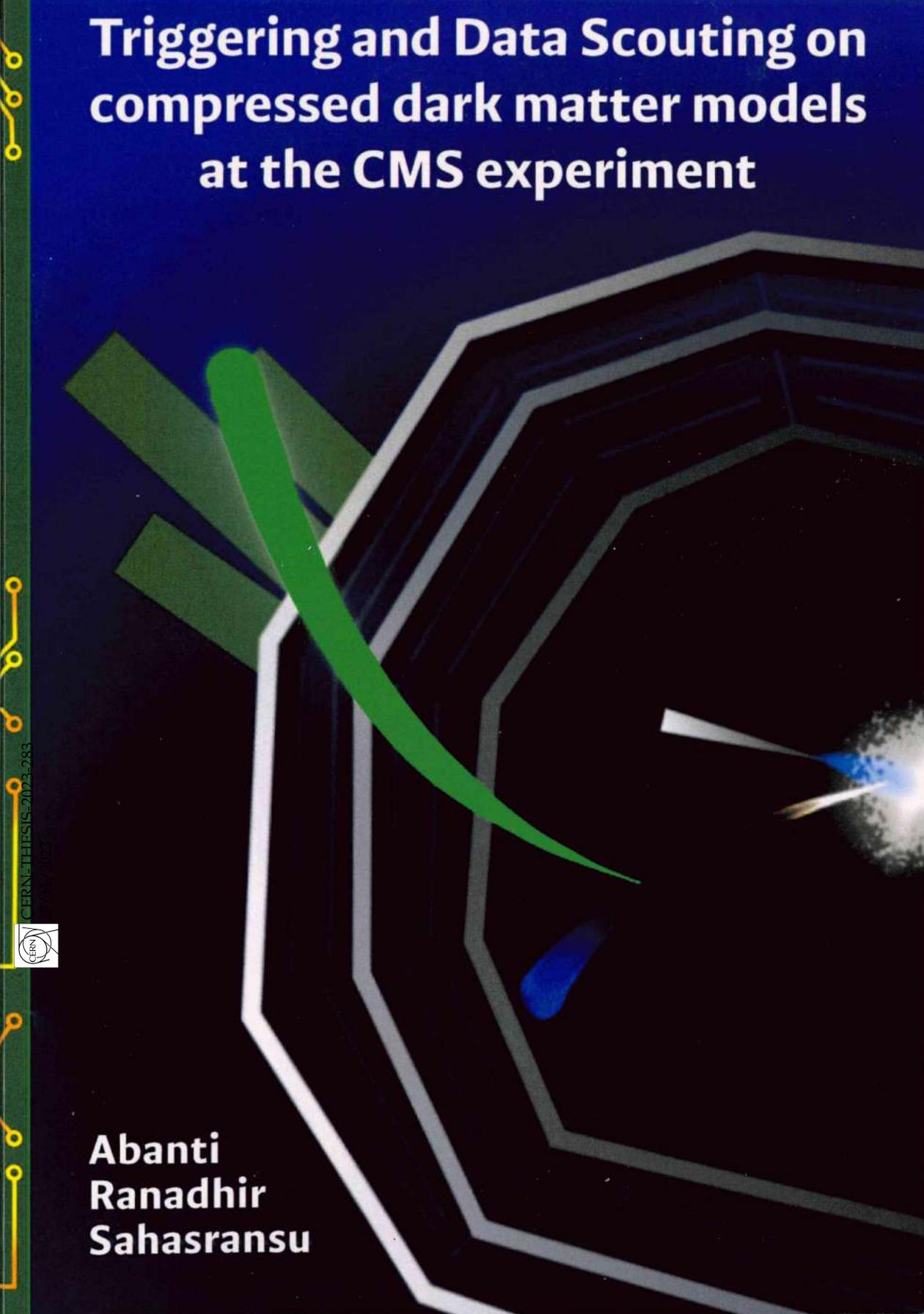


# Triggering and Data Scouting on compressed dark matter models at the CMS experiment

CERN-THESIS-2023-283



**Abanti  
Ranadhir  
Sahasransu**



Thesis submitted in fulfilment of the requirements for  
the award of the degree of Doctor in Sciences

```

1000101010100010101
0010010010010100110
1010010010101001010
1000101010101101011
1010101101101010101
1010010101101010101
0111001010101001110
    
```



9 789464 443875



VRIJE  
UNIVERSITEIT  
BRUSSEL



Thesis submitted in fulfilment of the requirements for the award of the degree of  
Doctor of Science

# TRIGGERING AND DATA SCOUTING ON COMPRESSED DARK MATTER MODELS AT THE CMS EXPERIMENT.

**Abanti Ranadhir Sahasransu**

27 November 2023

Promotor: Prof. Dr. Steven Lowette (IIHE, VUB)  
Co-promotor: Prof. Dr. Freya Blekman (DESY, University of Hamburg)  
Jury: Prof. Dr. Alberto Mariotti (VUB), chairman  
Prof. Sara Strandberg (Stockholm University),  
Prof. Laurent Thomas (ULB),  
Prof. Steven Goderls (VUB, Geochemistry)

**Faculty of Sciences and Bioengineering Sciences  
Department of Physics and Astrophysics**

Alle rechten voorbehouden. Niets van deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook, zonder voorafgaande schriftelijke toestemming van de auteur.

All rights reserved. No part of this publication may be produced in any form by print, photoprint, microfilm, electronic or any other means without permission from the author.

Printed by  
Crazy Copy Center Productions  
VUB Pleinlaan 2, 1050 Brussel  
Tel : +32 2 629 33 44  
crazycopy@vub.be  
www.crazycopy.be

Cover Page Design: Arunima Vadlamannati

ISBN : 9789464443875  
NUR CODE : 926  
THEMA : PHP

*Dedicated to the High Energy Physics researchers who chisel away at  
the wall that constrains human understanding...*

I shall be telling this with a sigh  
Somewhere ages and ages hence:  
Two roads diverged in a wood, and I—  
I took the one less traveled by,  
And that has made all the difference.

- From **The Road Not Taken** by Robert Frost



# Samenvatting

Modellen met gecompriëerde massaspectra met leptonen met een laag transvers momentum hebben een lage gevoeligheid voor directe detectie bij het CMS experiment vanwege de harde beperkingen bij de online gegevensselectie (trigger). In dit werk is aangetoond dat de significantie van de zoektocht wordt verkregen voor lepton verplaatsingen van  $\mathcal{O}(0.1 \text{ cm})$  tot  $\mathcal{O}(100 \text{ cm})$  door de  $p_T$  drempelwaarde voor het triggeren van de leptonen te verlagen tot 20 GeV. Het CMS experiment onderging een upgrade tijdens de LS2 van de LHC en was in voorbereiding voor Run 3. Er werden nieuwe triggers gemaakt voor standaard datacollectiestromen van het CMS experiment in LHC Run 3 gericht op verplaatste leptonen met een lage  $p_T$  drempelwaarde voor leptonen. De nieuwe triggers onderzoeken isolatie, vorm van de energiewater, en vertraging in de energieafzetting in de elektromagnetische calorimeter voor elektronen. De  $p_T$  drempelwaarde voor de eindtoestanden  $\mu\mu$  en  $\mu e$  is verlaagd tot 20 GeV, terwijl die voor  $e e$  is verlaagd tot respectievelijk 25 en 12 GeV voor de twee elektronen met de hoogste energie in het event. Bovendien kan voor gebeurtenissen met extra activiteit, zoals ontbrekende energie of de aanwezigheid van een jet, de drempelwaarde voor elektronen zo laag zijn als 10 GeV.

Daarnaast is de datascouting in het CMS experiment verbeterd in Run 3 van de LHC om elektronen ( $e$ ) en fotonen ( $\gamma$ ) met lage energie mee te nemen, waardoor we de  $p_T$  drempel kunnen verlagen zonder extra beperkingen. Gebeurtenissen met drempelwaarde voor een enkele  $e/\gamma$  van 30 GeV en asymmetrische di- $e/\gamma$  drempelwaarden van 16 GeV en 12 GeV werden opgeslagen. De standaardmodeldeeltjes  $J/\psi$  en  $\Upsilon$  meson werden waargenomen in de invariante massaverdeling van het di-elektron, samen met de piek voor het  $Z$ -boson. Dit is het eerste resultaat ooit in zijn soort binnen het CMS experiment en opent de mogelijkheid om lage energie zoekacties uit te voeren met de elektronen en fotonen in de scouting data verzameld door het CMS experiment in Run 3.



# Abstract

Compressed mass spectra models with low transverse momentum displaced leptons signature have low sensitivity for direct detection at the CMS experiment due to hard trigger constraints. It has been shown in this work that the search significance is obtained over lepton displacements of  $\mathcal{O}(0.1 \text{ cm})$  to  $\mathcal{O}(100 \text{ cm})$  by lowering the trigger  $p_T$  threshold on the leptons to 20 GeV. The CMS experiment was undergoing an upgrade during the LHC LS2 and was under preparation for Run 3. New triggers were created for standard data collection streams of the CMS experiment in LHC Run 3 targeting displaced leptons with low- $p_T$  threshold on leptons. The new triggers explore isolation, shower shape and delay in the energy deposition in the electromagnetic calorimeter for electrons. The  $p_T$  threshold on the final states  $\mu \mu$  and  $\mu e$  is lowered to 20 GeV while for  $e e$  is lowered to 25 and 12 GeV for the two highest energy electrons in the event respectively. Additionally, for events with additional activity such as missing energy or the presence of a jet, the threshold for electrons can be as low as 10 GeV.

Additionally, the data scouting in the CMS experiment has been enhanced in the Run 3 of the LHC to include low energy electrons ( $e$ ) and photons ( $\gamma$ ) that allow us to lower the  $p_T$  threshold without any additional constraints. Events with single  $e/\gamma$   $p_T$  threshold of 30 GeV and asymmetric di- $e/\gamma$   $p_T$  threshold of 16 GeV and 12 GeV have been stored. The standard model particles -  $J/\psi$  and  $\Upsilon$  meson were observed in the di-electron invariant mass distribution along with the peak for the  $Z$  boson. This is the first ever result of its kind in the CMS collaboration and opens the possibility to perform low energy searches with the electrons and photons in the scouting data collected by the CMS collaboration in Run 3.



# Author contributions

## Papers

1. F. Blekman, N. Desai, A. Filimonova, A.R. Sahasransu, and S. Westhoff. Soft displaced leptons at the LHC. *Journal of High Energy Physics* 2020, no. 11 (2020): 1-26.

My work made up most of the paper. I simulated the model and analysed its phenomenology. I simulated the background and extrapolated it. I trained the neural network and classified between signal and background.

2. (To be submitted) CMS collaboration. Enriching the physics program of CMS through data scouting and data parking. *Physics Reports*.

All the work in the scouting Run 3 electron and photons section of the paper were done by me. This work is elaborated in this thesis. I also wrote this section of the paper and helped to edit the chapter on Run 3 scouting.

3. (To be submitted) CMS collaboration. New long-lived triggers for LHC Run 3 in the CMS experiment. *Physics Reports*.

All the work on my triggers discussed in this thesis were part of the paper. I am also helping to write and edit this paper.

## Upcoming Conference Proceedings

1. (To be submitted) CMS Collaboration. CMS Triggers. International Conference on New Frontiers in Physics 2022. *International Journal of Modern Physics: Conference Series*. - Delayed due to COVID-19.

## Presentations at Conferences

1. Soft displaced leptons at the LHC. *Seventh workshop of the LHC LLP community*. 25 May 2020.
2. CMS Triggers. *XI International Conference in New Frontiers in Physics*. 7 September 2022.

## Teaching responsibilities

1. Sub-Atomic Physics II - 006510, 6 ECTS. *Vrije Universiteit Brussel*. Academic year 19-20, 20-21 and 21-22.
2. Bachelor Thesis. Strange quark tagging at future  $e^+ e^-$  machines by Lode Ward Vanhecke, 12 ECTS. *Vrije Universiteit Brussel*. Academic year 20-21.

# Preface

Our understanding of the formation of the universe has improved by leaps and bounds in the last millennium. In its march to progress, the human civilisation has always sought the answer to the most fundamental questions, among them the functioning of the universe we live in and its creation. In our quest for answers, in the last couple of millennia, an intricate understanding of our universe has been developed.

