs or uncertainties are assigned for the d_0-extrapolation based on the cosmic-ray results (*i. e.*, the η dependence is wholly determined by the measurement of the prompt SF).

$p_T^{\text{(DSA)}}$ scale and resolution

The $p_T^{(\text{DSA})}$ scale and resolution measurement defines relative q/p_T residuals (with q being the muon charge) similar to the online-p_T scale and resolution measurements presented in

Sec. 4.4.3 and 4.4.4. In contrast to the previous measurements, which used the offline $p_T^{(\mathrm{DSA})}$ as a reference, the results in this section attempt to measure the $p_T^{(\mathrm{DSA})}$ performance itself. Therefore, the notion of "true p_T" needs to be redefined.

Using cosmic-ray muons, one can leverage the fact that the upper and lower legs of a given cosmic-ray muon are tracks reconstructed from the *same* physical muon of a particular energy. This motivates the definition of the q/p_T residual $R_1^{(\mathrm{cosmic})}$, which treats the upper-leg $p_T^{(\mathrm{DSA})}$ as a reference for the probed lower-leg $p_T^{(\mathrm{DSA})}$.

$$R_1^{(\mathrm{cosmic})}(p_T, d_0) = \frac{1}{\sqrt{2}} \frac{(q/p_T^{(\mathrm{DSA})})^{(\mathrm{upper})} - (q/p_T^{(\mathrm{DSA})})^{(\mathrm{lower})}}{(q/p_T^{(\mathrm{DSA})})^{(\mathrm{lower})}} \tag{4.4.3}$$

However, the $R_1^{(\mathrm{cosmic})}$ definition does not account for muons from cosmic rays losing energy via material interactions when traversing the detector. Such energy losses result in intrinsic differences between the muon p_T values in the upper and lower detector hemisphere. Moreover, the CMS particle reconstruction assumes particles are propagating from the inner to the outer parts of the detector volume, but not in the opposite direction. An upper-leg reconstructed track coming from a typical cosmic-ray muon (hitting the detector from above) will traverse the detector in the opposite direction than the reconstruction algorithms assume. Any particle energy losses along the track will therefore appear as energy *gains* of an inside-out-reconstructed track – an unusual scenario from the reconstruction point of view that is not guaranteed to produce meaningful results consistently.

However, the energy loss in detector material is only relevant for $p_T \lesssim 30\,\mathrm{GeV}$, a regime in which the typical muon momenta correspond to relatively large energy loss (dE/dx). Hence, the measurement of $R_1^{(\mathrm{cosmic})}$ is exclusively performed in the momentum domain $p_T^{(\mathrm{DSA})} > 30\,\mathrm{GeV}$.

For lower-energy muons ($p_T^{(\mathrm{DSA})} < 30\,\mathrm{GeV}$), the upper-leg muons cannot serve as a valid reference anymore, and a different approach is chosen.

Correspondingly, low-p_T muons are measured by selecting lower-leg DSA muons in the lower hemisphere, which are also reconstructed in the CMS tracker. These tracker-based muon tracks (global muons with $\chi^2/\mathrm{ndof} < 10$ that require hits in six tracker layers) benefit from vastly superior $p_T^{(\mathrm{PAT})}$ resolution. Therefore, they are used as a proxy for the "true p_T" of the lower-hemisphere DSA muons, and the associated residual is defined as

$$R_2^{(\mathrm{cosmic})}(p_T, d_0) = \frac{(q/p_T^{(\mathrm{DSA})})^{(\mathrm{lower})} - (q/p_T^{(\mathrm{PAT})})^{(\mathrm{lower})}}{(q/p_T^{(\mathrm{PAT})})^{(\mathrm{lower})}} \,. \tag{4.4.4}$$

This method is used to study the $p_T^{(\mathrm{DSA})}$ resolution in the low-energy range of $10\,\mathrm{GeV} < p_T < 30\,\mathrm{GeV}$ with cosmic-ray data. However, since this strategy requires tracker-reconstructed versions of the DSA muons, the measurement is constrained to small muon displacements well within the tracker volume.

Like the previously discussed results, the overall measurement will be characterized by two distinct methods, one highly effective in the prompt-like regime ($R_2^{(\mathrm{cosmic})}$) and the other suited for the extrapolation of the prompt result to the non-prompt domain ($R_1^{(\mathrm{cosmic})}$).

Finally, a third residual definition, $R_2^{(\mathrm{MC})}$, is introduced for the measurement of $p_T^{(\mathrm{DSA})}$ scale and resolution in signal simulation, which has direct access to the "true p_T" via

TABLE 4.6: *Results of the σ_{extra} measurements in three p_T regimes, based on the relative residuals $R_2^{(cosmic)}$ and $R_2^{(MC)}$.*

p_T	$\sigma(R_2^{(\text{cosmic})})$	$\sigma(R_2^{(\text{MC})})$	$\sigma_{\text{extra}} \pm \Delta\sigma_{\text{extra}}$
15 GeV	0.189	0.151	0.114 ± 0.02
20 GeV	0.240	0.192	0.144 ± 0.02
50 GeV	0.271	0.270	0.023 ± 0.03

generator-level information:

$$R_2^{(\text{MC})}(p_T, d_0) = \frac{q/p_T^{(\text{DSA})} - q/p_T^{(\text{GEN})}}{q/p_T^{(\text{GEN})}} \,. \tag{4.4.5}$$

The results of all three measurement strategies are shown in Fig. 4.23. They are presented in terms of the fitted widths of the relative residuals, $\sigma(R_1^{(\text{cosmic})})$, $\sigma(R_2^{(\text{cosmic})})$, and $\sigma(R_2^{(\text{MC})})$, all as a function of $p_T^{(\text{DSA})}$. Around $p_T^{(\text{DSA})} \approx 30\,\text{GeV}$, a regime where all methods are applicable, the fitted widths range between 20% and 25%. At larger $p_T^{(\text{DSA})}$, the measurements are purely based on $R_1^{(\text{cosmic})}$ and $R_2^{(\text{MC})}$, and yield widths of up to 30% for $p_T^{(\text{DSA})} > 100\,\text{GeV}$. At smaller $p_T^{(\text{DSA})}$, $p_T^{(\text{DSA})} < 30\,\text{GeV}$, the $R_1^{(\text{cosmic})}$ measurement ceases to apply, and the $R_2^{(\text{cosmic})}$ and $R_2^{(\text{MC})}$ widths approach 15–20% at $p_T^{(\text{DSA})} \approx 10\,\text{GeV}$.

The small differences between $\sigma(R_2^{(\text{cosmic})})$ and $\sigma(R_2^{(\text{MC})})$ are used to define a $p_T^{(\text{DSA})}$ resolution uncertainty, $\sigma_{\text{extra}} = \sqrt{\sigma^2(R_2^{(\text{cosmic})}) - \sigma^2(R_2^{(\text{MC})})}$. The computed values of this resolution uncertainty are summarized in Tab. 4.6 in three p_T regimes.

The fitted widths of all three relative residuals are also studied as a function of d_0 for muons with $p_T^{(\text{DSA})} > 30\,\text{GeV}$ (see Fig. 4.23). The $R_2^{(\text{cosmic})}$ measurement is restricted to the tracker volume and therefore only provides a result for small d_0 values. The only method to directly assess the $p_T^{(\text{DSA})}$ performance for highly non-pointing muons in recorded data is the cosmic-ray measurement $R_1^{(\text{cosmic})}$. The width of $R_1^{(\text{cosmic})}$ varies between 0.18 and 0.15 in the range $d_0 < 100\,\text{cm}$.

Based on this variation, an additional non-prompt resolution uncertainty $\sigma_{\text{extrapolation}} = \sqrt{0.18^2 - 0.15^2} = 0.01$ is added to σ_{extra} to account for all d_0-dependent effects. However, since $\sigma_{\text{extra}} \gg \sigma_{\text{extrapolation}}$, the dominant source of uncertainty comes from the determination of σ_{extra}.

DSA dimuon vertexing performance

The upper and lower legs from a cosmic-ray muon can undergo the same common vertex (CV) fit as the dimuons targeted in the search (see also Sec. 4.3). It is therefore possible, at least in principle, to use cosmic-ray data also for the study of the offline dimuon vertexing behavior. In fact, about 80% of cosmic-ray muons with $d_0 < 250\,\text{cm}$ form a common dimuon vertex, as Fig. 4.24 (left) shows.

Common vertices formed from cosmic-ray muon legs span over a broad range of L_{xy} and easily cover typical displacements expected for CVs in signal processes. This argument is supported by Fig. 4.25 (left), which compares the dimuon L_{xy} distribution in cosmic-ray data (in different d_0 bins) with signal simulation. Unlike in signal processes, where usually $L_{xy} > d_0$ (due to the properties of the underlying model), the L_{xy} and d_0 is highly correlated

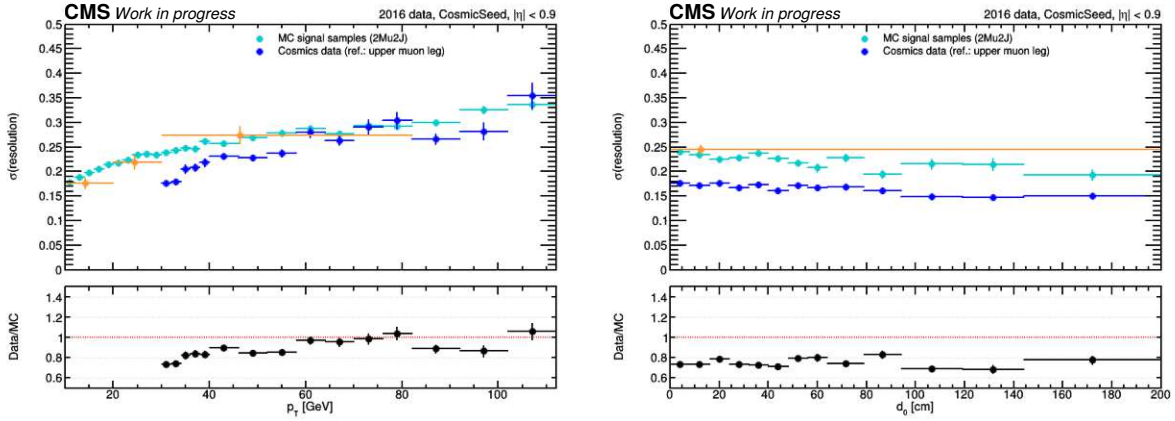


FIGURE 4.23: *Distributions of the fitted widths of the relative residuals* $R_1^{(cosmic)}$ (blue), $R_2^{(cosmic)}$ (orange), *and* $R_2^{(MC)}$ (teal), *presented as functions of* $p_T^{(DSA)}$ (left) *and* d_0 (right).

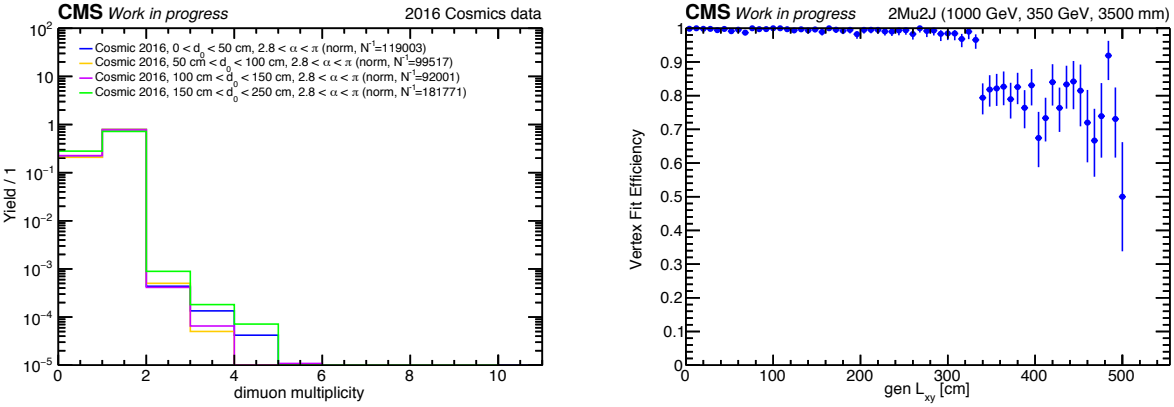


FIGURE 4.24: Left: *Multiplicity of dimuon vertices in cosmic-ray data in different* d_0 *selections.* Right: *Reconstruction efficiency of dimuon vertices in signal simulation as a function of the generator-level* L_{xy}.

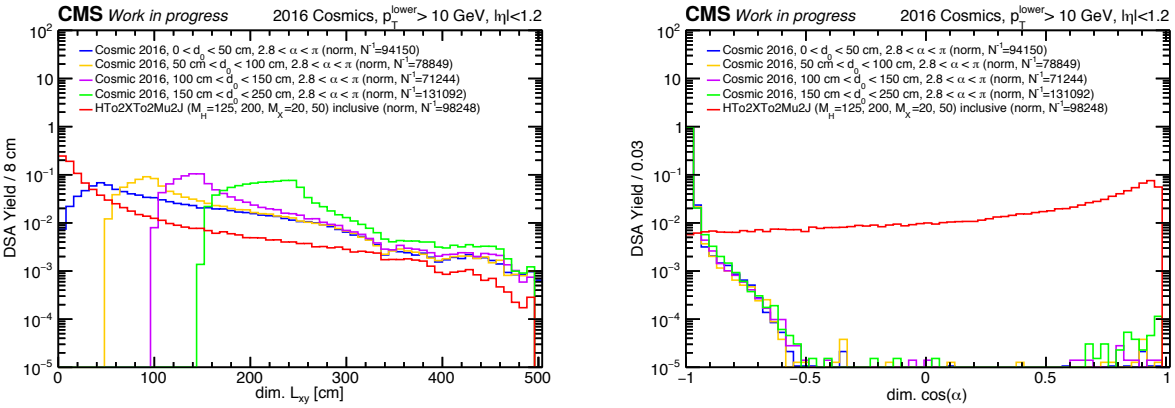


FIGURE 4.25: *Distributions of dimuons as functions of the common-vertex variables* L_{xy} (left) *and* $\cos\alpha$ (right) *in different* d_0 *bins. Cosmic-ray data are compared to signal simulation* (red).

in cosmic rays, with the majority of cosmic dimuons having $L_{xy} \approx d_0$ due to their near-vertical track alignment in the detector. A consequence of this L_{xy}–d_0 correlation is that the cosmic-ray L_{xy} distribution in Fig. 4.25 (left) appears truncated when the different d_0 cuts are applied.

Another essential difference between cosmic-ray dimuons and signal dimuons is manifest in the angular distribution between the dimuon legs (see Fig. 4.25 (right)). While $\cos\alpha$ peaks at -1 in cosmic-ray dimuons (reflecting their back-to-back nature), the signal $\cos\alpha$ distribution continuously increases towards $+1$. Since cosmic-ray muons require the common-vertex fitter to construct a CV along a pair of approximately (anti-)parallel muon tracks (unlike signal dimuons, which come in various $\cos\alpha$ configurations), a worse dimuon vertexing performance is generally expected for cosmic-ray data.

The differences in $\cos\alpha$ and the strong L_{xy}–d_0 correlation impose some limitations on the cosmic-ray dimuon measurements, which cannot be expected to reproduce signal-dimuon-like performance across all relevant signal topologies entirely faithfully. Still, cosmic-ray data provide a unique opportunity to probe the dimuon vertexing performance in a challenging topology and the otherwise inaccessible displacement domain beyond the inner tracker up to several meters.

Fig. 4.26 (left) shows the vertex χ^2/ndof distributions in cosmic-ray data (in different d_0 bins) and signal simulation. The behavior of the cosmic-ray dimuon vertexing is stable with d_0, up to $d_0 \approx 330\,\text{cm}$ (which roughly matches the radius of the solenoid magnet). Due to the L_{xy}–d_0 correlation in cosmic-rays, this stability statement is also true for L_{xy}. Given that the efficiency of the Run-2 signal triggers is essentially zero for $L_{xy} > 330\,\text{cm}$, no further attempts are made to validate the dimuon vertex performance in this highest-displacement regime.

Fig. 4.24 (right) shows the dimuon vertex efficiency as a function of L_{xy} in simulated signal. A rapid drop in efficiency at $L_{xy} > 330\,\text{cm}$ is observed, in accordance with the vertex χ^2/ndof performance measurement in cosmic-ray dimuons. The poor dimuon vertexing performance, too, generally discourages the search for displaced dimuons beyond $L_{xy} \gtrsim 330\,\text{cm}$.

Next, Fig. 4.26 (right) presents the invariant mass distribution of cosmic-ray dimuons (in different d_0 bins) and simulated signal dimuons. The reconstructed mass of dimuons formed from the legs of cosmic-ray muons depends on the energy of the incoming cosmic-ray

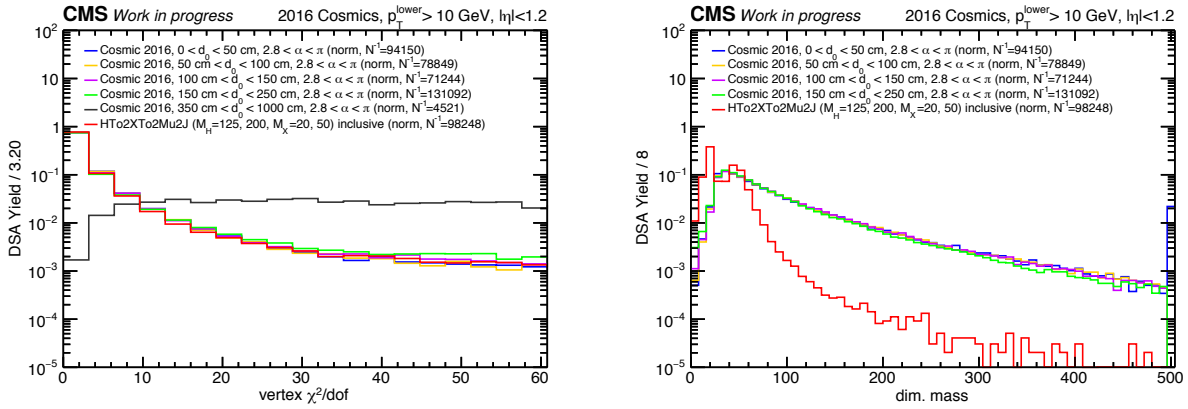


FIGURE 4.26: *distributions of dimuons as functions of the common-vertex variables* $\chi^2/ndof$ (left) *and dimuon invariant mass* (right) *in different* d_0 *bins. Cosmic-ray data are compared to signal simulation* (red).

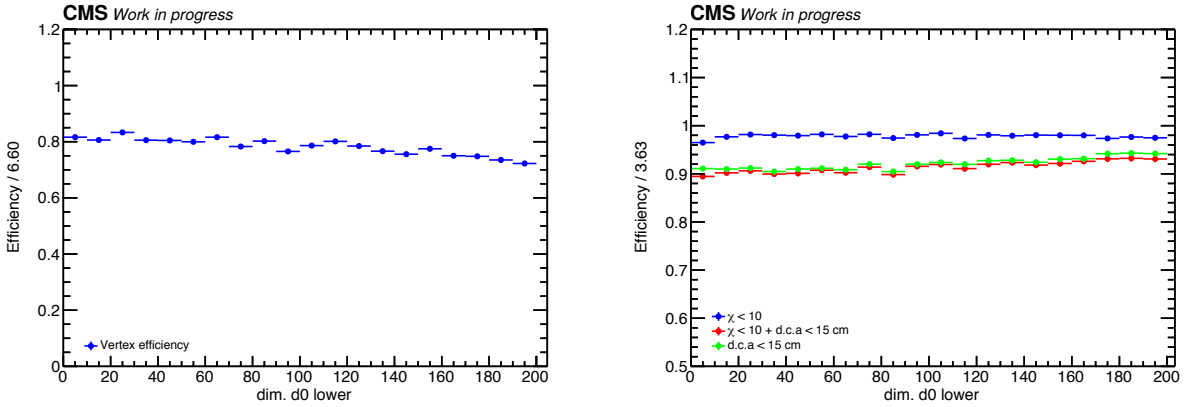


FIGURE 4.27: *Reconstruction efficiency* (left) *and identification efficiency* (right) *of dimuon vertices in cosmic-ray data as a function of the lower-leg* d_0.

muons but not on their displacement. Therefore, the mass spectrum obtained from cosmic-ray dimuons is expected to be independent of d_0 and different from the mass distribution from signal dimuons. The mass spectra in Fig. 4.26 (right) confirm these expectations. Moreover, these results are in agreement with the measured d_0 stability of the DSA muon resolution (see Sec. 4.4.6).

The dimuon vertex reconstruction efficiency as a function of the lower-leg d_0 in cosmic-ray data is shown in Fig. 4.27 (left). In accordance with the dimuon multiplicity measurement (Fig. 4.24), the efficiency is about 80% and virtually independent of d_0 in the regime relevant to the search.

Finally, Fig. 4.27 (right) also presents measurements of the displaced dimuon vertex identification efficiency in cosmic-ray data. Each of the cuts in the dimuon selection is probed separately. The vertex ID efficiency is approximately 90% when the cuts on DCA and vertex χ^2/ndof are applied. In accordance with the measurements in Fig. 4.26, the performance is again stable with d_0.

These results suggest that the study of dimuon vertex performance in the prompt or prompt-like regime can serve as a reasonable estimate of the vertexing performance for non-pointing dimuons (up to $d_0 \lesssim 330\,\text{cm}$).

4.5 Event selection

One of the core challenges of an analysis like the search for displaced dimuons is to find a set of uni-, bi-, and multivariate selection criteria that are capable of separating signal from background events in the high-dimensional phase space spanned by the analysis variables. The goal is to find a set of requirements that suppress any known background events efficiently while keeping the loss of signal at an acceptable rate.

This section describes the event selection criteria of the search, which were carefully optimized using signal and background MC samples as well as recorded data in dedicated signal-free control regions.

4.5.1 Event preselection

Events used for the analysis are filtered by the signal triggers described in Sec. 4.1.

Further, each event must contain at least two DSA muons with the following quality requirements:

- $N_{\text{stations}}(\text{CSC}) + N_{\text{stations}}(\text{DT}) > 1$
- $N_{\text{hits}}(\text{CSC}) > 12$

These loose quality criteria ensure properly reconstructed muons with reasonable p_T resolution and charge assignment.

Further, the pair(s) of HLT muons that triggered the event must match offline-reconstructed DSA muons with $p_T > 10\,\text{GeV}$ and $|\eta| < 2.0$. This HLT-DSA matching is performed in a cone, $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.4$, where the ΔR threshold was chosen with a good signal–background tradeoff in mind. Not only does this matching confirm the trigger decision on the level of the events, but it also helps to suppress events where muons unrelated to signal candidate events fired the trigger. Explicit HLT-DSA matching only serves the purpose of pre-selecting events and is not enforced in the later muon selection, thus allowing sensitivity to signals with larger dimuon multiplicity.

4.5.2 DSA-to-PAT muon association

One of the key characteristics of this search is the association of PAT muons (see Sec. 4.3) with DSA muons wherever possible and the subsequent partitioning of the dimuon event content into three distinct vertex categories: DSA-DSA, DSA-PAT, and PAT-PAT dimuons. The PAT-based dimuons profit from the superior tracking resolution in the CMS tracker, enabling the search to identify and reject prevalent SM backgrounds efficiently. DSA-DSA dimuons, on the other hand, benefit from the fact that the regime outside the tracker is virtually free from pp-collision background, with the only background events generally resulting from misreconstructed tracks. This early partitioning of dimuons allows for a holistic approach to CMS muon reconstruction and enables the best-possible detector performance.

The DSA–PAT muon association algorithm was designed with great care. It was optimized using signal and background simulations as well as collision data in signal-free domains (*e. g.*, at $|\Delta\Phi| > \pi/2$). The starting point of the association procedure are all preselected DSA muons. Subsequently, the algorithm tries to identify whether a given DSA muon can be associated with a muon that produced hits in the tracker (*i. e.*, a PAT muon). If such a PAT muon is found, the DSA muon is replaced by it, and the given DSA muon

is no longer part of the DSA-DSA dimuon analysis. The identification of associated PAT muons employs a combination of segment-based and proximity-matching criteria that have been optimized for finding PAT matches even for DSA muons with misreconstructed η and for being able to ensure matching to high-quality PAT muons while disambiguating between multiple PAT matches if needed. The DSA–PAT replacement procedure is described in full detail in [36].

In summary, the DSA–PAT association and replacement are highly efficient in suppressing background from pp collisions for DSA-DSA dimuons while preserving most true signal dimuons: Simulations show that more than 90% of signals outside the tracker ($L_{xy}{}^{(\mathrm{gen})} \gtrsim 65\,\mathrm{cm}$) are preserved while the background is reduced by a factor 5×10^4. [36]

Since each vertex category features dimuons with different tracking resolutions and different backgrounds, each category requires a dedicated analysis after the DSA–PAT replacement. This work targets the study of DSA-DSA dimuons, which are sensitive to the longest-lifetime LLP decays in a mostly background-free regime. However, the DSA-DSA analysis is part of a bigger project aiming to study PAT-based dimuons and ultimately combine all dimuon categories into a final and unified result. (The analysis of the PAT-based vertex categories, though, is beyond the scope of this work, and they will be presented in separate publications.)

4.5.3 DSA muon identification

DSA muons that remain after the DSA–PAT replacement procedure (see Sec. 4.5.2) – *i. e.*, DSA muons for which no PAT associations were found – are subject to further identification criteria. This ensures suppressing further any remaining backgrounds from misreconstructed (prompt-like, in truth) muons.

The DSA identification criteria presented in this section were optimized using simulated signal and background as well as collision data in signal-free regions (*e. g.*, $|\Delta\Phi| > \pi/2$, or dimuons with equally-charged legs).

First, a set of minimal quality criteria is enforced:

- $p_T > 10\,\mathrm{GeV}$
- $N_{\mathrm{hits}} > 12$.

Then, DSA muons are identified with the following criteria:

- $\sigma_{p_T}/p_T < 1.0$ (with σ_{p_T} being the uncertainty of the track fit): This ensures acceptable p_T resolution and charge measurement.
- Track-χ^2/ndof < 2.5: This minimum track quality requirement helps to further reduce backgrounds from poorly reconstructed tracks.
- $N_{\mathrm{hits}}(\mathrm{DT}) > 18$ or $N_{\mathrm{hits}}(\mathrm{CSC}) > 0$ (similar purpose as the track-χ^2/ndof requirements)
- Muon time with respect to the bunch-crossing time, $|t_{\text{inside-out}}| < 12\,\mathrm{ns}$: This cut serves multiple purposes (see below).
- Muon direction of travel must be "inside-out": a powerful handle to reduce cosmic-ray background (see below).

The DSA muon timing requirement serves three purposes.

First, a simple cut $|t_{\text{inside-out}}| < 12\,\text{ns}$ rejects events in which the L1 trigger *pre-fired*. A "normal" event contains two collision muons resulting in the times of the trigger firing and the hits in the tracker and muon stations all to be aligned. However, by accident, the pipelined trigger electronics may create a positive L1 decision that is 25 ns earlier than normal. In such a rare case, muon-system hits of promptly-produced muons are still recorded, but the tracker hit information is not (at least not in the given event). This results in the offline muon reconstruction to see a signature not unlike one expected from a typical LLP signal (*i. e.*, tracks in the muon system without any associated tracker hits). Fortunately, trigger-prefired muons are usually 25 ns late compared to "normal" muons since their timestamp comes from the positive L1 decision that happened 25 ns too early. Therefore, a simple cut on the muon time can effectively reject such rare pre-firing events.

Second, the DSA timing requirement vetoes muons from "out-of-time collision muons." These are muons that did not trigger the event but are still within the readout range of a triggered event. Similar to pre-firing events, such muons can produce false LLP signatures with muon-system-only muons.

Third, DSA muon time can be used to suppress cosmic-ray background. As explained in Sec. 4.4, a cosmic-ray muon traversing the CMS detector is reconstructed as two back-to-back DSA muons. Since cosmic-ray muons hit the detector uniformly, an off-center cosmic-ray muon can easily mimic a displaced DSA-DSA signal dimuon. Most cosmic-ray events are effectively rejected by dedicated criteria (see Sec. 4.5.5), but a small fraction can even pass these requirements. The used DSA muon timing cut further suppresses surviving cosmic-ray events by leveraging the fact that the time of at least one cosmic-ray leg will be late or early with respect to the bunch-crossing time.

In standard CMS software, timing information is only available for standalone muons (SA muons) but not for DSA muons. The first iteration of the 2016 DSA-DSA analysis [36] used the time of proximity-matched SA muons as a proxy for the DSA muon time (a strategy that has its limitations since not all DSA muons have SA matches). For this work, a custom implementation of DSA muon timing was undertaken, making the time information available for all reconstructed muons in the analysis.

Timestamps are obtained under the hypothesis that muons traverse the detector from the IP outwards. The corresponding time variable is therefore called $t_{\text{inside-out}}$. Fig. 4.28 shows that the time-at-vertex distributions have peaks near zero and core resolutions of 1.4–1.6 ns for the combined prompt-like and non-prompt signal dimuons from the simulated decay $H \to XX \to 2\mu\,2$. The non-prompt signals show marginally larger tails than the prompt ones, but in virtually all cases (99.9%), the signal dimuons have timestamps that fall within the $|t_{\text{inside-out}}| < 12\,\text{ns}$ cut. In contrast, 1 and 5 opposite-charge dimuons in recorded data at $|\Delta\Phi| > \pi/2$ are rejected in 2016 and 2018, respectively, which pass all other cuts.

Moreover, the timing measurement is used to define the muon direction of travel. There are two internal time measurements, $t_{\text{inside-out}}$ and $t_{\text{outside-in}}$, each based on a hypothesis of particle direction. The final muon direction is determined using the uncertainties, σ, of these respective time measurements:

- The "inside-out" direction is associated with $\sigma(t_{\text{inside-out}}) < \sigma(t_{\text{outside-in}})$ and encoded as the numerical value $+1$.
- The "outside-in" direction is associated with $\sigma(t_{\text{inside-out}}) > \sigma(t_{\text{outside-in}})$ and encoded as -1.

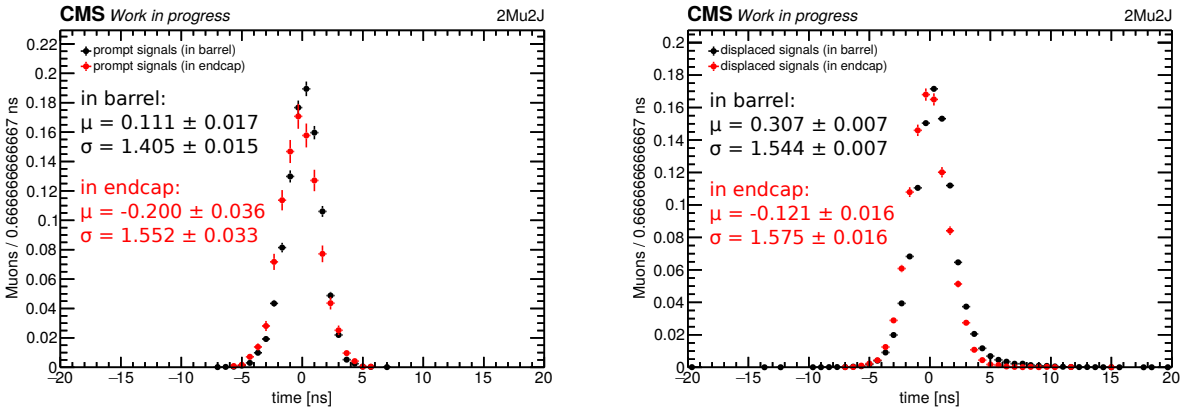


FIGURE 4.28: *Normalized distribution of the DSA time variable, $t_{\text{inside-out}}$, for the $H \to 2X \to 2\mu$ signal samples with the shortest* (left) *and the longest* (right) *lifetimes. Muons are categorized into barrel ($|\eta| < 0.9$) and endcap muons ($|\eta| > 1.2$) and are shown in black and red, respectively. The legends of each figure quote the mean and standard deviation (μ and σ) of Gaussian fits to the distributions.*

- The internal direction flag can also assume a third value (encoded as 0), indicating an insufficient number of internal timing measurements. However, this scenario occurs very rarely in signal samples (1 out of 1×10^5 DSA muons).

Again, the muon direction is a powerful handle to suppress cosmic-ray muons while preserving almost all of the signal. This circumstance is shown in Fig. 4.29, which compares the minimum differences between $\sigma(t_{\text{inside-out}})$ and $\sigma(t_{\text{outside-in}})$ in signal dimuons and in dimuons from events in a cosmic-ray-enriched selection (inverted $\cos\alpha$ cut and no timing cuts).

4.5.4 Criteria for muon pairing and dimuon vertex fitting

Dimuon candidates are formed from pairs of muons that pass the selection and identification criteria described in the previous sections. For this, all $n(n-1)/2$ possible muon pairings among n selected DSA and PAT muons are considered. Ultimately, the goal is to put criteria in place that correctly identify, out of all muon pairs in the event, those dimuons that correspond to the signal. Such a set of *pairing criteria*, however, is delicate to design, especially in events with more than one signal dimuon (*e. g.*, in events where two LLPs decay to a pair of muons each) and in the presence of muons that are not directly related to the signal (*e. g.*, pile-up muons or muons from other decays in the event).

The first step towards a well-performing set of pairing criteria is to filter the muon pairs by requiring a minimum distance of closest approach (DCA) of the two muon tracks helically extrapolated in the magnetic field (DCA $< 15\,\text{cm}$). This loose cut ensures that the common vertex fit in the next step is not attempted on a pair of widely separated tracks. It is between 92% and 99% efficient in signal, depending on the particle masses and lifetimes of the simulated decays, as Fig. 4.30 shows.

Next, a common vertex is attempted to be fitted [93, 110]. The efficiency of the common-vertex fit is close to 100% for dimuons with $L_{xy} < 320\,\text{cm}$, but decreases for L_{xy} values beyond. These inefficiencies for highly displaced muons are not a real problem for the analysis, given that the L1 trigger efficiency is essentially zero in this regime (see Fig. 4.7).

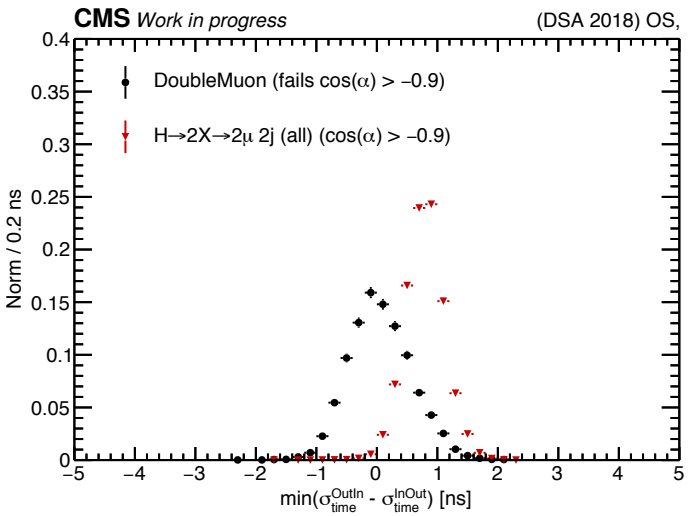


FIGURE 4.29: *Normalized distributions of the minimum difference between the uncertainties in the two different time measurements, $t_{inside\text{-}out}$ and $t_{outside\text{-}in}$ in dimuons of the combined $H \to 2X \to 2\mu$ signal samples* (red) *and a 2018 data sample that is enriched in cosmic-ray muons* (black)*. Positive values of the shown variable designate the "inside-out" direction measurement.*

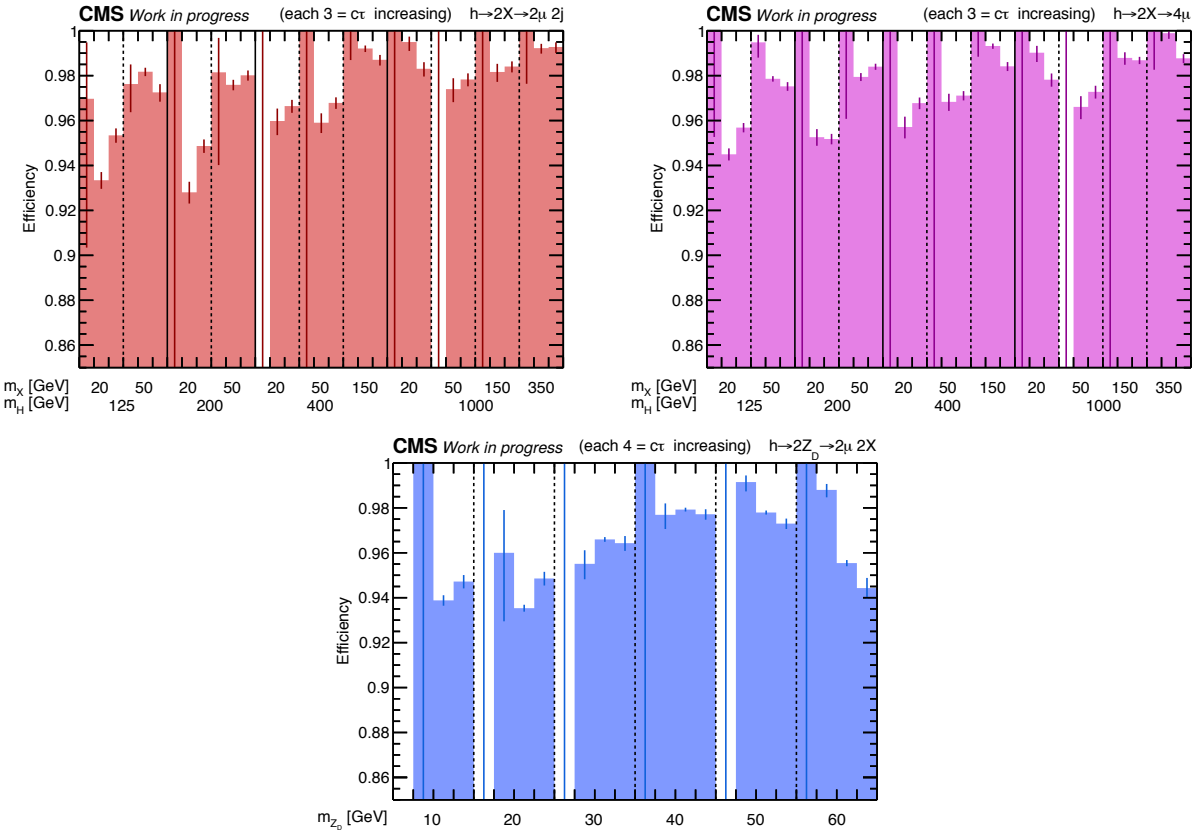


FIGURE 4.30: *Efficiencies of the DCA* $< 15\,\text{cm}$ *requirement in the* $H \to 2X \to 2\mu$ (top left), $H \to 2X \to 4\mu$ (top right), *and* $H \to 2Z_D \to 2\mu\,2X$ (bottom) *signal samples. The efficiency calculation is based on dimuons that pass the full analysis selection except for the DCA cut.*

The common-vertex fit modifies the input tracks such as to be consistent with originating from the displaced vertex. For $L_{xy} < 320\,\text{cm}$, such refitted tracks offer better p_T resolution than the original tracks. Therefore, unless otherwise specified, kinematic quantities used in the analysis (*e. g.*, p_T, dimuon invariant mass,...) are obtained from the refitted DSA tracks.