The Standard Model (SM) of Particle Physics encapsulates our best known understanding of the universe. It is an elegant combination of a deep understanding of theory and experiments accumulated over the history of humankind. It is the answer to the mysteries that captured the attention of and confused the best minds of the 20th century. It presents the interaction of particles, e.g. the electron and photons, and fields, e.g. the electromagnetic fields, in a single unified framework. Only the gravitational interactions lie beyond the scope of the SM.

The collider experiments at the Large Hadron Collider (LHC), CERN, e.g. the CMS and ATLAS, rigorously check the applicability of the SM to fundamental interactions. Till date, no deviation has been observed from the SM predictions. It perfectly models the interaction between all fundamental particles and forces and well-predicts the interaction and decay of composite particles. The SM theory has been so successful that it predicted a scalar interaction for the mass of particles, which was subsequently observed at the LHC. This was the Brout-Englert-Higgs boson observed in 2012 by the CMS and ATLAS experiments.

While the SM is immensely successful, our knowledge of the evolution of our universe is fundamentally lacking. This is exposed by experiments that look into deep space. There appear to be extended chunks of matter which have a significant gravitational presence and little to no interaction with particles and forces in the SM. This presence has been termed dark matter and forms an active branch of research in particle physics. In addition to filling our knowledge gaps in SM, the LHC offers an incredible opportunity to produce and discover dark matter. The discovery of dark matter at the LHC is marred by multiple technical challenges. Recent developments target relatively low energy interactions at the LHC for telltale signs of dark matter.

One challenge in collecting and analysing low-energy signatures at the colliders is the massive amount of data that is generated at these energies. A technical strategy called triggering selectively collects data at the CMS experiment. Often this is achieved by removing low-energy interactions which reduces sensitivity to discover low energy DM interactions. The work in this thesis substitutes energy-based selection with the quality and nature of detector data to improve sensitivity to dark matter discovery with low energy interactions. Additionally, a substitute

strategy called data scouting, where a reduced version of the data is collected, has been explored. The data scouting strategy has been improved to be sensitive to dark matter involving low-energy interactions with electrons and photons.

I have demonstrated that the new triggering mechanism is 5 to 30 times more sensitive to dark matter search at low energies than the previous triggers. I have also produced plots to validate data scouting by observing the decay of low-mass SM mesons in the data collected at the LHC in 2023 from proton collisions.

This thesis is structured as follows.

- **The Standard Model and Beyond** - Summarises the SM briefly and introduces the dark matter model with potential for discovery at the CMS detector.
- **The Compact Muon Solenoid Experiment** - The CMS experiment and the proton collisions at the LHC are discussed with special emphasis on the detector components involved in triggering and scouting.
- **Simulation and Physics Object Reconstruction** - The process to convert the detector data to comprehensible information for physics studies is covered in this chapter.
- **Soft and Displaced Leptons - Phenomenology** - The detector phenomenology of the dark matter model in the CMS detector is studied in detail in this chapter.
- **New HLT Paths for Soft Displaced Leptons** - This chapter utilises the phenomenological studies to develop new triggers so that the CMS experiment has improved physics sensitivity to discover low energy dark matter interactions.
- **Data Scouting** - This chapter discusses the challenges and the path taken to include low energy electrons and photons in scouting.
- **Outlook** - An overview of the scope of further work and potential for further physics explorations as a result of work done in this thesis is presented in this chapter.
- **Conclusion** - A full summary of the thesis work.

# Acknowledgement

Often, success is driven by a conducive environment. If not for the work of countless predecessors in experimental collider physicists before me, people working tirelessly to study the sublime, this thesis would not exist. Other than the meta, however, a significant number of people influenced this thesis directly.

I owe a great deal of gratitude to Prof. Freya Blekman for her constant support, supervision, foresight and motivation. There was a time when I felt lost in the dense body of research on dark matter. With her immense knowledge of the field and skill at supervision, she guided me to recover the momentum and motivation to continue experimental particle physics research.

I am very thankful to Prof. Steven Lowette for his guidance in completing this thesis, for recognising the value of my work and for supporting me with additional funding for the last few months. His expertise in the CMS experiment and search for exotic signatures were invaluable in providing the current structure of this thesis as well as the final results.

I thank Dr. Denise Müller, Dr. Anastasiia Filimonova, Prof. Susanne Westhoff and Prof. Nishita Desai for their direct contribution to the contents of this thesis through collaborations. Dr. Denise Müller actively supported correcting and commenting on the thesis text, adding an extra layer of comprehensibility in reading this thesis. Dr. Anastasiia Filimonova helped me understand the details of cosmology and model building in high-energy physics. In conjunction with Prof. Susanne Westhoff and Prof. Nishita Desai, their insight and software expertise were essential to complete the phenomenology research we published in the JHEP.

I also thank Prof. Alberto Mariotti and Prof. Ritesh Singh for the discussions, which helped me better grasp particle physics theory. I enjoyed taking the tutorials for the sub-atomic physics course taught by Prof. Alberto Mariotti. It was intellectually engaging and provided the opportunity to be creative while teaching. The contributions of the lecture series and online notes by Prof. David Tong to my understanding of theoretical particle physics and the Standard Model were essential as well.

The contribution of my many CMS collaborators is irreplaceable. Their technical expertise and prompt help were invaluable in finishing any work involving the CMS experiment. I cannot ask for a better team environment. I cannot mention all my collaborators, but some notable mentions include Dr. Sam James Harper, Prof. Swagata Mukherjee, Dr. Jakob Salfeld-Nebgen, Prof. David Sperka, Prof. Laurent Thomas, Dr. Karim el-Morabit, Dr. Elisa Fontanesi, Dr. Santiago R. P. Saenz, Bennett Paul Greenberg, Adelina E. Lintuluoto, and Soumya Dansana.

From egoistic claims to comic banters and lengthy discussions on the aspects of physics and

from politically charged discussions to shared frustration of searching for the elusive dark matter, all my IIHE friends will always have a special place in my life. I will always be on the lookout for more Belgian evenings with Komla Krampah Godwin, Pragati Mitra, Amandeep Kaur Kalsi, Orazio Zapparta, Tomas Kello, Indrani Jayam, Hugues Evard, Kunal Gautam, Emil Bols, Seth Moortgat, Eduardo Ploerer, Marta C. Molla, Martin Delcourt, Douglas J. P. Burns, Rijeesh Keeloth, Felix Schlueter, Nikos Stylianos, Nordin Breugelmans, Xander Nagels, Yarno Mercks, Aaron Rase, Jethro Stoffels, Katarina Simkova, Felix Heyen, Alexandre De Moor and Senne Van Putte.

The technical support of the IIHE IT staff Oliver Devroede, Denise Dutranois, Stephane Gerard, Adriano Scodrani, and Romain Rougny was precious. The support provided by the IIHE secretariat - Marleen Goeman, Sofie Van Den Bussche, Audrey Terrier and Nina Hindriks made my life free of administrative clutter. Their essential work helped me focus on research and is incredibly valued.

I am very thankful to my social circle of international friends for a good mental health. Hope to see Lennard, Gael, Alana, Manon, Emma, Elza, Rajesh, Anwesha, Avinash, Shreya, Marnick, Seevani, Chara, Efitxis, Pietro, Afea, Marouane, Faheem, Heidi, Miriam, Sofya, Lydia, Angie and Fien for some more volleyball smashes, summer barbecues and house parties. The COVID-19 isolation with Lennard, my first Halloween party with Gael, the warm country visits with Alana, Manon and Emma, on different occasions, will forever hold a special memory for me. I also thank Pietro and Soumya for being cosy housemates in Brussels. It was inevitable that we would remain lifelong friends after being house mates. I will never forget the house change from Ixelles to Schaerbeek with Soumya when we were so tired that the cold floor in our new unsettled home felt welcome. I was fortuitous to ever be friends with Godwin and share my birthday party with him on numerous occasions. It was great knowing the unique outlook of Hugues and Indrani always managed to make a delicious feast for when we had dinner at her place. Definitely looking forward to few more Belgian evenings with all my friends.

On a more personal note, a cherished mention of my parents, writers and social workers, Mrs Abanti Behera and Mr Ranadhir Jena, for all of their sacrifice, social teachings, life lessons and most importantly, love in building my life as it is today. Thank you for holding down the fort in Kendrapara during COVID-19 and for patiently bearing with my long absence for four years. I aspire to dedicate as much dedication to my research as you have to structure an educated, just and equal society. Lastly, thank you, Isabel Lushbough, for your support and encouragement in every aspect of my life. I hope to be worth your affection.

Abanti Ranadhir Sahasransu

# Natural units: From the very big to the very energetic

In addition to the commonly used system of units, this thesis will be using the *natural* system of units [1]. Mass, time and length are expressed in units of energy in natural units. This is different from the globally recognised standard international (SI) system of units where they are identified as fundamental quantities represented using  $[M]$ ,  $[T]$  and  $[L]$ , respectively. At the sub-atomic scale, the laws of physics allow us to construct equivalences among them.

The value of certain constants in SI units e.g. speed of light in the vacuum,  $c \sim 2.99792458 \times 10^8$  m/s and the reduced Planck's constant,  $\hbar \sim 1.0545718 \times 10^{-34}$  Js appear to be irrational and make calculations unpleasant. This is the result of a choice in the measure of a fundamental quantity in a particular system of units.

To simplify High Energy Physics (HEP) calculations, the choice of  $c = 1$  is made in natural units. The choice is based on a fundamental tenet of the special theory of relativity which states that the speed of light in the vacuum is a fundamental constant. Einstein's energy-mass equivalence relation,  $E = mc^2$ , thus changes to  $E = m$  in natural units; mass and energy can be expressed in units of energy. Choosing  $c = 1$  also allows us to put length and time on equivalent footing. In SI, the dimensions of  $c$  are  $[L][T]^{-1}$ . Choosing  $c = 1$ , is essentially using

$$[T] = \frac{[L]}{c}. \quad (1)$$

Length can be expressed in units of time and the same way time is expressed using units of length.

Planck's constant relates the energy of a photon to its frequency by the relation  $E = \hbar\omega$ , where  $\omega = 2\pi\nu$  and  $\nu$  is the photon frequency. In SI units,  $\hbar$  has the dimension  $[E][T]$ . Using a similar argument for  $\hbar$  as  $c$ , in natural units,  $\hbar = 1$  is chosen.

$$[E] = \frac{\hbar}{[T]}. \quad (2)$$

Intuitively, it relates the energy of a process to the typical order of time in which it happens. Thus, using Eq. 1, time and length can be expressed in units of inverse energy. The higher the energy of a process under consideration, the lower its length scale. Fig. 1 shows some practical physics processes to demonstrate this idea.

In natural units,  $c = \hbar = 1$ . Parameters in SI, expressed in dimensions of  $[MLT]$ , can be

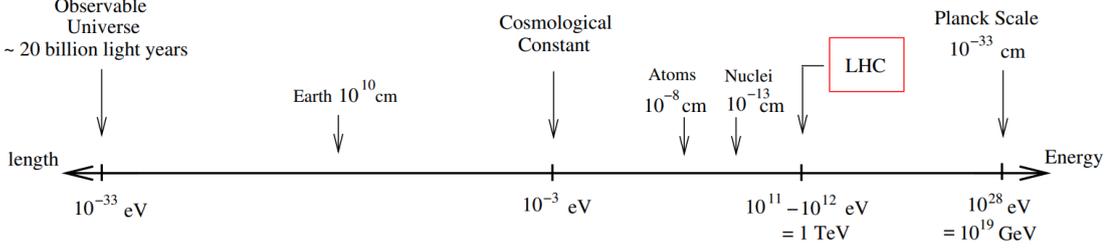


Figure 1: Energy scale of phenomena observed in the universe is shown. Eq. 3 gives the inverse relation for an equivalent length scale. The size of the universe is  $\sim 10^{28}$  cm, the maximum length that can be observed. On the other end, Newton's gravitational constant provides the lowest length scale for physical phenomena at  $\sim 10^{-33}$  cm. The ultimate goal is to build a theory in the confines of these limiting scales.

expressed in units of energy. Consequently, mass is expressed in units of energy while length and time are expressed in inverse energy units.

Electron-volt (eV) is chosen as the unit of energy in HEP. 1 eV is the energy gained by one electron across a potential difference of 1 volt. For larger units of energy, multiples of eV, such as mega electron-volt (MeV =  $10^6$  eV), giga electron-volt (GeV =  $10^9$  eV) and so on are used. The Large Hadron Collider (LHC) [2] at European Organization for Nuclear Research (CERN) operates at energy scales of tens of tera electron-volt (TeV =  $10^{12}$  eV).

### Scale of physics

Historically, the theoretical description of physical phenomena in the known universe has changed multiple times. As the energy scale of the physical process considered is increased, emergent physical phenomena require a new fundamental theory and the old description becomes an asymptotic approximation of the new fundamental theory. Using the natural units, an intuition can be developed for the energy or an equivalent length scale they apply. The classical theories of physics by Newton and Maxwell hold for a significant proportion of the observed phenomena. For a century, it successfully described classical mechanics, planetary motion and electromagnetism. However, as the experiments were able to reach higher in energy or equivalently smaller in length, the description of a physical system needed a more complex evaluation. Quantum mechanics developed as an answer for atomic scales but eventually, deviations were observed that demanded an evaluation. If the energy scale of a process is known, say  $E$ , its equivalent length scale can be obtained using Eq. 1 and Eq. 2

$$\lambda = \frac{\hbar c}{E} \quad (3)$$

Using Eq. 3 and a few heuristic arguments the energy scale for which a theory needs to be developed can be obtained. One of the weakest forces is gravitation. Newton's constant in gravitational force is given by

$$G_N \approx 6.67 \times 10^{-11} \text{m}^3 \text{kg}^{-1} \text{s}^{-2}$$

Newtons' constant has dimension  $[E]^{-2}$  in natural units. The corresponding energy and length

scales known as Planck mass ( $M_{pl}$ ) and Planck length ( $L_{pl}$ ),

$$\begin{aligned} M_{pl} &= \sqrt{\frac{\hbar c}{8\pi G}} \approx 2 \times 10^{18} \text{ GeV}, \\ L_{pl} &= \sqrt{\frac{8\pi\hbar G}{c^3}} \approx 8 \times 10^{-35} \text{ m} \end{aligned} \tag{4}$$

are the largest energy and the smallest length scale of the known universe respectively. On the other end, the largest length scale is the size of the known universe itself.

$$L_{universe} \approx 9 \times 10^{26} \text{ m} \tag{5}$$

The corresponding energy scale, also known as the Hubble constant, is,

$$H = 2 \times 10^{-33} \text{ eV} \tag{6}$$

This is the domain of the study of particle physics. Its goal is to develop a theory that can understand everything that happens in the energy scale  $10^{-33} \text{ eV} < E < 10^{27} \text{ eV}$  or equivalently in the length scale  $10^{26} \text{ m} > L > 10^{-34} \text{ m}$ . The concept of energy scale or length scale will be referred to multiple times in the thesis.



# Contents

<b>Samenvatting</b>	<b>iii</b>
<b>Abstract</b>	<b>v</b>
<b>Author contributions</b>	<b>vii</b>
<b>Preface</b>	<b>ix</b>
<b>Acknowledgement</b>	<b>xi</b>
<b>Units</b>	<b>xiii</b>
<b>1 The Standard Model and Beyond</b>	<b>1</b>
1.1 Gauge Fields in the SM . . . . .	2
1.2 Particles and Forces in the SM . . . . .	4
1.2.1 The Forces . . . . .	5
1.2.2 The Matter Particles . . . . .	9
1.3 Open Questions . . . . .	10
1.4 Dark Matter . . . . .	12
1.4.1 Gravitational Evidence of Dark Matter . . . . .	12
1.4.2 Theoretical Modelling of the DM . . . . .	14
1.4.3 Singlet-Triplet Higgs Portal DM model . . . . .	15
<b>2 The Compact Muon Solenoid Experiment</b>	<b>21</b>
2.1 The Large Hadron Collider (LHC) . . . . .	22
2.1.1 The CERN accelerator complex . . . . .	22
2.1.2 The LHC parameters . . . . .	25
2.2 Performance parameters of the LHC in Run 3 . . . . .	26
2.3 The Compact Muon Solenoid experiment . . . . .	26
2.3.1 Silicon tracker . . . . .	29
2.3.2 Electromagnetic calorimeter . . . . .	31
2.3.3 Hadronic calorimeter . . . . .	32
2.3.4 Solenoid magnet . . . . .	34
2.3.5 Muon system . . . . .	34
2.3.6 Triggering and data acquisition . . . . .	35

<b>3</b>	<b>Simulation and Physics Object Reconstruction</b>	<b>41</b>
3.1	Event Simulation . . . . .	42
3.2	Reconstruction of Physics Objects . . . . .	44
3.2.1	Tracking . . . . .	45
3.2.2	Primary Vertex . . . . .	46
3.2.3	Particle Flow . . . . .	46
3.2.4	Electron and Photon . . . . .	48
3.2.5	Muon . . . . .	53
3.2.6	Jet . . . . .	54
3.2.7	Missing Transverse Momentum . . . . .	54
3.3	Delphes Fast-Simulation Framework . . . . .	54
3.4	Machine Learning . . . . .	55
<b>4</b>	<b>Soft and Displaced Leptons - Phenomenology</b>	<b>59</b>
4.1	Existing LHC Searches for Displaced Leptons . . . . .	59
4.2	STHDM Model Benchmarks . . . . .	63
4.3	Simulation of Signal and Background Events . . . . .	64
4.3.1	Event pre-selections . . . . .	65
4.3.2	Background extrapolation to low momenta . . . . .	65
4.3.3	Event yields for signal and background . . . . .	67
4.4	Multi-variate analysis . . . . .	69
4.4.1	Kinematic distributions . . . . .	69
4.4.2	Neural network structure . . . . .	76
4.4.3	Classification performance . . . . .	77
4.5	Predictions for the LHC . . . . .	78
<b>5</b>	<b>New HLT Paths for Soft Displaced Leptons</b>	<b>83</b>
5.1	Factors Affecting New Trigger Inclusion . . . . .	85
5.1.1	Dataset Used for Trigger Studies . . . . .	86
5.1.2	Rate Measurement . . . . .	86
5.1.3	Time Measurement . . . . .	87
5.2	The $\mu\mu$ trigger . . . . .	87
5.3	The $\mu e/\gamma$ trigger . . . . .	93
5.3.1	Electron gen-matching . . . . .	96
5.3.2	Electron identification variables . . . . .	101
5.4	The $e/\gamma e/\gamma$ trigger . . . . .	105
5.4.1	Time delay in ECAL signature . . . . .	106
5.4.2	Minor second moment of the ECAL shower shape . . . . .	112
5.4.3	Gain in signal acceptance . . . . .	115
5.4.4	Observation in 2022 data and further changes . . . . .	116
5.5	Conclusion . . . . .	117
<b>6</b>	<b>Data Scouting</b>	<b>119</b>
6.1	Scouting strategy for electrons . . . . .	120
6.2	Lower the scouting $p_T$ threshold with new L1 triggers . . . . .	123
6.3	Identification of electrons in the scouting collection . . . . .	128
6.4	From seeded to unseeded reconstruction . . . . .	133
6.5	Identification of low-energy SM particles . . . . .	135
6.6	Observing the invariant mass peaks in data . . . . .	144

<i>CONTENTS</i>	xix
6.7 Trigger efficiency of scouting $e/\gamma$ paths. . . . .	148
6.8 Conclusion . . . . .	149
<b>7 Outlook</b>	<b>151</b>
<b>8 Conclusion</b>	<b>153</b>
<b>A Logarithmic distribution of <math>d_0</math></b>	<b>155</b>



# Chapter *1*

## The Standard Model and Beyond

*The Standard Model theory explains almost all processes in our solar system, roughly 17% of our galaxy and only about 5% of the matter-energy content of the Universe. In this chapter, I will briefly discuss the Standard Model and a particle-based dark matter theory as an expansion of the Standard Model. This will serve as the theoretical basis for our experimental explorations.*

In the last century, physicists developed the Standard Model (SM) of particle physics [3–7] to describe the fundamental interactions between elementary particles. It explained experimental results spanning the last century and continues to do so to an extremely good precision at the colliders [8]. An unlikely prediction of the SM was that the mass of fundamental particles comes from their interaction with a scalar particle. The discovery of the Brout-Englert-Higgs Boson at the LHC in 2012 [9,10], almost 50 years after its prediction [6,7], was the final piece to cement the foundations of the SM theory. Excepts for a few caveats, the SM governs almost all interactions in our Universe.

I have briefly discussed the SM and its constituents for completeness. The complete theory of SM can be read from many good books. [1, 11, 12] The SM unified the notion of particles, waves and forces. The classical understanding was intuitive. Forces act on non-rigid bodies to create motion. The electromagnetic force was understood to be the effect of classical fields that could propagate as waves [13]. Planck and Einstein’s observation of the quantised nature of electromagnetic waves [14], followed by De Broglie’s observation that the matter particles exhibit duality by behaving as waves, broke this deterministic behaviour and laid the foundation of Quantum Mechanics in the early 1920s [15]. Atomic and sub-atomic observations with protons, electrons and neutrons fit well into the description of quantum mechanics. Macroscopic observations such as diamagnetism were also explained [16]. However, the discovery of heavier particles in cosmic rays and the apparent symmetry among these particles [17, 18] pointed to a greater symmetry in the fundamental description of the sub-atomic particles. The phenomenon of parity violation in particle decays [19], analogous behaviour of particles and anti-particles except for their electric charge and identical description of particles from cosmological sources required a universal theory. The unification of the idea of fields and particles would pave the way.

The SM lay to rest an era of open questions. It is a theory of interacting fields. Particles, forces and waves appear as manifestations of an underlying quantum field, which I will describe in the following text. The nature of the manifestation depends on the energy scale of the physics under consideration. The mathematical construct of quantum field theory describes the interactions

between the fields.

The language of the SM is briefly described in Section 1.1 followed by a description of the constituents of the SM in Section 1.2 in terms of the conventional particles and forces. Section 1.3 focuses on the few open questions in the understanding of the Universe. In the final Section 1.4, I will discuss an extension to the SM theory which I have used to build experimental methods to search for physics beyond the SM [20].

## 1.1 Gauge Fields in the SM

Quantum Field Theory (QFT) [11] combines concepts from classical field theory, quantum mechanics and special relativity. The basic object in QFT is a field which can only take quantised energy values. A **field**,  $\phi_a(x^\mu)$  is a ubiquitous physical object which has a value at all points in space throughout the Universe. Here,  $a$  indexes the field when it is part of a group of fields and  $x^\mu$  is the space-time 4-vector. The value of a field can evolve with time. Any operation on the field, including what is conventionally referred to as a measurement, is done by operators  $\mathcal{O}$  acting on  $\phi_a(x^\mu)$ . Field dynamics are governed by the Lagrangian  $L(t)$  defined in Eq. 1.1. The Lagrangian is the spatial volume integral over the Lagrangian density  $\mathcal{L}$  which is a function of the field  $\phi_a(x^\mu)$  and its differential in the space-time 4-vector  $\partial_\mu\phi_a(x^\mu)$ .

$$L(t) = \int d^3x \mathcal{L}(\phi_a, \partial_\mu\phi_a, x^\mu) \quad (1.1)$$

Transformations that leave the Lagrangian density and, consequentially, the equations of motion invariant are called symmetries of the system. Based on the functional form of the transformation, a physical system can have local, global and internal symmetry. The symmetry is local if the transformation is a function of space-time coordinates. A global symmetry is a specific case where the transformation function does not vary with space-time coordinates. A system is said to have internal symmetry if the transformation of the field itself leaves the Lagrangian density invariant.

The fields in the SM enjoy multiple symmetries. Local symmetries of the SM result in interesting properties in particle interactions. Two of them are particularly important. The special theory of relativity requires physical systems to be invariant under the Lorentz sub-group of transformations [21].