If there are more than two sufficiently well-measured muons in an event, multiple dimuons with a common vertex can be formed, at least in principle. This motivates the definition of pairing criteria that aim to select the reconstructed dimuons corresponding to the signal dimuons.

The developed pairing criteria are described in detail in [36]. In short, they utilize a carefully selected combination of dimuon and muon-pair requirements to pick between zero and two dimuons in each dimuon category (DSA-DSA, PAT-PAT, and DSA-PAT dimuons). This means that, after the pairing criteria have been applied, an event can have up to six selected dimuons (two for each dimuon category). However, in practice, the number of selected dimuons in an event rarely exceeds two or three. The designed algorithm implementing the pairing criteria was found to be highly efficient in signal (99.6% in samples with two final-state muons and 91.1% in ones with four final-state muons [36]).

Finally, dimuons chosen by the pairing criteria must fulfill some minimal quality criteria to suppress background from signal-unrelated muons. Most notably, the analysis implements an upper cut on the vertex χ^2/ndof of the common-vertex fit, $\chi^2/\text{ndof} < 10$. This threshold is deliberately loose because distributions of χ^2/ndof are generally poorly reproduced in simulation, and a cut too stringent is prone to affect signals negatively.

The chosen $\chi^2/\text{ndof} < 10$ requirement keeps more than 96% of average signal but rejects 2 out of 5 background events (see Fig. 4.31) in the 2018 same-sign dimuon control region at $|\Delta\Phi| < \pi/4$ that would otherwise pass the full selection.

4.5.5 Rejection of cosmic-ray muons

Muons created in the upper layers of the atmosphere through interactions between cosmic radiation and nuclei of the atmosphere constitute a continuous flux of charged particles raining down to the Earth's surface. If a cosmic-ray muon crosses the CMS detector during data-taking, it is usually reconstructed as two back-to-back muon tracks, one in the upper and one in the lower detector hemisphere, as explained in Sec. 4.4.1. Generally, such a pair of cosmic-ray muon legs is naturally displaced from the PV and can mimic a signal from an LLP decay. Cosmic-ray muons, therefore, constitute an important source of background for the analysis.

Some of the cuts already described in previous sections help in reducing cosmic-ray background. In particular, the time and direction cuts (see Sec. 4.5.3) successfully reject most cosmic-ray muons thanks to the fact that (a) at least one of their legs is typically early or late in time with respect to the nominal bunch-crossing time and (b) one leg is facing in the "outside-in" direction. Further, the DCA requirement helps to suppress dimuons that are formed from the two legs of different cosmic-ray muons in events with cosmic-ray-induced muon showers.

A few cosmic-ray events, however, still pass all these requirements. Having studied cosmic-ray muons extensively (see Sec. 4.4), it is rather straightforward to reject the remaining cosmic-ray background efficiently by using an upper threshold on the 3D opening angle of the reconstructed muons, minimal quality requirements for the cosmic-ray muon legs, as well as further timing cuts.

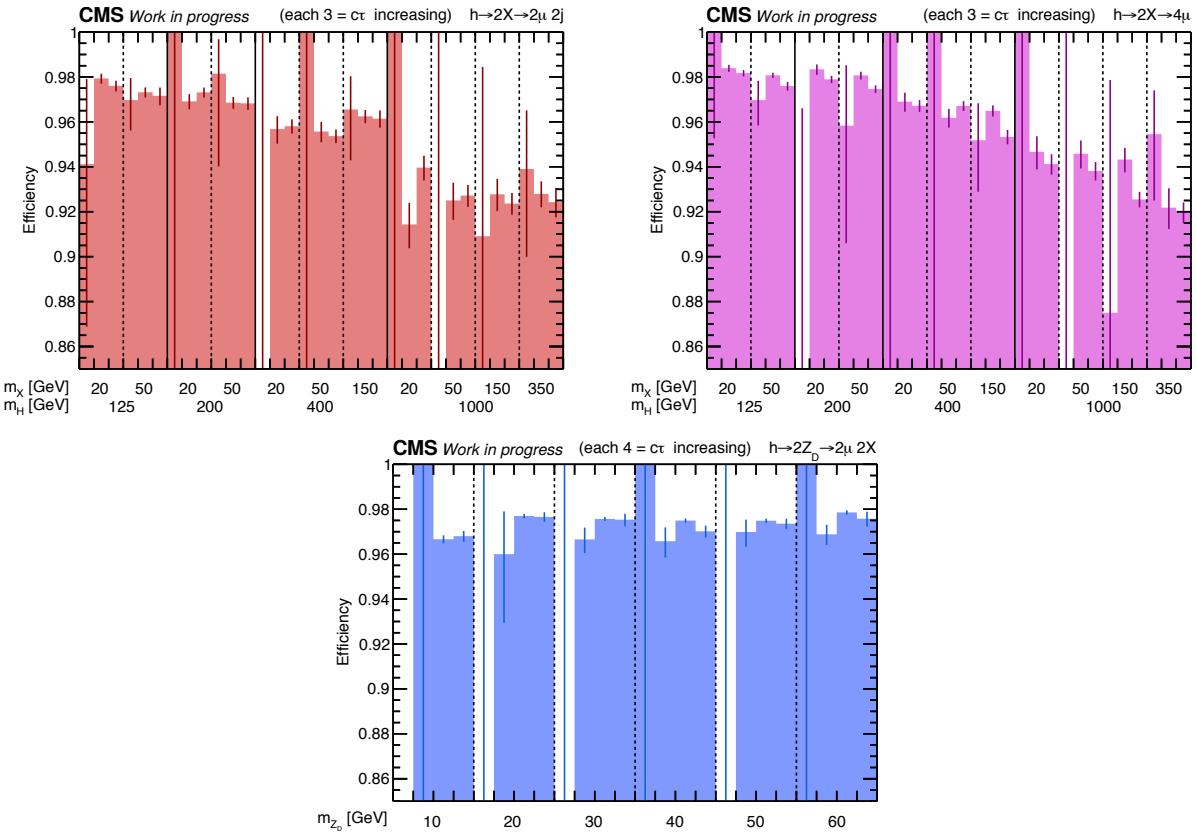


FIGURE 4.31: *Efficiencies of the vertex $\chi^2/ndof < 10$ requirement in the $H \to 2X \to 2\mu$ (top left), $H \to 2X \to 4\mu$ (top right), and $H \to 2Z_D \to 2\mu\,2X$ (bottom) signal samples. The efficiency calculation is based on dimuons that pass the full analysis selection except for the DCA cut.*

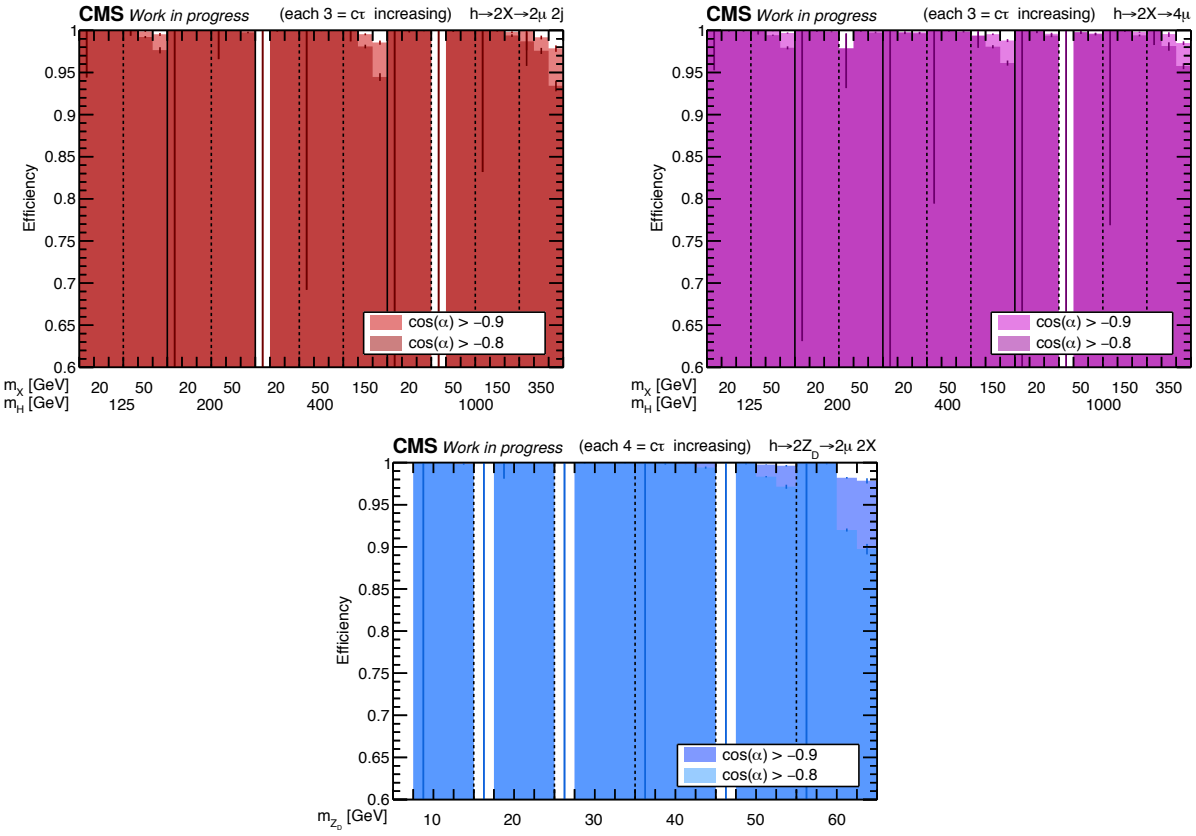


FIGURE 4.32: *Efficiencies of the two* $\cos\alpha$ *requirements applied in 2016 and 2018 data samples, measured in the 2018* $H \to 2X \to 2\mu$ (top left), $H \to 2X \to 4\mu$ (top right), *and* $h \to 2Z_D \to 2\mu\,2X$ (bottom) *signal samples. All dimuons that pass the full set of selection criteria except for* $\cos\alpha$ *are considered in the efficiency ratios.*

The application of the requirement $\cos\alpha > -0.8$ for 2016 data follows the equivalent HLT cut in that data-taking year (see Sec. 4.1) and suppresses back-to-back dimuons formed from cosmic-ray muons. For 2018 data, which were filtered by a trigger without angle requirements, a re-optimized threshold of $\cos\alpha > -0.9$ can be applied, which improves signal efficiency up to 8%, depending on the sample (see Fig. 4.32), while rejecting cosmic-ray background similarly efficiently.

Cosmic-ray muons that enter the detector diagonally (in contrast to the most common top-to-bottom muon direction) often result in the muon reconstruction producing multiple DSA muons, which, when paired, do not necessarily yield back-to-back muon pairs. This type of cosmic-ray background can easily mimic signal events and requires dedicated treatment. Muon tracks in such events are usually poorly measured and can therefore be rejected by requiring $N(\text{dimuon segments}) \geq 5$. This criterion leaves 99% of the signal intact. Generally, background from diagonal cosmic-ray muons almost exclusively appears in 2018 data filtered by the cosmic-seeded HLT path.

Another class of cosmic-ray background are events with multiple near-parallel muon tracks reconstructed in both detector hemispheres. Such events are caused by atmospheric showers. Naively, one would expect that any pairing of such parallel tracks would either result in back-to-back or collinear dimuons, which would be rejected by the $\cos\alpha$ requirement

or the $m_{\mu\mu} > 10\,\text{GeV}$ cut, respectively. However, it was found that the common-vertex fit can behave anomalously when fitting two (anti-)parallel muons with $d_0 > 100\,\text{cm}$. The fit can change the ϕ momentum direction of one of the two muons by about π, thus producing a dimuon with $\cos\alpha \lesssim 1$ and large $m_{\mu\mu}$. Such dimuons are neither rejected by the angle nor the mass requirement. To suppress background arising from atmospheric showers, the following three criteria have been designed.

First, each event must contain at least one reconstructed primary vertex with at least four associated tracks and with the transverse and longitudinal coordinates within 2 cm and 24 cm of the nominal collision point, respectively. This set of requirements constitutes the standard "good-vertex" event selection in CMS, and it helps to reject cosmic-only events unrelated to collision events. The effect of this requirement on signal efficiency is negligible.

Second, an upper cut on the number of parallel muons has been designed. [36] The quantity $N(\text{parallel pairs})$ is constructed by considering all pairings of DSA muons before the common-vertex fit with more than twelve valid CSC or DT hits and with $p_T > 5\,\text{GeV}$ and subsequently counting the number of pairs with $|\cos\alpha| \geq 0.99$. Most atmospheric showers produce events with $N(\text{parallel pairs}) \geq 6$, while this is true only for 0.2% of the $H \to XX \to 2\mu$ and 0.4% of the $H \to XX \to 4\mu$ signal. Thus, the requirement $N(\text{parallel pairs}) < 6$ is used to suppress events with cosmic-ray muon showers efficiently.

Third, any dimuons are vetoed that are accompanied by a third muon in the event that is back-to-back ($\cos\alpha < -0.9$, $p_T > 10\,\text{GeV}$) and has a time difference larger than 20 ns with respect to said dimuons. This criterion rejects events where a collinear dimuon is reconstructed from incident cosmic-rays in the same hemisphere. In such a scenario, there is detector activity opposite of the reconstructed dimuon, recorded at a time offset consistent with the typical crossing time of a cosmic-ray muon (about $\Delta t \approx 40\,\text{ns}$). This veto requirement also helps reject dimuon candidates reconstructed from one leg of a cosmic-ray muon and a muon from an overlapping pp collision. In total, the third-muon veto preserves 98–99% of the signal (see Fig. 4.33).

The above set of requirements suppresses any type of cosmic-ray muons efficiently and renders this background component negligible. To support this conclusion, the residual contamination of cosmic-ray muons in the analysis signal region is estimated with an auxiliary measurement that provides an upper bound for the number of dimuons passing the $\cos\alpha$ cut, N_p, via

$$N_p = N_f \frac{\varepsilon}{1-\varepsilon} \quad , \tag{4.5.1}$$

where N_f is the number of events that fail the nominal $\cos\alpha$ analysis cut, and ε is the efficiency to pass the $\cos\alpha$ cut obtained from a sample of cosmic-ray muons. This estimate is conservative in that it assumes *all* dimuons failing the $\cos\alpha$ requirement to be of cosmic nature. Therefore, the measurement result constitutes an upper limit on the cosmic-ray muon admixture in the analysis signal region. The numbers of dimuons failing the $\cos\alpha$ cut in 2016 and 2018 are 0 and 1, respectively. Together with the measured ε values (see Fig. 4.34), this results in upper limits[7] of 0.002 residual cosmic-ray events for 2016 and 0.003 for 2018.

4.5.6 Suppression of backgrounds from pp collisions

The main source of background in the analysis are SM events produced in pp collisions. Generally, four major types of SM background are relevant for the analysis of DSA-DSA

[7] Here, event yields of zero are substituted with the Poisson upper bound of 1.8 events.

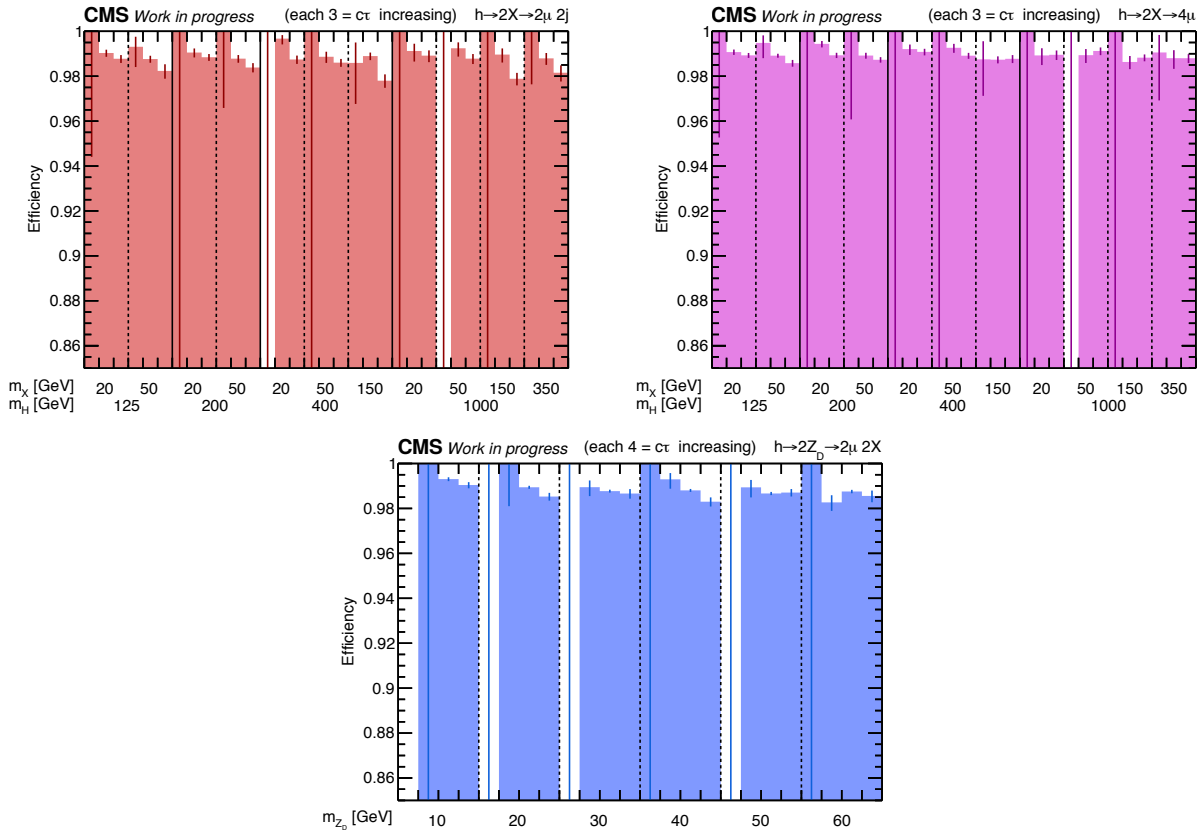


FIGURE 4.33: *Efficiencies of the "third-muon" veto requirement, measured in the 2018 $H \to 2X \to 2\mu$* (top left), *$H \to 2X \to 4\mu$* (top right), *and $h \to 2Z_D \to 2\mu\,2X$* (bottom) *signal samples. All dimuons that pass the full set of selection criteria except for the "third-muon" veto requirement are considered in the efficiency ratios.*

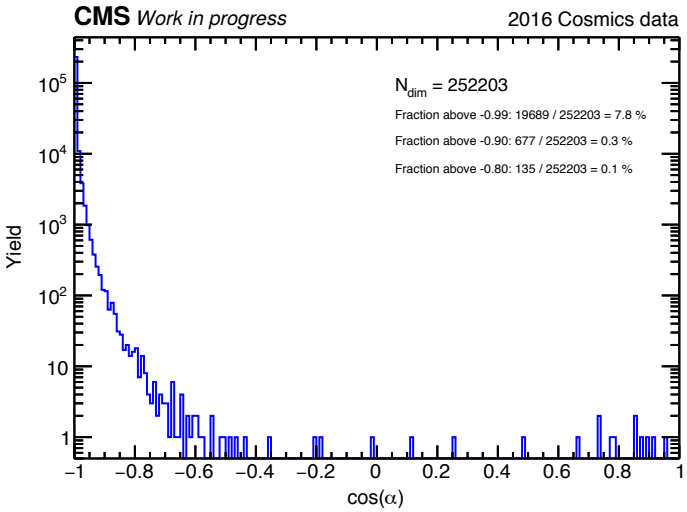


FIGURE 4.34: *Distribution of $\cos\alpha$ measured in a dedicated sample of cosmic-ray muons (see Sec. 4.2.4 for a description of the data sample).*

dimuons.

- Genuinely displaced dimuons, such as ones from decays of non-prompt low-mass resonances (like J/ψ) or ones formed from the products of the B meson "cascade" decays (like $b \to c\mu_1 X$, $c \to \mu_2 X$).

  The mass of such dimuons does not exceed a few GeV. Therefore, the simple requirement $m_{\mu\mu} > 10\,\text{GeV}$ suffices to suppress this background.

- Prompt dimuons that are mistakenly reconstructed as displaced dimuons due to detector or reconstruction imperfections.

  Here, one distinguishes between the following sub-classes of backgrounds:

  – $|\Delta\Phi|$-symmetric background (such as Drell-Yan $\mu\mu$ events): In such events, the dimuon momentum vector is uncorrelated with the L_{xy} vector and, as a consequence, the background is distributed symmetrically around $|\Delta\Phi| = \pi/2$. This contrasts signal processes, which produce distributions that are strongly peaked at $|\Delta\Phi| = 0$.

  While this background cannot be easily rejected, its $|\Delta\Phi|$ symmetry can be exploited by studying events in the signal-free regime of large $|\Delta\Phi|$. In fact, the region $3\pi/4 < |\Delta\Phi| \leq \pi$ will be called the *Drell-Yan control region* in the context of the analysis' background estimation (see Sec. 4.6.1).

  – $|\Delta\Phi|$-asymmetric background (such as QCD dijet and multijet events resulting in dimuons formed from either genuine or fake muons from different jets): Events of this type are expected to be asymmetric in $|\Delta\Phi|$ but symmetric in terms of the reconstructed dimuon charges. This means that events with equal-charge dimuons (commonly called *same-sign* (SS) events) can be used to split off a control region to study this class of background. This control region of SS events will be referred to as the *QCD control region* in the background estimation.

  – True low-mass dimuons ($m_{\mu\mu} < 10\,\text{GeV}$) that are wrongly reconstructed with higher mass in the signal region ($m_{\mu\mu} > 10\,\text{GeV}$) and with considerably large displacement: Such events are prevalent in 2018 data, where the $p_T^{(\text{L2})}$ thresholds are lower, and no mass cut is applied in the HLT paths. This class of events is studied in a special selection of dimuons, by using opposite-charge DSA-DSA dimuons (*i. e.*, *opposite-sign* (OS) dimuons) that pass the full analysis selection, including a cut $L_{xy} > 6$, which will be motivated shortly, but for which each muon leg is successfully associated with a PAT muon. Such PAT-associated DSA-DSA dimuons are excluded in the default DSA-DSA analysis as they are "moved" to the PAT-PAT dimuon category for separate study. This selection of dimuons revealed two common features of these events:

    1. DSA-DSA dimuons have a small opening in polar direction ($|\Delta\eta| < 0.1$) and large opening in azimuthal direction ($\Delta\phi > 0.1$), while the corresponding PAT-PAT-associated dimuons are low-mass dimuons with both $|\Delta\eta|$ and $|\Delta\phi|$ smaller than 0.1. Given that the track resolution of PAT muons is orders of magnitude better than that of DSA muons treating the PAT-PAT measurement as a proxy for the "truth" is justified in this context.
    2. One or both of the dimuon legs have no more than two or three muon segments and only a modest number of muon hits.

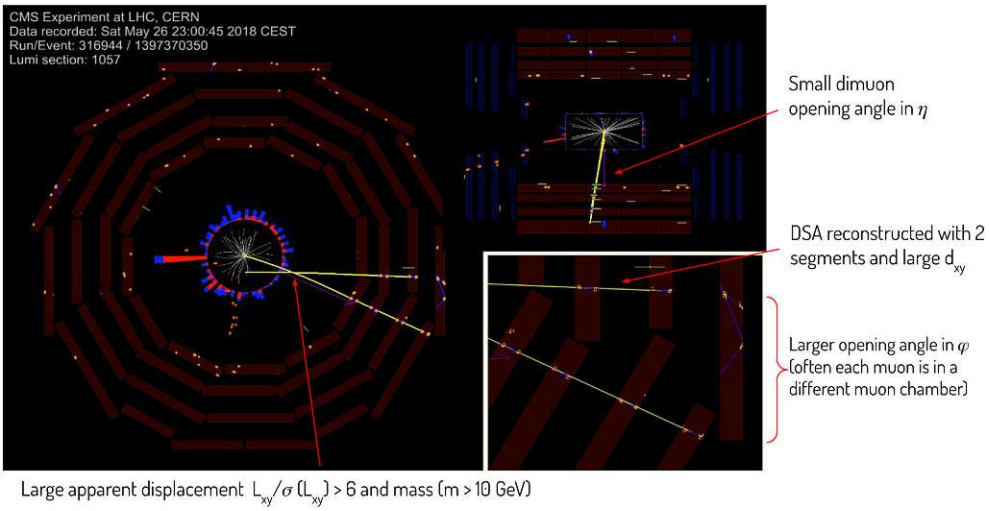


FIGURE 4.35: *Event display of a typical low-mass, equal-charge dimuon with* $L_{xy}/\sigma_{L_{xy}} > 6$. *The two distinct DSA muons forming the dimuon are produced in a pp collision and are close-by in the* η *direction (illustrated in the* R*–*z *plane in the top right view) but well separated in* ϕ *(illustrated in the* x*–*y *plane in the left and bottom right views). The nearly horizontal muon is measured by only two muon segments and, due to its poor track quality, gives rise to a displaced common vertex.*

In such events, dimuons with true small mass and initially small angular separation of their dimuon legs are opened in ϕ by the magnetic field as they propagate outwards through the detector (while they remain close-by in η direction). Then, because their muon tracks are usually reconstructed from a small number of hits, the reconstructed p_T of a muon can appear artificially large, resulting in muon tracks that appear straighter than they in fact are. This can give rise to dimuons both with an artificially displaced common vertex and overestimated $|\Delta\Phi|$, and, in further consequence, overestimated dimuon invariant mass, thus pushing a low-mass dimuon into the signal region at $m_{\mu\mu} > 10\,\text{GeV}$. An example of such an event is shown in Fig. 4.35. Most events with this characteristic are OS-dimuon events when measured in associated PAT-PAT dimuons. Therefore, it is expected that this type of background is predominantly seen in OS DSA-DSA dimuons (as opposed to SS DSA-DSA dimuons) and another peculiar background source for the analysis.

The rejection of such events happens by vetoing all DSA-DSA dimuons with $|\Delta\eta| < 0.1$ that have $N(\text{dimuon segments}) \leq 5$ or a barrel-only muon with $N(\text{DT hits}) \leq 24$. This set of requirements rejects 11 out of 16 SS events in 2018 data that pass all other selection criteria. In 2016 data, one out of three SS events is rejected. Depending on the combination of $(m_{\text{H}}, m_{\text{LLP}})$, between 90% and 99% of signal is preserved by these veto cuts, as Fig. 4.36 shows, with the largest efficiency losses observed in the most boosted signal topologies.

Finally, as yet another measure to suppress prompt backgrounds, dimuons are required to show a certain degree of displacement. Concretely, they have to pass the criterion $L_{xy}/\sigma_{L_{xy}} > 6$. This $L_{xy}/\sigma_{L_{xy}}$ threshold was determined by maximizing the expected statistical discovery significance using the Z_{Bi} figure of merit [111], which is an estimate of the

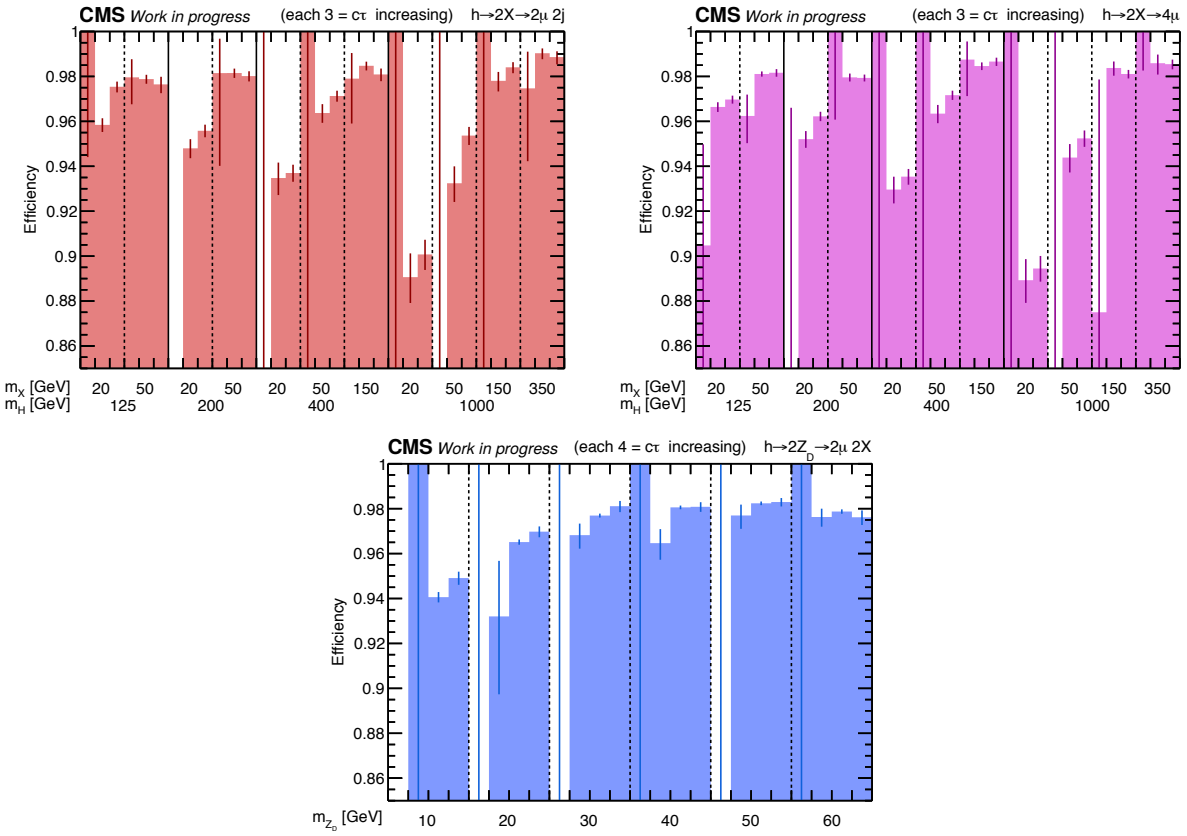


FIGURE 4.36: *Efficiencies of the extra $N(\text{dimuon segments}) + N(DT\ \text{hits})$ requirements for dimuons with $|\Delta\eta| < 0.1$ (see the text for details), measured in the 2018 $H \to 2X \to 2\mu$* (top left), *$H \to 2X \to 4\mu$* (top right), *and $h \to 2Z_D \to 2\mu\,2X$* (bottom) *signal samples. All dimuons that pass the full set of selection criteria except for the extra quality requirements are considered in the efficiency ratios.*

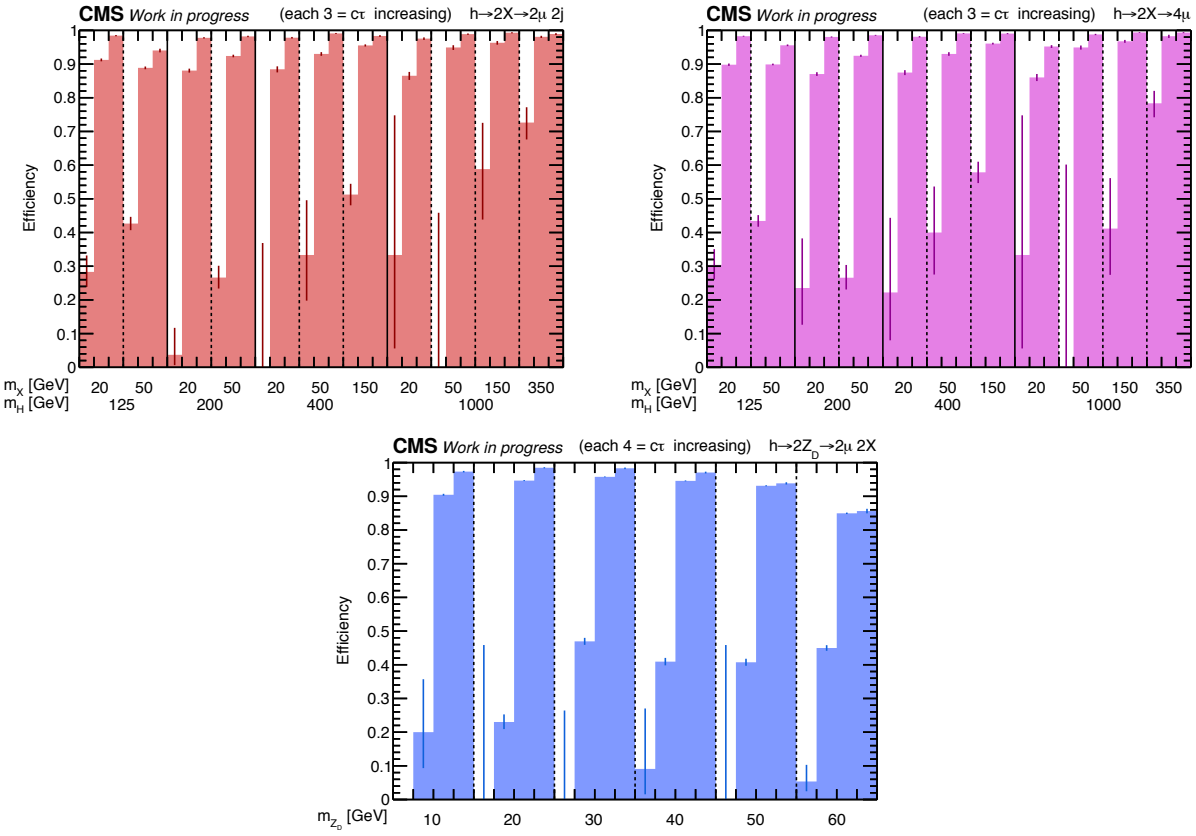


FIGURE 4.37: *Efficiencies of the $L_{xy}/\sigma_{L_{xy}}$ requirement, measured in the 2018 $H \to 2X \to 2\mu$* (top left), *$H \to 2X \to 4\mu$* (top right), *and $h \to 2Z_D \to 2\mu\,2X$* (bottom) *signal samples. All dimuons that pass the full set of selection criteria except for the $L_{xy}/\sigma_{L_{xy}}$ requirement are considered in the efficiency ratios.*

statistical significance of an excess of events in the signal region over the expected number of background events evaluated from a subsidiary measurement. More details on the performed Z_{Bi} optimization can be found in [36].

The net effect of the $L_{xy}/\sigma_{L_{xy}} > 6$ requirement is a rejection of 48 and 36 background events in the signal-free control regions at $3\pi/4 < |\Delta\Phi| < \pi$ in 2016 and 2018, respectively, while retaining 96–98% of signal dimuons on average (although the individual efficiencies depend on the particular signal sample, as Fig. 4.37 shows).

4.5.7 Invariant mass windows

All dimuons that pass the full set of selection criteria are additionally required to fall within certain intervals of dimuon invariant mass to test for the existence of an LLP with a given mass.

These mass intervals are chosen to be centered around the probed LLP mass, and their widths are determined by requiring that the respective interval contains more than 99% of the signal associated with the probed LLP mass. Fig. 4.38 shows examples of this procedure for $m(\text{LLP}) = 20\,\text{GeV}$ and $50\,\text{GeV}$. This choice results in mass windows that are typically 6-8 times larger than the DSA-DSA mass resolution. Given that this resolution does not strongly depend on the LLP lifetime or the BSM Higgs mass $m(\text{H})$, identical mass windows

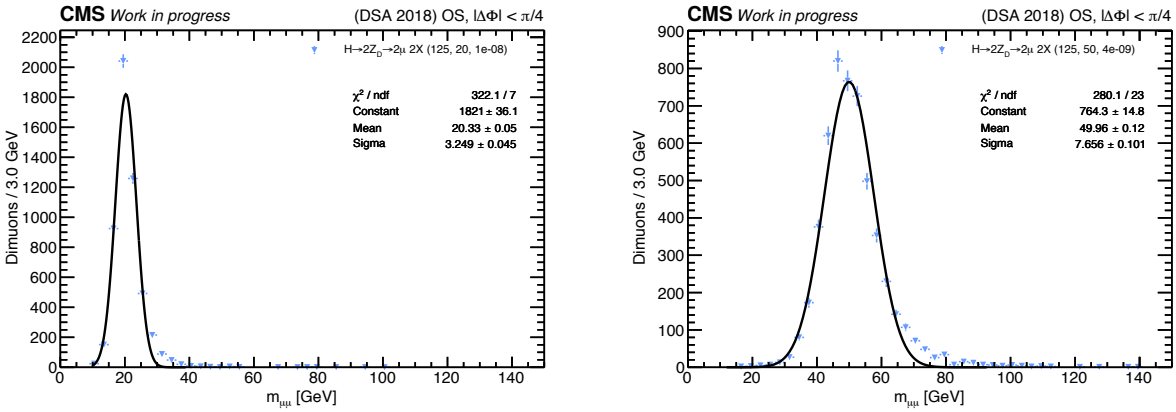


FIGURE 4.38: *Invariant mass distributions of dimuons passing the full set of selection criteria, measured in 2018 $h \to 2Z_D \to 2\mu\,2X$ samples with $m(Z_D) = 20\,\text{GeV}$ (left) and $m(Z_D) = 50\,\text{GeV}$ (right). The cores of the distribution peaks are fitted with Gaussian functions (overlaid in black color) and the resulting function parameters are shown in the legends.*

TABLE 4.7: *Mass windows associated with the LLP masses probed in the search.*

Probed LLP mass [GeV]	Mass window [GeV]
10, 20	$10 < m_{\mu\mu} < 32$
30	$15 < m_{\mu\mu} < 60$
40, 50	$20 < m_{\mu\mu} < 80$
60	$35 < m_{\mu\mu} < 120$
150	$35 < m_{\mu\mu} < 245$
350	$m_{\mu\mu} > 60$

are eventually used for all values of $c\tau$(LLP) and m(H) in the background prediction (see Sec. 4.6).

The resulting mass windows for the generated LLP masses (see Tab. 4.2 and 4.3) are listed in Tab. 4.7.

4.5.8 Summary of the event selection

The combined effect of the selection criteria described in this section is illustrated in Fig. 4.39, where the fraction of events passing the full set of event and (di)muon requirements out of all generated events defines the total signal efficiencies.

The efficiencies vary significantly as a function of the LLP lifetime. The smallest-$c\tau$ samples owe their low efficiency mainly to the DSA-to-PAT association, which removes muons from the pool of DSA muons in favor of the PAT-based analyses.[8] In contrast, the large-$c\tau$ samples give the highest efficiencies (around 1–10%) thanks to the DSA reconstruction allowing for muon reconstruction beyond the CMS tracker.

The main reason for higher efficiencies on average in 2018 data is the improved trigger in that year, which introduced a vertex-unconstrained muon reconstruction, looser trigger

[8] As mentioned at various places throughout this work, the PAT-based searches require dedicated analyses and will be presented in separate publications.

requirements (in terms of dimuon invariant mass and opening angle), as well as generally lower $p_T^{(\mathrm{L2})}$ thresholds.

The total efficiencies of the $H \to 2X \to 4\mu$ signal samples are computed as per-dimuon efficiencies to facilitate a more straightforward comparison with the $H \to 2X \to 2\mu$ samples of the BSM Heavy Scalar model. This shows that the efficiency for a dimuon to be selected is generally higher in events with two dimuons. The reason for this difference is the trigger: Events with *two* dimuons have a higher chance to pass the trigger requirements (and later have the non-triggering dimuon reconstructed and selected in the offline analysis) than events with only *one* dimuon. Especially at small m_H, where the trigger efficiency is usually low, these efficiency differences are sizeable.

Finally, the total efficiencies in the HAHM signal samples are presented as per-*event* efficiencies (*i. e.*, at least one dimuon must pass the selection criteria to count an event in the efficiency numerator). Again, a considerable increase of overall efficiencies in the 2018 setting is observed.

The total efficiencies presented in Fig. 4.39 are used in the statistical interpretation of the results (as per-event efficiencies, consistently) as discussed in Sec. 4.8.

4.6 Background estimation

In many physics analyses, the background in the signal region is accurately predicted with the help of MC simulations. Such an approach, however, is not necessarily adequate for this search. One of the reasons is that most of the expected background from pp collisions results from instrumental effects, such as reconstruction mistakes, and these are not guaranteed to be simulated in MC data in realistic form and rate. Moreover, none of the available MC simulations include the simulation of cosmic-ray showers or cosmic-ray muons overlapping with pp collision events. Finally, the limited statistical power of some of the available MC background samples further disfavors an MC-centric approach to background estimation.

The presented analysis chooses a background estimation strategy solely based on recorded data, thus overcoming the shortcomings of MC simulations. To this end, control regions are defined by inverting one or more selection criteria, which are enriched in certain types of background that can be studied in detail.

Events passing the full analysis selection and falling into the signal region of the search were "blinded" until the final stages of the analysis. This avoids any potential biases in the design of the event selection criteria.

In the most general terms – and as already outlined in Sec. 4.5.6 – the analysis expects background in two broad categories: $|\Delta\Phi|$-symmetric and $|\Delta\Phi|$-asymmetric background, like Drell-Yan or QCD processes, respectively. The remainder of this section will discuss each of them in detail.

4.6.1 Evaluation of Drell-Yan and other non-QCD backgrounds

The dimuon collinearity angle $|\Delta\Phi|$ – the angular separation of the dimuon momentum vector $p_T^{\mu\mu}$ and the L_{xy} vector – is a good discriminator for the identification of signal processes involving the production of LLPs and their decay to final-state dimuons. Such signal processes populate the domain of small $|\Delta\Phi|$ because their dimuon vectors are aligned with their L_{xy} vectors due to momentum conservation in the two-body decays. In contrast, $p_T^{\mu\mu}$

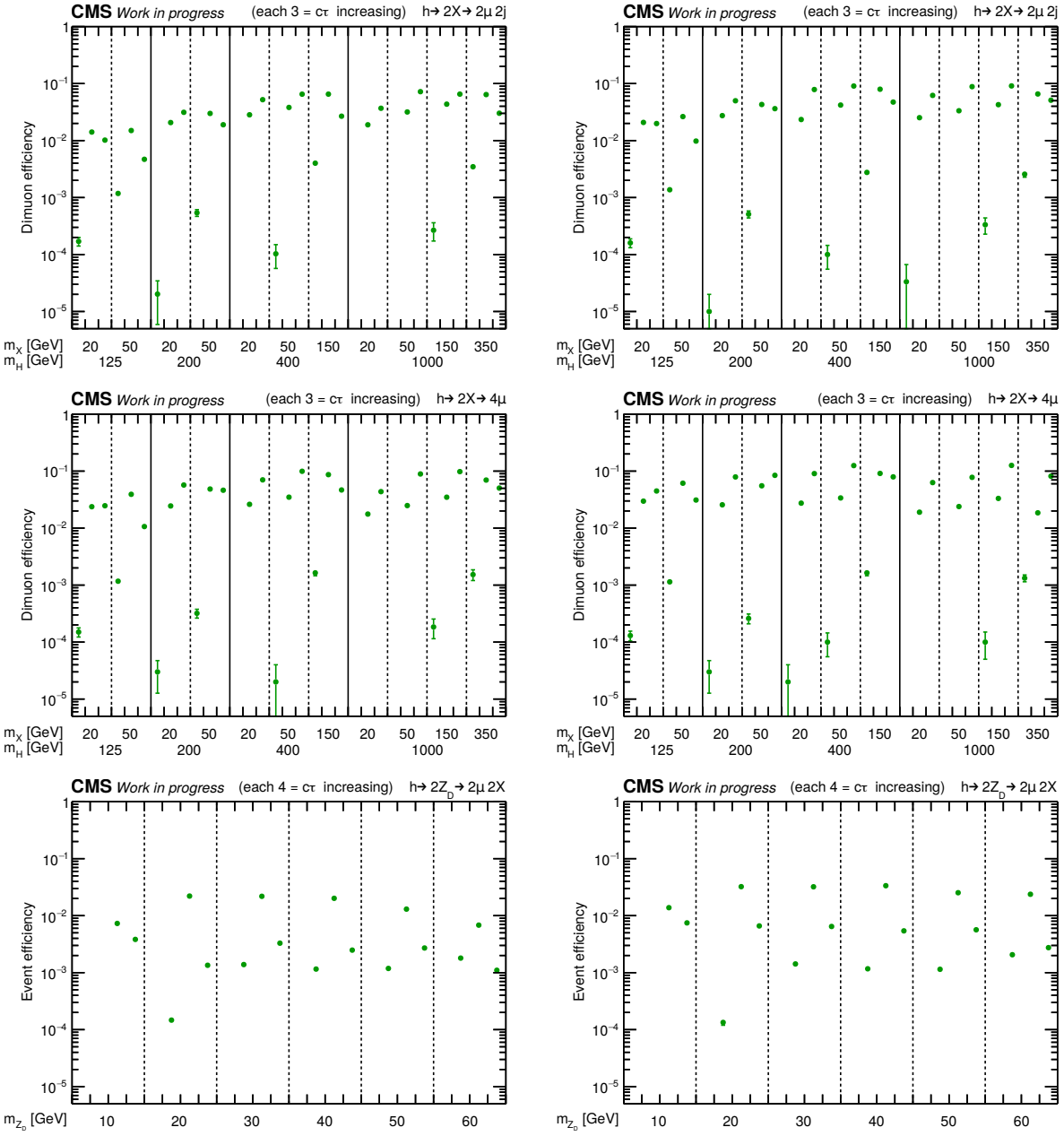


FIGURE 4.39: *Signal efficiencies for all generated* $(m_H, m_{LLP}, c\tau)$ *points of the various signal processes in 2016* (left) *and 2018* (right). *The efficiencies are defined as the fraction of events that pass all event and muon selection criteria.* Top: $H \to 2X \to 2\mu$, middle: $H \to 2X \to 4\mu$, bottom: $H \to 2Z_D \to 2\mu\,2X$.

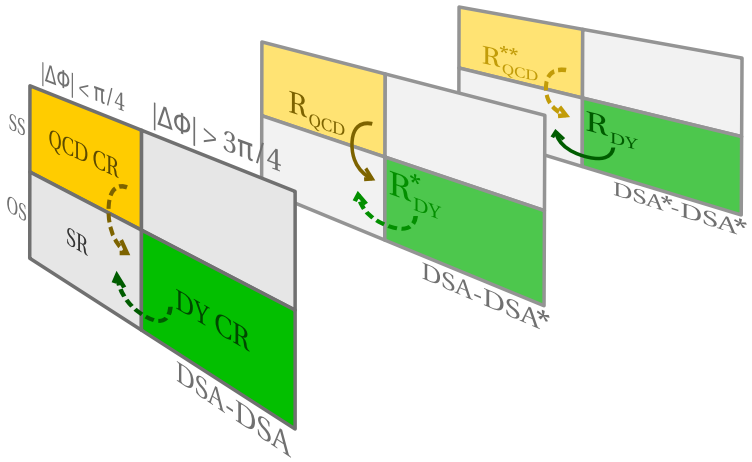


FIGURE 4.40: *Illustration of the background estimation strategy in terms of the various control regions and their interplay. The foremost layer contains the signal region (SR), the QCD control region (QCD CR), and the Drell-Yan control region (DY CR), which are defined via DSA-DSA dimuon charges (opposite- vs. same-sign dimuons, i. e., OS vs. SS) and* $|\Delta\Phi|$ *intervals (*$|\Delta\Phi| < \pi/4$ *vs.* $|\Delta\Phi| > 3\pi/4$*). Apart from these requirements, the dimuons in all regions must pass the full analysis selection. Data in the SR are blinded until the final stages of the analysis. The middle layer represents all DSA-DSA dimuons that have one DSA muon leg associated with a PAT muon (indicated by the asterisk in "DSA-DSA*") but otherwise are split equivalently according to dimuon charge and* $|\Delta\Phi|$*. The nominal QCD transfer factor,* $R_{QCD} = \frac{N(OS)}{N(SS)}$*, is measured in this DSA-DSA* selection. Finally, the layer in the back represents all DSA-DSA dimuons that have both their legs associated with PAT muons ("DSA*-DSA*"). In this selection, the nominal DY transfer factor,* $R_{DY} = \frac{N(|\Delta\Phi|<\pi/4)}{N(|\Delta\Phi|>3\pi/4)}$*, is obtained. Different requirements on the associated PAT muons in the DSA-DSA* and DSA*-DSA* selections apply depending on whether the DY or the QCD transfer factor is measured. In the context of the systematic uncertainty assessment, transfer factor equivalents are respectively measured in the neighboring layers,* R^{**}_{QCD} *and* R^{*}_{DY}*, and compared to the nominal* R_{QCD} *and* R_{DY} *measurements. Details on each of these background estimation aspects are described in this section.*

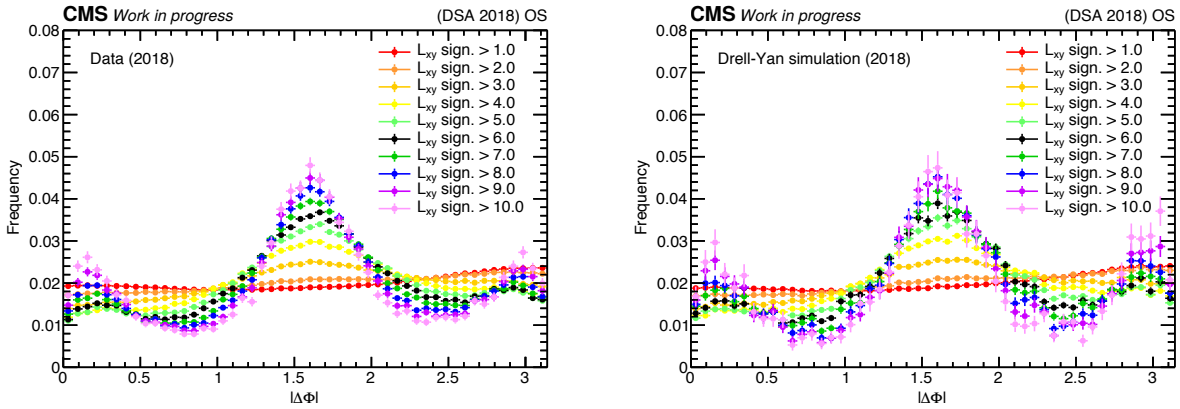


FIGURE 4.41: *Normalized distributions of* $|\Delta\Phi|$ *for PAT-associated DSA-DSA dimuons in the DY control region (see the text for details) obtained from 2018 collision data* (left) *and DY simulation* (right)*, each fulfilling sequentially more stringent* $L_{xy}/\sigma_{L_{xy}}$ *requirements.*

and L_{xy} are uncorrelated in many background processes, such as Drell-Yan (DY) interactions, and have a flat $|\Delta\Phi|$ spectrum in general.

The flatness of the $|\Delta\Phi|$ distribution in DY events motivates the definition of a dedicated region for their study, using events at $3\pi/4 < |\Delta\Phi| < \pi$ passing all other selection criteria. This large-$|\Delta\Phi|$ region is referred to as the *Drell-Yan control region* throughout this work.

An auxiliary sample (with comparatively large statistical power) for the further study of the $|\Delta\Phi|$ distribution shape in collision background is obtained by selecting events that have the DSA-to-PAT association (see Sec. 4.5.2) reversed, *i. e.*, DSA-DSA muons are selected that are both their DSA muon legs successfully associated with PAT muons. Additionally, the associated PAT-based dimuons must fulfill $L_{xy}^{(\mathrm{PAT})}/\sigma_{L_{xy}^{(\mathrm{PAT})}} < 1.0$ to ensure that they originate from promptly-produced background and not from signal processes. In other words, this sample gives access to dimuons that are removed from the original set of DSA-DSA dimuons in the nominal analysis selection because they are normally replaced by PAT-PAT or DSA-PAT dimuons and studied separately.

Thanks to the orthogonality of this selection to the signal-region configuration, this sample allows studying dimuons both with large and small $|\Delta\Phi|$.

To remove any potential admixture of QCD backgrounds, the PAT-associated dimuon legs are further required to be isolated. The PAT isolation criterion requires the scalar sum of the p_T of tracks in a cone of $\Delta R < 0.3$ around the muon divided by the muon p_T to be smaller than 0.05.

In accordance with Fig. 4.40, this selection will be dubbed the *DSA*-DSA* control region*, where the asterisks indicate that the selected DSA-DSA dimuons are doubly PAT-associated.

Fig. 4.41 shows the $|\Delta\Phi|$ distributions for DSA-DSA dimuons in the DSA*-DSA* control region, both for 2016 data and DY simulation. A scan over increasingly more stringent $L_{xy}/\sigma_{L_{xy}}$ cuts is performed, and each distribution is normalized to unit area to visualize the evolution of the distribution shape directly. In the case of $L_{xy}/\sigma_{L_{xy}} > 1.0$, the distributions are reasonably flat and symmetric around $|\Delta\Phi| = \pi/2$, both in data and MC. Once the $L_{xy}/\sigma_{L_{xy}}$ requirements are tightened, the distributions start to form a three-peaked structure with maxima near $|\Delta\Phi| \approx 0$, $|\Delta\Phi| \approx \pi/2$, and $|\Delta\Phi| \approx \pi$. Two mechanisms are

responsible for the formation of these peaks. [36] For dimuons with very small and large $|\Delta\Phi|$ ($|\Delta\Phi| \approx 0$ and $|\Delta\Phi| \approx \pi$), trigger efficiencies are favorable due to the pointing nature of the involved muons (see Fig. 4.7). Dimuons with $|\Delta\Phi| \approx \pi/2$ are oriented perpendicular to the L_{xy} vector, in which case the measurement of the uncertainty in the L_{xy} position, $\sigma_{L_{xy}}$, is more precise than in the case of dimuons that are aligned with the L_{xy} vector. The net effect is on average larger values of $L_{xy}/\sigma_{L_{xy}}$ for $|\Delta\Phi| \approx \pi/2$ dimuons that are favorably selected by increasing $L_{xy}/\sigma_{L_{xy}}$ cuts. However, neither of these two effects break the $|\Delta\Phi|$ symmetry around $|\Delta\Phi| = \pi/2$, and the resulting distributions are well reproduced in simulation. The definition of the DY control region, therefore, remains useful.

The ratio of events in the small-$|\Delta\Phi|$ and large-$|\Delta\Phi|$ regions slightly deviates from unity. To account for this, a correction factor ("transfer factor") is measured from the equivalent ratio in events of the DSA*-DSA* control regions, which is defined as

$$R_{\mathrm{DY}} = \frac{N^{\text{DSA*-DSA*}}(|\Delta\Phi| < \pi/4)}{N^{\text{DSA*-DSA*}}(|\Delta\Phi| > 3\pi/4)} \quad . \tag{4.6.1}$$

To remove any residual contamination from non-prompt, low-mass resonances in the R_{DY} measurement – a background which is asymmetric in $|\Delta\Phi|$ and evaluated separately – the additional requirement $m_{\mu\mu}^{\text{(PAT-PAT)}} > 15\,\mathrm{GeV}$ is applied to the DSA*-DSA* control region events. The measured R_{DY} is roughly constant as a function of $m_{\mu\mu}^{\text{(DSA-DSA)}}$, so the inclusively-measured R_{DY} is taken for all invariant mass windows.

Finally, the DY background in the signal region is estimated by

$$N(|\Delta\Phi| < \pi/4) = N(|\Delta\Phi| > 3\pi/4) \cdot R_{\mathrm{DY}} \; . \tag{4.6.2}$$

A visualization of the various control regions and their interplay in the context of the background estimation is sketched in Fig. 4.40.

As a further check, the possibility of measuring the correction factor R_{DY}^{*} in similar events where only *one* of the muons is associated with a PAT muon (*i. e.*, in the so-called *DSA-DSA* control region*, according to Fig. 4.40) is explored. The results of this measurement are compatible within statistical errors with the R_{DY} values obtained in the DSA*-DSA* control region in both data-taking years (see Tab. 4.8). Due to the larger number of control region events, the R_{DY} measurement in the DSA*-DSA* selection is chosen as the nominal method.

No events are observed in the DSA-DSA Drell-Yan control region once the full set of selection criteria is applied. This is true both for 2016 and 2018 data. The correction factor R_{DY} is 0.78 ± 0.01 (stat.) in 2016 and 0.76 ± 0.01 (stat.) in 2018.

The validation of the DY background prediction strategy and the associated systematic uncertainties will be discussed in what follows.

Validation of the Drell-Yan background estimation method

To test the validity of the DY background prediction method, the estimated DY background is compared to the observed background in a *validation region (VR)* that is enriched by this type of background. The region of the inverted $L_{xy}/\sigma_{L_{xy}}$ cut, $L_{xy}/\sigma_{L_{xy}} < 6$, is a natural candidate to facilitate this consistency check since events in this regime are expected to be produced promptly with $|\Delta\Phi|$ spectra being symmetric around $|\Delta\Phi| = \pi/2$. Indeed, the dimuons in this region have masses around the Z boson mass and are evenly spread

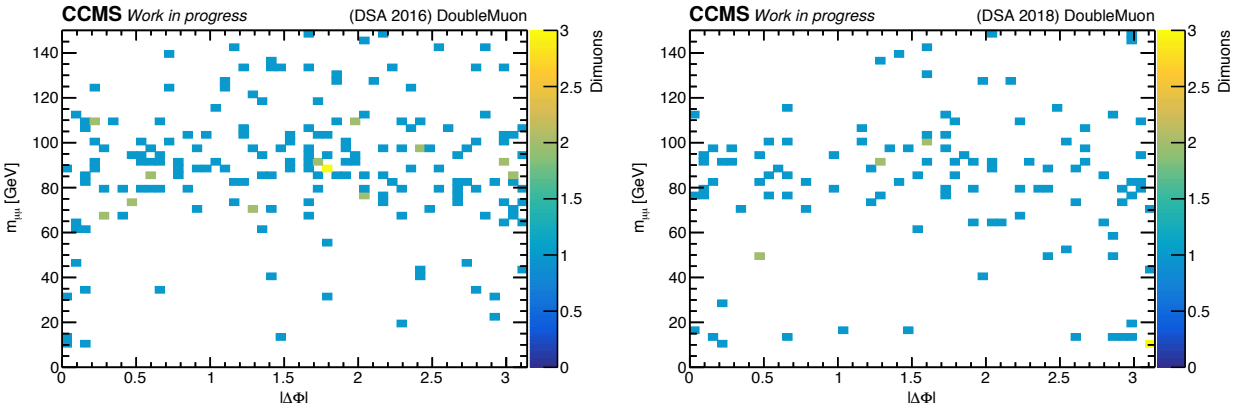


FIGURE 4.42: *Distributions of dimuons with* $L_{xy}/\sigma_{L_{xy}} \leq 6$ *passing all other selection criteria in the* $m_{\mu\mu}$–$|\Delta\Phi|$ *plane in 2016* (left) *and 2018* (right) *data.*

across the full $|\Delta\Phi|$ range, as Fig. 4.42 demonstrates. In the depicted $m_{\mu\mu}$–$|\Delta\Phi|$ plane, QCD events, if present, would cluster at small $|\Delta\Phi|$. Further, the amount of signal events in this VR is negligible, given the results of the CMS analyses with Run-1 data [65, 66].

The validation strategy is equivalent to the background estimation strategy illustrated in Fig. 4.40, with the only difference being that *all* depicted regions impose the inverted $L_{xy}/\sigma_{L_{xy}}$ criterion instead of the nominal analysis selection. In this scenario, what is called the signal region (SR) in Fig. 4.40, is referred to as a validation region that can be studied (*i. e.*, does not have to be blinded).

The validation is performed in bins of $L_{xy}/\sigma_{L_{xy}}$ to check for consistency of the estimation method across the full available range of dimuon displacement. The background prediction is computed in each bin separately, using Eq. 4.6.2, and subsequently compared to the observed validation region events in the same bin. Fig. 4.43 presents the results for each year. The prediction reproduces the observed data accurately in all bins, given the statistical uncertainties. (The statistical errors are dominated by the limited number of events at $|\Delta\Phi| > 3\pi/4$.) The datasets of both years yield similar values of R_{DY}.

As can be seen in Fig. 4.43, the total number of events is larger in 2016, although the larger integrated luminosity of the 2018 dataset should result in the opposite hierarchy of event yields. This apparent contradiction is explained by the lower tracking efficiency in parts of the 2016 data due to a known detector effect at that time (see Sec. 4.2.1). Poorer tracking efficiency causes fewer DSA muons to be replaced by PAT muons in the DSA-to-PAT association procedure, increasing the number of DSA-DSA dimuons in 2016 beyond what is expected from pure luminosity scaling. However, the background estimation method captures this effect as the tracking inefficiencies are independent of $|\Delta\Phi|$.

Systematic uncertainties in the Drell-Yan background prediction

The DY background estimation uncertainty is dominated by statistical uncertainties as a result of the vanishing number of events in the DSA-DSA Drell-Yan control regions. In contrast, the measurement of the DY transfer factor, R_{DY}, is performed in control regions with large statistical power due to the reversal of the DSA-to-PAT association for this purpose. While, therefore, the statistical uncertainties of R_{DY} are small, the R_{DY} measurement is prone to systematic uncertainties.

The stability of R_{DY} against modifications of the measurement region definition is eval-

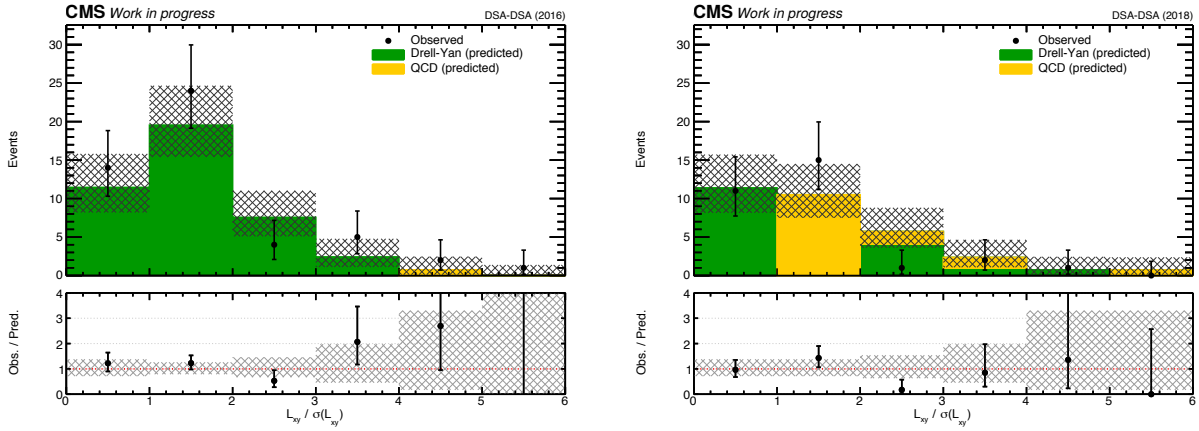


FIGURE 4.43: *Distributions of $L_{xy}/\sigma_{L_{xy}}$ of oppositely-charged dimuons with $|\Delta\Phi| < \pi/4$ in the DY validation region ($L_{xy}/\sigma_{L_{xy}} \leq 6$, shown in black color) compared to the predicted background in 2016* (left) *and* (right) *data. The shown uncertainties are of statistical nature only.*

TABLE 4.8: *Values of the DY transfer factor obtained in dimuon samples with either one or both DSA muons associated with PAT muons, together with the ratio of these two measurements.*

Definition of the DY transfer factor	2016	2018
R_{DY} (measured in DSA*-DSA*)	0.78 ± 0.01	0.76 ± 0.01
R^{*}_{DY} (measured in DSA-DSA*)	0.97 ± 0.23	0.71 ± 0.17
R^{*}_{DY} / R_{DY}	1.25 ± 0.30	0.93 ± 0.22

uated by comparing the nominal transfer factor measurement (*i. e.*, R_{DY} obtained in the DSA*-DSA* control region) to that obtained in the DSA-DSA* control region, R^{*}_{DY}. The two resulting transfer factors are compatible within their statistical uncertainties, which is shown in Tab. 4.8.

The R_{DY} is assumed to be mass-independent, and a mass-inclusive value is used for all signal mass windows of a given year. This assumption is verified by deriving R_{DY} for each mass bin individually and comparing it to the mass-inclusive measurement. Fig. 4.44 summarizes the results.

Based on this study, the R_{DY} measurement is assigned a 15% systematic uncertainty. This uncertainty has a negligible effect on the overall DY background prediction.

Lastly, the DY background predictions in the $L_{xy}/\sigma_{L_{xy}} \leq 6$ validation regions are compatible with the observed yields within the quoted statistical uncertainties (see again Fig. 4.43). Hence, no additional systematic uncertainty is derived from this comparison.

4.6.2 Evaluation of QCD background

Unlike the background evaluation strategy described in the previous section, this section concentrates on the study and prediction of background that is asymmetric around $|\Delta\Phi| = \pi/2$. More specifically, the focus lies on background that predominantly populates the domain of small $|\Delta\Phi|$, such as dimuon decays of non-prompt resonances at small mass (*e. g.*, J/ψ), cascade decays of B mesons ($b \to c\mu_1 X$, $c \to \mu_2 X$), or dimuons that are

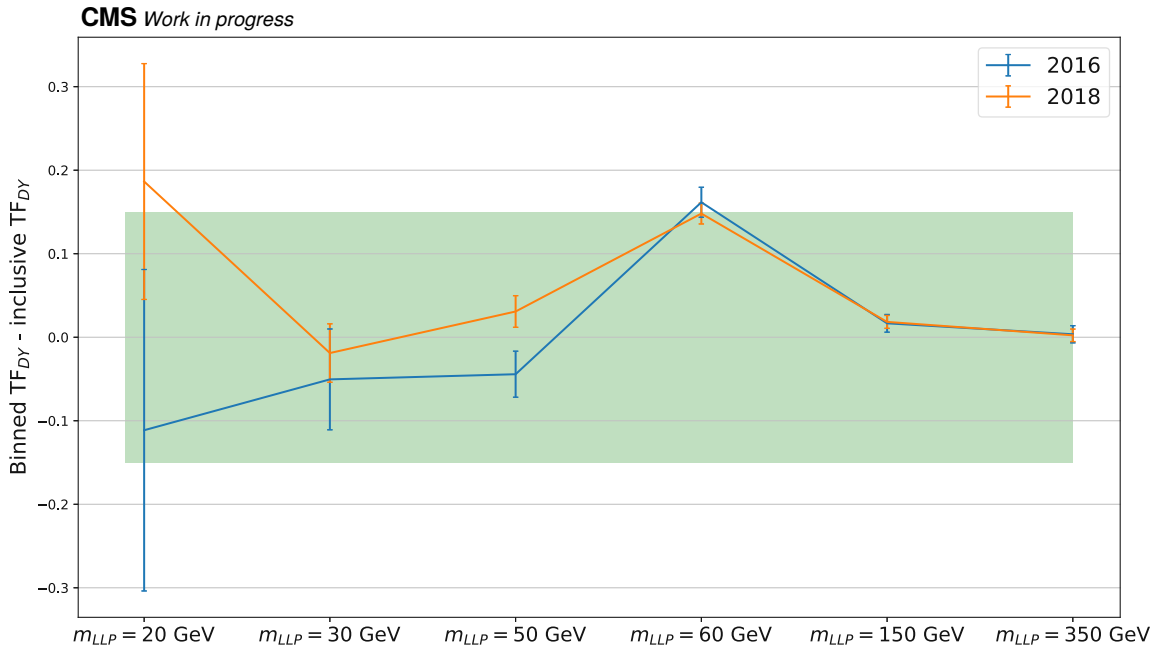


FIGURE 4.44: *Differences between a mass-differential and the mass-inclusive R_{DY} in 2016* (blue color) *and 2018* (orange color). *The mass-differential R_{DY} measurements are performed in the mass windows corresponding to the indicated m_{LLP} values. A 15% systematic uncertainty in the mass-inclusive R_{DY} measurement is overlaid in the form of a green band.*

formed from unrelated, non-prompt muons of the same or of different jets. The collection of $|\Delta\Phi|$-asymmetric background is uniformly referred to as *QCD background* in the context of this analysis.

These QCD background processes produce dimuons with masses below the $m_{\mu\mu} = 10\,\text{GeV}$ threshold and should therefore fail the event selection criteria. However, due to the limited resolution of the DSA muon reconstruction, such background events can leak into the signal region above $m_{\mu\mu} > 10\,\text{GeV}$ because of reconstruction mistakes. There, they can easily mimic signal candidates near the 10-GeV mass threshold thanks to their small $|\Delta\Phi|$ values and signal-like $L_{xy}/\sigma_{L_{xy}}$ values.

Hence, careful studies of the QCD background are vital for the accurate prediction of this background component in the signal region. Similar to the DY estimation scenario, two control regions are defined that are tailored to the (QCD) background under study. The first control region sample is obtained by inverting the DSA-to-PAT association criterion for *one* of the DSA-DSA dimuon legs (*i. e.*, the DSA-DSA* control region selection in Fig. 4.40). The other one is obtained by inverting the criterion for *both* legs (*i. e.*, the DSA*-DSA* control region selection). In either region, to exclude any non-QCD background sources as well as potential signal, each associated PAT muon is required to be non-isolated (with the scalar sum of the track p_T values in a cone of $\Delta R < 0.3$ around that muon divided by the muon's p_T required to be greater than 0.1).

The $|\Delta\Phi|$ distribution resulting from this selection is shown in Fig. 4.45. Progressively more stringent $L_{xy}/\sigma_{L_{xy}}$ requirements reveal stronger and stronger signal-like peaks at $|\Delta\Phi| \approx 0$. To further support the argument that this background is complimentary to the $|\Delta\Phi|$-symmetric backgrounds discussed previously, equivalent distributions are shown for the $m_{\mu\mu}$ and ΔR variables (see Fig. 4.46). Again, they start to form more and more pronounced structures as the $L_{xy}/\sigma_{L_{xy}}$ requirements are tightened. Most notably, they

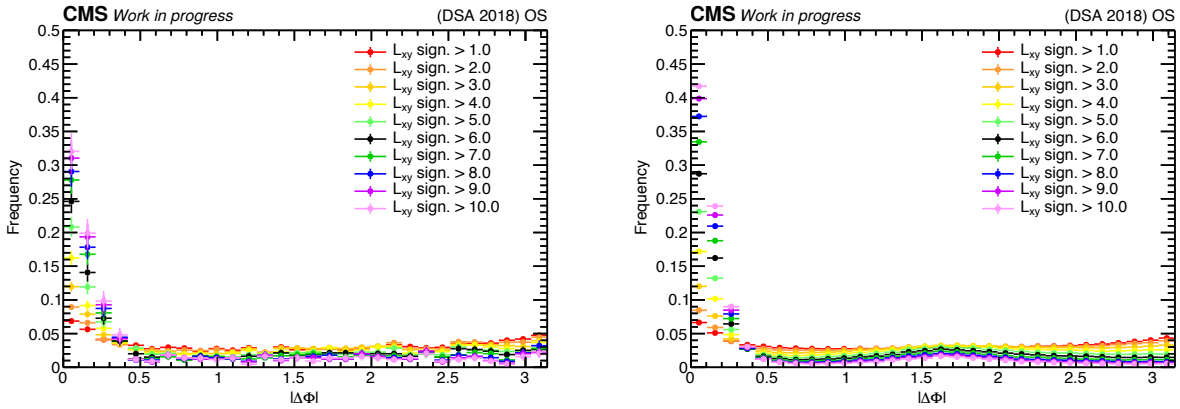


FIGURE 4.45: *Normalized distributions of* $|\Delta\Phi|$ *for PAT-associated DSA-DSA dimuons in the QCD control region (see the text for details), each fulfilling sequentially more stringent* $L_{xy}/\sigma_{L_{xy}}$ *requirements. The measurements are performed with DSA-DSA dimuons having either two PAT-associated legs* (left) *or one PAT-associated leg* (right).

develop clear peaks at small $m_{\mu\mu}$ and small ΔR. This is contrasted by similar distributions obtained from DY-enriched samples, which feature peaks at $m_{\mu\mu} \approx m(Z)$ and $\Delta R \approx \pi$ (see Fig. 4.47).

QCD background processes that produce DSA-DSA dimuons with opposite charge signs of their muon legs (OS dimuons) above $m_{\mu\mu} > 10\,\text{GeV}$ already suffer from a certain degree of misreconstruction. It is therefore expected that these samples also come with a certain degree of muon charge misassignment. Together with the fact that many QCD backgrounds are charge-symmetric in the first place, a certain fraction of DSA-DSA dimuons with the same charge sign (SS dimuons) is expected to be formed from QCD backgrounds.

Therefore, the chosen control region for evaluating the QCD background (the *QCD control region*) is represented by SS dimuons at $|\Delta\Phi| < \pi/4$. The QCD contribution to the signal region is then estimated via

$$N(OS) = N(SS) \cdot R_{\text{QCD}} \quad , \tag{4.6.3}$$

where the transfer factor R_{QCD} between the two QCD regions is obtained from the ratio of OS to SS dimuons in the selection with the DSA-PAT association inverted in one dimuon leg (*i. e.*, the DSA-DSA* control region):

$$R_{\text{QCD}} = \frac{N^{\text{DSA-DSA}^*}(OS)}{N^{\text{DSA-DSA}^*}(SS)} \quad . \tag{4.6.4}$$

Again, a visualization of the various control regions and their interplay in the context of the background estimation can be found in Fig. 4.40.

Depending on whether one or both dimuon legs have their PAT association criterion inverted, the resulting DSA-DSA* and DSA*-DSA* control regions produce different values for the QCD transfer factor, with the DSA*-DSA* selection giving the largest R^{**}_{QCD}. These differences are particularly significant for dimuons with $m_{\mu\mu} < 35\,\text{GeV}$.

The fact that the transfer factors depend on the number of PAT-associated muons per dimuon suggests a correlation between the DSA-PAT replacement probability and the

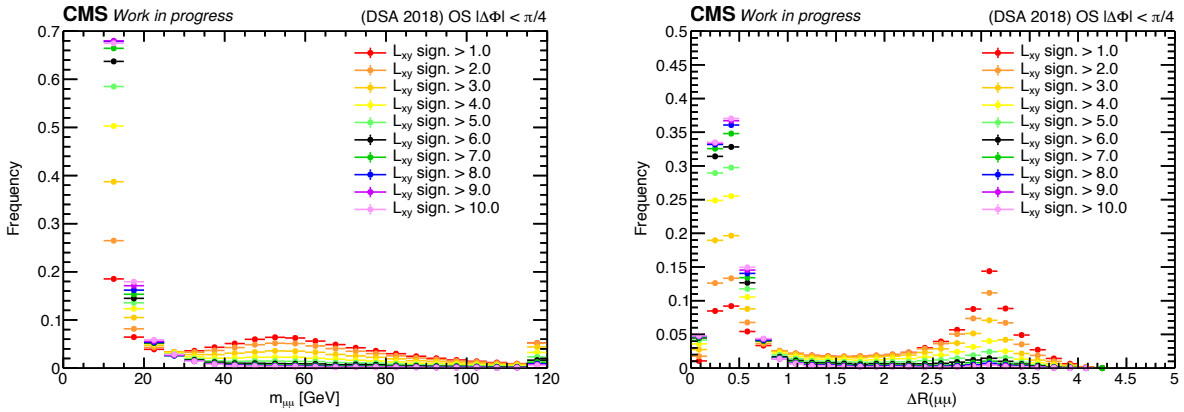


FIGURE 4.46: *Normalized distributions of* $m_{\mu\mu}$ (left) *and* ΔR (right) *for DSA-DSA dimuons with one PAT-associated leg in the QCD control region (see the text for details), each fulfilling sequentially more stringent* $L_{xy}/\sigma_{L_{xy}}$ *requirements.*

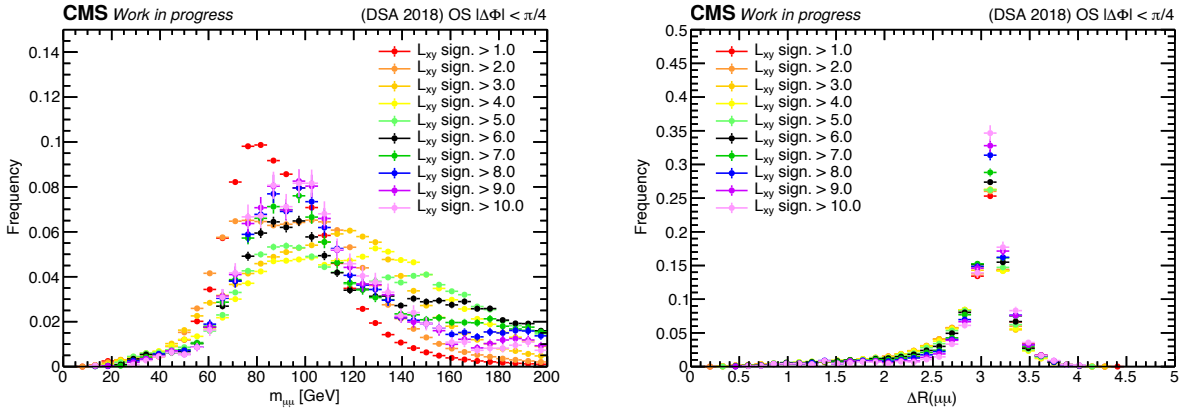


FIGURE 4.47: *Normalized distributions of* $m_{\mu\mu}$ (left) *and* ΔR (right) *for PAT-associated DSA-DSA dimuons in the DY control region (see the text for details), each fulfilling sequentially more stringent* $L_{xy}/\sigma_{L_{xy}}$ *requirements.*

DSA charge assignment. Two-dimensional projections of PAT-vs-DSA masses for different configurations of PAT-vs-DSA charges (see Fig. 4.48) reveal two major classes of dimuons. The domain of small DSA-DSA mass (up to roughly 40 GeV) is dominated by low-mass resonances that leak into the $m_{\mu\mu}^{\text{(DSA-DSA)}} > 10\,\text{GeV}$ region due to misreconstruction effects. These events mainly populate the SS control region due to charge misassignment. At large DSA-DSA mass (beyond 40 GeV), one finds dimuons mainly from unrelated non-prompt muons. Both classes of events mix in the intermediate-mass range.