Additionally, SM fields are **gauge invariant**. This means they have an internal symmetry under a unitary complex valued transformation called gauge transformation.

$$\phi_a(x) \rightarrow \phi'_a(x) = U(x)\phi_a(x) \quad (1.2)$$

where  $U$  is the transformation operator. The unitary requirement is to ensure that the field's norm remains unchanged after the gauge transformation. A gauge transformation is also a local transformation.

In the context of symmetry, one of the most important contributions was from Emmy Noether. Noether's theorem states that every differentiable symmetry of the Lagrangian, Eq. 1.1, corresponds to a conserved current [22]. The translational and rotational invariance of the action in the Poincaré group of transformations give rise to energy-momentum and angular momentum conservation laws, respectively. The various forms of gauge invariance give rise to multiple conserved charges in the SM.

SM fields are quantum fields. A quantum field has discrete energy levels. The possible energy values are eigenvalues of the Hamiltonian operator,  $H = \int \mathcal{H}d^3x$ , where  $\mathcal{H}$  is called the

Hamiltonian density.  $\mathcal{H}(\phi_a(x^\mu), \pi_a(x^\mu), x^\mu)$  is a function of the space-time coordinates, field  $\phi_a(x^\mu)$  and its conjugate momentum  $\pi_a(x^\mu)$ , where  $\pi_a = \partial\mathcal{L}/\partial\dot{\phi}_a$ . It is obtained from the Lagrangian density given by Eq. 1.3.

$$\mathcal{H}(\phi_a(x), \pi_a(x), x) = \pi\dot{\phi} + \pi^\dagger\dot{\phi}^\dagger - \mathcal{L}(\phi_a(x), \pi_a(x), x) \quad (1.3)$$

Particles are created and destroyed from these quantised gauge fields with quantised energy values.

The QFT formalism allows for some important properties in the SM. The observables in the SM are obtained by defining an energy scale cut-off where the SM theory re-normalises. The values for the observables would diverge in the absence of this re-normalising energy scale leading to inconsistent theory. The re-normalising energy scale of the SM is an important aspect which comes from the QFT formalism. In addition, the description of the Universe in terms of an ever fluctuating ubiquitous field explains the universal identical nature of particles. The fluctuations in these vacuum fields were first theorised by H. Casimir and were later measured by S. Lamoreaux [23].

Electrons ( $e$ ), protons ( $p$ ) and neutrons ( $n$ ) were the first three sub-atomic particles to be discovered. The electromagnetic force, expressed using Maxwell's equations, was the first classical field theory formulated. In the SM, the matter particles like  $e$ ,  $p$  and  $n$  and force particles like the photons ( $\gamma$ ) are all quantisation of their fundamental fields. The low energy scale manifestation as matter particles or force particles is parameterised by the intrinsic spin of the field. The matter particles and the force particles are the quantisations of Dirac's theory of spin-half fields and Yang-Mills theory of spin-one fields respectively. The difference in the spin consequentially results in a difference in the statistics they follow. The matter particles are called fermions, named after Enrico Fermi, and the force particles are called bosons, named after Satyendra Nath Bose.

The Dirac field [24] is a quantised complex-valued spinor field. The Lagrangian density for a spinor with mass  $m$  is given by,

$$\mathcal{L}_D = \bar{\psi}(x)(i\zeta^\mu\partial_\mu - m)\psi(x) \quad (1.4)$$

where the  $\zeta^\mu$  matrices follow the Clifford algebra [24]. The field describes the dynamics of the spin-half particles. The Dirac Lagrangian is Lorentz invariant. However, it doesn't remain invariant on local gauge transformation. Under a gauge transformation, like Eq. 1.2, that depends on space-time coordinates,

$$\psi_D \rightarrow \psi'_D = U(x)\psi_D, \quad (1.5)$$

The Dirac Lagrangian transforms as,

$$\begin{aligned} \mathcal{L}_D &\rightarrow \mathcal{L}'_D = \bar{\psi}'_D (i\zeta^\mu\partial_\mu - m)\psi'_D \\ &= i\zeta^0\psi_D^\dagger U^\dagger(x)\zeta^\mu\partial_\mu[U(x)\psi_D] - m\zeta^0\psi_D^\dagger U^\dagger(x)U(x)\psi_D \\ &= i\zeta^0\psi_D^\dagger U^\dagger(x)\zeta^\mu U(x)\partial_\mu[\psi_D] + i\zeta^0\psi_D^\dagger U^\dagger(x)\zeta^\mu[\partial_\mu U(x)]\psi_D - m\zeta^0\psi_D^\dagger\psi_D \\ &= \mathcal{L}_D + i\zeta^0\bar{\psi}_D U^\dagger\zeta^\mu[\partial_\mu U]\psi_D. \end{aligned} \quad (1.6)$$

Violation of the local gauge invariance, as shown in Eq. 1.6, implies that charges in a localised interaction are not conserved. This leads to a theory inconsistent with experimental results. The local gauge invariance for fermionic fields is restored on interaction with the force particles.

A necessary component for theoretical consistency with experiments, the local conservation of charge, requires a minimal theory of interacting spin-half and spin-one fields.

Yang-Mills (YM) theory of fields involves co-interacting vector-valued fields [25]. The YM fields that co-interact form a closed group and are canonically referred to be part of the same force. Mathematically, the co-interaction between the YM fields is governed by a specific Lie Algebra. Representing a field by  $A_\mu^a$ , where  $\mu$  denotes the four space-time components of the vector-valued fields,  $a$  is the index of the field in a group of co-interacting fields and  $T^a$  denotes the generators of the Lie Algebra, the YM field potential connects the individual fields to an effective field potential given by

$$A_\mu = A_\mu^a T^a. \quad (1.7)$$

The field potential at a point defines the measure of interaction with the force. The number of independent fields is fixed by the number of generators of the algebra. A limiting case with a single field and hence no co-interaction applies to the well-known electromagnetic force [26]. The dynamics of the co-interacting fields is governed by the Lagrangian which is a function of the field strength defined as,

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + ig[A_\mu, A_\nu], \quad (1.8)$$

where  $g$  is called the coupling constant. The Lagrangian density for the YM fields is given by,

$$\mathcal{L}_{YM} = -\frac{1}{2} \text{tr.}(F^{\mu\nu} F_{\mu\nu}), \quad (1.9)$$

where  $\text{tr.}()$  stands for the trace of a matrix. The Lagrangian density is Lorentz invariant to preserve the conservation of energy and momentum. It is also manifestly invariant under a gauge transformation in the group defined by the algebra in the group of co-interacting fields. This implies that the charge that arises due to this group of fields is conserved in their co-interactions. The YM fields in the SM theory follow  $SU(N)$  algebra where  $N$  defines the number of independent charges that remain conserved in an interaction with the force.

The transformation of the YM field potential Eq. 1.7 under local gauge transformation, see Eq. 1.5, is defined in such a way that the additional term in Eq. 1.6 cancels and local gauge invariance is restored. For a local gauge transformation  $U(x)$ , the YM field potential transforms as given below.

$$A'_\mu = U A_\mu U^{-1} - \frac{i}{g} U [\partial_\mu U^{-1}] \quad (1.10)$$

Thus, the local gauge invariant theory of fermions minimally includes an interaction with a vector-valued force field of the form  $-g\bar{\psi}\zeta^\mu A_\mu\psi$ , justifying the choice of name *coupling constant* for  $g$ ,

$$\mathcal{L}_D = -\frac{1}{2} \text{tr.}(F^{\mu\nu} F_{\mu\nu}) + \bar{\psi}(x)(i\zeta^\mu D_\mu - m)\psi(x). \quad (1.11)$$

where  $D_\mu = \partial_\mu - igA_\mu$ .

In the upcoming section, a closer look at the particles and forces in the SM is given.

## 1.2 Particles and Forces in the SM

The SM is composed of 12 fundamental spin-1/2 fermions, 12 fundamental spin-1 gauge bosons and 1 fundamental spin-0 scalar particle. There are two basic forces: the strong force and the electroweak (EW) force. The strong force interacts following the  $SU(3)$  algebra while the EW force uses the  $SU(2)_L \otimes U(1)$ . Thus, the Lie Algebra group of the SM interactions is

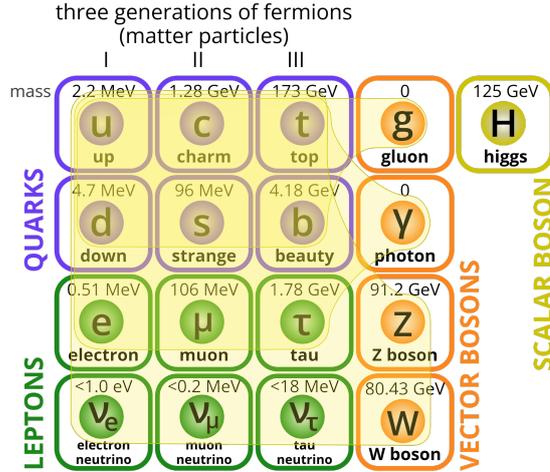


Figure 1.1: A graphical representation of the Standard Model of elementary particles. The classifications between the particles and their masses are shown. The yellow-shaded regions group the particles to the possible interactions they can have. Each fundamental particle is a quantised expression of the underlying fundamental field. [3, 6, 7, 27, 28]

$SU(3) \otimes SU(2)_L \otimes U(1)$  [27]. Based on the forces that the fermions interact with they are grouped into leptons and quarks. The EW force violates parity symmetry. The chiral nature of the electroweak force prevents direct mass terms of the form  $-m\bar{\psi}\psi$  and  $-MA^\mu A_\mu$  in the SM Lagrangian. However, the massive gauge bosons gain mass from the symmetry-breaking mechanism proposed by R. Brout, F. Englert and P.W. Higgs, called the Brout Englert Higgs (BEH) mechanism. The BEH mechanism included a spin-0 scalar field as the final component of the SM. The fermions obtain mass through Yukawa interactions with this scalar field [3].

### 1.2.1 The Forces

Interactions in the SM happen in two ways. Either the fermionic fields interact via the exchange of a boson or the bosonic fields that are part of the same group interact among each other. Canonically speaking, matter particles can interact via the exchange of a force particle and the force particles can interact among each other if they are part of the same force. Richard Feynman simplified the cross-section calculations by introducing the Feynman diagrams for quantum electrodynamics [29] which were subsequently generalised to include all other forces.

There are four fundamentally distinct forces: the electromagnetic force, the strong force, the weak force and gravitation. Gravity lies outside the purview of SM and no experimentally-verified quantum theory exists for it, so it shall not be the topic of our discussion. The electromagnetic (EM) and the weak force are part of the SM electroweak (EW) force. The EM force manifests itself as independent force because, as we will see, in the mass eigenstate basis one of the gauge bosons remains massless. This is the photon, force carrier in EM. Its massless nature allows it to be effective in long-length scales and manifest itself as an independent force at the macroscopic level. The quantised theory for EM is called quantum electrodynamics and for the strong force

is called quantum chromo-dynamics.

### Quantum Electro-Dynamics (QED)

QED [26, 28] is the theory of a YM field of the  $U(1)$  gauge group. This is not the  $U(1)$  gauge group in the SM which is parity-violating. The QED  $U(1)$  gauge group arises from the mixing of the  $SU(2)_L$  and the  $U(1)$  SM gauge groups. This is described in more detail later in this section. The gauge boson in QED is the photon ( $\gamma$ ) which is massless. Particles which feel the effect of QED are electrically charged.

### Quantum Chromo-Dynamics (QCD)

QCD is the theory of the Yang-Mills field of the  $SU(3)$  gauge group. The Lie-Algebra of  $SU(3)$  requires eight generators. [17] Corresponding to the generators, there are 8 spin-1 fields which are the force mediators of QCD, called gluons. The  $SU(3)$  gauge group implies three conserved charges, red, blue and green, collectively called the colour charge.<sup>1</sup> Fermionic particles that feel the effect of QCD carry colour charge and fermionic anti-particles carry anti-colour charge. The gluons are massless and are composites of colour and anti-colour charge.