A detailed study based on DSA-vs-PAT charge comparisons (see Appendix A) confirms that, indeed, the charge mismeasurement rate is higher for non-PAT-associated muons than for associated ones. The effect of this correlation is that the PAT association tends to decrease the denominator of the R_{QCD} measurement, thus leading to overestimation of the QCD background. In principle, this affects both control regions defined via the inverted PAT association, but the effect is smaller in the sample with only *one* DSA muon PAT-associated. This region is more representative of the DSA-DSA signal region and therefore chosen as the nominal measurement region for the QCD transfer factor R_{QCD}. Moreover, the results of the DSA-vs-PAT charge comparison study are used to assign systematic uncertainties to the R_{QCD} measurement, which will be described at the end of this section.

The R_{QCD} measurement at high mass suffers from large statistical fluctuations, which are avoided using an inclusive transfer factor for $m_{\mu\mu} > 35\,\text{GeV}$. This approach is justified because R_{QCD} variations are expected to be much smaller than statistical uncertainties in this regime, which is confirmed by the measurements presented in Fig. 4.49.

Finally, the QCD control region with SS dimuons contains 2 events with $m_{\mu\mu} = 13.5\,\text{GeV}$ and $m_{\mu\mu} = 32.0\,\text{GeV}$ in 2016 data and 3 events with $m_{\mu\mu} = 14.8\,\text{GeV}$, $m_{\mu\mu} = 16.2\,\text{GeV}$, and $m_{\mu\mu} = 18.2\,\text{GeV}$ in 2018 data.

Validation of the QCD background estimation method

The strategy for testing the QCD background prediction method is similar to the DY validation procedure of Sec. 4.6.1 in that dedicated validation regions are defined, which are enriched in the type of background under study. In the QCD case, two such validation regions are constructed.

First, the nominal requirement for the dimuon invariant mass ($m_{\mu\mu} > 10\,\text{GeV}$) is inverted and dimuons in the low-mass range $6\,\text{GeV} < m_{\mu\mu} < 10\,\text{GeV}$ are studied.[9] The validation strategy, then, is equivalent to the background estimation procedure illustrated in Fig. 4.40, but where all depicted regions apply the small-mass requirement. For brevity, this validation region is called *VR-IMASS*.

Second, the quality criteria $N(\text{dimuon segments}) > 5$ and $N(\text{DT hits}) > 24$ are inverted for dimuons with $|\Delta\eta| < 0.1$. Again, the conceptual image of Fig. 4.40 holds, but this time with the inverted quality criteria applied. The resulting validation region is referred to as *VR-IQDETA*. Compared to VR-IMASS, this validation region has the advantage that it can be probed across the full mass range, including the mass domain of the search region.

In both validation regions, the observed events cluster at small $|\Delta\Phi|$ values, indicating that the contribution of $|\Delta\Phi|$-symmetric backgrounds – the ones validated in Sec. 4.6.1 –

[9] One cannot go much lower than $m_{\mu\mu} \simeq 6\,\text{GeV}$ for validation purposes as the fraction of properly reconstructed OS dimuons increases disproportionally towards the lowest mass bins. This manifests itself in rapidly growing transfer factors (*i. e.*, OS/SS events) and, consequently, in considerable overprediction of low-mass background. In other words, the regime at $m_{\mu\mu} < 6\,\text{GeV}$ is no longer representative of the search region at $m_{\mu\mu} > 10\,\text{GeV}$ and a validation of the background estimation is not adequate there.

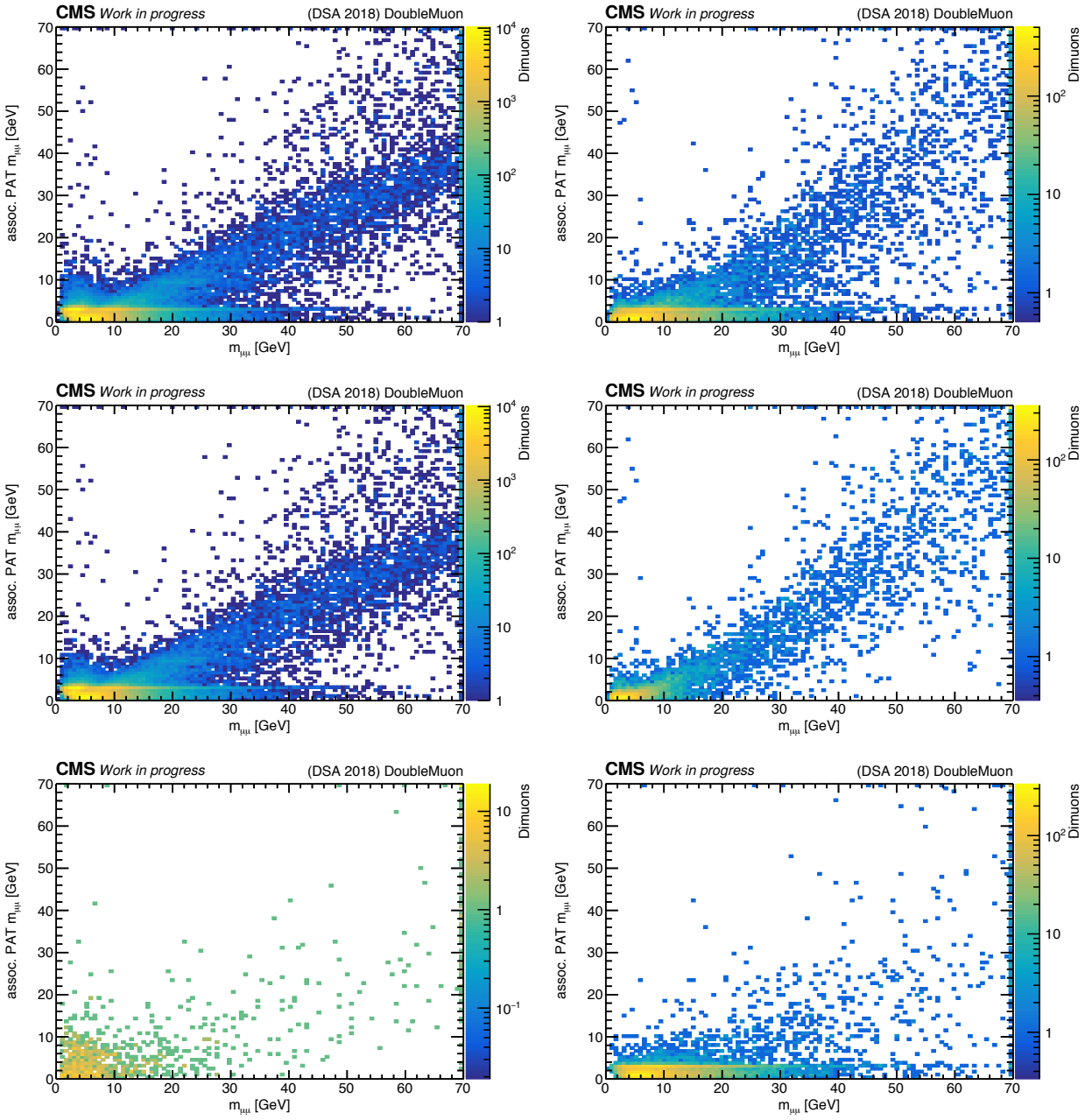


FIGURE 4.48: *Comparisons of* $m_{\mu\mu}^{(DSA)}$ *(horizontally) and* $m_{\mu\mu}^{(PAT)}$ *(vertically) for DSA-DSA dimuons with both legs associated with PAT muons. The left column shows oppositely-charged dimuons, the right columns equally-charged dimuons.* Top row: *The full control sample,* middle row: *a subsample with matching DSA and PAT charges,* bottom row: *the complementary sample with at least one mismatch between the DSA and associated PAT muons.*

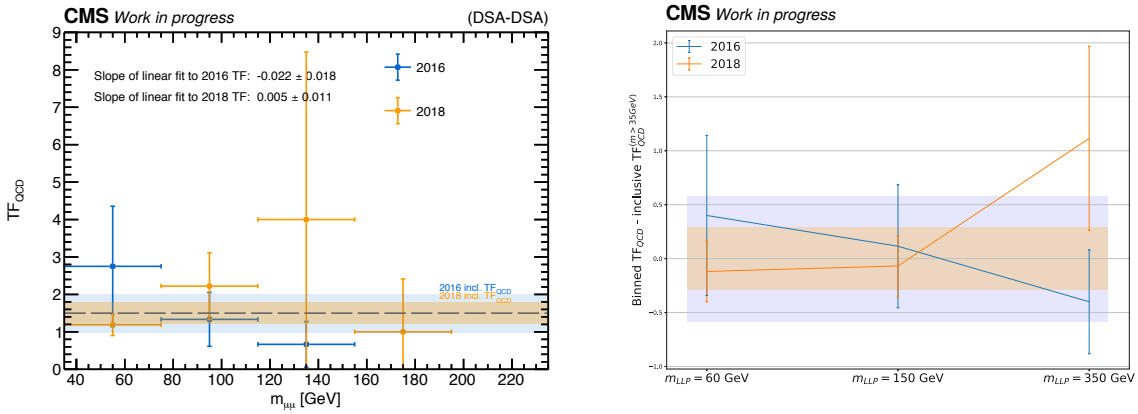


FIGURE 4.49: Left: *Differential measurements of R_{QCD} compared to the inclusive R_{QCD} measurement for $m_{\mu\mu} > 35\,\text{GeV}$ (dashed line) and its statistical uncertainty (blue and orange bands). Linear fits to the measurements of each year are consistent with mass independence of R_{QCD} for $m_{\mu\mu} > 35\,\text{GeV}$, as indicated by the text overlays.* Right: *Deviation of mass-differential R_{QCD} values (obtained in individual mass windows corresponding to m_{LLP} from the mass-inclusive value (for $m_{\mu\mu} > 35\,\text{GeV}$), together with bands indicating the statistical uncertainties of the mass-inclusive value for each year.*

is negligible (see Fig. 4.50).

The background prediction in either validation region is performed according to Eq. 4.6.3 using a transfer factor R_{QCD} (see Eq. 4.6.4) measured in DSA-DSA* dimuons, which have *one* PAT-associated dimuon leg. The prediction is contrasted with the observed yield in the respective validation region. Since R_{QCD} shows a strong mass dependence, particularly towards small mass, the validation is performed in bins of dimuon invariant mass.

Fig. 4.51 presents the results of the validation in the VR-IMASS region in 2018 data. Predicted and observed numbers of events agree well within their statistical uncertainties. The relatively large statistical uncertainties are caused by the small number of events in the DSA-DSA SS control region, $N(SS)$, which are magnified because of $R_{\text{QCD}} > 1$. However, given that the statistical uncertainties in the search regions are of comparable (or even

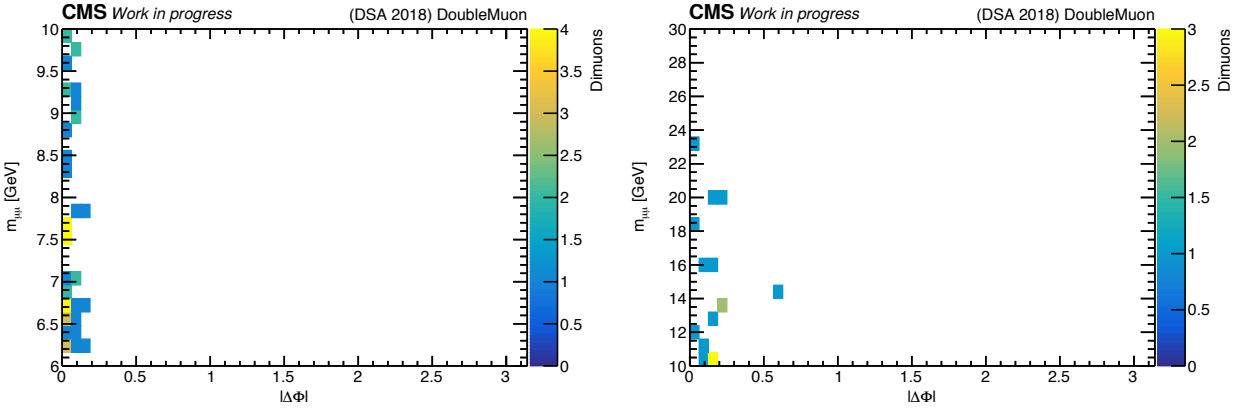


FIGURE 4.50: *Distributions of dimuons in the $m_{\mu\mu}$–$|\Delta\Phi|$ plane in the 2018 VR-IMASS* (left) *and VR-IQDETA* (right) *validation regions.*

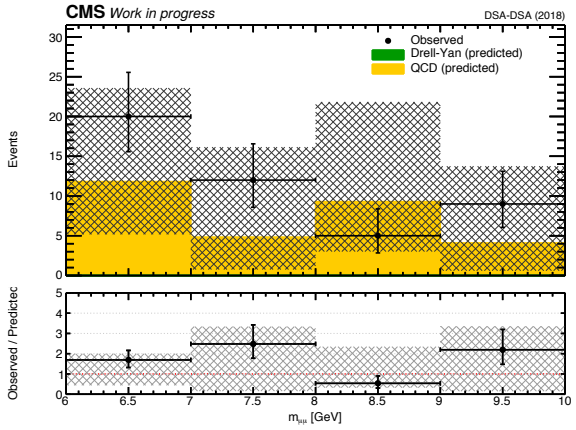


FIGURE 4.51: *Distributions of $m_{\mu\mu}$ of oppositely-charged dimuons with $|\Delta\Phi| < \pi/4$ in the VR-IMASS QCD validation region ($m_{\mu\mu} \leq 10\,\text{GeV}$, shown in black color) compared to the predicted background. All errors are of statistical nature only.*

larger) size, the presented validation method offers enough confidence in the background estimation procedure to be adequate. The VR-IMASS validation cannot be performed in 2016 data due to the HLT requirement $m_{\mu\mu}^{(\text{HLT})} > 10\,\text{GeV}$ in that year.

Fig. 4.52 summarizes the results of the VR-IQDETA validation in the search mass windows corresponding to the smallest probed LLP masses. The results in the larger mass windows ($m_{\mu\mu} > 35\,\text{GeV}$) are not explicitly shown in this figure since both the predicted and the observed yields are zero. The top panel in Fig. 4.52 visualizes the results obtained in 2018 data. 2016 data, as well as the subset of 2018 events that are triggered by the pp-seeded HLT path,[10] are shown in the bottom panel. Again, good agreement between predicted and observed background event yields is established in all scenarios. The expected background in the analysis search region is well represented by the VR-IQDETA events in terms of the expected background, its statistical uncertainties, and the typical size of R_{QCD}.

Systematic uncertainties in the QCD background prediction

The uncertainty of the QCD background prediction is dominated by statistical uncertainties originating from a small number of events in the QCD control region. On the other hand, the transfer factor R_{QCD} is measured in separate control regions, which have large statistical powers due to their reversal of the DSA-to-PAT association but are more impacted by systematic uncertainties.

To determine the systematic uncertainties associated with the R_{QCD} measurement, the stability of the transfer factor values against modifications of the control region definitions is assessed, and the level of agreement between the background prediction and observation in the validation regions is cross-checked.

The QCD background predictions in the VR-IMASS and VR-IQDETA validation regions (see Fig. 4.51 and 4.52, respectively) are compatible with the observed yields in those regions. No systematic uncertainty of R_{QCD} is derived from these checks.

[10] Since one of the main differences between the 2016 and 2018 signal triggers is the addition of cosmic-seeded paths in 2018 (see Sec. 4.1 for details), a 2016 sample filtered by the pp-seeded HLT version is enriched in the type of background that is expected in 2016 data.

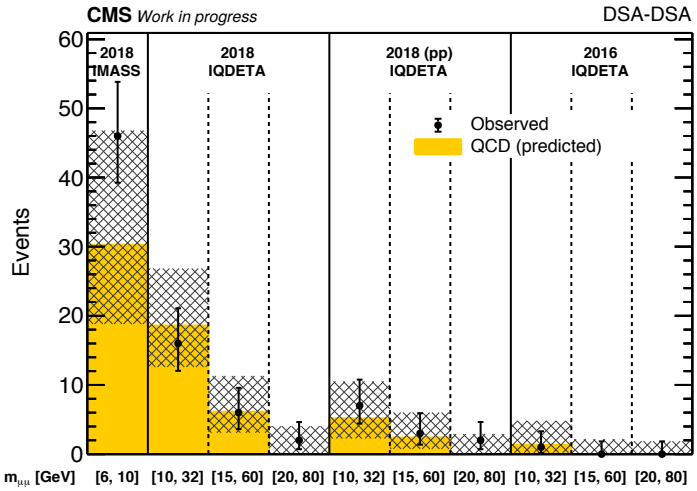


FIGURE 4.52: *Comparisons of predicted (yellow bars) and observed (black markers) numbers of events in the VR-IMASS and VR-IQDETA validation regions as functions of dimuon invariant mass windows. (Mass windows that are not shown here contain zero predicted and observed events.) The two left-hand sectors show the full 2018 data set in the VR-IMASS and VR-IQDETA validation regions, the third sector shows the subset of 2018 data in the VR-IQDETA selection that is filtered by the pp-seeded trigger, and the right-hand sector shows VR-IQDETA data in 2016. The shown uncertainties are of statistical nature only.*

As described in Sec. 4.6.2, a correlation between the DSA-to-PAT association and the correct charge assignment is observed, which potentially biases the measurement towards larger R_{QCD} values. The effect of this correlation is minimized in the QCD background prediction by using DSA-DSA* dimuons, which have only *one* PAT-associated dimuon leg (instead of the DSA*-DSA* selection with two PAT associations). However, some correlation bias remains even in this measurement approach, particularly for small values of dimuon invariant mass.

To take into account the remaining correlation bias in the R_{QCD} measurement region, a one-sided systematic uncertainty is defined using R_{QCD} values that are predicted using DSA charge mismeasurement probabilities (described in Appendix A). The final systematic uncertainties are chosen to be the difference between the predicted and the measured R_{QCD} values. Tab. 4.9 lists them, measured in the different signal-region mass windows. The systematic uncertainties range between 0% and −35%, depending on the mass window and the year.

Since the transfer factors vary with DSA-DSA mass, they are applied separately for each mass window in the QCD background estimation of the analysis (according to the R_{QCD} values in Tab. 4.9).

4.6.3 Background predictions and observed events

The signal region of the search is divided into several overlapping mass windows, which cover the entire mass spectrum starting from $m_{\mu\mu} > 10\,\mathrm{GeV}$ (see Sec. 4.5.7). In each mass window, the background is estimated as the sum of the $|\Delta\Phi|$-symmetric (DY and other

TABLE 4.9: *Relative systematic uncertainties in the R_{QCD} measurement for the different search region mass windows.*

Mass window [GeV]	Uncertainty
$10 < m_{\mu\mu} < 32$	-8%
$15 < m_{\mu\mu} < 60$	-0%
$20 < m_{\mu\mu} < 80$	-13%
$m_{\mu\mu} > 35$	-20%

TABLE 4.10: *Predicted background events in 2016 data, in dimuon invariant mass windows. The second and third columns show the control region yields and transfer factors in the DY background estimation, and the fourth and fifth columns contain the corresponding values for the QCD background estimation. The two background predictions are combined for the full estimate in the last column.*

Mass window [GeV]	$N(OS; \lvert\Delta\Phi\rvert > 3\pi/4)$	R_{DY}	$N(SS, \lvert\Delta\Phi\rvert < \pi/4)$	R_{QCD}	N_{est}
$10 < m_{\mu\mu} < 32$	0	0.78 ± 0.01	1	1.24 ± 0.28	$1.2^{+3.2}_{-1.1}$
$15 < m_{\mu\mu} < 60$	0	0.78 ± 0.01	1	1.08 ± 0.29	$1.1^{+2.9}_{-0.9}$
$20 < m_{\mu\mu} < 80$	0	0.78 ± 0.01	1	1.55 ± 0.44	$1.6^{+3.9}_{-1.4}$
$35 < m_{\mu\mu} < 120$	0	0.78 ± 0.01	0	1.50 ± 0.52	$0.0^{+3.1}_{-0.0}$
$35 < m_{\mu\mu} < 245$	0	0.78 ± 0.01	0	1.50 ± 0.52	$0.0^{+3.1}_{-0.0}$
$m_{\mu\mu} > 60$	0	0.78 ± 0.01	0	1.50 ± 0.52	$0.0^{+3.1}_{-0.0}$

non-QCD backgrounds as described in Sec. 4.6.1) and the $|\Delta\Phi|$-asymmetric backgrounds (QCD processes as described in Sec. 4.6.2).

The full background prediction is quantified in Tab. 4.10 for 2016 and Tab. 4.11 for 2018. In addition to the estimated number of background events, N_{est}, these tables also contain the event yields of the DY and QCD control regions, $N(OS; |\Delta\Phi| > 3\pi/$ and $N(SS; |\Delta\Phi| < \pi/4$, as well as the DY and the QCD transfer factors, R_{DY} and R_{QCD} for each mass window.

The main uncertainties in the transfer factors are of systematic nature. However, since the event yields in the control regions are small, the uncertainties of the *total* background prediction are ultimately driven by statistical uncertainties.

The residual background is composed of QCD events, and its contribution decreases towards larger dimuon invariant mass. At the lowest-probed masses, in the first and second mass windows, this background is larger in magnitude in 2018 than in 2016. This is due to the larger integrated luminosity of the 2018 data set and the less stringent HLT requirements in that data sample. In both data-taking years, the search is virtually background-free at masses beyond $m_{\mu\mu} > 35\,\text{GeV}$.

The expected and observed numbers of events are eventually interpreted as single-bin counting experiments in each mass window (see Sec. 4.8).

TABLE 4.11: *Predicted background events in 2018 data, in dimuon invariant mass windows. The second and third columns show the control region yields and transfer factors in the DY background estimation, and the fourth and fifth columns contain the corresponding values for the QCD background estimation. The two background predictions are combined for the full estimate in the last column.*

Mass window [GeV]	$N(OS; \lvert\Delta\Phi\rvert > 3\pi/4)$	R_{DY}	$N(SS, \lvert\Delta\Phi\rvert < \pi/4)$	R_{QCD}	N_{est}
$10 < m_{\mu\mu} < 32$	0	0.76 ± 0.01	3	2.30 ± 0.22	$6.9^{+6.9}_{-3.8}$
$15 < m_{\mu\mu} < 60$	0	0.76 ± 0.01	2	1.55 ± 0.19	$3.1^{+4.3}_{-2.0}$
$20 < m_{\mu\mu} < 80$	0	0.76 ± 0.01	0	1.25 ± 0.19	$0.0^{+2.7}_{-0.0}$
$35 < m_{\mu\mu} < 120$	0	0.76 ± 0.01	0	1.50 ± 0.29	$0.0^{+3.1}_{-0.0}$
$35 < m_{\mu\mu} < 245$	0	0.76 ± 0.01	0	1.50 ± 0.29	$0.0^{+3.1}_{-0.0}$
$m_{\mu\mu} > 60$	0	0.76 ± 0.01	0	1.50 ± 0.29	$0.0^{+3.1}_{-0.0}$

4.7 Signal-related systematic uncertainties and scale factors

The rates of expected signal are subject to systematic uncertainties.[11] Further, MC simulation often does not reproduce the characteristics of recorded data fully accurately, which one usually compensates for by applying *scale factors* (*i. e.*, event or dimuon weights) to the simulation.

This section describes the relevant scale factors and systematic uncertainties. Generally, sources of systematic uncertainties are considered to be uncorrelated between the data-taking years unless stated otherwise.

In the following, all systematic uncertainties and scale factors affecting the rate of expected signal are described.

4.7.1 Luminosity

The integrated luminosity measurement is associated with an uncertainty of 1.2% in 2016 [112] and 2.5% in 2018 [113], which directly affects the normalization of the signal rate.

4.7.2 Collision pile-up

The inelastic pp cross-section of 69.2 mb used in the pile-up (PU) reweighting is varied by ±5%. This results in variations of the signal efficiencies smaller than 2% for all signal models and LLP masses and lifetimes. Therefore, a systematic uncertainty of 2% (fully correlated between running periods) is assigned for the evaluation of signal efficiencies due to pile-up.

[11] The background evaluation methods, too, are subject to systematic uncertainties. These were already discussed in Sec. 4.6.1 and 4.6.2.

4.7.3 DSA muon reconstruction and identification

The signal efficiencies measured in simulated samples are corrected by reconstruction and identification scale factors for prompt DSA muons, which are extrapolated to the non-prompt domain for muons with $d_0 > 0\,\text{cm}$.

The prompt scale factors are measured using a "tag-and-probe" approach with muons from J/Ψ and Z decays.[12] They depend on p_T and η of the muons.

The d_0 dependence of the reconstruction and identification performance was studied with cosmic-ray muons and discussed in Sec. 4.4.6 and 4.4.6, respectively.

As was shown in these sections and in Fig. 4.16 in particular, the DSA reconstruction scale factor does not depend on d_0. Therefore, the prompt reconstruction scale factors do not need any further corrections when extrapolated to larger d_0. The d_0 extrapolation is merely associated with an additional systematic uncertainty of 2%.

The DSA muon identification scale factor, on the other hand, shows a mild linear dependence on d_0. Thus, the prompt identification scale factor needs to be corrected by a linear function that decreases it by 3.5% at $d_0 = 100\,\text{cm}$ (with respect to its value at $d_0 = 0\,\text{cm}$), as shown in Fig. 4.21. An additional 2% systematic uncertainty is assigned to the d_0 extrapolation of the identification scale factor.

The resulting efficiency corrections vary between 0.83 and 0.88, depending on the signal sample, and represent the largest data-driven signal rate correction in the analysis.

4.7.4 Trigger efficiency

Similar to the strategy of the DSA muon reconstruction and identification scale factors and systematic uncertainties, the corresponding quantities for the trigger efficiency are evaluated in the prompt-like regime using Z decays and extrapolated to the non-prompt regime using auxiliary measurements with cosmic-ray muons.

The prompt-like scale factors ($d_0 < 10\,\text{cm}$) are obtained from a tag-and-probe measurement in Z decays, and they are generally p_T- and η-dependent.[13]

Non-pointing muons ($d_0 > 10\,\text{cm}$), which typically appear in displaced decays, suffer from d_0-dependent trigger inefficiencies due to the L1-intrinsic beam-spot constraint (see Sec. 4.4.3). The trigger efficiencies as well as the muon reconstruction and p_T resolution at both L1 and HLT stages are validated using cosmic-ray data. Among others, the result of these studies are the d_0-dependent trigger scale factors presented in Sec. 4.4.5.

The trigger scale factors do not show any clear dependence on d_0. The variations of the measurements with respect to the prompt scale factor increase with growing d_0 values, which is not only due to statistical fluctuations in the underlying data but also due to potential systematic effects intrinsic to the cosmic-ray data sample.

Therefore, the full trigger scale factors are determined by the prompt scale factor – the product of the single-muon trigger scale factors – and are extrapolated to larger d_0 values by adding a systematic uncertainty that increases linearly with d_0 according to the (orange) bands shown in Fig. 4.14.

[12] This measurement was carried out by other analyzers in the context of the CMS "Muon Physics Objects Group" (Muon POG) and constitutes a general recommendation for analyses using muons in CMS. Therefore, further details on these measurements are omitted here.

[13] The scale factor measurement relies on trigger efficiencies that are based on single-muon analogs of the double-muon signal triggers used in this search. These single-muon triggers were introduced in the 2018 trigger configuration for the purpose of performance studies. Again, this measurement is provided as a standard CMS reference by the Muon POG, and further details are omitted here.

TABLE 4.12: *Efficiency of the DSA-to-PAT association for DSA muons from $Z \to \mu\mu$ decays in collision data and Drell-Yan simulation. All values are given in percent.*

Year	Collision data	DY simulation
2016	99.65 ± 0.01	99.87 ± 0.01
2018	99.76 ± 0.01	99.88 ± 0.01

4.7.5 DSA muon p_T resolution

Imperfections in the DSA muon p_T resolution can lead to modified signal yields when applying any of the mass window requirements that are described in Sec. 4.5.7. However, these mass windows are designed to contain more than 99% of the signal and correspond to about ± 3 times the invariant mass resolution in simulation. The impact of DSA muon p_T mis-modeling on the signal rates is therefore expected to be small.

Nevertheless, to quantify this impact, the DSA muon p_T is smeared in simulated signal events using Gaussian distributions with $\mu = 1$ and $\sigma = \sigma_{\text{extra}}$, where σ_{extra} is applied in three bins of p_T according to Tab. 4.6. The σ_{extra} values vary from about half the expected p_T resolution in the lowest-p_T bin to less than 10% of the resolution in the highest-p_T bin. In each bin, the σ_{extra} variations with d_0 are within the statistical uncertainty, $\Delta\sigma_{\text{extra}}$, so there is no further need to introduce any d_0 dependencies in the measurement.

The variations in signal rate as a result of the DSA muon p_T smearing are less than 2% in all cases except the (boosted) signals $(m(\text{H}), m(\text{LLP})) = (1000, 20)\,\text{GeV}$, $(400, 20)\,\text{GeV}$, and $(125, 10)\,\text{GeV}$. For these exceptional cases, the respective variations -7.0%, -3.3%, and $+22.2\%$ are measured in 2016 samples and -2.6%, $< 2\%$, and $+49.7\%$ in 2018 samples.

Systematic uncertainties of 2% are thus assigned to all signals, while the three exceptional signal samples are allowed to have larger (asymmetric) uncertainties in the direction of their respective measurements.

4.7.6 DSA-to-PAT muon association

The accuracy of the developed DSA-to-PAT muon association algorithm (see Sec. 4.5.2) was studied using a sample of DSA-DSA dimuons from $Z \to \mu\mu$ decays. Apart from the explicit requirement $L_{xy}/\sigma_{L_{xy}} < 6$, these dimuons must fulfill the remaining set of selection criteria, and their invariant masses must fall in a mass window of $\pm 10\,\text{GeV}$ around m_Z.

Subsequently, all constituent DSA muons with $p_T > 10\,\text{GeV}$ are fed into the PAT-association algorithm, and an "association efficiency" is computed as the fraction of PAT-matched DSA muons out of all selected DSA muons. The resulting efficiencies are obtained for collision data and DY simulation (see Tab. 4.12). They data–MC differences are about 0.2% in 2016 and 0.1% in 2018. The overall effect on the analysis results, therefore, is negligible.

4.7.7 Dimuon selection

The main potential sources for systematic uncertainties in the dimuon selection originate in the vertex $\chi^2/\text{ndof} < 10$ and the DCA $< 15\,\text{cm}$ requirements. Comparing the selection efficiency in dimuons from Drell-Yan events in collision data and MC simulation allows the modeling accuracy of these variables to be evaluated. (The inversion of the DSA-to-PAT

TABLE 4.13: *Vertex χ^2/ndof and DCA selection efficiencies in $Z \to \mu\mu$ events in data and simulation. Statistical uncertainties are negligible and therefore omitted.*

Year	vertex χ^2/ndof < 10 $\varepsilon^{(\text{data})}$	$\varepsilon^{(\text{MC})}$	SF	DCA < 15 cm $\varepsilon^{(\text{data})}$	$\varepsilon^{(\text{MC})}$	SF
2016	0.98	0.98	1.0	0.98	0.99	0.99
2018	0.98	0.98	1.0	0.98	0.99	0.99

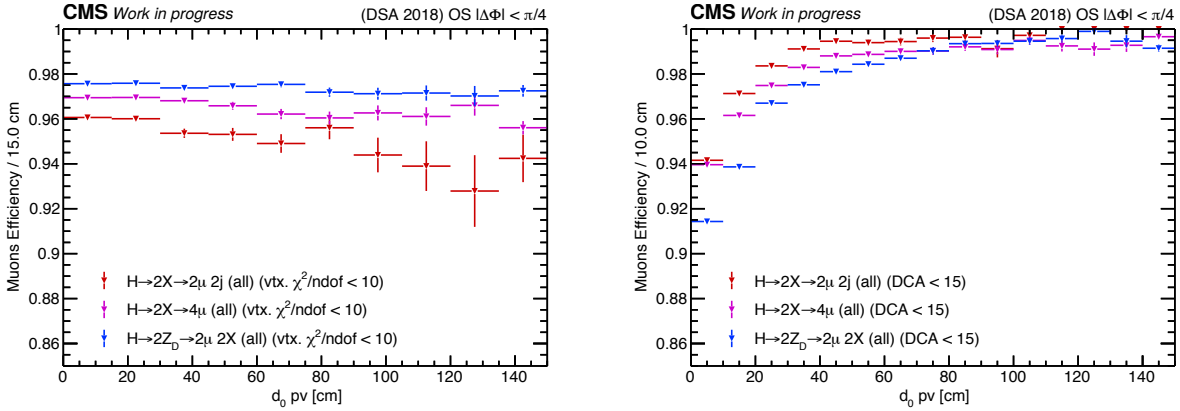


FIGURE 4.53: *Efficiencies of the vertex χ^2/ndof* (left) *and DCA* (right) *requirements measured in the 2018 signal samples. All dimuons that pass the full set of selection criteria except for the vertex χ^2/ndof and DCA cuts, respectively, are considered in the efficiency ratios.*

association and the omission of the requirement of $L_{xy}/\sigma_{L_{xy}} > 6$ enriches the samples in $Z \to \mu\mu$ events.)

Tab. 4.13 shows the χ^2/ndof and DCA efficiencies in data and simulation for each year. The values range from 0.98 to 0.99 and the corresponding (prompt) scale factors are close to 1.

To extrapolate these measurements to the non-prompt regime of the search region, cosmic-ray muons again serve as the measurement basis when studying the d_0 dependence of the efficiencies. These studies are described in detail in Sec. 4.4.6. Particularly, Fig. 4.27 shows that the vertex χ^2/ndof and DCA efficiencies exhibit only a mild dependence on d_0. The vertex χ^2/ndof efficiency is constant up to $d_0 \approx 200\,\text{cm}$, and the DCA efficiency increases by 2% when reaching this d_0 value.

Following these measurements, a systematic uncertainty of 2% is assigned for the dimuon selection as a result of the uncertainties in the vertex χ^2/ndof and DCA selection efficiencies.

Additionally, the efficiencies of the vertex χ^2/ndof and DCA requirements are checked in simulated signals. As the results in Fig. 4.53 show, the signal efficiencies are reasonably stable with growing d_0.

Small differences between vertex χ^2/ndof efficiencies in various signal samples reflect a mild performance dependence on the muon p_T, which is convoluted with differences in muon spectra among the samples.

A mild drop in the DCA cut efficiencies at small d_0 values is observed, which does not match the efficiency measured in Drell-Yan events. This efficiency drop is caused by a class

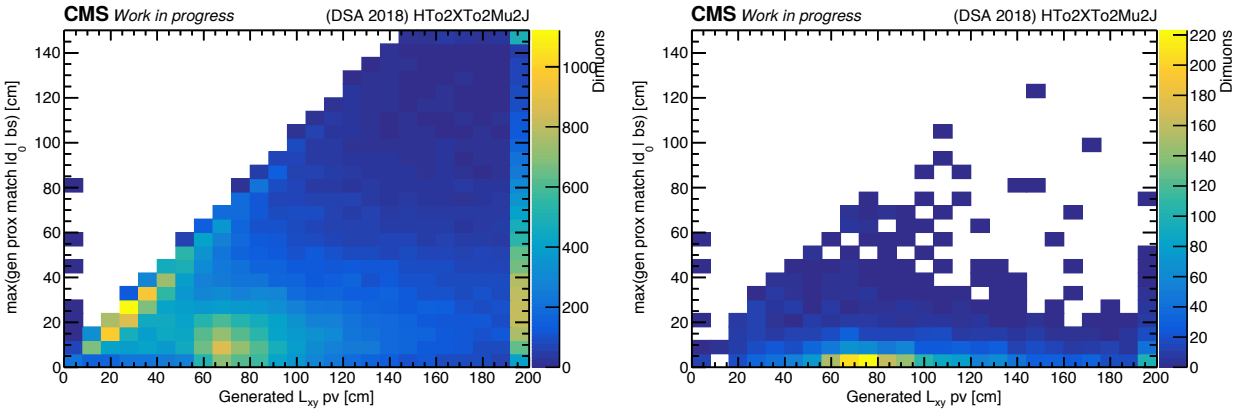


FIGURE 4.54: *Distributions of $H \to 2X \to 2\mu$ signal events in the $\max(d_0{}^{(GEN)})$–$L_{xy}{}^{(GEN)}$ plane.* Left: *Dimuons passing all cuts but the DCA requirement.* Right: *Dimuons failing the DCA requirement. The phase space of $L_{xy} \lesssim 60\,\text{cm}$ and small d_0 is depleted as a consequence of the DSA-to-PAT muon replacement.*

of signal events with very specific kinematic properties, as detailed investigations revealed. Fig. 4.54 illustrates that, in events failing the DCA cut, both muons have relatively small $d_0^{(\text{GEN})}$ values (up to only a few centimeters) while the decay vertex is situated at large $L_{xy}^{(\text{GEN})}$ values (tens of centimeters). Because such muon pairs have a small azimuthal opening (*i. e.*, small $\Delta\phi$) and relatively small p_T, the extrapolated trajectories of the two muons are more likely to be separated from each other by more than 15 cm if their p_T values are not accurately reconstructed. The DCA cut inefficiencies inherent to this class of events can neither be tested with cosmic-ray muons (due to their back-to-back configuration) nor with collision data (due to different kinematics and small L_{xy}). Given that the size of the inefficiency is relatively small (varying between 1% and 8%, depending on the signal sample, as shown in Fig. 4.30), an additional systematic uncertainty equal to the DCA cut efficiency of each respective sample is assigned.

4.7.8 $L_{xy}/\sigma_{L_{xy}}$ resolution

The modeling accuracy of the $L_{xy}/\sigma_{L_{xy}}$ resolution is checked using DSA-DSA dimuons that have associated PAT-PAT dimuons. Subsequently, the $L_{xy}^{(\text{DSA})}/\sigma_{L_{xy}^{(\text{DSA})}} - L_{xy}^{(\text{PAT})}/\sigma_{L_{xy}^{(\text{PAT})}}$ distributions obtained after dropping the $L_{xy}/\sigma_{L_{xy}} > 6$ requirement are used as a proxy for the $L_{xy}/\sigma_{L_{xy}}$ resolution.

The distributions for 2018 data and Drell-Yan simulation are shown as an example in Fig. 4.55. Gaussian fits to the core of the respective distributions yield the widths, $\sigma^{L_{xy}/\sigma_{L_{xy}}}$, and the mean, $\mu^{L_{xy}/\sigma_{L_{xy}}}$, listed in Tab. 4.14.

The results from simulation are generally similar to the ones from data, with the core resolution in the 2018 simulation being slightly optimistic. Given the exponential decrease of the L_{xy} distribution in signal, this observation indicates that the signal simulation cannot overestimate the efficiency of the $L_{xy}/\sigma_{L_{xy}} > 6$ requirement. Therefore, as a conservative measure, the $L_{xy}/\sigma_{L_{xy}}$ efficiency predicted by the signal simulation is used, and no additional systematic uncertainties are applied.

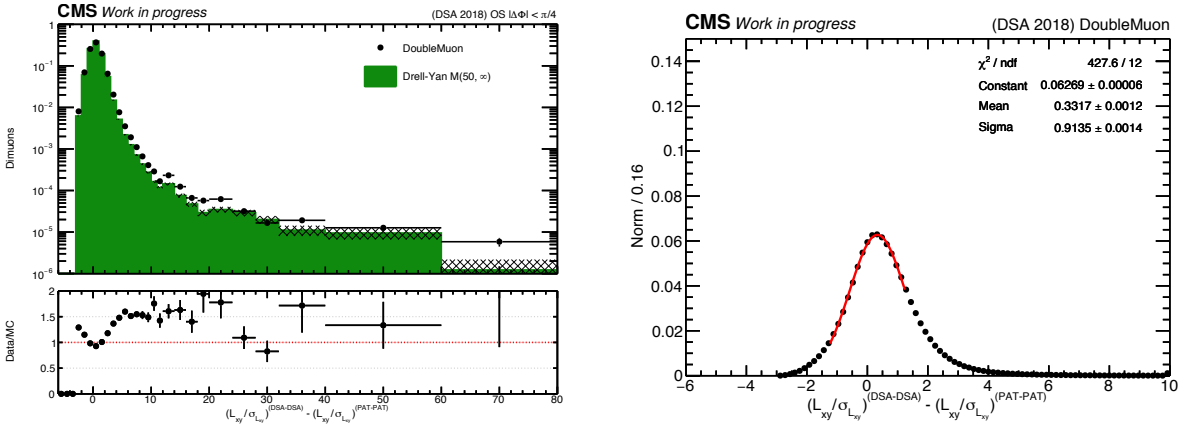


FIGURE 4.55: Left: *Distribution of* $L_{xy}^{(DSA)}/\sigma_{L_{xy}^{(DSA)}} - L_{xy}^{(PAT)}/\sigma_{L_{xy}^{(PAT)}}$ *obtained from PAT-associated DSA-DSA dimuons with* $L_{xy}/\sigma_{L_{xy}} \geq 0$ *in 2018 collision data (black color) and DY simulation (green color).* Right: *Gaussian fit to the core of the collision data distribution from the left-hand figure, with the fitted function parameters in the legend.*

TABLE 4.14: *Values of mean and width* ($\mu^{L_{xy}/\sigma_{L_{xy}}}$ *and* $\sigma^{L_{xy}/\sigma_{L_{xy}}}$*, respectively) of Gaussian fits to the cores of the* $L_{xy}^{(DSA)}/\sigma_{L_{xy}^{(DSA)}} - L_{xy}^{(PAT)}/\sigma_{L_{xy}^{(PAT)}}$ *distributions in collision data and DY simulation, obtained from PAT-associated DSA-DSA dimuons with* $L_{xy}/\sigma_{L_{xy}} \geq 0$.

	$\mu^{L_{xy}/\sigma_{L_{xy}}}$		$\sigma^{L_{xy}/\sigma_{L_{xy}}}$	
Year	Data	MC	Data	MC
2016	0.541 ± 0.002	0.521 ± 0.003	0.819 ± 0.002	0.829 ± 0.003
2018	0.332 ± 0.001	0.340 ± 0.001	0.914 ± 0.001	0.853 ± 0.002

4.8 Statistical analysis and results

For the first time in this work, this section will reveal the events observed in the signal region of the search after unblinding. After a brief overview of the observed signal-region dimuons and a comparison with the expected background in Sec. 4.8.1, the results will be presented as expected vs. observed upper limits on the signal production cross-section (Sec. 4.8.2) in the two signal benchmark models, *i. e.*, the BSM Heavy Scalar model (Sec. 4.8.3) and the HAHM scenario (Sec. 4.8.4).

4.8.1 Events passing the full set of selection criteria after unblinding

The analysis "signal region" (SR) contains all oppositely-charged dimuons at $|\Delta\Phi| < 3\pi/4$. This region was blinded until the final stages of the analysis. Only after having finalized the full event selection and the background prediction strategy were the signal-region events observed for the first time.[14]

The SR data were unblinded in three steps (10%, 30%, 100% of the full dataset). At each step, the observed events were scrutinized, and if any anomalies originating from SM backgrounds had been found, further optimization of the analysis would have been considered. However, it should be emphasized that the threshold for any modifications of the event selection or background estimation was increasingly higher at each unblinding step and only considered as a last resort.

As a matter of fact, no such anomalies were identified at any of the steps, and no modifications of the original analysis were undertaken.

In total, two signal-region events were found in 2016 data and 16 events in 2018. These observations are compatible within statistical uncertainties with the expectations for either year, both in terms of the overall prediction and the per-mass-window estimates. An illustration of expected-vs-observed events as a function of (non-overlapping) mass bins is given in Fig. 4.56, which also contains the integrated event yields and their statistical uncertainties. The results for the nominal, overlapping mass windows (as described in Sec. 4.5.7) are summarized in Tab. 4.15 and 4.16. Finally, the observed events and their values of a selection of key variables are listed in Tab. 4.17 and 4.18. Event displays of all SR events are collected in Appendix B.

After careful inspection of the individual SR events, and given the compatibility of expected and observed events in all search-region mass bins, the conclusion is drawn that there is no statistically significant excess of signal(-like) events beyond the expected background from SM processes.

4.8.2 Limit setting procedure

The results in this section are presented as upper limits on the product of signal production cross-section and branching ratio to two final-state muons. These quantities are obtained both for the BSM Heavy Scalar benchmark model and the Hidden Abelian Higgs ("dark photon") model and presented in Sec. 4.8.3 and 4.8.4, respectively.

[14] The 2016 search, after having been unblinded for the purpose of a previous thesis [36], was re-blinded before the 2018 dataset was studied in detail.

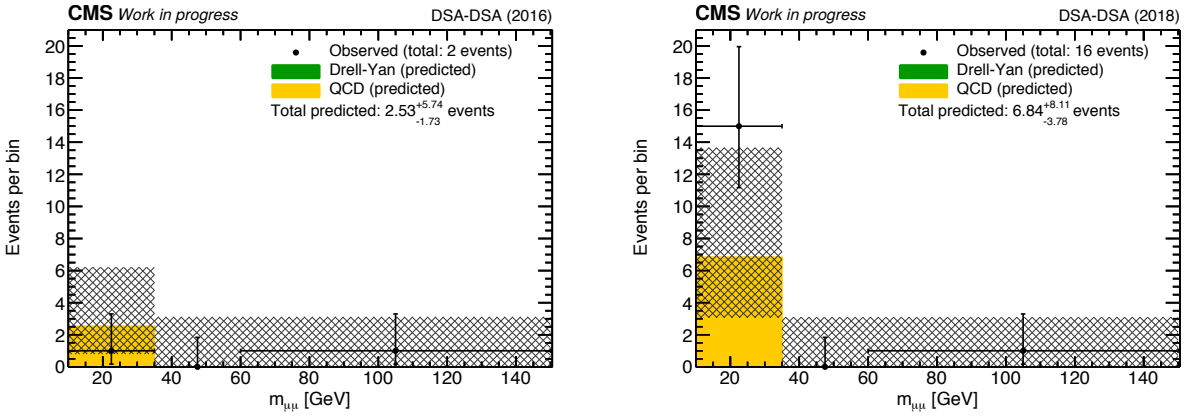


FIGURE 4.56: *Expected (yellow) vs. observed (black) events as a function of non-overlapping mass bins in the 2016* (left) *and 2018* (right) *datasets. Additionally, the integrated event yields are quoted in the legends. All uncertainties are purely statistical.*

TABLE 4.15: *Expected vs. observed numbers of events in the 2016 dataset after unblinding of the signal region. All uncertainties are purely statistical.*

Mass window [GeV]	N_{est}	N_{obs}
$10 < m_{\mu\mu} < 32$	$1.2^{+3.2}_{-1.1}$	1
$15 < m_{\mu\mu} < 60$	$1.1^{+2.9}_{-0.9}$	1
$20 < m_{\mu\mu} < 80$	$1.6^{+3.9}_{-1.4}$	1
$35 < m_{\mu\mu} < 120$	$0.0^{+3.1}_{-0.0}$	0
$35 < m_{\mu\mu} < 245$	$0.0^{+3.1}_{-0.0}$	1
$m_{\mu\mu} > 60$	$0.0^{+3.1}_{-0.0}$	1

TABLE 4.16: *Expected vs. observed numbers of events in the 2018 dataset after unblinding of the signal region. All uncertainties are purely statistical.*

Mass window [GeV]	N_{est}	N_{obs}
$10 < m_{\mu\mu} < 32$	$6.9^{+6.9}_{-3.8}$	15
$15 < m_{\mu\mu} < 60$	$3.1^{+4.3}_{-2.0}$	3
$20 < m_{\mu\mu} < 80$	$0.0^{+2.7}_{-0.0}$	1
$35 < m_{\mu\mu} < 120$	$0.0^{+3.1}_{-0.0}$	0
$35 < m_{\mu\mu} < 245$	$0.0^{+3.1}_{-0.0}$	1
$m_{\mu\mu} > 60$	$0.0^{+3.1}_{-0.0}$	1

TABLE 4.17: *Properties of the observed signal-region dimuons in 2016 data.*

Run, LS, event numbers	Dimuon quantities				Per-muon quantities			
	$m_{\mu\mu}$ [GeV]	L_{xy} [cm]	$L_{xy}/\sigma_{L_{xy}}$	$\lvert\Delta\Phi\rvert$	p_T [GeV]	η	ϕ	d_0 [cm]
278308, 390, 635154343	184.06	6.22	8.81	0.38	76.42, 46.60	1.82, 0.07	−0.93, 1.53	6.49, 1.05
276935, 522, 801995839	24.01	33.91	6.76	0.13	42.08, 51.12	0.45, 0.04	1.48, 1.88	4.58, 12.92

TABLE 4.18: *Properties of the observed signal-region dimuons in 2018 data.*

Run, LS, event numbers	Dimuon quantities				Per-muon quantities			
	$m_{\mu\mu}$ [GeV]	L_{xy} [cm]	$L_{xy}/\sigma_{L_{xy}}$	$\lvert\Delta\Phi\rvert$	p_T [GeV]	η	ϕ	d_0 [cm]
322492, 988, 1745559180	12.02	219.08	97.00	0.12	31.44, 43.22	0.36, 0.37	1.17, 0.98	-58.75, -2.63
320917, 1776, 2808532235	14.49	212.41	100.10	0.27	35.59, 21.86	0.08, 0.05	3.11, -2.82	89.85, 1.07
321475, 659, 1057092687	16.56	213.55	105.09	0.12	88.85, 45.97	0.16, 0.13	-3.00, -2.82	42.09, -4.30
319337, 2154, 2167493931	13.17	64.96	11.05	0.06	21.99, 39.67	1.29, 1.32	0.91, 0.51	-12.39, 13.99
317182, 1252, 1782530391	10.07	137.48	13.57	0.02	234.84, 36.28	0.14, 0.14	0.19, 0.13	-3.99, 7.59
317661, 516, 746797820	10.45	163.42	32.41	0.05	67.74, 21.93	0.70, 0.64	-0.79, -0.64	16.25, -17.47
316240, 1188, 1674517098	11.28	204.51	37.70	0.01	125.32, 12.05	0.99, 1.01	1.30, 1.37	1.95, -35.76
321818, 318, 530726864	151.40	16.87	10.19	0.41	270.48, 273.03	-0.18, 0.04	0.21, 0.24	-5.61, -9.35
320040, 190, 303920363	26.49	203.98	55.76	0.19	25.60, 380.73	-0.42, -0.41	1.13, 1.30	4.20, -40.50
320038, 421, 631745850	12.35	70.11	7.14	0.14	57.80, 33.42	1.20, 0.99	0.30, 0.46	14.29, 0.30
323790, 265, 423394858	11.69	225.62	64.48	0.25	76.31, 11.30	-0.85, -0.72	0.05, 0.17	62.79, 8.10
321961, 128, 234644931	15.85	111.94	24.03	0.19	17.57, 16.77	1.26, 1.45	0.96, 0.14	-26.10, 67.19
323696, 129, 249675640	11.12	215.50	58.98	0.10	79.96, 32.26	0.44, 0.44	-1.57, -1.68	-32.09, 3.25
316876, 112, 137324273	11.03	206.07	57.73	0.10	12.98, 22.51	-1.73, -1.39	3.11, 2.83	-76.36, 5.67
322355, 121, 148061811	10.45	153.11	30.74	0.07	31.71, 30.54	-0.09, 0.004	0.36, 0.59	31.81, -11.30
321933, 213, 327330844	11.01	179.73	18.29	0.08	371.66, 29.49	0.13, 0.13	-2.79, -2.82	-16.04, -3.97

The limit calculation is based on a modified frequentist approach [114, 115] and uses the Combine software package [116], which was originally developed for statistically combining the results of the Higgs boson searches [117]. The multi-bin counting experiment likelihood is used, taking into account the number of observed events in the signal region, the expected number of background events, the expected number of signal events, and the systematic uncertainties. For the expected signal, the latter are treated as nuisance parameters in the likelihood using log-normal distributions. The prediction of the background in the search region is implemented using a multi-bin likelihood, which is a product of Poisson distributions corresponding to the signal region and the control regions. Each background contribution (*i. e.*, DY and QCD background) is estimated separately and eventually combined in the final search-region prediction. Because the number of expected background events is small, the limits are computed by generating toy experiments to obtain the distribution of the test statistic. The full configuration of Combine parameters, as well as an example Combine input file, is given in Appendix C.

If a genuine signal were present in the recorded data, the number of signal events with at least one selected dimuon would be

$$N_S = \mathcal{L}\sigma_{\mathrm{LLP}} \cdot \left[2\mathcal{B}\left(1-\mathcal{B}\right)\varepsilon_{2\mu} + \mathcal{B}^2\varepsilon_{4\mu}\right] \ , \tag{4.8.1}$$

where $\mathcal{L}$ is the integrated luminosity of the dataset, $\sigma_{\mathrm{LLP}} = \sigma(\mathrm{H})\mathcal{B}(\mathrm{H} \to 2\,\mathrm{LLP})$ is the signal production cross-section, $\mathcal{B} = \mathcal{B}(\mathrm{LLP} \to \mu\mu)$ is the branching fraction of the LLP decay to a muon pair, and $\varepsilon_{2\mu}$ and $\varepsilon_{4\mu}$ are the respective fractions of 2μ and 4μ events in which at least one dimuon candidate passes all selection criteria. The signal efficiencies, ε_i, are determined with MC simulation and corrected with the data-to-simulation scale factors described in Sec. 4.7. They are obtained separately for each data-taking year, signal model, and mass window.

In the BSM Heavy Scalar benchmark model, the signal cross-section, σ_{LLP}, is the product of the BSM Higgs gluon-fusion cross-section, $\sigma(\mathrm{H})$, and the branching fraction of the BSM Higgs boson to a pair of long-lived particles, $\mathcal{B}(\mathrm{H} \to XX)$. The branching fraction of X to a pair of final-state muons is given by $\mathcal{B}(X \to \mu\mu)$, and it controls both the total yields and relative fractions of $\mathrm{H} \to XX \to 2\mu$ and $\mathrm{H} \to XX \to 4\mu$ events. To provide results that are as model-independent as possible, both σ_{LLP} and $\mathcal{B}(X \to \mu\mu)$ are treated as free parameters.

Thus, the number of expected events in the BSM Heavy Scalar benchmark samples scales according to

$$N_S = \mathcal{L}\sigma(\mathrm{H})\mathcal{B}(\mathrm{H} \to XX) \cdot \left[2\mathcal{B}(X \to \mu\mu)\left(1-\mathcal{B}(X \to \mu\mu)\right)\varepsilon_{2\mu} + \left(\mathcal{B}(X \to \mu\mu)\right)^2\varepsilon_{4\mu}\right] \ . \tag{4.8.2}$$

In the Hidden Abelian Higgs Model, σ_{LLP} is given as the product of the SM Higgs gluon-fusion cross-section, $\sigma(\mathrm{h})$, and the branching fraction of the SM Higgs boson to a pair of long-lived dark photons, $\mathcal{B}(\mathrm{h} \to Z_D Z_D)$. The predicted value of $\mathcal{B}(\mathrm{h} \to Z_D Z_D)$ depends on the parameters of the hidden sector according to Eq. 2.2.6. On the other hand, the dark photon branching fraction $\mathcal{B}(Z_D \to \mu\mu)$ is a function only of $m(Z_D)$, and its values range from 10% to 16% in the probed mass range (see Tab. 4.2). In contrast to the BSM Higgs benchmark samples, the dark photon samples were produced with the predicted Z_D-to-$\mu\mu$ branching fractions, and so the measured signal efficiencies, ε_{Z_D}, already contain the appropriate mix of 2μ and 4μ signal dimuons.

The expected signal yield in the HAHM model can be expressed as

$$N_S = \mathcal{L}\sigma(\mathrm{h})\mathcal{B}(\mathrm{h} \to Z_D Z_D)\mathcal{B}(Z_D \to \mu\mu) \cdot \left[2 - \mathcal{B}(Z_D \to \mu\mu)\right] \cdot \varepsilon_{Z_D} \ . \tag{4.8.3}$$

Results for lifetimes that do not lie on the grid of simulated $c\tau_{\text{LLP}}$ values (see Tab. 4.2 and 4.3) are obtained by reweighting the signals from a generated lifetime point.

4.8.3 Limits in the BSM Heavy Scalar model

The expected and observed 95% confidence level (CL) upper limits on $\sigma(\text{H} \to XX)\mathcal{B}(X \to \mu\mu)$, with $\mathcal{B}(X \to \mu\mu) \equiv 1$, are presented as a function of the LLP lifetime, $c\tau_X$, in the BSM Higgs benchmark model. They are obtained from 2016 and 2018 data and from the full Run-2 dataset (2016 + 2018 data). The results are shown for $m(\text{H}) = 125\,\text{GeV}$ in Fig. 4.57, for $m(\text{H}) = 200\,\text{GeV}$ in Fig. 4.58, for $m(\text{H}) = 400\,\text{GeV}$ in Fig. 4.59, and for $m(\text{H}) = 1000\,\text{GeV}$ in Fig. 4.60.

The relative sensitivity reached with 2016 and 2018 data varies significantly with the parameters of the underlying signal model. At small LLP mass, the signal efficiency is larger in 2018 than in 2016 as a result of more inclusive triggers. However, the predicted background levels are also larger in 2018. The net effect are comparable sensitivities between the years.

For larger LLP masses, the expected background is reduced to a negligible rate in both years. As a consequence, the larger sensitivity in 2018 data comes to light, further enhanced by the larger integrated luminosity of the 2018 data compared to 2016 data. More concretely, the 2018 sensitivity at $c\tau_X \leq 10\,\text{cm}$ benefits significantly from the cosmic-seeded trigger introduced in this year.

The combined 95% CL upper limits (2016 + 2018) are most stringent for $c\tau_X = 10$–$100\,\text{cm}$, where they can exclude values of $\sigma(\text{H}) \to XX)\mathcal{B}(X \to \mu\mu)$ as low as 0.3 fb for large $m(X)$. For small LLP mass, $m(X) \approx 20\,\text{GeV}$, the trigger and selection efficiencies decrease, and the best upper limits are of the order of a few femtobarns in this case.

The results obtained via DSA-DSA dimuons are not sensitive to the smallest lifetimes, as dimuons in this regime are usually replaced by PAT-PAT or DSA-PAT dimuons and analyzed separately.[15] On the other end of the lifetime spectrum, at large $c\tau_X$ values, the sensitivity is limited by the trigger efficiency (particularly at the L1 trigger) and the fact that an increasing fraction of dimuons is produced outside the CMS detector.

4.8.4 Limits in the Hidden Abelian Higgs Model

The expected and observed 95% CL upper limits in the context of the HAHM dark-photon model are presented in Fig. 4.61 in a similar manner to the results of the previous section. The limits are set on $\sigma(h \to Z_D Z_D)\cdot\mathcal{B}(Z_D \to \mu\mu)$ for different $m(Z_D)$ hypotheses. The value of $\mathcal{B}(Z_D \to \mu\mu)$ decreases from = 0.1538 at $m(Z_D) = 10\,\text{GeV}$ to 0.1069 at $m(Z_D) = 60\,\text{GeV}$ (see also Tab. 4.2).

For the same reasons as discussed in the context of the BSM Heavy Scalar benchmark limits, the sensitivity differences between 2016 and 2018 data varies considerably with $m(Z_D)$. At small Z_D masses ($m(Z_D) = 10\,\text{GeV}$ and $20\,\text{GeV}$), both the signal efficiencies and the predicted background levels are larger in 2018 than in 2016, producing comparable sensitivities for the two datasets. In the intermediate mass range ($m(Z_D) = 30\,\text{GeV}$), the 2018 dataset has larger sensitivity thanks to its higher signal efficiencies and larger integrated luminosity, despite the higher background levels. At the highest-probed dark

[15] As mentioned before, the analysis of the PAT-based dimuon categories is beyond the scope of this work and a parallel effort of a collaborating group. Results of these studies, including a combination of the results of all dimuon categories, will be published in the near future.

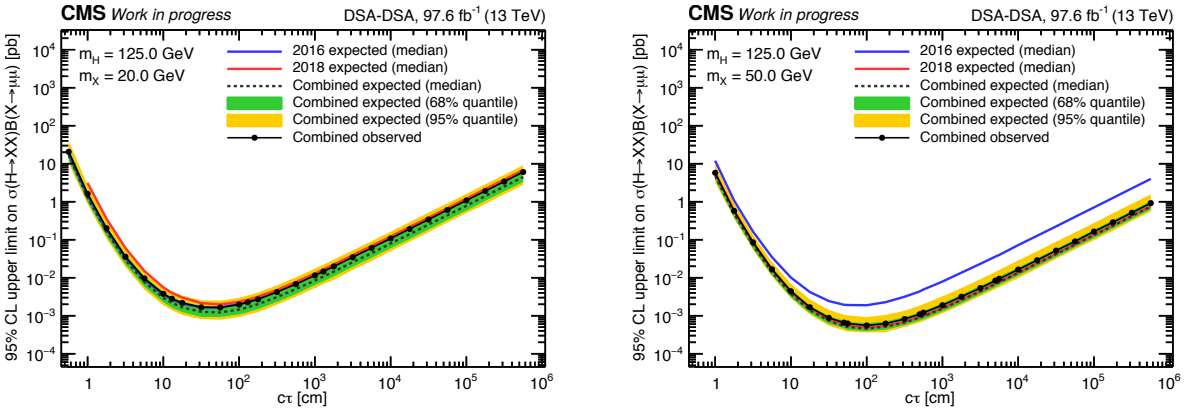


FIGURE 4.57: *Expected and observed 95% CL upper limits on* $\sigma(H \to XX)\mathcal{B}(X \to \mu\mu)$ *as a function of* $c\tau_X$ *for* $m(H) = 125\,\text{GeV}$ *and* $m(X) = 20\,\text{GeV}$ (left) *and* $m(X) = 50\,\text{GeV}$ (right). *The solid blue, solid red, and dashed black lines correspond to the limits obtained with 2016, 2018, and with the combined 2016–2018 data, respectively. The green and yellow bands respectively show the 68% and 95% quantiles for the expected combined limits.*

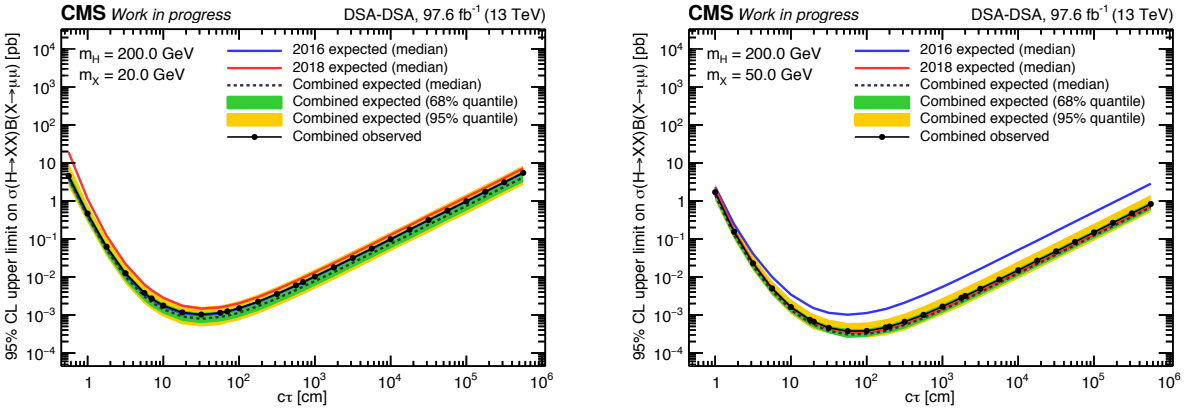


FIGURE 4.58: *Expected and observed 95% CL upper limits on* $\sigma(H \to XX)\mathcal{B}(X \to \mu\mu)$ *as a function of* $c\tau_X$ *for* $m(H) = 200\,\text{GeV}$ *and* $m(X) = 20\,\text{GeV}$ (left) *and* $m(X) = 50\,\text{GeV}$ (right). *The solid blue, solid red, and dashed black lines correspond to the limits obtained with 2016, 2018, and with the combined 2016–2018 data, respectively. The green and yellow bands respectively show the 68% and 95% quantiles for the expected combined limits.*

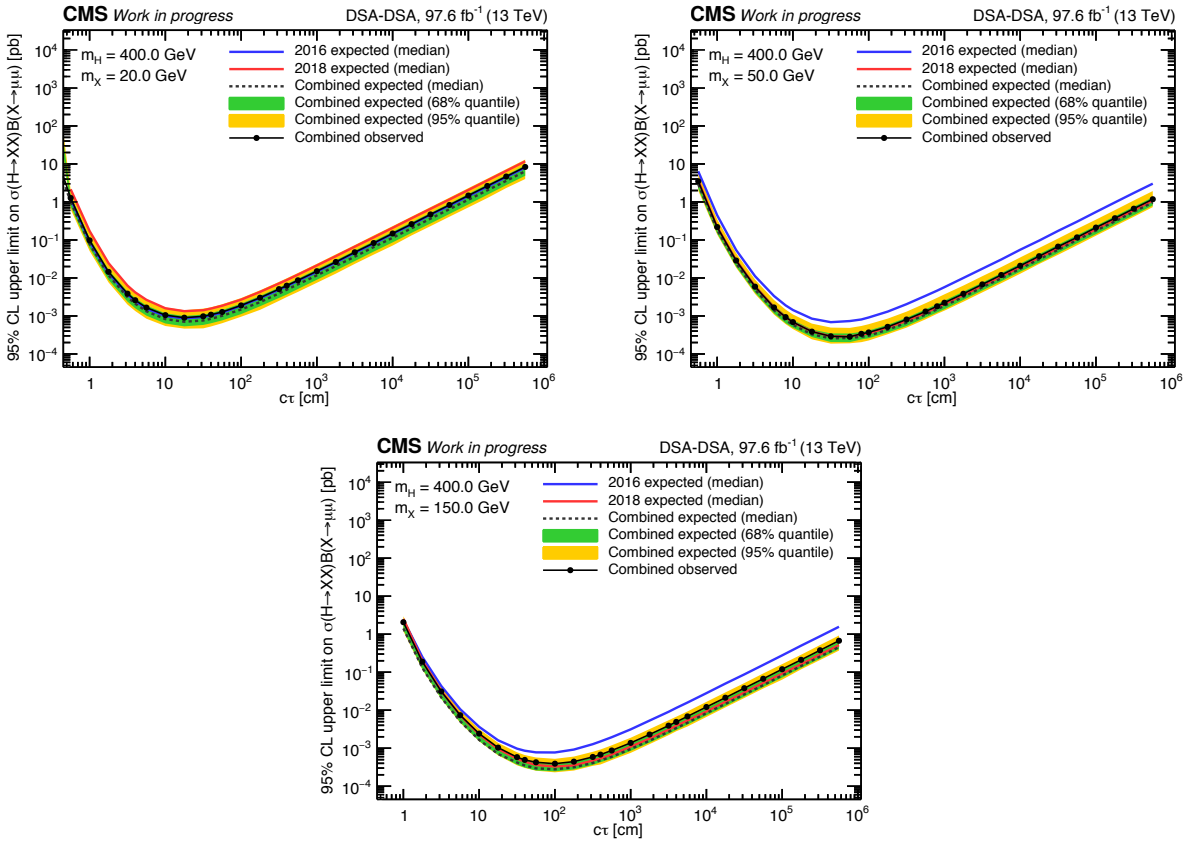


FIGURE 4.59: *Expected and observed 95% CL upper limits on* $\sigma(H \to XX)\mathcal{B}(X \to \mu\mu)$ *as a function of* $c\tau_X$ *for* $m(H) = 400\,\text{GeV}$ *and* $m(X) = 20\,\text{GeV}$ (top left) *and* $m(X) = 50\,\text{GeV}$ (top right) *and* $m(X) = 150\,\text{GeV}$ (bottom). *The solid blue, solid red, and dashed black lines correspond to the limits obtained with 2016, 2018, and with the combined 2016–2018 data, respectively. The green and yellow bands respectively show the 68% and 95% quantiles for the expected combined limits.*

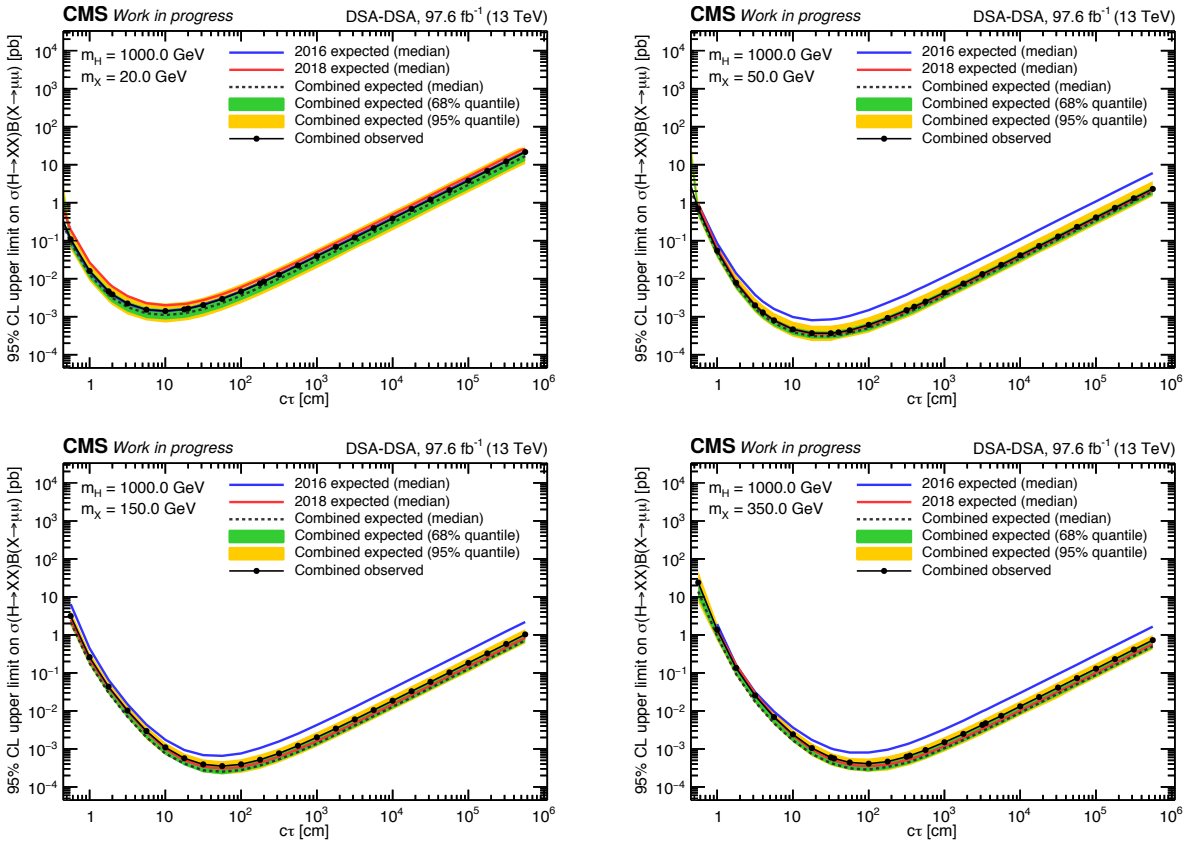


FIGURE 4.60: *Expected and observed 95% CL upper limits on* $\sigma(H \to XX)\mathcal{B}(X \to \mu\mu)$ *as a function of* $c\tau_X$ *for* $m(H) = 1000\,\text{GeV}$ *and* $m(X) = 20\,\text{GeV}$ (top left) *and* $m(X) = 50\,\text{GeV}$ (top right) *and* $m(X) = 150\,\text{GeV}$ (bottom left) *and* $m(X) = 350\,\text{GeV}$ (bottom right). *The solid blue, solid red, and dashed black lines correspond to the limits obtained with 2016, 2018, and with the combined 2016–2018 data, respectively. The green and yellow bands respectively show the 68% and 95% quantiles for the expected combined limits.*

photon masses ($m(Z_D) \geq 40\,\text{GeV}$), both 2016 and 2018 searches are virtually background-free, and the 2018 search can set significantly more stringent constraints, again, as a result of larger signal efficiencies and luminosity.

The expected 2016 limits at $m(Z_D) = 20\,\text{GeV}$, 40 GeV, and 60 GeV can be compared with the expected limits set in the ATLAS search for displaced dimuons with 2016 data [64]. These ATLAS results are overlaid in Fig. 4.61. One can see that this analysis of CMS data is more sensitive than the ATLAS search at large lifetimes. The improvements over the ATLAS limits become more pronounced as $m(Z_D)$ increases, which is likely due to a requirement of $\Delta R(\mu\mu) < 0.5$ in the signal trigger used by ATLAS, resulting in decreasing efficiencies as $m(Z_D)$ approaches $m(\text{h})/2$.

The upper limits on $\sigma(\text{h} \to Z_D Z_D) \cdot \mathcal{B}(Z_D \to \mu\mu)$ in Fig. 4.61 are compared to the theoretical predictions for a set of $\mathcal{B}(\text{h} \to Z_D Z_D)$ values between 10% and 0.05%. At the dark-photon lifetimes where the search is most sensitive (*i. e.*, beyond $c\tau_{Z_D} > 10\,\text{cm}$, depending on $m(Z_D)$), the combined 2016+2018 analysis is expected to exclude $\mathcal{B}(\text{h} \to Z_D Z_D)$ of more than 0.1% for $m(Z_D) = 10\,\text{GeV}$ and almost 0.001% for $m(Z_D) = 60\,\text{GeV}$ at 95% CL in the absence of signal. These constraints on rare SM Higgs boson decays are tighter than those derived from searches for invisible Higgs boson decays [118] and than indirect constraints from measurements of the SM Higgs boson couplings [14].

The expected 95% CL exclusion contours in the $m(Z_D)$–$c\tau_{Z_D}$ plane are shown in Fig. 4.62 *(left)* for several representative values of $\mathcal{B}(\text{h} \to Z_D Z_D)$. They are directly derived from the expected 2016+2018 upper limits of Fig. 4.61. These results can also be interpreted as limits on the kinetic mixing parameter ε in the HAHM model. The corresponding 95% CL exclusion contours are presented in the $m(Z_D)$–ε plane and shown in Fig. 4.62 *(right)*. For $\mathcal{B}(\text{h} \to Z_D Z_D) = 1\%$, the search excludes almost two orders of magnitude of ε values at $m(Z_D) = 60\,\text{GeV}$ and more than an order of magnitude at $m(Z_D) = 10\,\text{GeV}$. A comparison with the ATLAS results for $\mathcal{B}(\text{h} \to Z_D Z_D) = 1\%$ and 10% [64] shows that the sensitivity of the current CMS analysis extends to much smaller values of kinetic mixing ε.