The gluons can self-interact. The strength of all the interactions in QCD is determined by an equivalent form of the coupling constant in the Eq. 1.10, called the strong coupling constant  $g_s$ . The value of  $g_s$  depends on the re-normalisation energy scale cut-off of the theory. It decreases with an increase in the energy of the interaction probed. Or equivalently,  $g_s$  increases with an increase in length scale.

The energy required to increase the spatial separation between two coloured particles is directly proportional to the distance between them. At length scales larger than the proton size ( $\approx 10^{-15}$  m), the coloured particles have sufficient energy to pull other coloured particles from the surrounding quantum field so that it is colour neutral. Hence, quasi-stable particles are always colour-neutral. This phenomenon is called colour confinement. In a high-energy environment, where colour-neutral particle bound states can be shattered, the process of hadronisation gives rise to a jet of colour-neutral particles. The phenomenology of hadronisation is an active field of research [30].

### The Electro-Weak (EW) Theory

The electroweak theory involves interacting YM fields with the gauge group  $SU(2)_L \otimes U(1)$ . [3] The  $SU(2)_L$  component, has three generators, with three corresponding gauge fields  $W_\mu^I$ ,  $I = 1, 2, 3$ . The conserved charge for the weak iso-spin gauge field is called the weak iso-spin ( $W$ ). The  $U(1)$  component has one gauge field  $B_\mu$ . The corresponding conserved charge is called the weak hypercharge, denoted by  $Y$ .

It was first discovered in the Wu experiment that beta decay, which proceeds through weak interactions, was parity violating in nature [19]. The EW theory was developed with the idea that the left chiral fermions form a weak iso-spin doublet while the right-handed fermions are singlets under the weak iso-spin. The weak iso-spin component mixes with the weak hypercharge component in such a way that the resulting electromagnetic force, which also has  $U(1)$  symmetry, does not violate parity.

The  $W$  takes half-integer values for the left-handed fermions and is  $W = 0$  for the right-handed fermions. The left-handed fermions in the weak iso-spin doublet are distinguished by the third component of the weak iso-spin field  $W_3$ . The  $Y$  takes different values depending on the fermion

---

<sup>1</sup>Conventionally these are named after the primary colours, however, these are just representative of the fundamental quantised property. This concept is similar to the spin property of point-like particles. The particles do not actually undergo the physical motion of spinning. It is just a representation of a quantised property of the field.

such that the electric charge for the said fermion  $Q = W_3 + (1/2)Y$ . The individual charges for the fermions will be discussed in detail when I discuss the different flavours of fermions. The coupling constants for the  $SU(2)_L \otimes U(1)$  gauge group are denoted by  $g_W$  and  $g_Y$  respectively.

The preferential interaction of the EW theory to left-handed particles and the right-handed anti-particles prevents direct mass terms in the SM Lagrangian. Specifically, if Dirac spinors were considered in the chiral basis,

$$\psi = (\psi_L \quad \psi_R), \bar{\psi} = \zeta^0 \psi^\dagger = \begin{pmatrix} \psi_R^* \\ \psi_L^* \end{pmatrix} \quad (1.12)$$

where  $\psi_L$  and  $\psi_R$  are the left and the right-handed component of the spinor respectively, the mass term  $-m\bar{\psi}\psi = -m(\psi_R^*\psi_L + \psi_L^*\psi_R)$  is no longer admissible in the Lagrangian. Under a local gauge transformation of the weak group,  $\psi_L$  and  $\psi_R$  transform differently. A direct mass term of the form  $-m\bar{\psi}\psi$  for fermions and  $-MB^\mu B_\mu$  for the gauge fields will violate the SM local gauge invariance.

The observed fermions and bosons, however, with the exception of the  $\gamma$  and the gluons, do have mass. The theory would seem inconsistent with experimental observations. Robert Brout, Francois Englert and Peter Higgs proposed an electroweak symmetry-breaking mechanism by introducing a scalar Brout-Englert-Higgs (BEH) field [6, 7]. Massive gauge bosons and fermions are a result of interaction with the BEH field.

### Electroweak Symmetry Breaking and the BEH Boson

Mass of the vector bosons and the fermions arise as interactions with a complex scalar field  $\phi$ . In the minimal representation of the  $SU(2)_L$  gauge group,  $\phi$  is a doublet and has hypercharge  $Y = 1$ . The dynamics of the field  $\mathcal{L}_\phi$  and its potential  $V(\phi)$  are given below.

$$\begin{aligned} \mathcal{L}_\phi &= (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi) \\ \text{with } V(\phi) &= -\mu^2 \phi^\dagger \phi + \frac{1}{2} \lambda (\phi^\dagger \phi)^2 \end{aligned} \quad (1.13)$$

where  $D_\mu = \partial_\mu + ig_W W_\mu + ig_Y B_\mu$  is the co-variant derivative of the EW Lagrangian with the corresponding iso-spin and hypercharge for the scalar field. The potential for the scalar is minimised for values  $\phi^\dagger \phi = \mu^2/\lambda$ . The value of  $\phi$  that minimises the BEH potential is called the vacuum expectation value (VEV),  $v$ . It is given by

$$v = \sqrt{\frac{\mu^2}{\lambda}}. \quad (1.14)$$

The significance of VEV is that the BEH field takes a non-zero value in absence of all interactions in vacuum.

The scalar field,  $\phi$ , has four degrees of freedom. The particular structure of the BEH field, shown in Fig. 1.2, has no fundamental explanation within SM. In the unitary gauge representation, the value of  $\phi$  is chosen as,

$$\phi(x) = \begin{pmatrix} 0 \\ v + \frac{1}{\sqrt{2}} h_0(x) \end{pmatrix}. \quad (1.15)$$

The electroweak fields  $W_\mu^1, W_\mu^2, W_\mu^3$  and  $B_\mu$  acquire mass from the non-zero VEV. These fields are not observable fields, however. To obtain the gauge bosons from the electroweak fields, a

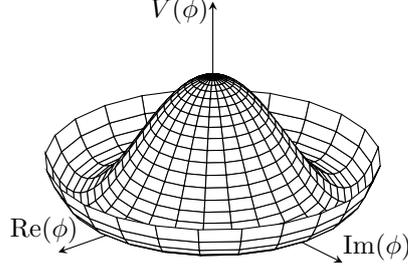


Figure 1.2: The shape of the BEH potential. In the vacuum, the potential sits at its minimum value which is the continuous depression in the Mexican hat shape.

In the unitary gauge representation, the BEH field is chosen to be real only.

Table 1.1: The standard model bosons and their properties. The electroweak bosons are massive and the BEH boson is charged under the weak interactions.

Force	Boson	Mass [GeV]	Electric Charge	Colour Charge (k)	Weak Isospin	Hyper- charge
QED	Photon	0	0	0	0	0
QCD	8 gluons	0	0	$k\bar{k}^2$	0	0
EW	$W^\pm$	80.4	$\pm 1$	0	$\pm 1$	0
EW	Z	91.2	0	0	0	0
BEH	H	125	0	0	-1/2	+1

change to mass eigenstate is required. The electroweak fields mix via the Weinberg electroweak angle  $\theta_{EW}$  defined in Eq. 1.16.

$$\theta_{EW} = \tan^{-1} \frac{g_Y}{g_W} \quad (1.16)$$

The fields in the mass eigenstate with observable bosons are given by

$$\begin{aligned} A_\mu &= \sin\theta_{EW} W_\mu^3 + \cos\theta_{EW} B_\mu \\ Z_\mu^0 &= \cos\theta_{EW} W_\mu^3 - \sin\theta_{EW} B_\mu \\ W_\mu^\pm &= \sqrt{1/2}(W_\mu^1 \mp iW_\mu^2) \end{aligned} \quad (1.17)$$

The  $W_\mu^+$ ,  $W_\mu^-$  [31] and the  $Z_\mu^0$  [32] bosons, are parity-violating and mediate the weak interactions. The photon ( $A_\mu$ ), force-mediator of the QED, does not violate parity. The photon is massless. The electrically charged  $W$ -bosons have a mass of 80.36 GeV and the electrically neutral  $Z$ -boson has a mass of 91.2 GeV.

The fermions gain mass via ad-hoc Yukawa interaction with the BEH field with terms of the form  $-\lambda_\psi \bar{\psi} \phi \psi$  in the SM Lagrangian. The Lagrangian also has the terms  $-\lambda v^2 (h_0)^2$  which gives mass to the scalar field itself. The resulting boson, called the BEH boson, was a prediction of the EW symmetry-breaking mechanism and was discovered 50 years later at the Large Hadron Collider [9, 10].

Tab. 1.1 lists the different force carriers with their corresponding mass and electric charge.

### 1.2.2 The Matter Particles

All matter particles are fermions. They are described by their corresponding Dirac field. Following the nature of the Dirac field, each particle has its corresponding anti-particle with an opposite electric charge.

The fermions form all observable matter that makes up the Universe around us. The fermions come in three generations, with each successive generation more massive than the last. Historically, this was influential to their discovery. The heaviest fermion to be discovered has a mass equal to that of the gold nucleus at  $\approx 173 \text{ GeV}$  [33] and was discovered by the experiments D0 and CDF at the Tevatron in 1995 [34, 35]. Each generation is composed of a charged lepton, the corresponding neutrino, an up-type quark and a down-type quark.

The fermions are classified based on whether they feel the effect of the strong force and the EM force. All fermions interact via the weak force. The table 1.2 shows the values of the quantum charge for the different groups of fermions.

Table 1.2: The standard model fermions with different charges for their chiral states. The electric charge is opposite for particles from anti-particles.

Chiral Fermion	Electric Charge ( $Q$ )	Weak Isospin ( $W_3$ )	Hyper-charge ( $Y$ )	Colour Charge	
$u_L$	+2/3	+1/2	+1/3	r/b/g	
$d_L$	-1/3	-1/2	-1/3		
$\bar{u}_R$	-2/3	-1/2	-1/3		
$\bar{d}_R$	+1/3	+1/2	+1/3		
$u_R$	+2/3	0	+4/3		
$d_R$	-1/3	0	-2/3		
$\bar{u}_L$	-2/3	0	-4/3		
$\bar{d}_L$	+1/3	0	+2/3		
$e_L$	-1	-1/2	-1		0
$\nu_L$	0	+1/2	+1		
$\bar{e}_R$	+1	+1/2	+1		
$\bar{\nu}_R$	0	-1/2	-1		
$e_R$	-1	0	+1		
$\bar{e}_L$	+1	0	-1		

#### The Leptons

Leptons are fermions with zero colour charge. Hence, they do not undergo QCD interaction. Every fermionic generation includes a specific flavour of a lepton, its anti-particle and their neutrino and anti-neutrino counterpart. The three generations, represented by the green particles in Fig. 1.1, are the electron ( $e$ ), electron-type neutrino ( $\nu_e$ ) in the first generation, the muon ( $\mu$ ),

<sup>2</sup> $k$  here stands for the colour charge of QCD. The gluons are combinations of colour and anti-colour charge. 3 colours make 9  $k\bar{k}$  possibilities. The 9  $k\bar{k}$  are linearly combined to 8 independent combinations which are the colour charge of the gluons.

muon-type neutrino ( $\nu_\mu$ ) in the second generation and the tau ( $\tau$ ), tau-type neutrino ( $\nu_\tau$ ) in the third generation. Conventionally, the particles ( $e, \mu$  and  $\tau$ ) have electric charge of  $-1$ , the anti-particles ( $\bar{e}, \bar{\mu}$  and  $\bar{\tau}$ ) have electric charge of  $+1$  and the neutrinos and anti-neutrinos are electrically neutral. Except for the mass, the leptons in successive generations have identical physical properties in the SM. This phenomenon is called lepton universality.

The neutrinos are massive particles, as has been observed from the neutrino oscillation experiments [36–38]. However, in the SM they are assumed mass less. A massive neutrino in the SM would entail the existence of right handed neutrino which has not been observed till date.

The left-handed particles and the right-handed anti-particles form a weak iso-spin doublet. The  $W$ -bosons couples the lepton(anti-lepton) to its corresponding neutrino(anti-neutrino). The third component of the weak iso-spin,  $W_3$ , carries the value of  $-1/2$  for the leptons and  $+1/2$  for the neutrinos. The values are flipped for the anti-particles. The right-handed particles and left-handed anti-particles are singlets under the weak iso-spin. The weak coupling allows for lepton flavour change between the various generations. Both tau and muon can decay to electrons because of their heavier mass via the weak interaction. Analogously, the tau can decay to a muon.