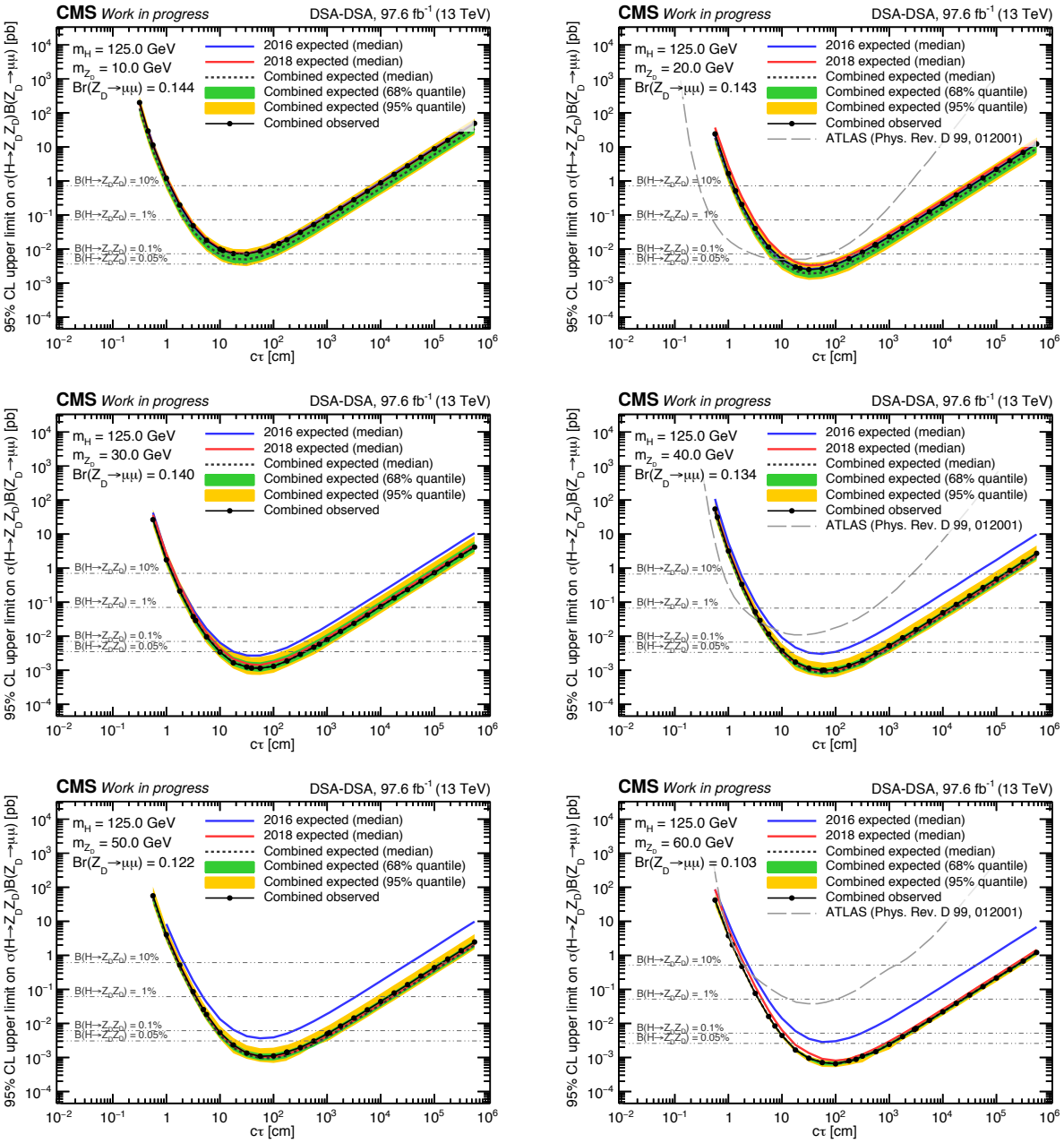


FIGURE 4.61: *Expected and observed 95% CL upper limits on* $\sigma(h \to Z_D Z_D)\mathcal{B}(Z_D \to \mu\mu)$ *as a function of* $c\tau_{Z_D}$ *ranging from* $m(Z_D) = 10\,\text{GeV}$ (top left) *to* $m(Z_D) = 60\,\text{GeV}$ (bottom right). *The solid blue, solid red, and dashed black lines correspond to the limits obtained with 2016, 2018, and with the combined 2016–2018 data, respectively. The green and yellow bands respectively show the 68% and 95% quantiles for the expected combined limits. The gray horizontal lines correspond to the theoretical predictions for values of* $\mathcal{B}(h \to Z_D Z_D$ *indicated next to the lines. The scenarios for* $m(Z_D)$ *of* 20, 40, *and* 60 GeV *are compared to the expected limits obtained in the ATLAS search for displaced dimuons with 2016 data (dashed gray curves) [64].*

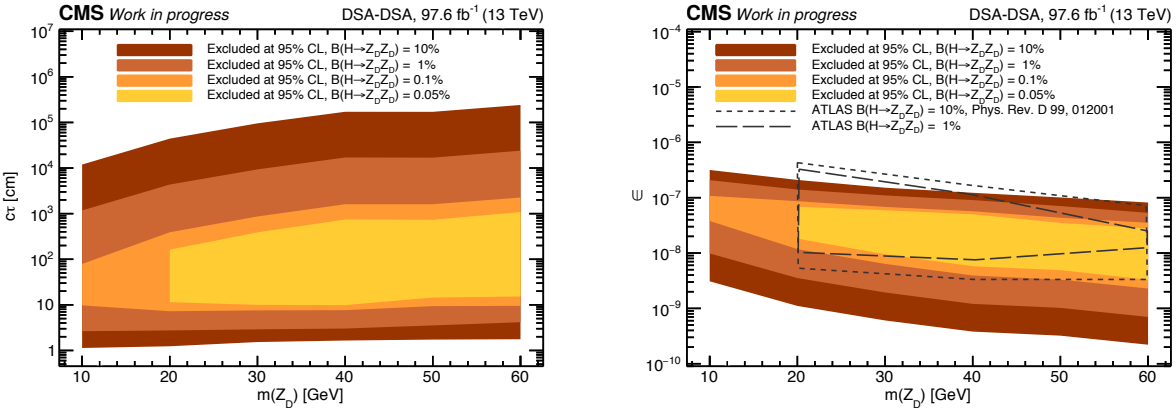


FIGURE 4.62: *Expected 95% CL exclusion contours in the* $(m(Z_D), c\tau_{Z_D})$ (left) *and* $(m(Z_D), \varepsilon)$ (right) *planes. The contours correspond to several representative values of* $\mathcal{B}(h \to Z_D Z_D)$ *ranging from* 0.05% *to* 10%*. The results of the measurements at the generated six* Z_D *mass points are connected by straight lines. The results in the* $(m(Z_D), \varepsilon)$ *plane are compared with the observed excluded regions for* $\mathcal{B}(h \to Z_D Z_D)$ *of* 1% *and* 10% *obtained in the ATLAS search for displaced dimuons with 2016 data [64].*

Summary and concluding remarks

Physics beyond the Standard Model of elementary particles (SM) remains a mystery. Despite the shortcomings of the SM and the potential of long-lived particles (LLPs) to address some of them, the generic and inclusive search for displaced dimuons presented in this work did not discover new particles or decay chains beyond the expectations from the known SM processes.

However, a great many insights have been made in the quest for searching for displaced dimuons in pp-collisions at $\sqrt{s} = 13\,\text{TeV}$ recorded in 2016 and 2018.

Extensive muon object studies were performed with cosmic-ray muons, *i. e.*, the only known natural source of muons displaced in the detector. In doing so, properties specific to cosmic-ray muons were exploited to gain a detailed understanding of the muon reconstruction performance at all relevant stages of the CMS muon reconstruction, from the *online* reconstruction algorithms at the L1 and HLT triggers to the *offline* ones.

These studies identified and measured a large source of muon reconstruction inefficiency in CMS, the so-called *L1 bias* for muons with non-zero transverse impact parameter d_0 due to the implicit beam-spot constraint in the L1 muon reconstruction algorithms.

Further, measurements with cosmic-ray muons demonstrated the superiority of the "cosmic-seeded" online muon reconstruction in the regime of large muon d_0, which was introduced in the 2018 data-taking period.

Careful decoupling of the various reconstruction effects revealed that the *major* source of inefficiency originates in the *L1 bias*, with the rest of the muon reconstruction pipeline working reasonably efficiently over a broad range of transverse impact parameters thanks to the inclusion of the cosmic-seeded L2 reconstruction algorithm.

The introduction of the cosmic-seeded HLT path for 2018 data did not only bring increased levels and new types of cosmic-ray background but also additional pp-collision background because of its generally less stringent muon pair requirements. This allowed for richer control regions, detailed studies of the novel kinds of background in 2018 data, and, ultimately, the design of effective event selection criteria.

The presented search for displaced dimuons was *blinded* until the final stages of the analysis, *i. e.*, the events in the search regions (oppositely-charged dimuons with small collinearity angle, $|\Delta\Phi| < \pi/4$) were not looked at before the analysis strategy, event selection, and background estimation were finalized.

One of the key characteristics of the analysis was the prediction of SM backgrounds based on recorded detector data alone, thus circumventing the reliance on MC simulation for this purpose entirely (which, after all, cannot be trusted to model all the relevant instrumental effects sufficiently faithfully).

In the end, this data-driven approach made it possible to estimate the SM background content of the search regions within the predicted statistical uncertainties, as the comparison with the data observed after unblinding demonstrated. No excess beyond the expectations from SM processes was seen in any of the search regions.

The results were eventually interpreted in the context of two common signal benchmark models, the BSM Heavy Higgs and the HAHM scenarios, where upper limits on the LLP production cross-sections were calculated. Large volumes of yet-unexplored phase space were thus excluded, with the presented search showing its strength, particularly in the domain of large LLP lifetimes as a consequence of the chosen muon-system-only reconstruction approach.

Looking forward, a few immediate predictions for the near future of displaced muon searches can be made based on the insights gained in the presented analysis.

First, except for the purpose of background rejection, the work at hand did not use any tracker-based muons (called *PAT muons* in this context). Instead, it relied on objects reconstructed with muon-system-only hit information (*i. e.*, DSA muons). The inclusion of dimuon vertex categories based on PAT muons (*i. e.*, PAT-PAT and DSA-PAT dimuons) will complement the DSA-DSA analysis and provide strong upper limits, especially in the regime of small lifetimes. An ultimate combination of all three dimuon categories in the statistical interpretation of the results will give even more competitive upper limits. Each of these dimuon categories has its own analysis and requires dedicated event selection optimization, studies of backgrounds, and assessment of systematic uncertainties. The work on the PAT-based displaced dimuon searches is ongoing, and these results will be presented in future publications, together with a combined result for all three dimuon vertex categories.

Finally, as this work demonstrated, the biggest gain in CMS muon reconstruction efficiency can be achieved by optimizing the L1 muon reconstruction for muons with non-zero transverse impact parameter d_0. Concretely, the removal of the implicit beam-spot constraint at L1 will give a significant boost to future displaced-muon searches. This very effort has already been started, and a considerable amount of work is ongoing behind the scenes, aiming for a vastly improved online muon reconstruction in the LHC Run 3 and beyond.

This work is merely the beginning of LLP searches for displaced dimuons. Significantly improved experimental and analytical methods, together with vastly larger datasets, promise a bright future for LLP searches beyond the SM.

Appendices

DSA charge-misassignment analysis in the QCD background evaluation

In the QCD background estimation of the primary analysis, the transfer factor R_{QCD} is measured as the ratio of oppositely- to equally-charged DSA-DSA dimuons that have one of their muon legs associated with a PAT muon (by virtue of the DSA–PAT association algorithm discussed in Sec. 4.5.2).

In principle, the R_{QCD} measurement could also be performed with doubly-PAT-associated DSA-DSA dimuons, but while this configuration provides a useful handle for the studies of this section, the conclusions drawn from these very studies disfavor this option for the R_{QCD} measurement in the nominal background estimation.

The goal of this section is to assess whether there is a correlation between the DSA charge misassignment and the DSA-to-PAT muon association.

To this end, singly-PAT-associated DSA-DSA muons are studied, and the DSA charges are compared to their PAT-charge counterparts. The PAT muon charges are treated as a proxy for the true muon charges, which is sufficiently justified given that the PAT measurement is generally known to be reliable. This sample can be decomposed into four distinct subsamples, which are split according to their DSA-DSA charges and whether there is a charge-flip between DSA and associated PAT muon:

$$N^{\mathrm{OS}}_{q^{\mathrm{DSA}}=q^{\mathrm{PAT}}} = (1-P_{\mathrm{assoc}})(1-P_{\text{non-assoc}})N^{\mathrm{OS}}_{\mathrm{true}} + (1-P_{\mathrm{assoc}})P_{\text{non-assoc}}N^{\mathrm{SS}}_{\mathrm{true}} \tag{A.1}$$

$$N^{\mathrm{OS}}_{q^{\mathrm{DSA}}\neq q^{\mathrm{PAT}}} = P_{\mathrm{assoc}}P_{\text{non-assoc}}N^{\mathrm{OS}}_{\mathrm{true}} + P_{\mathrm{assoc}}(1-P_{\text{non-assoc}})N^{\mathrm{SS}}_{\mathrm{true}} \tag{A.2}$$

$$N^{\mathrm{SS}}_{q^{\mathrm{DSA}}=q^{\mathrm{PAT}}} = (1-P_{\mathrm{assoc}})P_{\text{non-assoc}}N^{\mathrm{OS}}_{\mathrm{true}} + (1-P_{\mathrm{assoc}})(1-P_{\text{non-assoc}})N^{\mathrm{SS}}_{\mathrm{true}} \tag{A.3}$$

$$N^{\mathrm{SS}}_{q^{\mathrm{DSA}}\neq q^{\mathrm{PAT}}} = P_{\mathrm{assoc}}(1-P_{\text{non-assoc}})N^{\mathrm{OS}}_{\mathrm{true}} + P_{\mathrm{assoc}}P_{\text{non-assoc}}N^{\mathrm{SS}}_{\mathrm{true}}\ , \tag{A.4}$$

where P_{assoc} and $P_{\text{non-assoc}}$ are the charge misassignment probabilities of the PAT-associated and PAT-non-associated DSA muon, respectively, and $N^{\mathrm{OS}}_{\mathrm{true}}$ and $N^{\mathrm{SS}}_{\mathrm{true}}$ are the respective numbers of true opposite- and same-sign dimuons in the sample. The indices $q^{\mathrm{DSA}} = q^{\mathrm{PAT}}$ or $q^{\mathrm{DSA}} \neq q^{\mathrm{PAT}}$ indicate the agreement or disagreement, respectively, between the DSA and PAT charges in the DSA muon leg associated with a PAT muon.

From the above equations, the DSA charge misassignment probabilities can be determined as

$$P_{\mathrm{assoc}} = \frac{N^{\mathrm{OS}}_{q^{\mathrm{DSA}}\neq q^{\mathrm{PAT}}} + N^{\mathrm{SS}}_{q^{\mathrm{DSA}}\neq q^{\mathrm{PAT}}}}{N^{\mathrm{OS}}_{\mathrm{true}} + N^{\mathrm{SS}}_{\mathrm{true}}} \tag{A.5}$$

$$P_{\text{non-assoc}} = \frac{R_{\mathrm{true}} - R_{q^{\mathrm{DSA}}=q^{\mathrm{PAT}}}}{(R_{\mathrm{true}}-1)(R_{q^{\mathrm{DSA}}=q^{\mathrm{PAT}}}+1)}\ , \tag{A.6}$$

where

$$R_{\text{true}} = \frac{N^{\text{OS}}_{\text{true}}}{N^{\text{SS}}_{\text{true}}} \tag{A.7}$$

$$R_{q^{\text{DSA}}=q^{\text{PAT}}} = \frac{N^{\text{OS}}_{q^{\text{DSA}}=q^{\text{PAT}}}}{N^{\text{SS}}_{q^{\text{DSA}}=q^{\text{PAT}}}} \,. \tag{A.8}$$

From the singly-PAT-associated DSA-DSA dimuon sample alone, it is impossible to determine R_{true}, because the true charge of the non-associated DSA muon is unknown. However, if one assumes that the physical composition of the singly-PAT-associated sample is similar to the composition of the doubly-PAT-associated sample, one can use for R_{true} the corresponding ratio from the doubly-PAT-associated sample (where it can be measured using the PAT charges of both muon legs). This assumption is supported by the similarity of the distributions in Fig. 4.45 and 4.46.

If P_{assoc} and $P_{\text{non-assoc}}$ were equal, Eq. A.1 would indicate no correlation between the DSA charge assignment and the DSA-to-PAT association. However, the measurements yield $P_{\text{assoc}} \neq P_{\text{non-assoc}}$ (see Tab. A.1 and A.2), which suggests such a correlation if the above assumptions hold.

Assuming further that the similarity in sample composition also holds for the nominal DSA-DSA sample with *none* of the DSA muons PAT-associated and that the two muons can be treated independently, one can the QCD transfer factor R_{QCD} from the charge-misassignment probability $P_{\text{non-assoc}}$:

$$R_{\text{QCD}} = \frac{(1-x)R^{(\text{doubly-PAT-assoc})}_{\text{true}} + x}{(1-x) + xR^{(\text{doubly-PAT-assoc})}_{\text{true}}} \tag{A.9}$$

$$x = 2 \cdot P_{\text{non-assoc}}(1 - P_{\text{non-assoc}}) \tag{A.10}$$

The discussion around Fig. 4.48 indicates that the QCD-enriched samples (*i. e.*, the ones used for the R_{QCD} measurements) contain two dominant types of events. At small mass, the sample is primarily composed of mismeasured low-mass resonances. At high mass, instead, the dominant source of dimuons is formed from two independent non-prompt muons from heavy-flavor meson decays.

If a correlation between DSA charge misassignment and DSA-to-PAT association exists, the measured transfer factors tend to be overestimated (because they are measured in a sample that is selected via a DSA-to-PAT-association requirement). In fact, the degree of overestimation strongly depends on whether *one* or *both* DSA muons are required to be PAT-associated in the transfer factor measurement.

This suggests using the *singly*-PAT-associated DSA-DSA dimuon selection for the R_{QCD} measurement instead of the double-PAT-associated one in order to reduce the correlation bias.

Given that several assumptions went into Eq. A.9, the choice has been made not to use the R_{QCD} values computed via charge misassignment probabilities for the R_{QCD} central values but to use them to assign a downwards-asymmetric systematic uncertainty for the nominal R_{QCD} measurement of the primary analysis (see Tab. 4.9).

Finally, Tab. A.1 and A.2 summarize the numerical results in 2016 and 2018 data, respectively. They also contain a comparison of the R_{QCD} values obtained using Eq. A.9 and the ones obtained via the nominal transfer factor measurement using a sample of singly-PAT-associated DSA-DSA dimuons. The first line in Tab. A.2 addresses the low-mass validation region (*VR-IMASS*), which is only available in 2018 due to the trigger

TABLE A.1: *Ingredients and values of R_{QCD} calculated from Eq. A.9 in 2016 data compared to R_{QCD} obtained from singly-PAT-associated DSA-DSA dimuons (nominal R_{QCD} used in the analysis). The last column shows the relative systematic uncertainty on the nominal R_{QCD}, which results from its difference to R_{QCD} measured from $P_{non\text{-}assoc}$. All uncertainties are statistical.*

Mass [GeV]	$R_{\text{true}}^{(\text{doubly-PAT-assoc})}$	P_{assoc}	$P_{\text{non-assoc}}$	R_{QCD} from $P_{\text{non-assoc}}$	Nominal R_{QCD}	Uncertainty
$10 < m_{\mu\mu} < 32$	21.3	0.09 ± 0.03	0.40 ± 0.07	1.1 ± 0.1	1.2 ± 0.3	−8%
$15 < m_{\mu\mu} < 60$	8.1	0.09 ± 0.04	0.41 ± 0.09	1.1 ± 0.1	1.1 ± 0.3	0%
$20 < m_{\mu\mu} < 80$	5.2	0.20 ± 0.19	0.32 ± 0.10	1.2 ± 0.2	1.6 ± 0.5	−13%
$m_{\mu\mu} > 35$	4.0	0.29 ± 0.28	0.30 ± 0.14	1.2 ± 0.3	1.5 ± 0.5	−20%

TABLE A.2: *Ingredients and values of R_{QCD} calculated from Eq. A.9 in 2018 data compared to R_{QCD} obtained from singly-PAT-associated DSA-DSA dimuons (nominal R_{QCD} used in the analysis). The last column shows the relative systematic uncertainty on the nominal R_{QCD}, which results from its difference to R_{QCD} measured from $P_{non\text{-}assoc}$. All uncertainties are statistical.*

Mass [GeV]	$R_{\text{true}}^{(\text{doubly-PAT-assoc})}$	P_{assoc}	$P_{\text{non-assoc}}$	R_{QCD} from $P_{\text{non-assoc}}$	Nominal R_{QCD}	Uncertainty
$6 < m_{\mu\mu} < 10$	43.3	0.039 ± 0.004	0.12 ± 0.01	3.4 ± 0.2	5.9 ± 0.2	
$10 < m_{\mu\mu} < 32$	29.4	0.030 ± 0.007	0.27 ± 0.02	1.5 ± 0.1	2.3 ± 0.2	−35%
$15 < m_{\mu\mu} < 60$	8.3	0.019 ± 0.008	0.36 ± 0.04	1.1 ± 0.1	1.6 ± 0.2	−31%
$20 < m_{\mu\mu} < 80$	4.9	0.011 ± 0.008	0.41 ± 0.05	1.0 ± 0.1	1.3 ± 0.2	−23%
$m_{\mu\mu} > 35$	3.7	0.018 ± 0.013	0.32 ± 0.08	1.2 ± 0.2	1.5 ± 0.3	−20%

configurations. (This very trigger configuration also affects the mass spectrum within the $10\,\text{GeV} < m_{\mu\mu} < 32\,\text{GeV}$ mass window of the nominal analysis, which contributes to the difference in transfer factors in this mass bin between the two years.)

Event displays of observed signal-region events

Event displays of all observed signal-region events in 2016 and 2018 are respectively shown in Fig. B.1–B.2 and Fig. B.3–B.18 (see also Tab. 4.17 and 4.18).

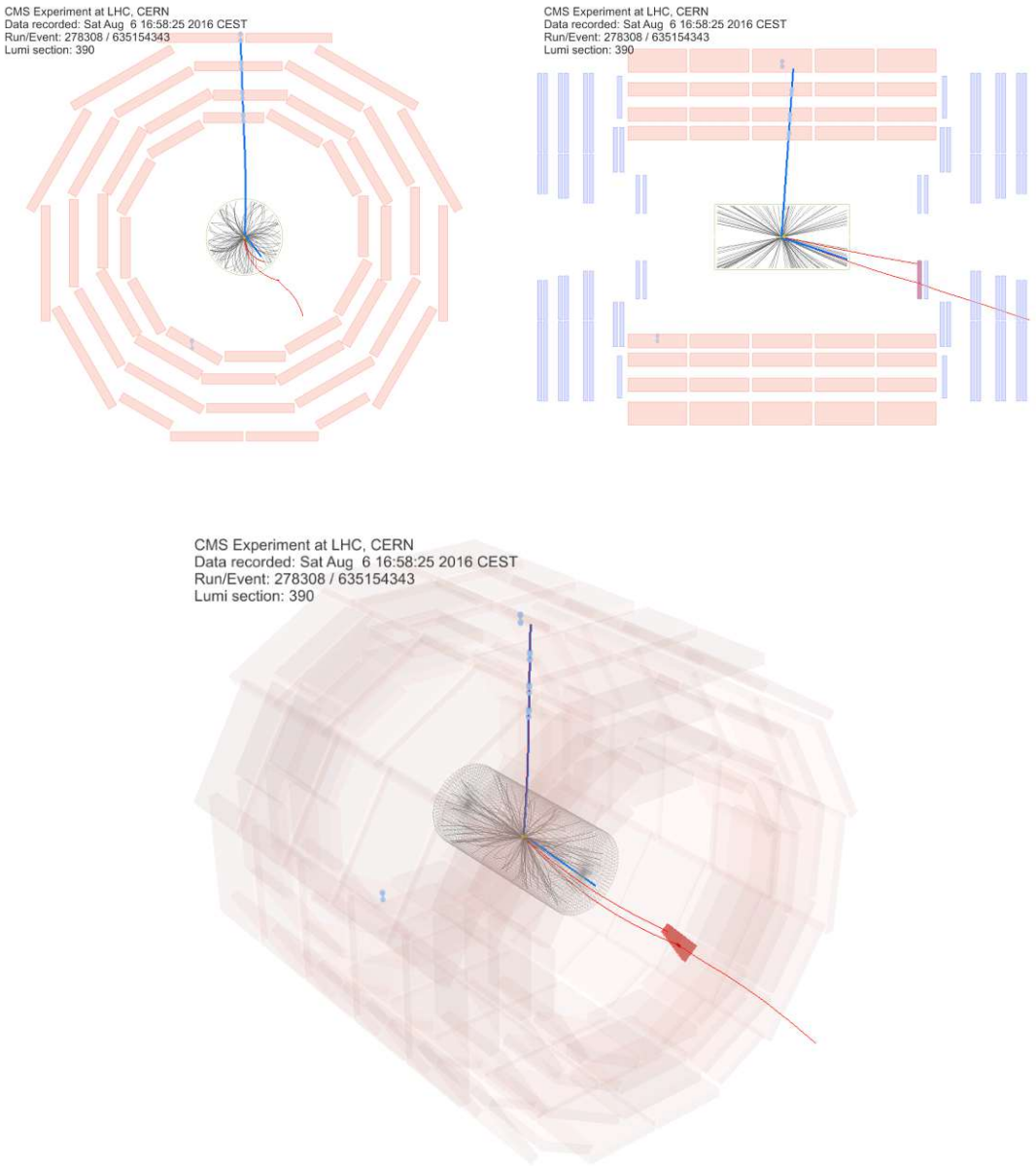


FIGURE B.1: *Display of the event 635154343 (run 278308, lumi-section 390) observed in the 2016 signal region. Blue tracks indicate DSA muons and red tracks PAT muons.* Top left: *ρ–ϕ view,* top right: *ρ–z view,* bottom: *3D view.*

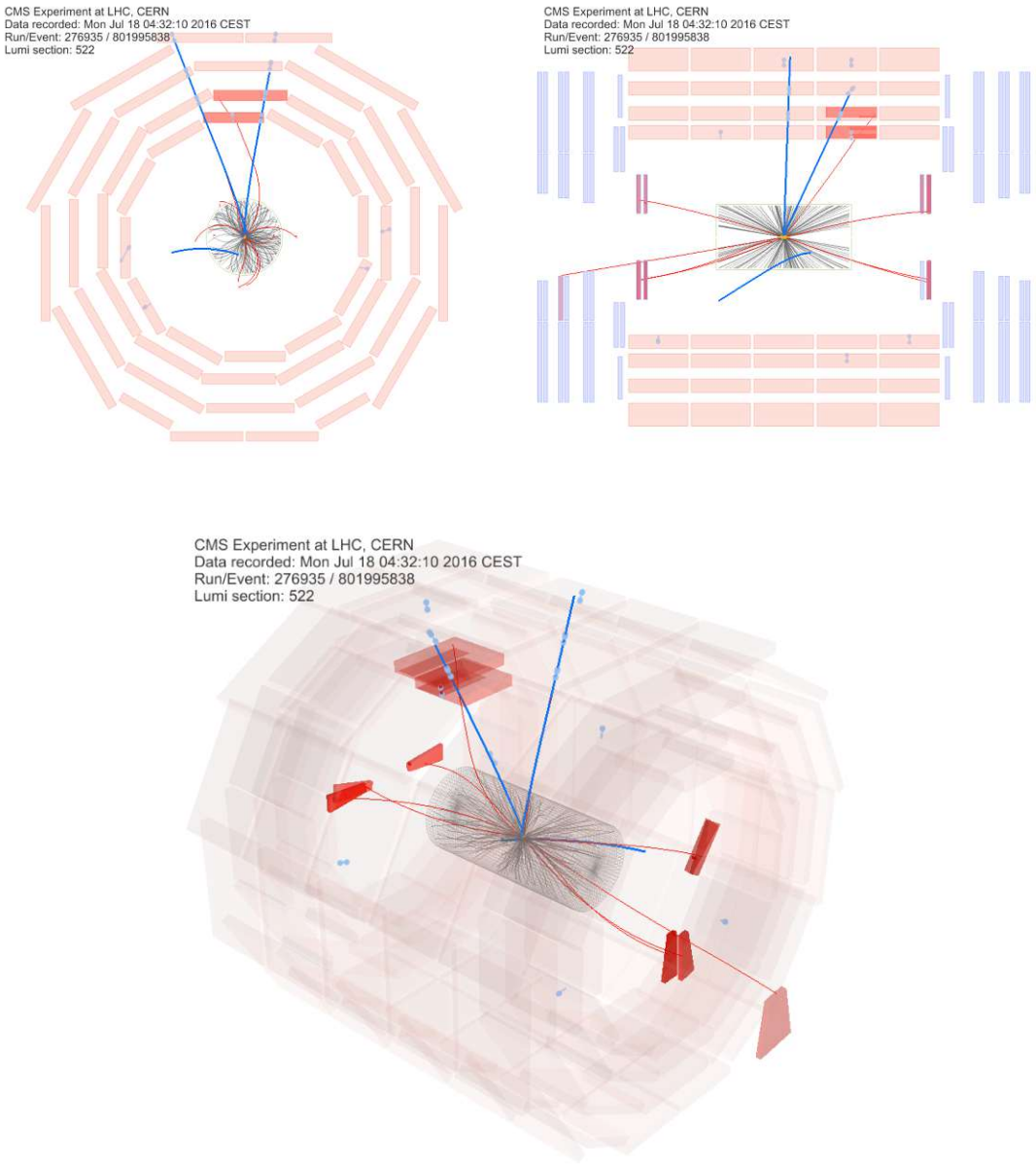


FIGURE B.2: *Graphical display of the event 635154343 (run 278308, lumi-section 390) observed in the 2016 signal region. Blue tracks indicate DSA muons and red tracks PAT muons.* Top left: *ρ–ϕ view,* top right: *ρ–z view,* bottom: *3D view.*

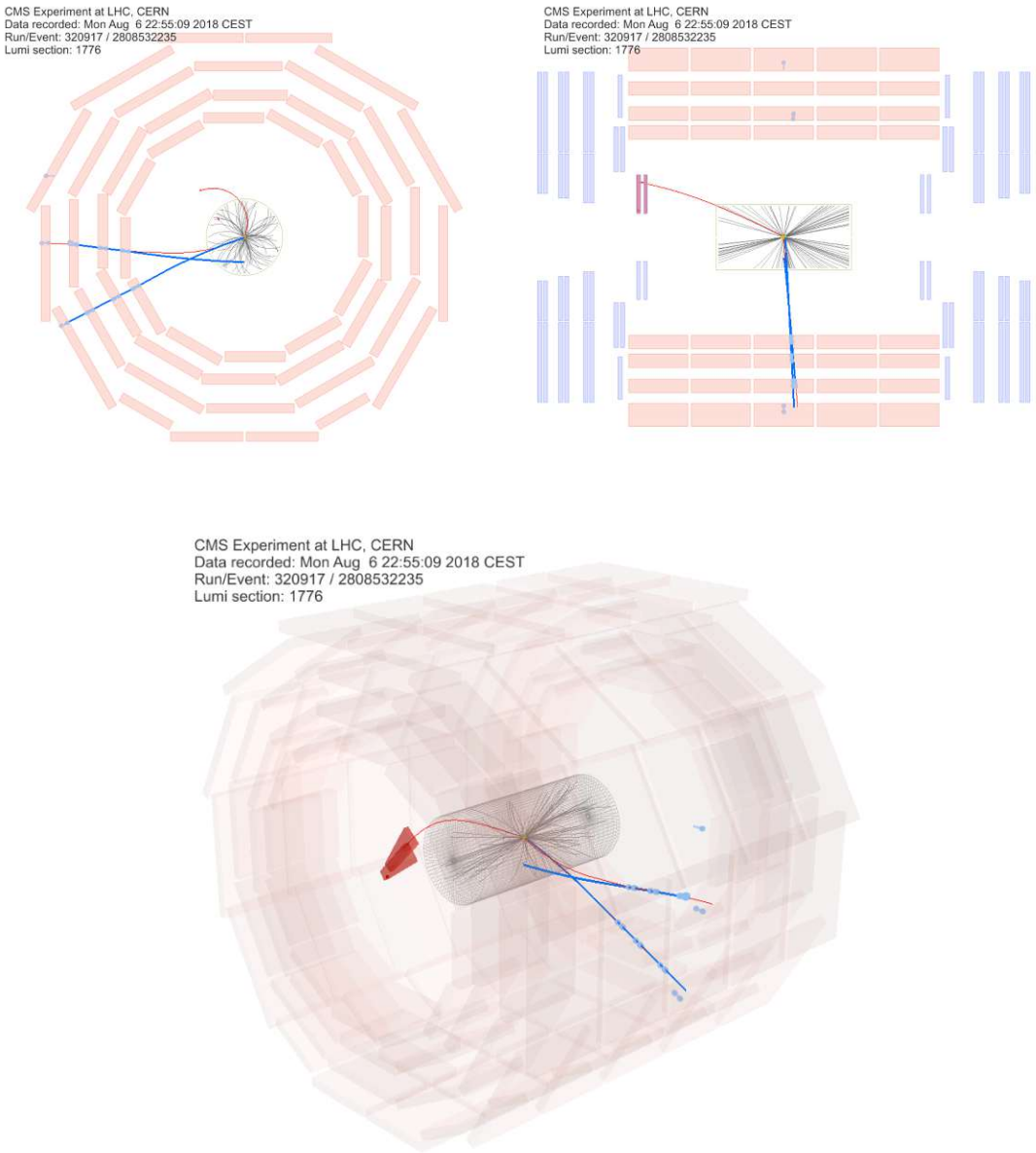


FIGURE B.3: *Graphical display of the event 2808532235 (run 320917, lumisection 1776) observed in the 2018 signal region. Blue tracks indicate DSA muons and red tracks PAT muons.* Top left: *ρ–ϕ view,* top right: *ρ–z view,* bottom: *3D view.*

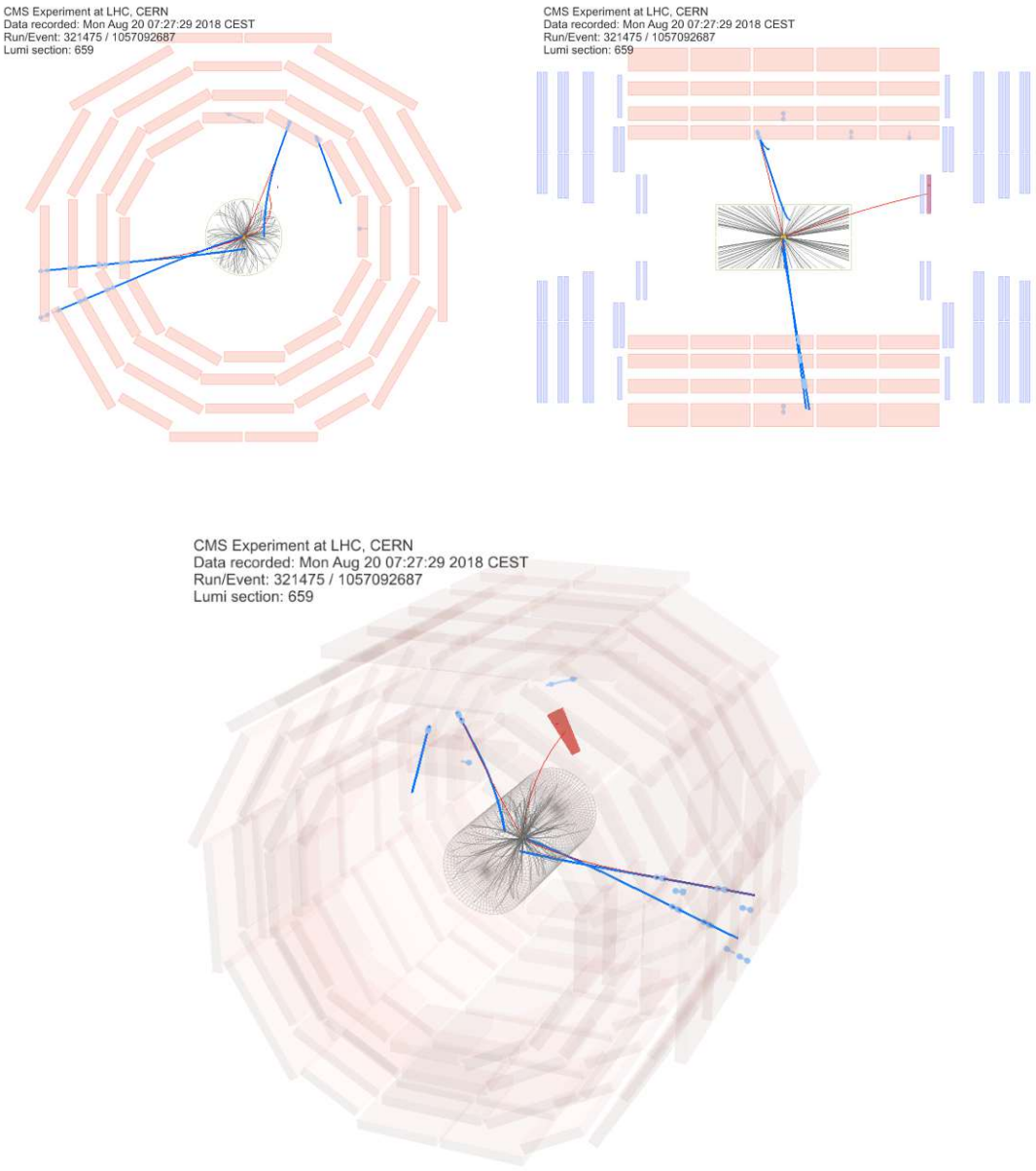


FIGURE B.4: *Graphical display of the event 1057092687 (run 321475, lumi-section 659) observed in the 2018 signal region. Blue tracks indicate DSA muons and red tracks PAT muons.* Top left: *ρ–ϕ view,* top right: *ρ–z view,* bottom: *3D view.*

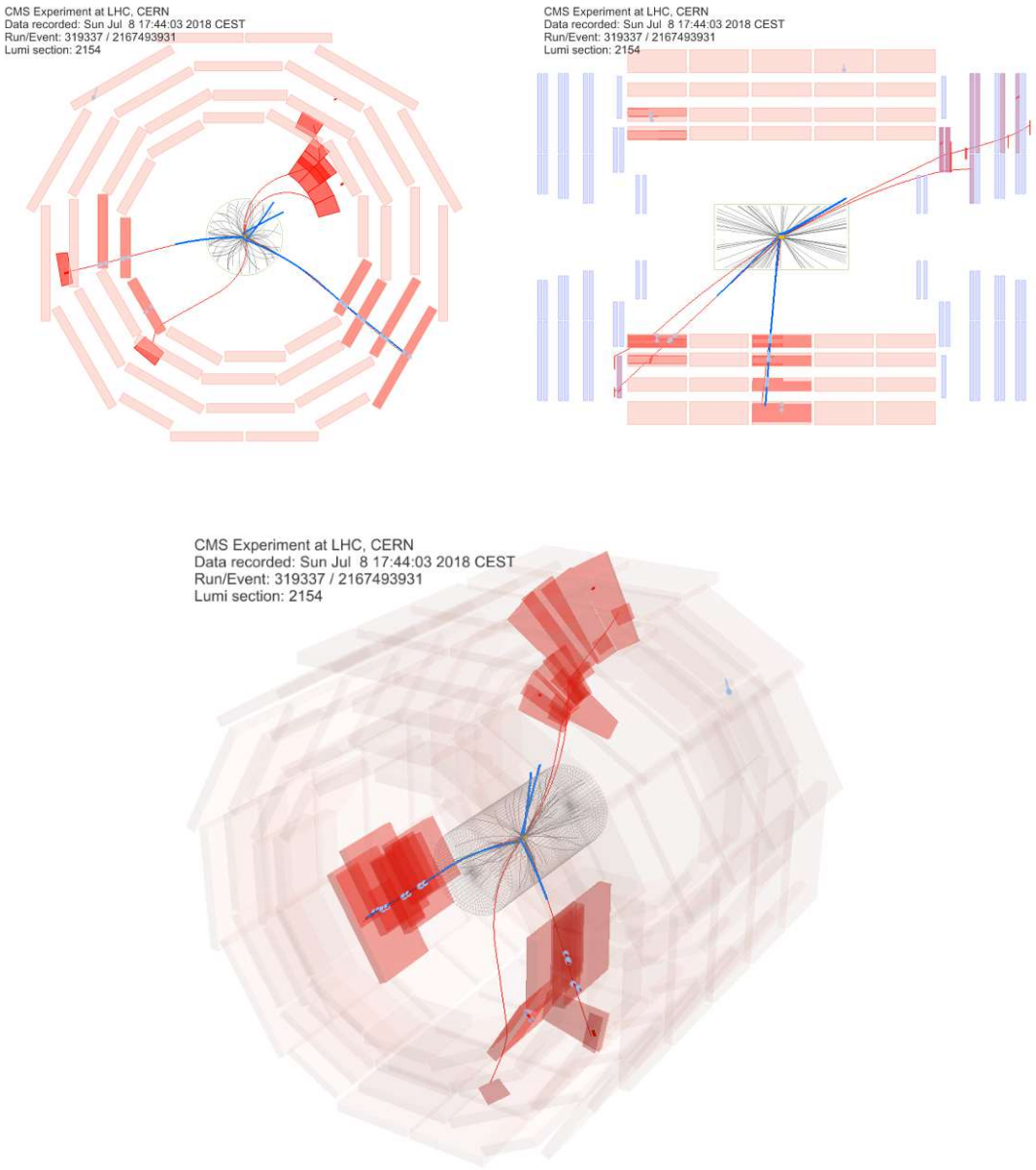


FIGURE B.5: *Graphical display of the event 2167493931 (run 319337, lumi-section 2154) observed in the 2018 signal region. Blue tracks indicate DSA muons and red tracks PAT muons.* Top left: *ρ–ϕ view,* top right: *ρ–z view,* bottom: *3D view.*

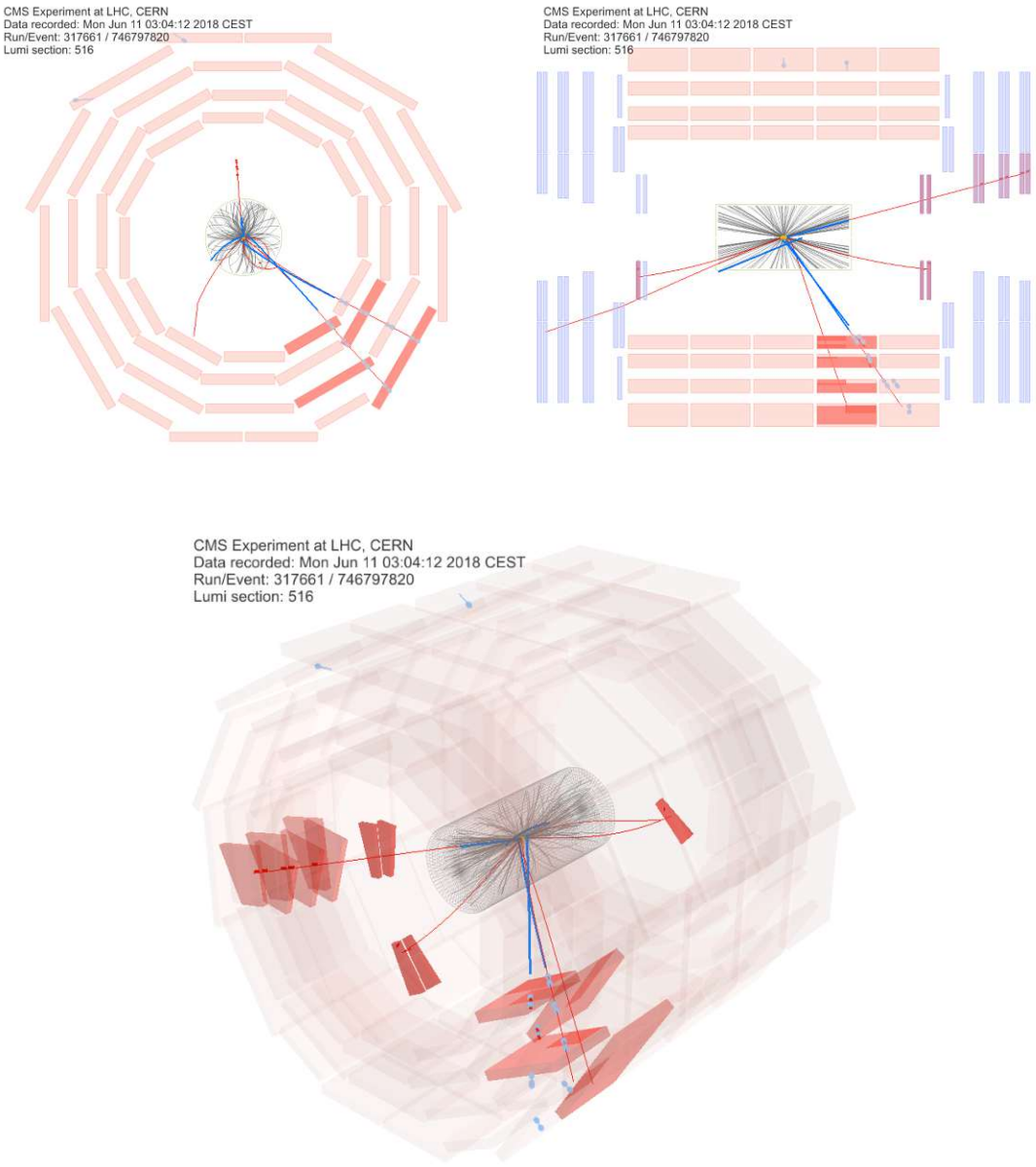


FIGURE B.6: *Graphical display of the event 746797820 (run 317661, lumi-section 516) observed in the 2018 signal region. Blue tracks indicate DSA muons and red tracks PAT muons.* Top left: *ρ–ϕ view,* top right: *ρ–z view,* bottom: *3D view.*

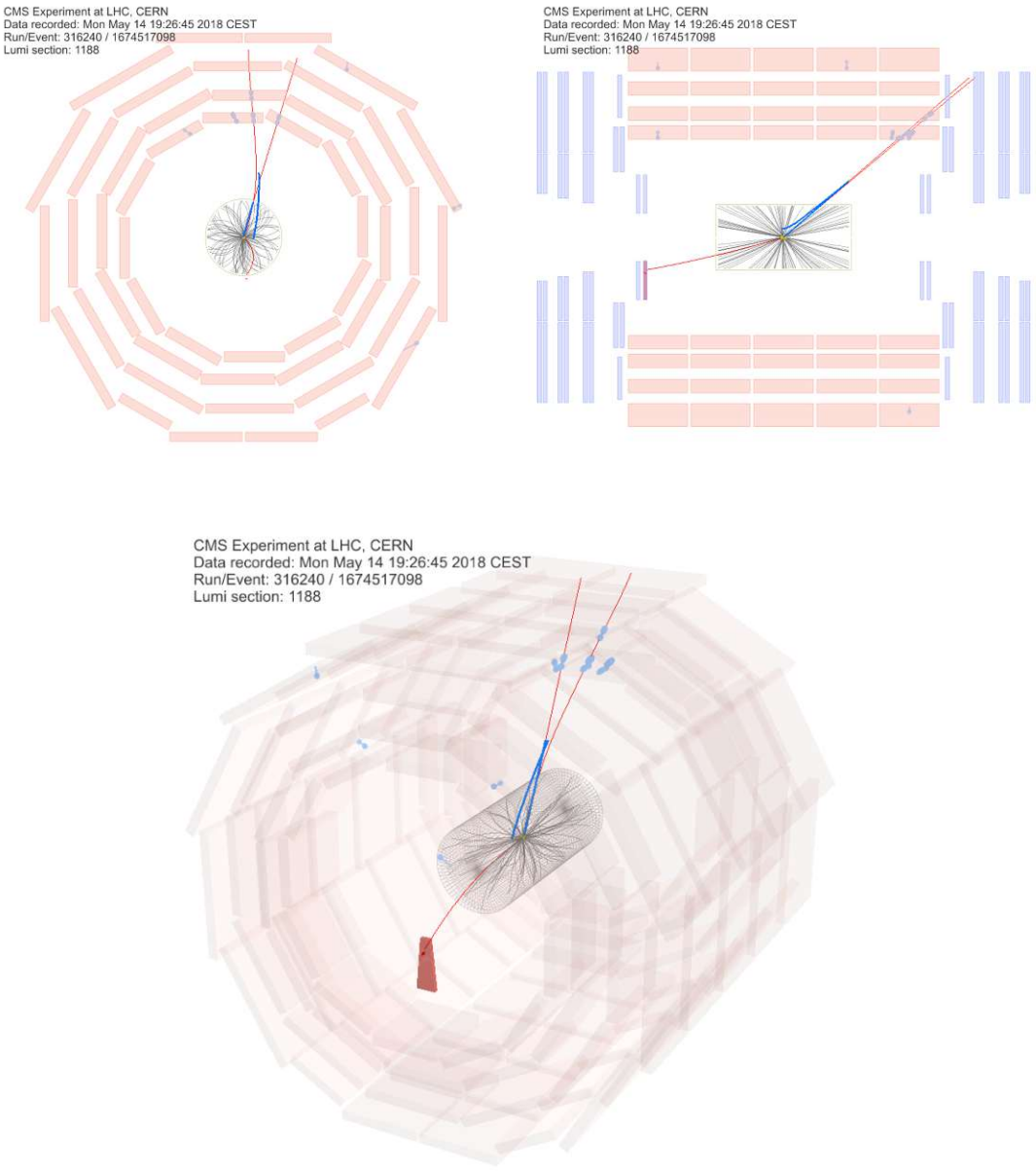


FIGURE B.7: *Graphical display of the event 1674517098 (run 316240, lumi-section 1188) observed in the 2018 signal region. Blue tracks indicate DSA muons and red tracks PAT muons.* Top left: *ρ–ϕ view,* top right: *ρ–z view,* bottom: *3D view.*

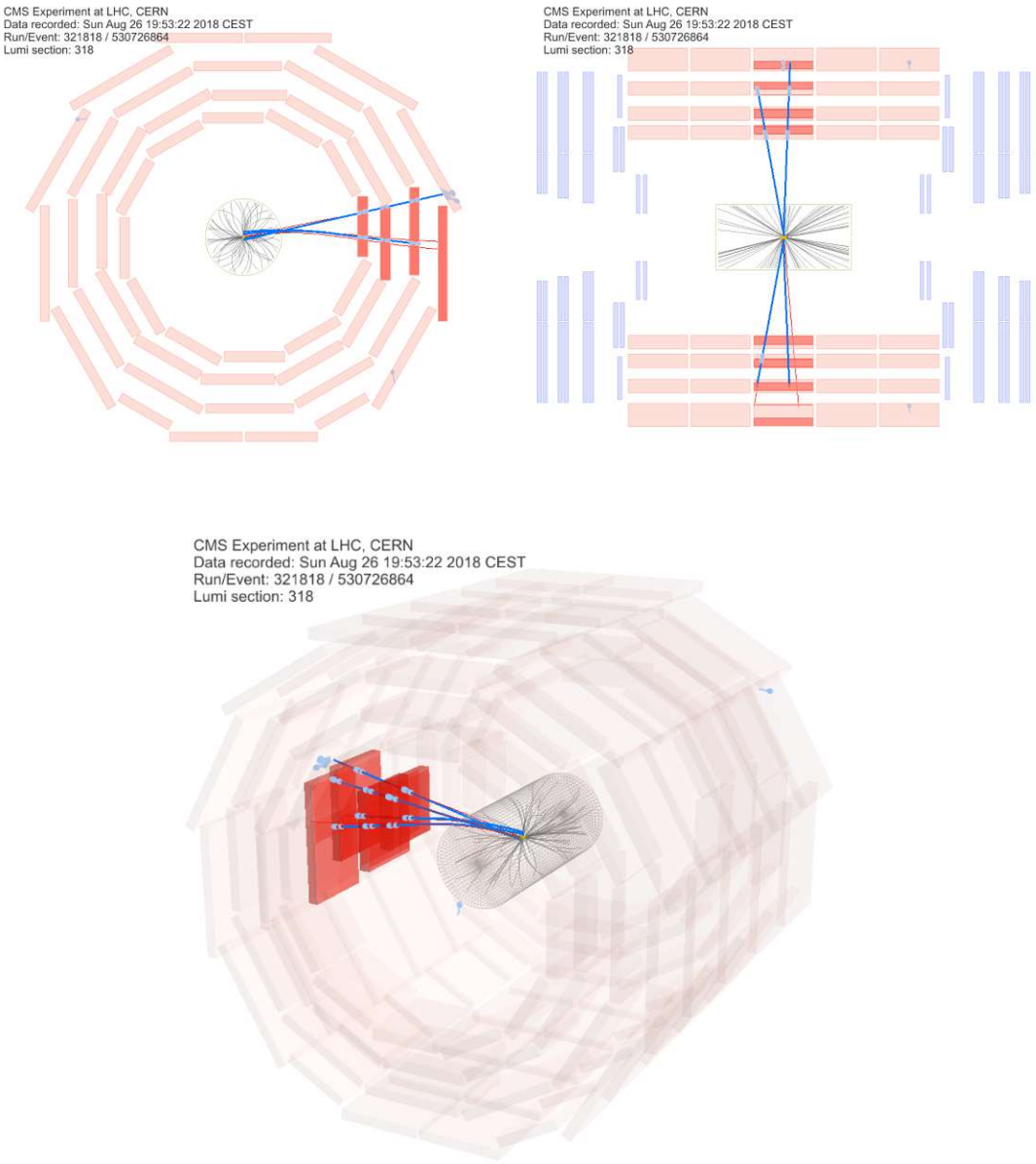


FIGURE B.8: *Graphical display of the event 530726864 (run 321818, lumi-section 318) observed in the 2018 signal region. Blue tracks indicate DSA muons and red tracks PAT muons.* Top left: *ρ–ϕ view,* top right: *ρ–z view,* bottom: *3D view.*

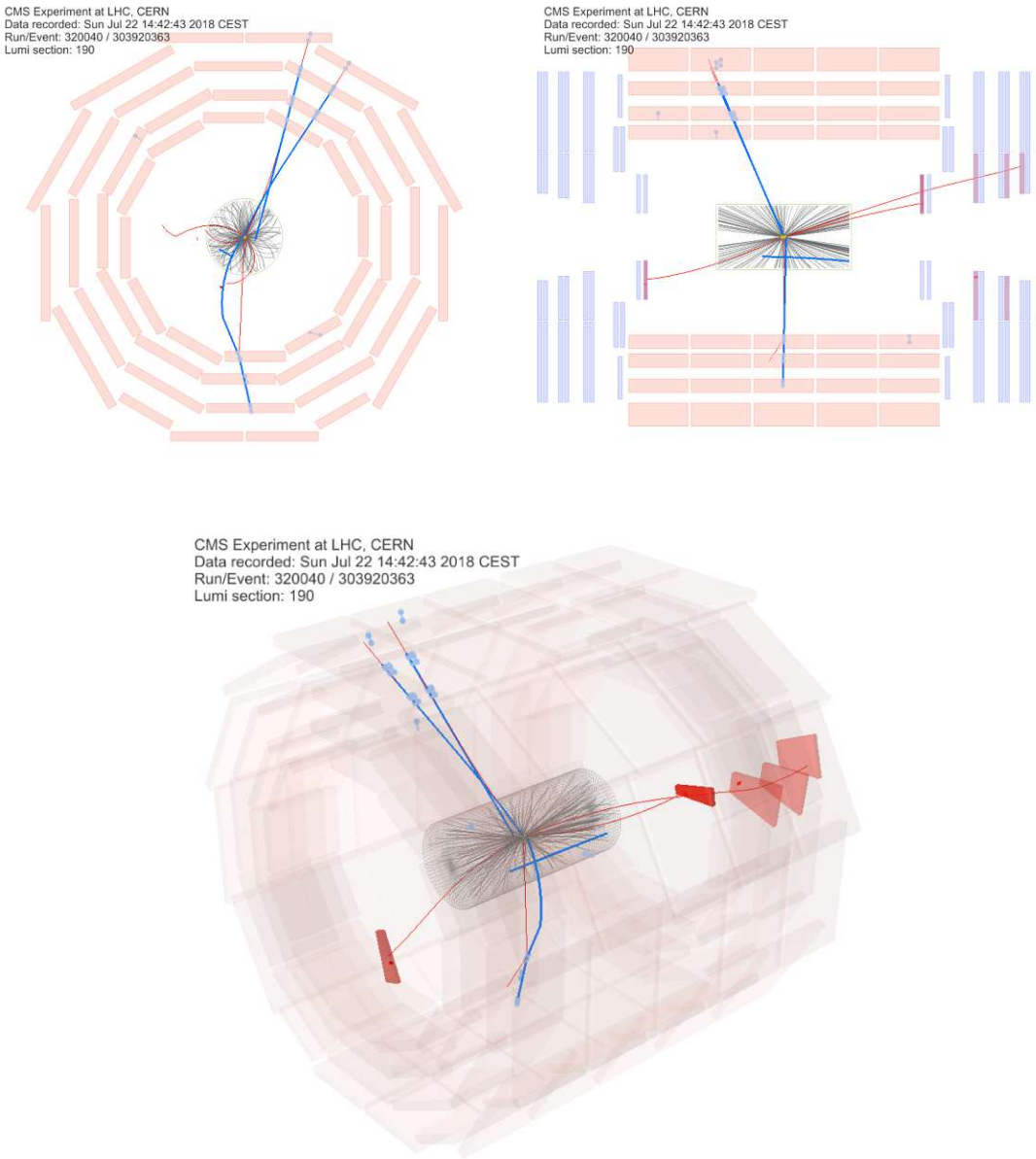


FIGURE B.9: *Graphical display of the event 303920363 (run 320040, lumi-section 190) observed in the 2018 signal region. Blue tracks indicate DSA muons and red tracks PAT muons.* Top left: *ρ–ϕ view,* top right: *ρ–z view,* bottom: *3D view.*

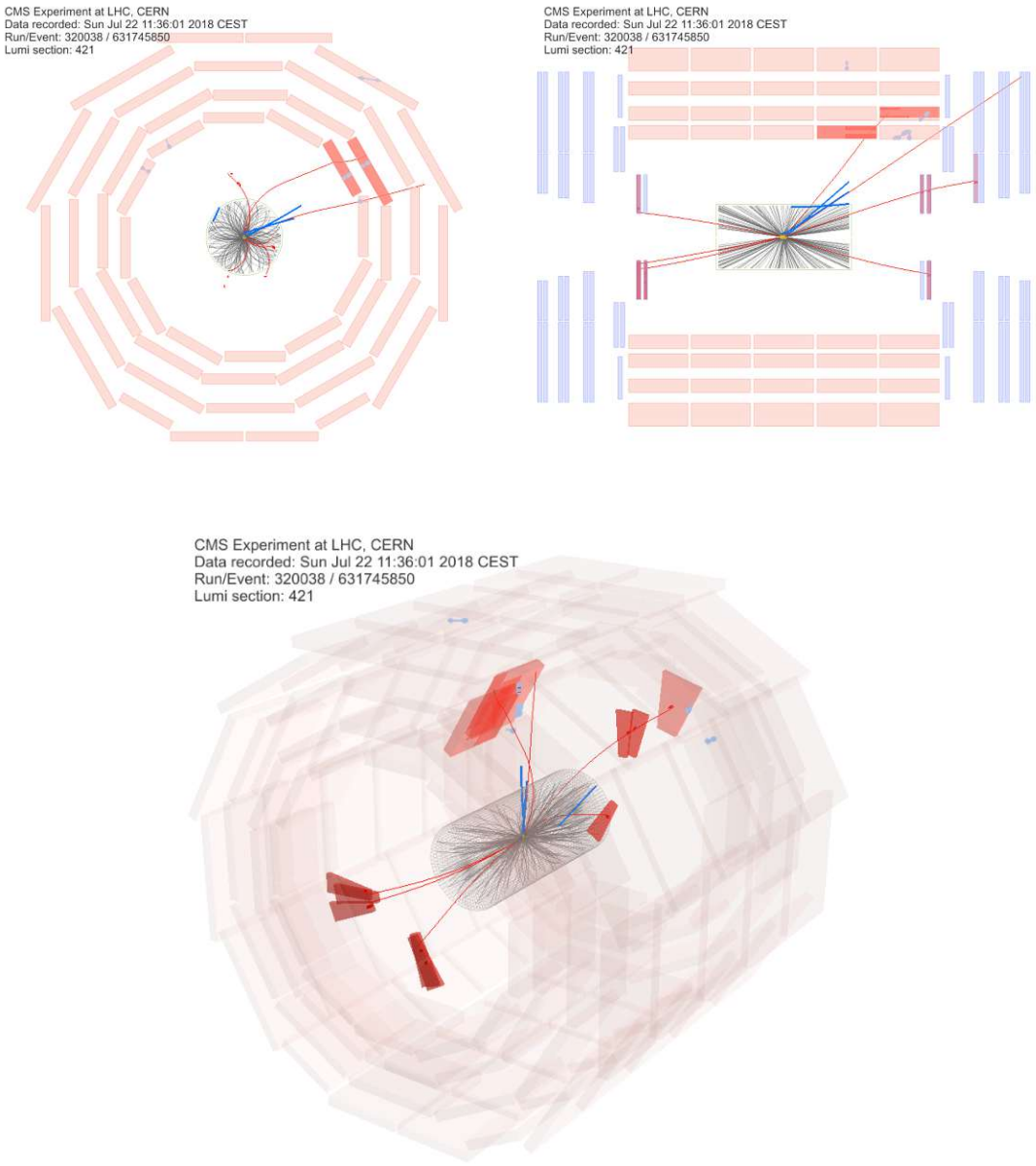


FIGURE B.10: *Graphical display of the event 631745850 (run 320038, lumi-section 421) observed in the 2018 signal region. Blue tracks indicate DSA muons and red tracks PAT muons.* Top left: *ρ–ϕ view,* top right: *ρ–z view,* bottom: *3D view.*

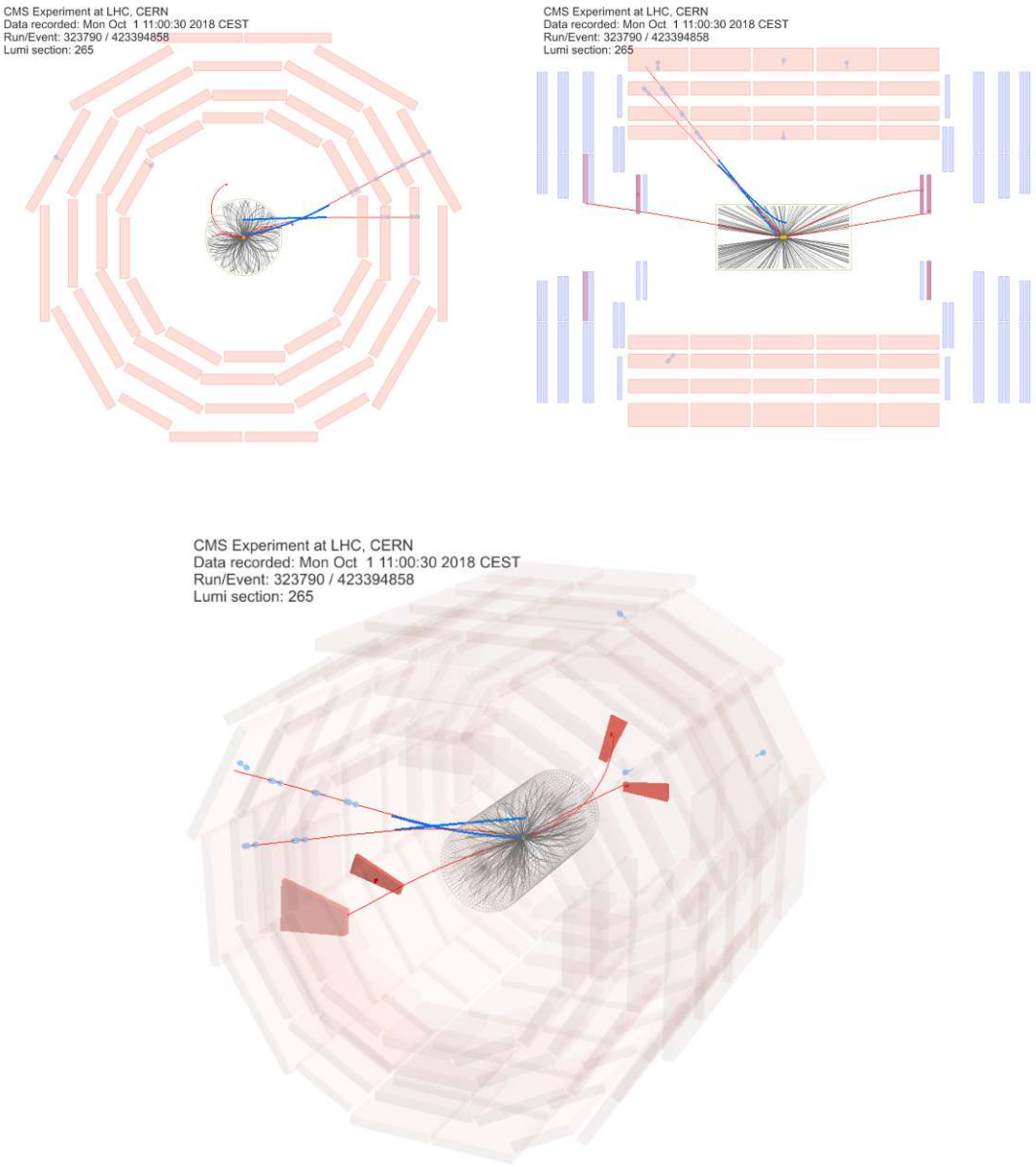


FIGURE B.11: *Graphical display of the event 423394858 (run 323790, lumi-section 265) observed in the 2018 signal region. Blue tracks indicate DSA muons and red tracks PAT muons.* Top left: *ρ–ϕ view,* top right: *ρ–z view,* bottom: *3D view.*

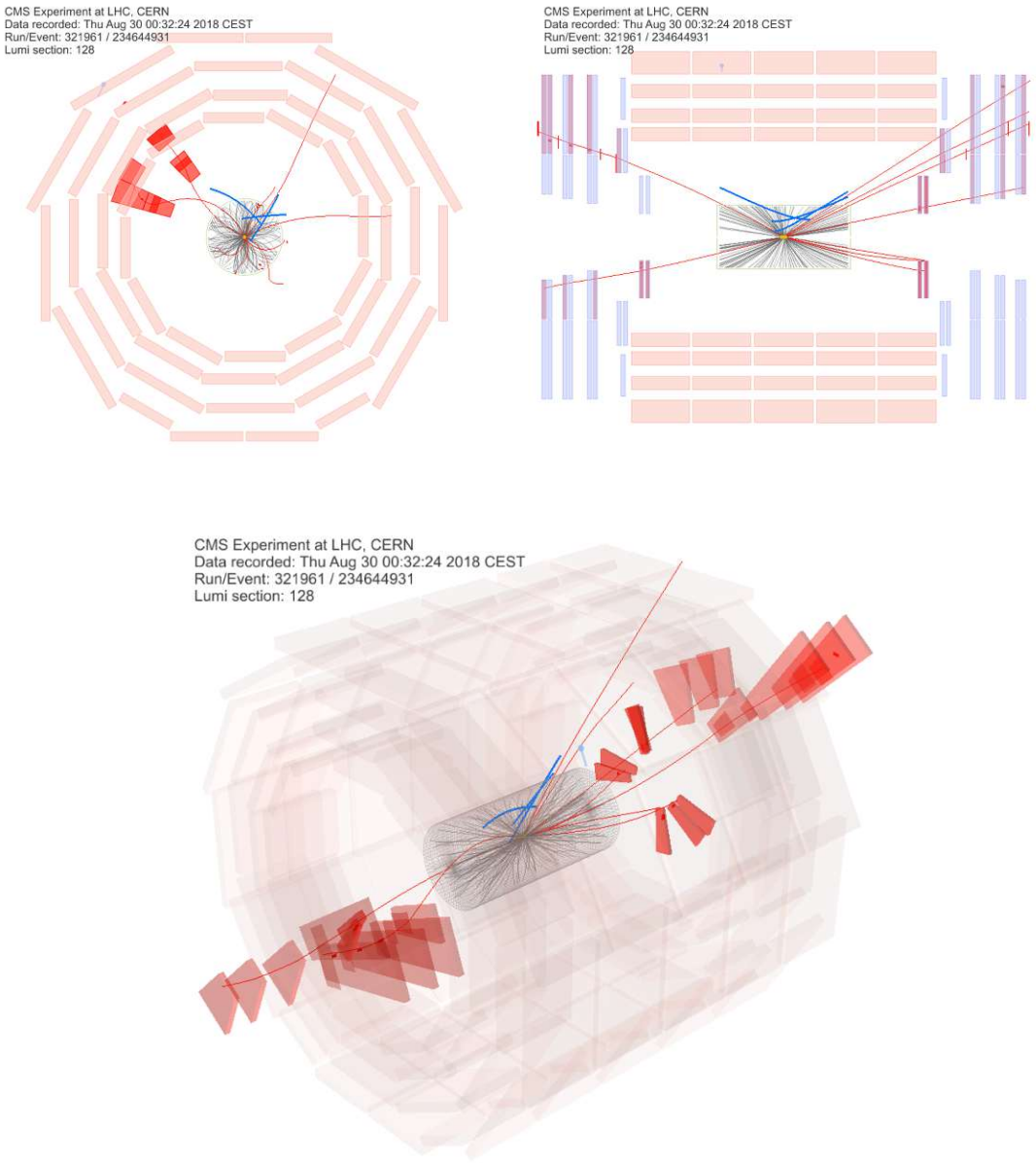


FIGURE B.12: *Graphical display of the event 234644931 (run 321961, lumi-section 128) observed in the 2018 signal region. Blue tracks indicate DSA muons and red tracks PAT muons.* Top left: *ρ–ϕ view,* top right: *ρ–z view,* bottom: *3D view.*

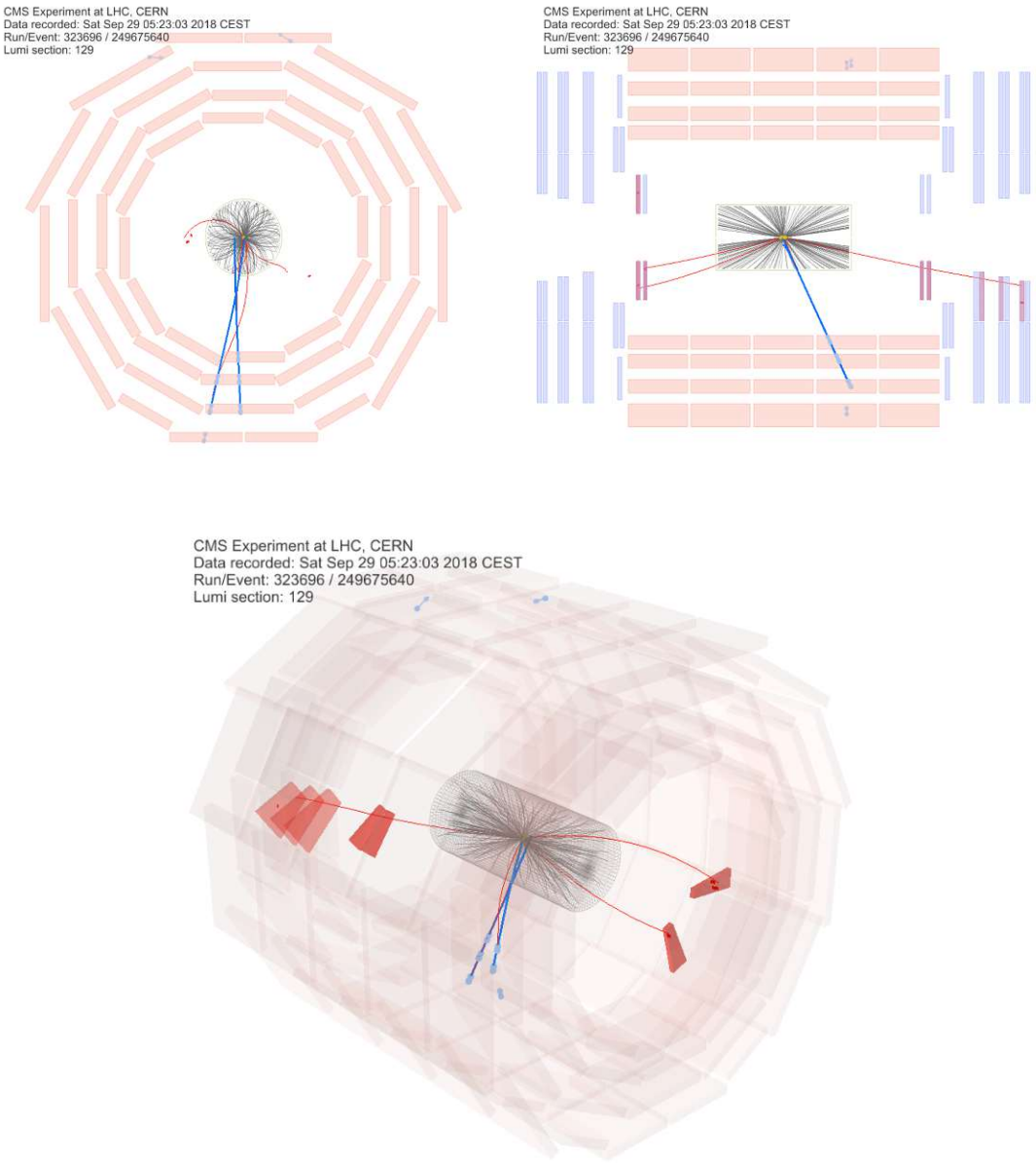


FIGURE B.13: *Graphical display of the event 24967564 (run 323695, lumi-section 129) observed in the 2018 signal region. Blue tracks indicate DSA muons and red tracks PAT muons.* Top left: *ρ–ϕ view,* top right: *ρ–z view,* bottom: *3D view.*

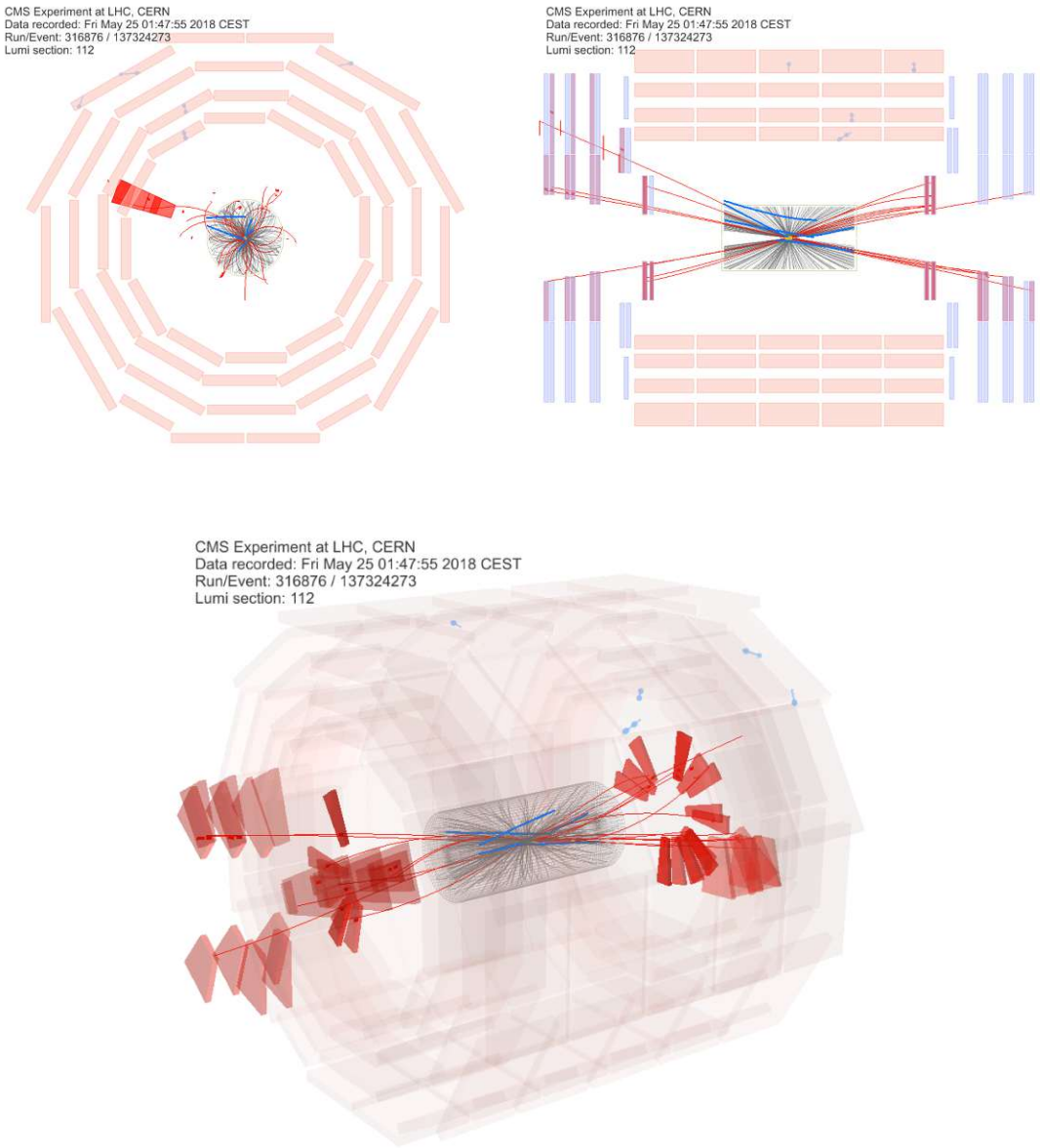


FIGURE B.14: *Graphical display of the event 137324273 (run 316876, lumi-section 112) observed in the 2018 signal region. Blue tracks indicate DSA muons and red tracks PAT muons.* Top left: *ρ–ϕ view,* top right: *ρ–z view,* bottom: *3D view.*