### The Quarks

In addition to the electric charge and the weak charge, the quarks have colour charge. Like the leptons, the left-handed quarks form a doublet in the weak  $SU(2)_L$ . The right-handed quarks on the other hand are singlets in the weak force. Each of the three generations has an analogue to the up quark ( $u$ ) with an electric charge of  $+2/3$ , and the other analogue to the down quark ( $d$ ) with an electric charge of  $-1/3$ . Along with the first generation  $u$  and  $d$ , there is charm ( $c$ ) and strange ( $s$ ) in the second generation and top ( $t$ ) and beauty ( $b$ )<sup>3</sup> in the third generation.

The quarks are subject to strong force via interaction with the gluons. In Fig. 1.1, they are in the blue boxes. Each of the quarks in addition to having an anti-particle also has 2 other colour counterparts. Due to colour confinement [39], any naturally occurring particle has to be colour neutral. Individual quarks combine to form colour-neutral particle states, called hadrons. Depending on whether one forms a bound state of three quarks or quarks with anti-quarks, one gets baryons or mesons, respectively. The proton is a colourless combination of 1 down and 2 up quarks. It is the lightest baryon possible. Hence, it is also the only stable baryon [40]. In the scope of SM, the proton should not decay.

Quarks are all massive particles with mass given in Tab. 1.1. The BEH field produces mass terms for the quarks. The quarks can mix through the BEH field. Obtaining the mass through interaction with the BEH scalar field results in quark mixing between the generations. The (Cabibo-Kobayashi-Masakawa) CKM matrix parameterises this [3, 41, 42].

## 1.3 Open Questions

The SM is a remarkable theory that unifies many seemingly unrelated phenomena. The calculations from the SM show consistency with high-precision tests at colliders [8]. Major deviations come from cosmological observations and neutrino oscillations. These observations can be broadly grouped into three categories. There are some unexplained aspects in the SM itself that require

---

<sup>3</sup>The  $b$  quark is referred to as bottom or beauty quark. In deference to my colleagues who have worked tirelessly to create neural networks to identify this particle up to 90% efficiency with less than 10% fake rate, I will call this a beauty quark in this text.

a more fundamental explanation. The SM completely ignores gravity. The final category comes from astrophysical observations.

The SM assumes neutrinos to be massless, and thus considers only left-handed neutrinos in the lepton multiplet. This is partly owing to the fact that no experiment has ever reported observation of right-handed neutrinos. Neutrino oscillations [36–38], however, require neutrinos to be necessarily massive. This implies the neutrinos in the SM mix through the BEH field and minimally the SM Lagrangian needs to be modified to include this.

The coupling constants of the three forces depend on the energy scale of the process in consideration. A striking feature of running the values for the coupling constants with energy scale is that at a higher energy scale,  $\sim 10^{16}$  GeV, the coupling constants for the hypercharge, weak and strong forces seem to unify to the same order of value. The explanation for this is unknown and leads to the idea that at higher energies, the Universe has a fundamental symmetry that manifests itself as the three known forces. There have been attempts to unite the three forces to one grand unified theory (GUT) [43–45].

A rather unexpected behaviour is observed in the mass of the BEH boson. The observed mass of the BEH boson is related with its self-coupling and is an input to the SM. It also happens to be the only fundamental parameter in the SM which is not dimensionless. As the physics at higher energy scales with respect to the BEH boson mass is probed, it is expected that the BEH boson mass should obtain corrections from higher order terms in perturbative calculations. However, no such correction is observed on the bare mass of the BEH boson. There appears to be some process that varies in such a way as to exactly cancel the contributions from vacuum fluctuations at all energy scales to leave a residual mass of 125 GeV for the Higgs boson. This forms the hierarchy problem [46]. In the same context, although the BEH mechanism provides mass to the particles, the masses of particles are determined experimentally. The reason for some clear patterns observed in the mass of particles is not given by the SM [47]. The SM also does not provide any significant reason for why the number of generations is limited to three.

There have been attempts to construct a quantum theory of gravity and integrate gravity to the SM [48, 49]. However, this is impeded by two factors. The first is that gravity is such a weak effect that it is entirely inconsequential to particle physics experiments. The other is the unreasonable effectiveness of classical gravity with no data that requires or hints at quantised gravity. If the gravitational effects were significant in particle physics experiments, these effects could be measured at colliders. There have been attempts to develop a spin-2 field [50] but these are not renormalisable. All the predictions are necessarily at the Planck length scale, 15 orders of magnitude higher than is possible at the most-energetic man-made experiment at the LHC. Naturally occurring phenomena e.g black holes [51] could help probe further but the relevant information is largely behind the event horizon where it has not been possible to access information from yet [52].

There are observations from cosmological sources that the SM does not explain. The only force effective on large-length scales is gravity. Gravitational force is attractive in nature. However, experimental evidence suggests that the Universe’s expansion is accelerating [53]. This must come from some energy pressure. This is the dark energy component of the  $\Lambda$ -CDM model of the Universe [54].

The asymmetry observed in the dominance of matter contained in the Universe over anti-matter content is another contentious issue as there is no known theory with experimental evidence that suggests such asymmetry. I will look at some astrophysical observations which are more local in the upcoming section. These suggest the presence of a non-luminous massive form of matter field, the other properties of which are unknown.

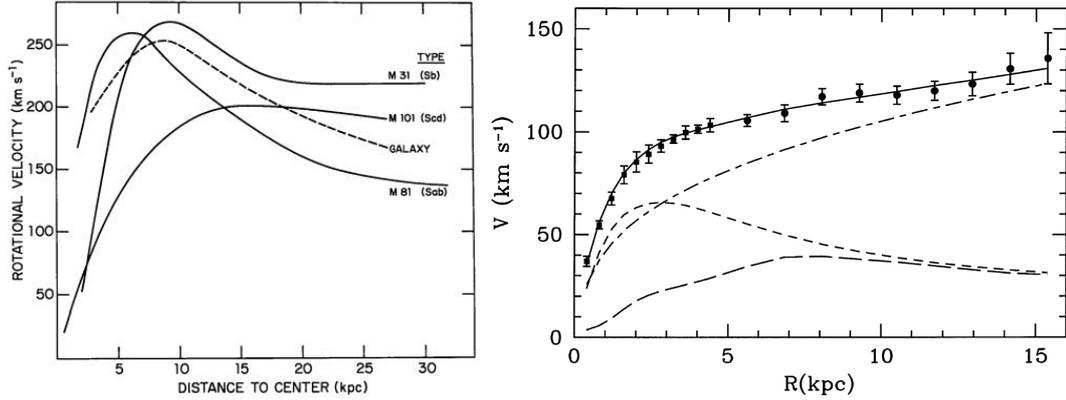


Figure 1.3: Left: Galaxy rotation curve for multiple galaxies. The milky way galaxy is dashed.<sup>4</sup> Right: Galaxy rotation curve of the M33 galaxy showing contribution from disk (dashed), gas (dotted) and dark matter halo (dashed-dotted) fit to the data (dot with error bars) [55]

## 1.4 Dark Matter

Cosmological and astrophysical observations have yielded results that deviate from SM and classical gravity expectations. The results have been established over multiple sources ranging over all of the energy and length scales of the observable universe from visible to microwave electromagnetic waves, high-energy photon bursts, neutrinos and other cosmic ray particles. The observations are self-consistent. They point towards a missing mass in the Universe that exerts a gravitational influence on astronomical length scales and yet couples very weakly with the SM fields. This perplexing phenomenon has been termed dark matter (DM), owing to its non-luminous nature.

I have briefly elaborated on some of these observations in the following sub-section.

### 1.4.1 Gravitational Evidence of Dark Matter

#### Galaxy Rotation Curves

The radial dependence of the angular velocity of objects rotating around the centre of a galaxy, see Fig. 1.3 left, has been measured repeatedly across multiple galaxies [56,57]. The radial mass distribution of each galaxy from its centre can be estimated from these rotation curves. The flat tail in the angular velocity distribution implies linearly increasing mass,  $M(r) \propto r$  on moving radially outward in the outer edges of the galaxies. This is in direct conflict with the luminous mass of these galaxies. Towards the outer galaxy edge the luminous mass accounts for a tiny portion of the mass of the galaxies. A good fit to the angular velocity can be obtained by adding a DM halo contribution, see Fig. 1.3, to the luminous mass of the galaxies [55]. Dark matter dominates the outer reaches of the galaxies.

#### Missing Mass in Galaxy Clusters

In 1933, from the red-shift of the EM spectrum, see Fig. 1.4, Fritz Zwicky showed that galaxies in the Coma cluster were moving too fast to be bound together as a cluster by its visible mass [58]. Rood et. al. [59] analysed over 1000 galaxies in the Coma cluster. They reaffirmed the observation that the overall mass in the Coma cluster is about 300 times larger than its luminous mass. In

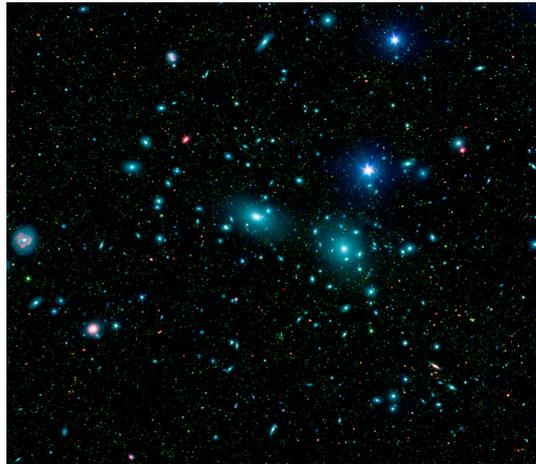


Figure 1.4: Left: A Sloan Digital Sky Survey/Spitzer Space Telescope mosaic of the Coma Cluster in long-wavelength infrared (red), short-wavelength infrared (green), and visible light. The many faint green smudges are dwarf galaxies in the cluster [60].



Figure 1.5: Left: Lensing effect in the galaxy cluster SMACS J0723 [61]. Right: Two sub-clusters collided in the bullet cluster separating the luminous matter, hot gases (pink) and dark matter (blue) [62].

addition, they also pointed out that this mass is from an unknown source and cannot be explained by low luminosity sources such as dwarf galaxies. The cause of this missing mass seems to arise from a dark source with extensive gravitational effects.

### Gravitational Lensing

The trajectory of light bends when it passes close to a massive body [63]. The presence of a massive body in the line of sight to a bright object from our solar system acts as a focus lens [64]. It can cause the image to appear as an arch, or a ring or cause extended distortion e.g. that of an extended cluster, see Fig. 1.5 left. This has been used to compute the mass distribution along

<sup>4</sup>Source: <https://ned.ipac.caltech.edu/level5/Sept16/Bertone/Bertone4.html>.

the line of sight and estimate the total mass in a galaxy cluster [65]. The observations effectively agree with a non-luminous extended mass distribution with weak interactions with the SM.

An analysis of the bullet cluster [66], Fig. 1.5 right, showed the separation of three distinct sources of mass in a galaxy cluster. Two sub-clusters of galaxies had undergone a collision. The luminous mass in the galaxies, such as stars, had passed through without interaction. This is expected as far-spaced stars in galaxies have a low probability of interacting. The hot gas in the clusters, observed in the x-ray and shown in pink in the image, had interacted strongly and sat out of the individual clusters. The blue massive matter, reconstructed from the gravitational lensing effect only, seemed to have completely passed through with bare minimum interaction. The best theory that supports such an observation is a non-luminous (dark) matter theory that interacts at most weakly with itself and with ordinary matter [66].

### The Cosmic Microwave Background Spectrum

The Cosmic Microwave Background (CMB) is an almost perfect black-body spectrum from the epoch of the Universe when the transition occurred from radiation dominated to an electrically neutral and transparent Universe [67]. The tiny temperature fluctuations of the relic photons provide information about the composition of the Universe. The best fit to the  $\Lambda$ CDM cosmological model by the Planck collaboration [68] shows that the Universe's energy budget is composed of about 5% of SM baryonic matter, 27% of DM and 68% of an unknown substance which causes the accelerated expansion of the Universe. The unknown nature of the 68% of the Universe is termed the dark energy component.