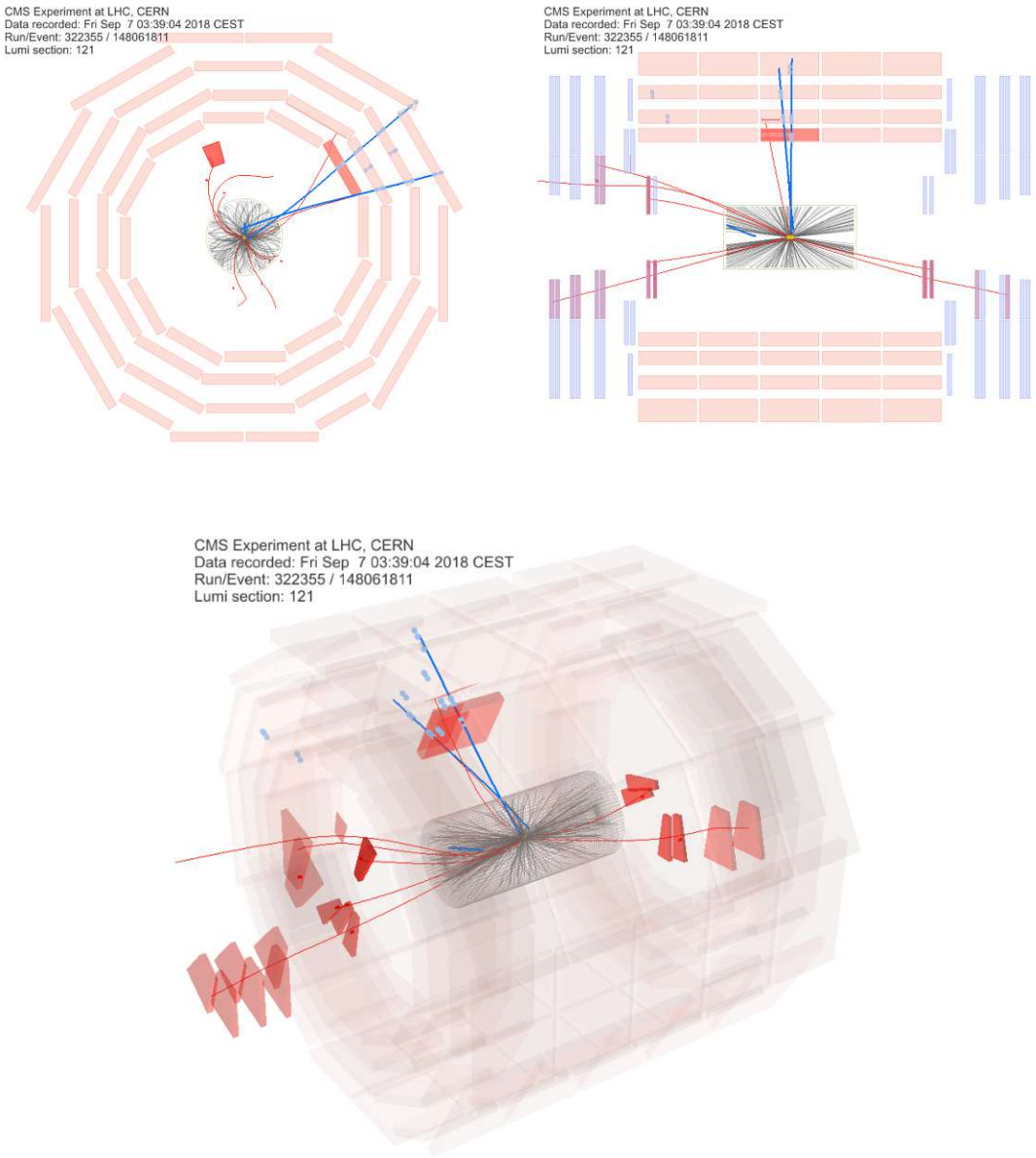


FIGURE B.15: *Graphical display of the event 148061811 (run 322355, lumi-section 121) observed in the 2018 signal region. Blue tracks indicate DSA muons and red tracks PAT muons.* Top left: *ρ–ϕ view,* top right: *ρ–z view,* bottom: *3D view.*

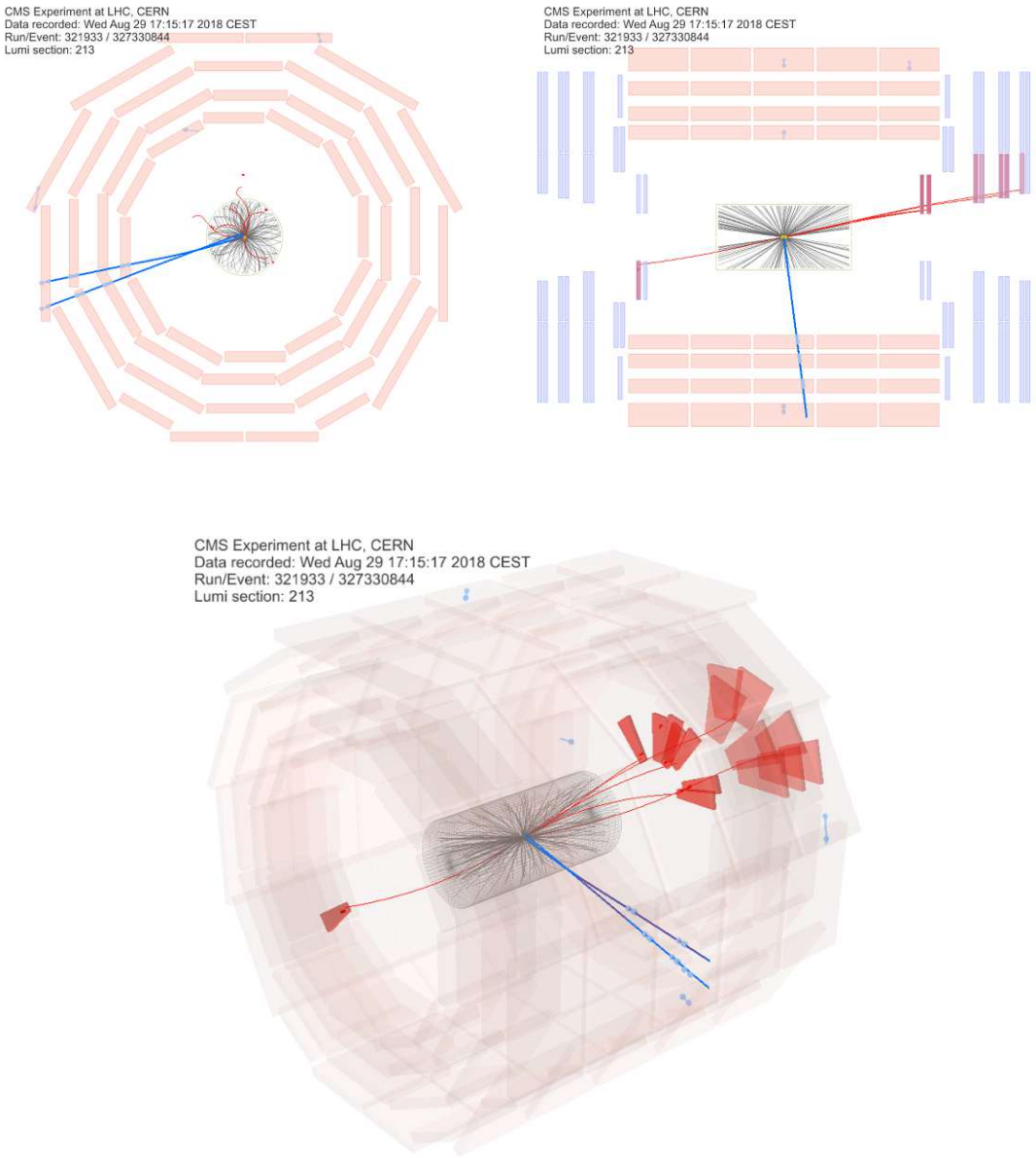


FIGURE B.16: *Graphical display of the event 327330844 (run 321933, lumi-section 213) observed in the 2018 signal region. Blue tracks indicate DSA muons and red tracks PAT muons.* Top left: *ρ–ϕ view,* top right: *ρ–z view,* bottom: *3D view.*

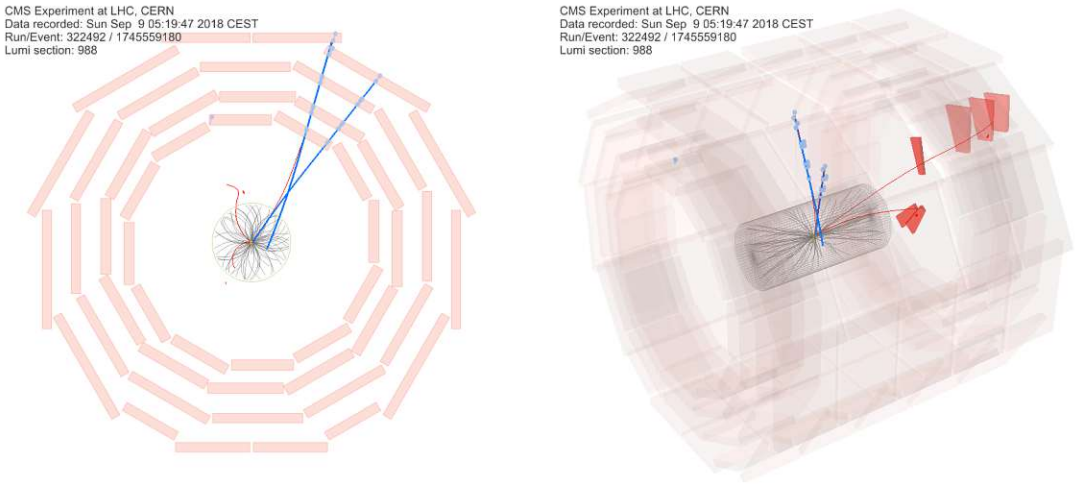


FIGURE B.17: *Graphical display of the event 1745559180 (run 322492, lumisection 988) observed in the 2018 signal region. Blue tracks indicate DSA muons and red tracks PAT muons.* Left: *ρ–ϕ view,* right: *3D view.*

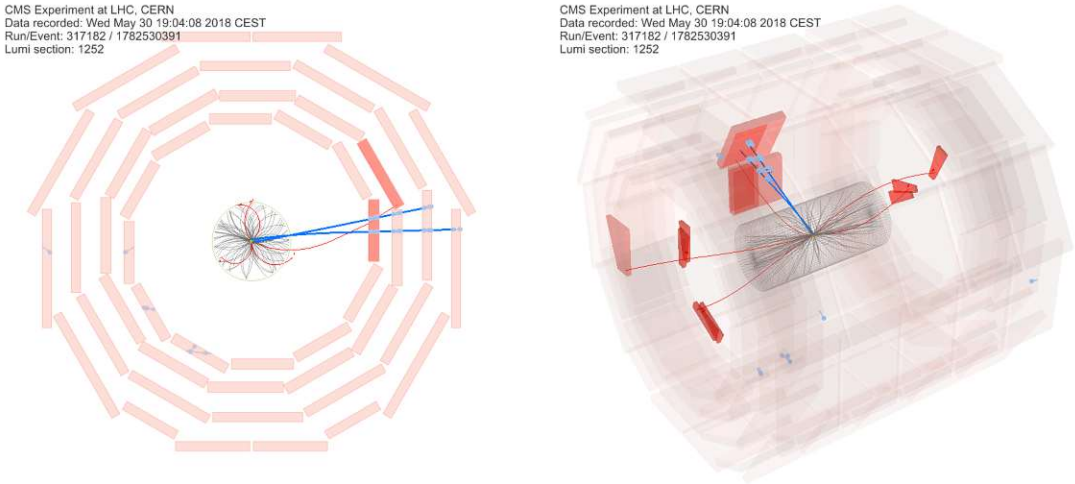


FIGURE B.18: *Graphical display of the event 1782530391 (run 317182, lumisection 1252) observed in the 2018 signal region. Blue tracks indicate DSA muons and red tracks PAT muons.* Left: *ρ–ϕ view,* right: *3D view.*

Technical configuration of the Higgs Combine tool

The following shows an example of an input *data card* for the Higgs Combine tool [116], which is used to compute the upper limits on the LLP production cross-section presented in Sec. 4.8.

Each data card defines the expected signal rate, the observed number of events, the nuisances, and the event yields in the involved control regions applying to the particular $(m_H, m_{\text{LLP}}, c\tau)$ configuration. The full background estimation is performed during the Combine call by combining the control region yields in the "ABCD" fashion defined via so-called *rate parameters* (`rateParam`). More information on data card setup and layout is documented in [116].

```
#
imax 7
jmax 2
kmax *
---------------------------------------------------------------------------------------
bin                          A       BQCD    CQCD    DQCD    BDY     CDY     DDY
observation                  15      3       372     162     0       13965   18316
---------------------------------------------------------------------------------------
bin                          A           A         BQCD    CQCD    DQCD    A       BDY     CDY     DDY
process                      2Mu         QCD       QCD     QCD     QCD     DY      DY      DY      DY
process                      0           1         1       1       1       2       2       2       2
rate                         3.661552677 1         1       1       1       1       1       1       1
---------------------------------------------------------------------------------------
lumi-2018           lnN      1.025       -         -       -       -       -       -       -       -
pileup-2018         lnN      1.020       -         -       -       -       -       -       -       -
dsa-stat-2Mu-2018   lnN      1.009       -         -       -       -       -       -       -       -
dsa-syst-QCD-2018   lnN      -           0.65/1.0  -       -       -       -       -       -       -
dsa-syst-DY-2018    lnN      -           -         -       -       -       1.150   -       -       -
dsa-syst-Vtx-2018   lnN      1.072       -         -       -       -       -       -       -       -
dsa-syst-Trg-2018   lnN      1.213       -         -       -       -       -       -       -       -
dsa-syst-ID         lnN      1.040       -         -       -       -       -       -       -       -
dsa-syst-Rec        lnN      1.040       -         -       -       -       -       -       -       -
dsa-syst-Res-2018   lnN      1.04        -         -       -       -       -       -       -       -

rate-dsa-A-QCD-2018 rateParam   A     QCD (@0*@1/@2) rate-dsa-B-QCD-2018,rate-dsa-C-QCD-2018,rate-dsa-D-QCD-2018
rate-dsa-B-QCD-2018 rateParam   BQCD  QCD 3.000000 [0.33786,10.97748]
rate-dsa-C-QCD-2018 rateParam   CQCD  QCD 372      [324.19830,424.62395]
rate-dsa-D-QCD-2018 rateParam   DQCD  QCD 162      [131.09410,197.76106]
rate-dsa-A-DY-2018  rateParam   A     DY (@0*@1/@2) rate-dsa-B-DY-2018,rate-dsa-C-DY-2018,rate-dsa-D-DY-2018
rate-dsa-B-DY-2018  rateParam   BDY   DY 0.000001 [0.00000,5.29832]
rate-dsa-C-DY-2018  rateParam   CDY   DY 13965    [13662.48348,14272.28399]
rate-dsa-D-DY-2018  rateParam   DDY   DY 18316    [17969.27446,18667.49164]
```

LISTING C.1: *Example of a Combine data card for 2018 data.*

For each configuration of m_H, m_{LLP}, and $c\tau$, the Combine tool is used as follows (with `<datacard>` being the data-card input file corresponding to the respective (m_H, m_{LLP}, $c\tau$) setting):

```
combine -d <datacard>  \
    -M HybridNew --toysH 5000 \
    -H AsymptoticLimits \
    --LHCmode LHC-limits \
    --expectedFromGrid=<quantile>
    -s -1 \
    --compile --fork 10  \
```

Here, the provided options invoke the following behavior:

- `-M HybridNew --toysH 5000`: generates 5000 toy experiments to obtain the distribution of the test statistic instead of using the asymptotic frequentist approximations (which are only valid for a large number of involved events)
- `-H AsymptotitcLimits`: estimates the upper limit on the signal strength using asymptotic frequentist methods to inform the range of signal strength over which to search when running the toy experiments
- `--LHCmode LHC-limits`: computes the distribution of the LHC-style test statistic; profiles the nuisance parameters and thus treats them in a frequentist fashion; the nuisance parameters are fixed to their post-fit values for each toy experiment to make the method sensitive to observations in data
- `--expectedFromGrid=<quntile>`: computes a certain quantile for the expected upper limits; `<quantile>` has a value of 0.5 for the median of the distribution under the background-only hypothesis, 0.16 or 0.84 for the lower or upper edges of the 68% quantile, or 0.025 or 0.975 for the lower and upper edges of the 95% quantile
- `-s -1`: obtains a random seed for the toy Monte-Carlo
- `--compile --fork 10`: technical configuration (mainly to control memory consumption and optimize the command runtime)

Eventually, the results for the upper limits computed by Combine are rescaled before plotting to account for the injected cross-section times branching ratio.

List of figures

List of tables

Acronyms

APD avalanche photodiodes

APV Analogue Pipeline (Voltage mode)

BMTF barrel muon track finder

BS beam spot

CERN European Organization for Nuclear Research

CL confidence level

CMS Compact Muon Solenoid Experiment

CSC cathode strip chamber

CV common vertex

DAQ data acquisition

DCA distance of closest approach

DM Dark Matter

DSA muon displaced standalone muon

DT drift tube

DY Drell-Yan

EB ECAL Barrel

ECAL Electromagnetic Calorimeter

EE ECAL Endcap

EMTF endcap muon track finder

EOM equations of motion

ES ECAL Preshower Detector

FPGA field-programmable gate array

FWHM full width at half maximum

HAHM Hidden Abelian Higgs Model

HB HCAL Barrel

HCAL Hadron Calorimeter

HE HCAL Endcap

HF HCAL Forward

HLT High-Level Trigger

HO HCAL Outer

IP interaction point

IP5 LHC Interaction Point 5

L1 Level-1 Trigger

L1 muon L1 muon

L2 muon L2 muon

L3 muon L3 muon

LHC Large Hadron Collider

LLP long-lived particle

MC Monte Carlo

MET missing transverse energy

OMTF overlap muon track finder

PAT Physics Analysis Tools

PAT muon PAT muon

PF Particle Flow

PS Proton Synchrotron

PS prescale

PSB Proton Synchrotron Booster

PU pile-up

PV primary vertex

QCD Quantum Chromodynamics

RPC resistive plate chamber

RSA muon refitted standalone muon

SA muon standalone muon

SF scale factor

SL superlayer

SM Standard Model

SPS Super Proton Synchrotron

TEC Tracker Endcaps

TIB Tracker Inner Barrel

TID Tracker Inner Disk

TOB Tracker Outer Barrel

TP trigger primitive

uGMT global muon trigger

VPT vacuum phototriodes

VR validation region

VR-IMASS low-mass validation region

VR-IQDETA validation region with inverted quality requirements for small-$|\Delta\eta|$ dimuons

Bibliography

[1] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579, 1961. doi: 10.1016/0029-5582(61)90469-2.

[2] S. Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264, 1967. doi: 10.1103/PhysRevLett.19.1264.

[3] A. Salam. Weak and Electromagnetic Interactions. *Conf. Proc. C*, 680519:367, 1968. doi: 10.1142/9789812795915_0034.

[4] O. S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock. LHC design report vol. 1: The LHC main ring. Technical report, CERN, June 2004. URL `https://cds.cern.ch/record/782076`.

[5] O. Buning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock. LHC Design Report Vol. 2: The LHC infrastructure and general services. *CERN*, November 2004. doi: 10.5170/CERN-2004-003-V-2.

[6] M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl. LHC Design Report Vol. 3: The LHC injector chain. *CERN*, December 2004. doi: 10.5170/CERN-2004-003-V-3.

[7] L. Evans and P. Bryant. LHC machine. *J. Instrum.*, 003:S08001, 2008. doi: 10.1088/1748-0221/3/08/s08001.

[8] P. W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508, 1964. doi: 10.1103/PhysRevLett.13.508.

[9] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321, 1964. doi: 10.1103/PhysRevLett.13.321.

[10] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1, 2012. doi: 10.1016/j.physletb.2012.08.020.

[11] CMS Collaboration. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30, 2012. doi: 10.1016/j.physletb.2012.08.021.

[12] CMS Collaboration. Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 11:185, 2018. doi: 10.1007/JHEP11(2018)185.

[13] CMS Collaboration. Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s}$=13 TeV. *JHEP*, 11:047, 2017. doi: 10.1007/JHEP11(2017)047.

[14] CMS Collaboration. Combined measurements of Higgs boson couplings in protonproton collisions at $\sqrt{s}$=13 TeV. *Eur. Phys. J. C*, 79(5):421, 2019. doi: 10.1140/epjc/s10052-019-6909-y.

[15] CMS Collaboration. Observation of Higgs boson decay to bottom quarks. *Phys. Rev. Lett.*, 121(12):121801, 2018. doi: 10.1103/PhysRevLett.121.121801.

[16] CMS Collaboration. Observation of $t\bar{t}h$ production. *Phys. Rev. Lett.*, 120(23):231801, 2018. doi: 10.1103/PhysRevLett.120.231801.

[17] CMS Collaboration. Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector. *Phys. Lett. B*, 779:283–316, 2018. doi: 10.1016/j.physletb.2018.02.004.

[18] Jakob Schwichtenberg. *Physics from Symmetry*. Springer, 2015.

[19] M. Strassler. The Hierarchy Problem. Personal blog. URL `https://profmattstrassler.com/articles-and-posts/particle-physics-basics/the-hierarchy-problem/`. Last accessed: Jul 5$^{\text{th}}$, 2021.

[20] K. A. Tomilin. Natural Systems of Units. To the Centenary Anniversary of the Planck System. In *Proceedings Of The XXII Workshop On High Energy Physics And Field Theory*, 1999.

[21] S. P. Martin. A Supersymmetry primer. *Adv. Ser. Direct. High Energy Phys.*, 18:1–98, 1998. doi: 10.1142/9789812839657_0001.

[22] A. Zee. *Quantum field theory in a nutshell*, volume 7. Princeton University Press, 2010.

[23] M. Drees, R. Godbole, and P. Roy. Theory and phenomenology of sparticles: An account of four-dimensional N=1 supersymmetry in high energy physics. *Hackensack, USA: World Scientific (2004) 555 p*, 2004.

[24] J. Alimena, J. Beacham, M. Borsato, Y. Cheng, X. C. Vidal, G. Cottin, D. Curtin, A. De Roeck, N. Desai, J. A. Evans, et al. Searching for long-lived particles beyond the Standard Model at the Large Hadron Collider. *J. Phys. G: Nucl. Part. Phys.*, 47(9):090501, 2020. URL `https://iopscience.iop.org/article/10.1088/1361-6471/ab4574/meta`.

[25] Planck Collaboration. Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.*, 594:A13, 2016. doi: 10.1051/0004-6361/201525830.

[26] D. Clowe, A. Gonzalez, and M. Markevitch. Weak lensing mass reconstruction of the interacting cluster 1E0657-558: Direct evidence for the existence of dark matter. *Astrophys. J.*, 604:596–603, 2004. doi: 10.1086/381970.

[27] M. Markevitch, A. H. Gonzalez, D. Clowe, A. Vikhlinin, L. David, W. Forman, C. Jones, S. Murray, and W. Tucker. Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1E0657-56. *Astrophys. J.*, 606:819–824, 2004. doi: 10.1086/383178.

[28] M. Bradac, S. W. Allen, T. Treu, H. Ebeling, R. Massey, R. G. Morris, A. von der Linden, and D. Applegate. Revealing the properties of dark matter in the merging cluster MACSJ0025.4-1222. *Astrophys. J.*, 687:959, 2008. doi: 10.1086/591246.

[29] B. Holdom. Two U(1)'s and Epsilon Charge Shifts. *Phys. Lett.*, 166B:196–198, 1986. doi: 10.1016/0370-2693(86)91377-8.

[30] M. Pospelov, A. Ritz, and M. B. Voloshin. Secluded WIMP Dark Matter. *Phys. Lett.*, B662:53–61, 2008. doi: 10.1016/j.physletb.2008.02.052.

[31] LHCb Collaboration. Search for $A' \to \mu^+\mu^-$ decays. *LHCB-PAPER-2019-031*, 2019. URL `https://cds.cern.ch/record/2693834`.

[32] CMS Collaboration. Search for a narrow resonance lighter than 200 GeV decaying to a pair of muons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 124:131802, December 2019. doi: 10.1103/PhysRevLett.124.131802. URL `https://cds.cern.ch/record/2703964`.

[33] D. Curtin, R. Essig, S. Gori, and J. Shelton. Illuminating dark photons with high-energy colliders. *J. High Energy Phys.*, 2015(2):157, 2015.

[34] James D. Wells. How to Find a Hidden World at the Large Hadron Collider. March 2008. URL `https://arxiv.org/abs/0803.1243`.

[35] J. M. Strassler and K. M. Zurek. Discovering the Higgs through highly-displaced vertices. *Phys. Lett. B*, 661(4):263–267, 2008. ISSN 0370-2693. doi: https://doi.org/10.1016/j.physletb.2008.02.008. URL `https://www.sciencedirect.com/science/article/pii/S0370269308001846`.

[36] 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, September 2019. URL `https://cds.cern.ch/record/2690341`.

[37] P. A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020. doi: 10.1093/ptep/ptaa104.

[38] E. Noether. Invariante Variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1918:235–257, 1918. URL `http://eudml.org/doc/59024`.

[39] FT2. Spontaneous symmetry breaking. Wikimedia Commons (CC BY-SA 3.0). URL `https://commons.wikimedia.org/wiki/File:Spontaneous_symmetry_breaking_(explanatory_diagram).png`. Accessed: 6.4.2021.

[40] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental Test of Parity Conservation in Beta Decay. *Phys. Rev.*, 105:1413–1415, February 1957. doi: 10.1103/PhysRev.105.1413. URL `https://link.aps.org/doi/10.1103/PhysRev.105.1413`.

[41] M. Gell-Mann. Symmetries of Baryons and Mesons. *Phys. Rev.*, 125:1067, 1962. doi: 10.1103/PhysRev.125.1067.

[42] H. Fritzsch, M. Gell-Mann, and H. Leutwyler. Advantages of the color octet gluon picture. *Physics Letters B*, 47(4):365, 1973. doi: https://doi.org/10.1016/0370-2693(73)90625-4.

[43] D. J. Gross and F. Wilczek. Asymptotically free gauge theories. II. *Phys. Rev. D*, 9:980, 1974. doi: 10.1103/PhysRevD.9.980.

[44] S. Weinberg. Non-Abelian Gauge Theories of the Strong Interactions. *Phys. Rev. Lett.*, 31:494, 1973. doi: 10.1103/PhysRevLett.31.494.

[45] H. D. Politzer. Reliable Perturbative Results for Strong Interactions? *Phys. Rev. Lett.*, 30:1346, 1973. doi: 10.1103/PhysRevLett.30.1346.

[46] D. J. Gross and F. Wilczek. Ultraviolet Behavior of Non-Abelian Gauge Theories. *Phys. Rev. Lett.*, 30:1343, 1973. doi: 10.1103/PhysRevLett.30.1343.

[47] V. M. Abazov et al. Evidence for a $B_s^0\,\pi^\pm$ state. *Phys. Rev. Lett.*, 117(2):022003, 2016. doi: 10.1103/PhysRevLett.117.022003.

[48] R. Aaij et al. Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\Psi K^- p$ Decays. *Phys. Rev. Lett.*, 115:072001, 2015. doi: 10.1103/PhysRevLett.115.072001.

[49] B. Shuve. Theory Overview of Long-Lived Particles at the LHC, April 2017. URL `https://indico.cern.ch/event/607314/contributions/2542308/attachments/1447888/2231430/LHC-LLP_Shuve.pdf`.

[50] M. J. Ramsey-Musolf. Long Lived Particles at the LHC, October 2017. URL `https://indico.cern.ch/event/649760/contributions/2689321/attachments/1542431/2419468/MJRM_LLP_Trieste_17.pdf`.

[51] H. Russell. An experimental introduction to long-lived searches at the LHC, April 2017. URL `https://indico.cern.ch/event/607314/contributions/2542309/attachments/1447873/2231444/20170424_LLPs.pdf`.

[52] Z. Chacko, H.-S. Goh, and R. Harnik. The Twin Higgs: Natural electroweak breaking from mirror symmetry. *Phys. Rev. Lett.*, 96:231802, 2006. doi: 10.1103/PhysRevLett.96.231802.

[53] G. Burdman, Z. Chacko, H.-S. Goh, and R. Harnik. Folded supersymmetry and the LEP paradox. *JHEP*, 02:009, 2007. doi: 10.1088/1126-6708/2007/02/009.

[54] N. Craig and K. Howe. Doubling down on naturalness with a supersymmetric twin Higgs. *JHEP*, 03:140, 2014. doi: 10.1007/JHEP03(2014)140.

[55] N. Craig, S. Knapen, and P. Longhi. Neutral Naturalness from Orbifold Higgs Models. *Phys. Rev. Lett.*, 114(6):061803, 2015. doi: 10.1103/PhysRevLett.114.061803.

[56] G. Burdman, Z. Chacko, R. Harnik, L. de Lima, and C. B. Verhaaren. Colorless Top Partners, a 125 GeV Higgs, and the Limits on Naturalness. *Phys. Rev. D*, 91(5):055007, 2015. doi: 10.1103/PhysRevD.91.055007.

[57] V. Silveira and A. Zee. Scalar phantoms. *Phys. Lett. B*, 161(1-3):136–140, 1985.

[58] M. Pospelov, A. Ritz, and M. B. Voloshin. Secluded WIMP Dark Matter. *Phys. Lett. B*, 662:53–61, 2008. doi: 10.1016/j.physletb.2008.02.052.

[59] D. Feldman, B. Kors, and P. Nath. Extra-weakly Interacting Dark Matter. *Phys. Rev. D*, 75:023503, 2007. doi: 10.1103/PhysRevD.75.023503.

[60] D. E. Morrissey and M. J. Ramsey-Musolf. Electroweak baryogenesis. *New J. Phys.*, 14:125003, 2012. doi: 10.1088/1367-2630/14/12/125003.

[61] M. J. Strassler and K. M. Zurek. Echoes of a hidden valley at hadron colliders. *Phys. Lett. B*, 651:374–379, 2007. doi: 10.1016/j.physletb.2007.06.055.

[62] LHCb Collaboration. Search for A' to decays. *Phys. Rev. Lett.*, 124:041801, Oct 2019. doi: 10.1103/PhysRevLett.124.041801. URL `https://cds.cern.ch/record/2693834`.

[63] D. Curtin, R. Essig, S. Gori, P. Jaiswal, A. Katz, T. Liu, Z. Liu, D. McKeen, J. Shelton, M. Strassler, et al. Exotic decays of the 125 GeV Higgs boson. *Phys. Rev. D*, 90 (7):075004, 2014.

[64] ATLAS Collaboration. 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. D*, 99(1):012001, 2019.

[65] CMS Collaboration. 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(5):052012, 2015.

[66] 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. URL `https://cds.cern.ch/record/2005761`.

[67] CMS Collaboration. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008. doi: 10.1088/1748-0221/3/08/S08004. URL `https://doi.org/10.1088/1748-0221/3/08/S08004`.

[68] CMS Collaboration. CMS Physics: Technical Design Report Volume 1: Detector Performance and Software. *CERN*, 2006. URL `https://cds.cern.ch/record/922757`.

[69] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08002, 2008. doi: 10.1088/1748-0221/3/08/S08003. URL `https://cds.cern.ch/record/1129811`.

[70] ALICE Collaboration. The ALICE experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08002–S08002, 2008.

[71] LHCb Collaboration. The LHCb detector at the LHC. *JINST*, 3(08):S08005, 2008.

[72] Elias Métral. LHC: status, prospects and future challenges. *PoS*, LHCP2016:002, 2016. doi: 10.22323/1.276.0002. URL `https://pos.sissa.it/276/002`.

[73] E. Mobs. The CERN accelerator complex - 2019. Complexe des acclrateurs du CERN - 2019. July 2019. URL https://cds.cern.ch/record/2684277.

[74] CMS Collaboration. Public CMS luminosity information. *Public CMS twiki*, 2021. URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. Last accessed: Jun 22$^{\text{th}}$, 2021.

[75] T. Sakuma and T. McCauley. Detector and Event Visualization with SketchUp at the CMS Experiment. *J. Phys. Conf. Ser.*, 513:022032, 2014. doi: 10.1088/1742-6596/513/2/022032.

[76] CMS Collaboration. *The CMS magnet project: Technical Design Report*. Technical design report. CMS. CERN-LHCC-97-010, Geneva, 1997. URL https://cds.cern.ch/record/331056.

[77] CMS Collaboration. Status of the construction of the CMS magnet. *IEEE Trans. Appl. Supercond.*, 14(2):542–547, 2004. doi: 10.1109/TASC.2004.829715.

[78] CMS Collaboration. The CMS detector magnet. *IEEE Transactions on Applied Superconductivity*, 10(1):389–394, 2000. doi: 10.1109/77.828255.

[79] CMS Collaboration. *The CMS tracker system project: Technical Design Report*. Technical design report. CMS. CERN-LHCC-98-006, Geneva, 1997. URL https://cds.cern.ch/record/368412.

[80] CMS Collaboration. *The CMS tracker: addendum to the Technical Design Report*. Technical design report. CMS. CERN-LHCC-2000-016, Geneva, 2000. URL https://cds.cern.ch/record/490194.

[81] CMS Collaboration. CMS Technical Design Report for the Pixel Detector Upgrade. *CERN*, September 2012. doi: 10.2172/1151650. URL https://www.osti.gov/biblio/1151650.

[82] CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014. URL https://iopscience.iop.org/article/10.1088/.

[83] CMS Collaboration. CMS Tracking POG Performance Plots for year 2016, 2016. URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackingPOGPlots2016.

[84] CMS Collaboration. Primary vertex resolution in 2016. *CERN-CMS-DP-2016-041*, July 2016. URL https://cds.cern.ch/record/2202968.

[85] CMS Collaboration. The CMS electromagnetic calorimeter project: Technical design report. *Technical Design Report CMS. CERN, Geneva*, 47, 1997. URL https://cds.cern.ch/record/349375.

[86] CMS Collaboration. The CMS hadron calorimeter project: Technical Design Report. 1997. URL https://cds.cern.ch/record/357153.

[87] CMS Collaboration. *The CMS muon project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997. URL https://cds.cern.ch/record/343814.

[88] CMS collaboration. The CMS trigger system. *JINST*, 12(01):P01020, 2017. doi: 10.1088/1748-0221/12/01/P01020.

[89] A. Tapper and D. Acosta. CMS Technical Design Report for the Level-1 Trigger Upgrade. Technical Report CERN-LHCC-2013-011. CMS-TDR-12, June 2013. URL `https://cds.cern.ch/record/1556311`.

[90] CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. *Journal of Instrumentation*, 12(10):P10003–P10003, October 2017. ISSN 1748-0221. doi: 10.1088/1748-0221/12/10/p10003. URL `http://dx.doi.org/10.1088/1748-0221/12/10/P10003`.

[91] CMS Collaboration. Performance of the CMS muon trigger system in proton-proton collisions at $\sqrt{s}$ = 13 TeV. Technical Report arXiv:2102.04790. CMS-MUO-19-001-003, CERN, Geneva, February 2021. URL `https://cds.cern.ch/record/2751483`. Submitted to JINST. All figures and tables can be found at http://cms-results.web.cern.ch/cms-results/public-results/publications/MUO-19-001 (CMS Public Pages).

[92] CMS Collaboration. Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s}$= 13 TeV. *Journal of Instrumentation*, 15, 2020.

[93] Frühwirth, R. 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(2-3):444–450, 1987.

[94] CMS Collaboration. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$ = 13 TeV. *JINST*, 13(06):P06015, 2018. doi: 10.1088/1748-0221/13/06/P06015.

[95] CMS Collaboration. Performance of CMS Muon Reconstruction in pp Collision Events at $\sqrt{s}$ =7 TeV. *JINST*, 7:P10002, 2012. doi: 10.1088/1748-0221/7/10/P10002.

[96] CMS Collaboration. Performance of CMS muon reconstruction in cosmic-ray events. *JINST*, 5(03):T03022, 2010.

[97] CMS Collaboration. Search for long-lived particles that decay into final states containing two muons, reconstructed using only the CMS muon chambers. Technical Report CMS-PAS-EXO-14-012, CERN, Geneva, 2015. URL `https://cds.cern.ch/record/2005761`.

[98] CMS Collaboration. Muon Reconstruction and Identification Improvements for Run-2 and First Results with 2015 Run Data. *CERN*, July 2015. URL `https://cds.cern.ch/record/2037372`.

[99] Giovanni Abbiendi. The CMS muon system in Run2: preparation, status and first results. *The European Physical Society Conference on High Energy Physics*, July 2015. URL `https://arxiv.org/abs/1510.05424`.

[100] J. Alwall et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014. doi: 10.1007/JHEP07(2014)079.

[101] J. Butterworth et al. PDF4LHC recommendations for LHC Run II. *J Phys G*, 43: 023001, 2016. doi: 10.1088/0954-3899/43/2/023001.

[102] T. Sjöstrand et al. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191: 159–177, 2015. doi: 10.1016/j.cpc.2015.01.024.

[103] CMS Collaboration. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur Phys J C*, 76(3):155, 2016. doi: 10.1140/epjc/s10052-016-3988-x.

[104] CMS Collaboration. Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements. *Eur. Phys. J. C*, 80(1):4, 2020. doi: 10.1140/epjc/s10052-019-7499-4.

[105] GEANT4 Collaboration. GEANT4: A Simulation Toolkit. *Nucl. Instrum. Meth. A*, 506(3):250–303, 2003.

[106] CMS Collaboration. *CMS computing: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 2005. URL `https://cds.cern.ch/record/838359`.

[107] M. Bahr et al. Herwig++ Physics and Manual. *Eur. Phys. J. C*, 58:639–707, 2008. doi: 10.1140/epjc/s10052-008-0798-9.

[108] CMS Collaboration. The Performance of the CMS Muon Detector in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV at the LHC. *JINST*, 8:P11002, 2013. doi: 10.1088/1748-0221/8/11/P11002.

[109] CMS Collaboration. Measurement of the Charge Ratio of Atmospheric Muons with the CMS Detector. *Phys. Lett. B*, 692:83–104, 2010. doi: 10.1016/j.physletb.2010.07.033.

[110] T. Speer, K. Prokofiev, R. Frühwirth, W. Waltenberger, and P. Vanlaer. Vertex Fitting in the CMS Tracker. Technical report, CERN, Geneva, February 2006. URL `https://cds.cern.ch/record/927395`.

[111] R. D. Cousins, J. T. 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. *Nucl. Instrum. Meth. A*, 595(2):480–501, 2008. doi: 10.1016/j.nima.2008.07.086.

[112] CMS Collaboration. Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS. April 2021.

[113] CMS Collaboration. CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV. 2019.

[114] A. L. Read. Presentation of search results: The CL_s technique. *J Phys G*, 28:2693–2704, 2002. doi: 10.1088/0954-3899/28/10/313.

[115] T. Junk. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth. A*, 434:435–443, 1999. doi: 10.1016/S0168-9002(99)00498-2.

[116] CMS Collaboration. Documentation of the Higgs Combine Tool. URL http://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/part4/usefullinks.html#citations.

[117] The ATLAS and CMS Collaborations and The LHC Higgs Combination Group. Procedure for the LHC Higgs boson search combination in Summer 2011. Technical report, CERN, August 2011. URL https://cds.cern.ch/record/1379837.

[118] CMS Collaboration. 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: 520–551, 2019. doi: 10.1016/j.physletb.2019.04.025.