### Large Scale Structure of the Universe

Given the uniformity of the early Universe from the CMB spectrum, the large-scale structures observed in today's Universe need seeding by gravitational wells [69]. The gravitational wells should be formed by a field that decoupled during the radiation-dominated epoch of the Universe. In addition, this field should be electrically neutral and non-relativistic. Electrical neutrality is required due to the non-luminous nature of the DM. It has to be non-relativistic in the matter-dominated epoch so that it doesn't have enough energy to break out of the early gravitational structure of the Universe.

DM causes phenomena observed on very different length scales, ranging from dwarf galaxies to the observed universe. Multiple theories have been developed trying to explain these interwoven sets of observations. Over time, however, theories with additional quantum field(s) (and hence new particles) have survived without strong experimental bounds. The singlet-triplet Higgs portal DM model based on a re-interpreted phenomenology by Filimonova and Westhoff [20] provides a new phase space where collider experiments can search for DM. In this thesis, I have worked on the phenomenology of this model and have contributed to increasing its search significance at the LHC collider.

## 1.4.2 Theoretical Modelling of the DM

Multiple approaches have been used to explain the DM observations. Theories that target DM phenomenology come from distinctly different sources. Most of them include additions to the current paradigm of the SM theory and classical gravity. The absence of a theory for quantised gravity leaves space for models which assume the breakdown of classical gravity at lower length scales. Modified gravitational theories such as the MOND theory [70] and TeVeS model [71] are heavily constrained by experimental bounds. The data leans towards the  $\Lambda$ CDM model of the Universe [72–74].

There have been a few approaches to explain DM within the purview of the SM. DM phe-

nomenology through the SM neutrinos required a lower bound to the mass,  $m > 300\text{--}400\text{ eV}$  [75], which directly contradicts the upper bound from oscillation and direct detection experiments [76]. The above approaches are strongly constrained by experimental searches.

DM manifestation by additional fields beyond the SM is a common approach [77]. In such scenarios, the relic density is usually assumed to be set by a freeze-out mechanism in which the DM was in equilibrium with SM in the early universe and then decoupled as the universe cooled down. If the DM candidate couples to the SM, experimentally verifiable signatures can be searched for at colliders e.g. LHC. Specific properties can be attributed to these fields based on earlier observations. The DM candidate should be massive and electrically neutral. It should be stable on cosmological time scales. During the cosmological structure formation epoch of the Universe, when matter separated from the photons by the cooling down of the Universe, it should have been non-relativistic. This is necessary to explain the gravitational seeding for the observed non-homogeneity in the Universe today [69]. It should make up roughly 27% of the energy budget of the Universe. Most theories assume that the DM candidate was created non-relativistic, when it decoupled from the big bang plasma, in what are so-called cold DM models.

It is to my experimental interest that I rely on an effective field theory (EFT) approach to DM phenomenology. Historically, there have been approaches to building complete mathematically consistent theories at a higher energy scale. An example of such an approach is the minimal supersymmetric SM. [78] Low energy predictions of such theories have been tested at the colliders. Since no significant evidence was obtained, constraints were set on the viability of such theories [8]. These approaches require a lot of effort in model building for the full theory before a small set of parameters can be experimentally tested. EFTs, in a diametrically opposite approach, extend the SM with additional fields that can be re-interpreted as the low-energy limitation of a high-energy UV complete theory. This approach enables generic predictions for classes of new theories with similar low-energy structures. The new fields will effectively affect observables which can be tested as deviations from the SM in terms of the experimentally possible final states [79]. Fermi's theory and the corresponding EW theory at the energy scale of the W boson mass is an example of an EFT success [80].

New fields must be added in an EFT to the SM Lagrangian in a way that preserves the SM's gauge invariance. One way this can be achieved is by adding higher dimensional operators that are functions of the SM fields. The SM field in this context defines the *portal* to the DM [81] phenomenology. Minimalist SM portals to DM have been studied and constrained gradually leading to the sophisticated realisation of effective portals and searches at the LHC [82].

The section below along with Chapter 4, has been published [83] in the Journal of High Energy Physics. Significant overlap is expected in the writing and the ideas between the published paper and these sections. The work was carried out in collaboration with Freya Blekman, Nishita Desai, Anastasiia Filimonova and Sussanne Westhoff.

### 1.4.3 Singlet-Triplet Higgs Portal DM model

The singlet-triplet higgs portal dark matter (STHDM) model [20] extends the SM Lagrangian with two Majorana fermion fields. One of the fields is a singlet under the SM gauge group, and the other is a triplet under the SM EW gauge group with zero hypercharge. The fields acquire mass through an effective interaction with the scalar BEH field. UV-complete realisations of this model could include a fermion doublet [84] or a scalar triplet [85]. The scalar scenario of STHDM has been extensively studied at the LHC. [85, 86] The phenomenology of the pseudo-

scalar scenario of this model is considered here.

After the EW symmetry breaking, the charged triplet states remain unaffected by the mixing while the neutral singlet and the neutral triplet mix through interaction with the BEH field to form neutral mass eigenstates. The normal mass hierarchy of the dark states is the heavy neutral field, followed by the two charged fields, followed by the light neutral state. The neutral mass eigenstate with the lightest mass is the DM candidate. It is possible for loop corrections to result in reverse mass hierarchy by adding to the mass of the charged dark state and making it heavier than both the neutral eigenstates. [87] This results in a collider phenomenology which is discussed in Ref. [20, 88], and has been left out in this thesis due to the sake of simplicity. The reverse mass hierarchy contributes the third term in the denominator of Eq. 1.25.

The relic density of DM is driven by the stable neutral dark matter candidate  $\chi^0$  with mass  $m_0$  and the charged dark partner  $\chi^\pm$  with mass  $m_c$ . Mass splitting in the dark sector is defined by Eq. 1.18.

$$\Delta m = m_c - m_0 \ll m_0. \quad (1.18)$$

### Co-annihilating and Co-scattering Dark Matter

The abundance of  $\chi^0$  was suppressed at the moment of its separation from the SM particles if its annihilation had been enhanced in the early Universe through its charged dark partner. The DM states would have been in equilibrium with the SM particles in the early Universe and departed from this equilibrium when the Universe expanded and cooled down. Freeze-out could have been achieved for  $\chi^0$  depending on the strength of its coupling with its charged partner.

Considering non-relativistic  $\chi^0$  during freeze-out, the number densities for dark states around the weak scale,  $n_0$  for  $\chi^0$  and  $n_c$  for  $\chi^\pm$ , follow an exponentially decaying distribution with the freeze-out temperature  $T_f$  [88]. The relative number density of  $\chi^\pm$  and  $\chi^0$  is thus dependent on  $\Delta m$  as

$$\frac{n_c}{n_0} \sim \exp(-\Delta m/T_f). \quad (1.19)$$

Around the freeze-out temperature  $T_f \approx 0.2 m_0$  [20], both dark states should be abundant to permit efficient DM interactions. This results in a (model-dependent) upper bound on the mass difference. Assuming weak scale masses for  $m_0$ , freeze-out through co-annihilation and co-scattering predicts a mass difference of about

$$10 \text{ GeV} \lesssim \Delta m \lesssim 40 \text{ GeV}. \quad (1.20)$$

The small mass splitting is known as the compressed mass scenario.

The coupling of  $\chi^0$  to  $\chi^\pm$  ( $g_\chi$ ) determines the dominant process that sets the relic abundance upon freeze-out. For large  $g_\chi$ , pair annihilation  $\chi^0\chi^0 \rightarrow ff$  is efficient around  $T_f$ . As the value for  $g_\chi$  is lowered, the co-annihilation process  $\chi^0\chi^\pm \rightarrow ff$  provides a more efficient mechanism. Co-scattering of DM with SM particles  $f$  followed by the annihilation of the charged partner sets the relic abundance for the dark sector at lowest values for  $g_\chi$  where co-annihilation becomes inefficient.

$$\chi^0 f \rightarrow \chi^\pm f' \quad \longrightarrow \quad \chi^\pm \chi^\mp \rightarrow ff. \quad (1.21)$$

Freeze-out for  $\chi^0$  typically occurs in the range around the SM weak coupling strength.

$$10^{-3} > g_\chi > 10^{-8}. \quad (1.22)$$

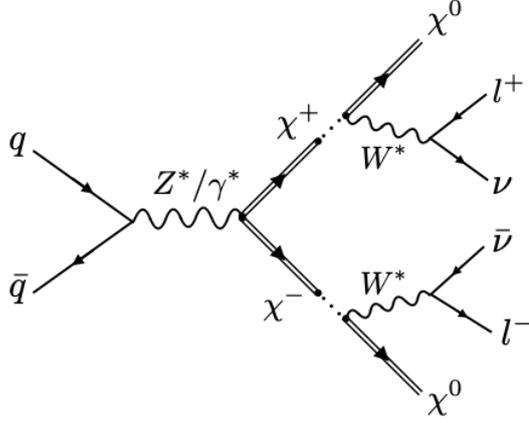


Figure 1.6: Feynman diagram for a signal with soft displaced leptons from decays of dark states at hadron colliders.

DM does not reach thermal equilibrium with SM particles below this range of coupling and a different mechanism has to be invoked to explain the observed relic abundance.

### Signal Characteristics

The dark partner carries an electroweak charge enabling its production at colliders through the weak neutral current. Subsequent decay of the dark partner leads to signatures like

$$pp \rightarrow \chi^+ \chi^- \rightarrow (\chi^0 f)(\chi^0 f). \quad (1.23)$$

Provided that the dark partner is produced at a moderate boost, the SM decay products  $f$  will carry little momentum due to the small mass splitting. Compared to the usual LHC energies, SM processes with low transverse momentum,  $p_T < 50$  GeV, are referred to as soft. Depending on the decay,  $f$  can be one or several particles, at least one carrying an electric charge. Well-known examples of such a process are supersymmetric charginos decaying into neutralinos and leptons or jets [89–92].

Dark states with EW couplings result in final states with soft displaced leptons and soft missing energy via

$$pp \rightarrow Z^*/\gamma^* \rightarrow \chi^+ \chi^- \rightarrow (\chi^0 \ell^+ \nu)(\chi^0 \ell^- \bar{\nu}). \quad (1.24)$$

The corresponding Feynman diagram is shown in Fig. 1.6.

The charged dark state,  $\chi^\pm$ , decays via the weak interactions which can result in detector signatures with leptons and hadrons. The branching fraction to each particle is governed by its gauge quantum number. The CKM matrix is assumed to be diagonal for the calculation of branching fraction. Based on the allowed range for  $\Delta m$ , decay to the third generation quark pair is forbidden. The decay width to lepton and each quark in the normal mass ordering scenario is the same. An additional contribution from the decay of  $\chi^\pm$  to the heavier neutral dark portal particle is possible via production of a pion in the reverse mass ordering scenario. The branching fraction to a lepton,  $\ell$  is given by,

$$\mathcal{B}(\chi^+ \rightarrow \chi^0 \ell \nu) \sim \frac{\Gamma_\ell}{3\Gamma_\ell + 2N_c \Gamma_\ell + \Gamma_\pi} \quad (1.25)$$

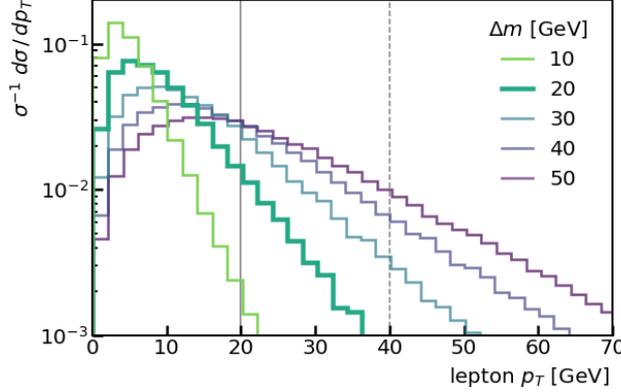


Figure 1.7: Transverse momentum distribution of the lepton  $\ell^\pm$  in  $pp \rightarrow \chi^+\chi^- \rightarrow (\chi^0\ell^+\nu)(\chi^0\ell^-\bar{\nu})$  at  $\sqrt{s} = 13$  TeV for  $m_c = 324$  GeV and various mass splittings  $\Delta m$ , normalised to the total cross section. The distribution at  $\Delta m = 20$  GeV (thick green) is typical for weak-scale dark matter. The dashed line shows the lepton momentum threshold in [93, 94]; the solid line indicates the threshold in our analysis. Based on event generation at parton level using MadGraph5\_aMC@NLO.

where  $\Gamma_\ell$  stands for the partial decay width into leptons and quarks in the normal mass ordering scenario and  $\Gamma_\pi$  stands for the partial decay width to pions in the reverse mass ordering scenario. The number of colours for each quark is represented by  $N_c = 3$ .

Small mass splitting implies the charged dark partner decays via an off-shell  $W$  boson. Three variables - the mass of the  $\chi^\pm$ , the mass splitting and the nominal decay length of the  $\chi^\pm$ , determine the rate and kinematics of this signature.

$$(m_c, \Delta m, c\tau_c), \quad (1.26)$$

$c\tau_c$  is dependent on the lifetime of  $\chi^\pm$  which in turn depends on  $g_\chi$ . The partner's mass  $m_c$  determines the  $\chi^+\chi^-$  production rate at the LHC. The mass splitting  $\Delta m$  is set by requiring the observed relic abundance.

In Fig. 1.7 the transverse momentum distribution of the lepton  $\ell^\pm$  for  $m_c = 324$  GeV and various  $\Delta m$  is shown, without applying any kinematic cuts. At this mass scale, co-annihilation or co-scattering predict  $\Delta m = 20$  GeV. The lepton transverse momentum distribution peaks at  $p_T(\ell) \approx 7$  GeV. The mass scale of the dark partner has little impact on the spectrum at low momenta. In the relevant mass range  $m_c = 100 - 500$  GeV, it only mildly affects the tail of the distribution. Only, three body decays to leptons are considered here. It is possible for two body decay scenarios with this model due to the effect of electroweak corrections, which gives rise to pion signatures with even lower transverse momentum, and has not been used due to a foresight of detector constraints.

A large fraction of events is rejected at the trigger level because the transverse momentum of the leptons is much lower than the threshold of conventional lepton triggers used by ATLAS and CMS [95–97], as indicated by the dashed line. The main goal of this thesis is to lower this threshold to  $p_T > 20$  GeV (black line) to be sensitive to compressed mass spectrum DM scenarios.

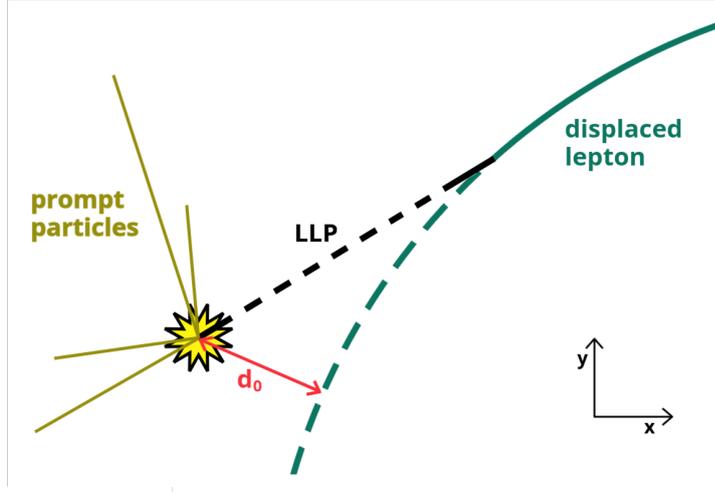


Figure 1.8: The track’s closest approach to the  $p$ - $p$  collision point is  $d_0$ . The yellow star points to the  $p$ - $p$  collision vertex. For lepton displacements in the region close to the  $p$ - $p$  beam,  $d_0$  will have a low value compared to if the displaced lepton was in the direction perpendicular to the  $p$ - $p$  beam.

The nominal decay length  $c\tau_c$  depends on the dark matter coupling  $g_\chi$ . The three-body partial decay width can be estimated for  $\Delta m \ll m_0 \sim m_W$  using EW interaction models as [89] in Eq. 1.27. The parameters for weak-scale co-scattering have been chosen as a reference assuming proximity to the EW interaction as a first-order approximation.

$$\Gamma_c \approx \frac{g_\chi^2 G_F}{30\sqrt{2}\pi^3} \frac{(\Delta m)^5}{m_W^2} = 1.6 \times 10^{-14} \text{ GeV} \left( \frac{g_\chi}{6 \times 10^{-5}} \right)^2 \left( \frac{\Delta m}{20 \text{ GeV}} \right)^5, \quad (1.27)$$

Assuming that no other decay channels are accessible, the proper lifetime and the nominal decay length of the charged dark partner are given by

$$\tau_c = \frac{1}{\Gamma_c} \approx 41 \text{ ps}, \quad c\tau_c \approx 1 \text{ cm}. \quad (1.28)$$

Low  $g_\chi$  values with pre-dominant co-scattering scenarios predict nominal decay lengths of  $\mathcal{O}(m)$ . The decay length scales roughly with the inverse square of the coupling parameter. From annihilation scenarios, with larger coupling in the dark sector, the decay length is shorter. The lab frame decay length can be obtained from the nominal decay length by boosting the distance with Lorentz boost of the charged dark partner, see Eq. 1.29.

$$d = (\beta\gamma) c\tau_c, \quad (1.29)$$

Due to the exponential decay probability, the number of dark partners decaying within a sphere of radius  $d$  around the production point is given by

$$N(d) = N(0) \int_0^d \frac{dr}{d} \exp\left(-\frac{r}{d}\right), \quad (1.30)$$

where  $N(0)$  is the number of dark states produced at the collision point.

The transverse impact parameter,  $d_0$ , parameterises the displacement of the daughter lepton

from the decay of the charged dark partner.  $d_0$  is the closest approach of the lepton track to the collision point, see Fig. 1.8.

The range of nominal decay lengths  $c\tau_c$  that can be probed within this range of  $d_0$  depends on the overall boost of the dark states, which determines the decay length  $d$  (see Eq. (1.29)), and on the transverse component of the boost. For  $m_c < 500$  GeV, most particles are produced with boosts within  $0.2 < (\beta\gamma) < 5$ , with a peak around  $(\beta\gamma) \approx 1$ . Highly boosted, *i.e.*, light-dark states tend to be emitted along the beamline, leading to a smaller transverse decay length  $d_0 < d$ . In a typical scenario with  $m_c = 220$  GeV and  $\Delta m = 20$  GeV, the dark partner has a transverse momentum of  $p_T(\chi^\pm) \approx 100$  GeV and the decay lepton carries  $p_T(\ell) \approx 10$  GeV. In this case the observable range of  $d_0$  from Eq. (1.29) roughly corresponds to

$$1 \text{ mm} \lesssim c\tau_c \lesssim 20 \text{ cm}. \quad (1.31)$$

This is within the fiducial acceptance of the compact muon solenoid (CMS) detector around the LHC as will be discussed in Chapter 2. Lower values of  $m_c$  for the charged dark states will result in boosted lepton with low  $d_0$ , outside the fiducial acceptance of the CMS detector. The CMS detector is well suited to probe co-scattering compressed dark matter, as well as co-annihilation with sufficiently small dark matter couplings provided the trigger threshold on transverse momenta of leptons is lowered.

# Chapter 2

## The Compact Muon Solenoid Experiment

*The LHC and the experiments surrounding it, including the CMS experiment, are the success story of a global effort in high energy physics. I will discuss the structure of the LHC and the CMS detector in this chapter, with an emphasis on the general triggering mechanism followed at the CMS experiment. New triggers built to target novel physics phase-space are an integral part of this thesis.*

Technological advancements in particle detectors and particle colliders made a crucial contribution to the in-depth understanding of the Universe. Electrons were discovered in 1897 by J.J. Thomson [98] in cathode ray tubes. Subsequently, protons and neutrons were discovered by scattering radiation from radioactive materials incident on fixed targets [99,100]. Further discoveries were achieved with cyclotrons and subsequently synchrotrons [101] enabling high energy particle collisions at energy scales larger than 1 GeV. The particle detection technology saw similar improvements from cloud chambers [102] and scintillators to composite structures with layers of materials dedicated to enable efficient reconstruction of the energy-momentum of sub-atomic particles. [103,104].

The LHC [2] runs at an unprecedented centre-of-momentum (c.o.m.) energy of 13.6 TeV and collides proton-proton ( $p-p$ ) bunches around 40 million times in a second. A single collision creates hundreds to thousands of particles [105], whose interaction in the detector materials varies. The charged leptons<sup>1</sup> interact electromagnetically, while the hadrons also interact through the strong force. Particles like muons are stable in the length scale of the detectors, while particles like the neutral pion decay instantaneously. All neutrinos pass through the detector without any substantial interaction. Detecting and studying the particles that are created during these collisions requires sophisticated detectors. The CMS experiment [106] uses a multi-layered detector structure with each layer designed to detect a specific aspect of the particles in every  $p-p$  collision. The information from multiple sub detectors is combined to reconstruct the particles that resulted from the  $p-p$  collision.

The data generated by the LHC is selectively stored using the trigger mechanism, allowing the usage of available computing resources efficiently. A downside to triggering is that the discarded

---

<sup>1</sup>Conventionally, I refer to the particles with electric charge as charged particles. Particles which feel the strong force effect are called coloured. Charges under the influence of other forces will be explicitly mentioned.

data of a rejected  $p$ - $p$  collision is irreversible. Hence, triggering becomes a critical aspect of the CMS experiment.

In the first section, I will discuss the governing principles of the LHC and its design parameters. In the second section, the CMS experiment and its sub systems are briefly described. In the third and final section, we look at the new run parameters of the LHC as it restarted the  $p$ - $p$  collisions in 2022.

## 2.1 The Large Hadron Collider (LHC)

The Large Hadron Collider (LHC) is a synchrotron-based accelerator [107–109]. It is a 26.7 km long ring of super-conducting dipole magnets cooled down to  $-271.3$  °C. It is situated in a cavern 50 to 170 m below the ground level. It accelerates two charged beams of protons or heavy ions moving in opposite directions before colliding them at specific desired points, called interaction points, in the accelerator where the physics experiments are located. The dipole magnets reach a maximum field strength of 8.3 T to bend the particle beams. A vacuum of  $10^{-10}$  mbar is maintained in the beam pipe to minimise energy loss from multiple scattering.

Colliding high-energy beams create an intense quark-gluon plasma that resembles the Universe’s early stages after it expanded in the inflationary period. Under the LHC energies, massive particle can be created by exciting fields with larger mass parameters. These massive particles are extremely short-lived and subsequently decay. The properties of the massive particles can be measured from the relatively stable sub-atomic constituents of its decay process. The decay constituents are detected by the experiments located around the LHC. It is possible to produce and thus measure the properties of fundamental sub-atomic particles at the LHC.

The LHC undergoes periodic phases of collisions and shutdowns<sup>2</sup>. The shutdowns are necessary to maintain, upgrade and test the LHC accelerator components as well as the detector systems of the experiments. The collision phases, during which data is taken, are called runs. At the start of my doctoral degree studies in September 2018, the Run 2 of the LHC was ending. It was an ideal moment to improve the CMS experiment’s physics prospects for the next LHC run by designing and adding triggers to search for exotic physics processes. The Run 3 of the LHC began in mid-July 2022 and is scheduled to continue until 2025.

### 2.1.1 The CERN accelerator complex

The LHC is a part of the accelerator complex of Conseil Européen pour la Recherche Nucléaire (CERN) at Meyrin, Switzerland. CERN has been one of the most important centres of sub atomic physics research since its foundation in 1954. It houses powerful accelerators, each with a rich history of physics studies, which have subsequently become pre-accelerators to more powerful accelerators (see Fig. 2.1) at the end of their lifetime.

Synchrotrons cannot accelerate a particle from rest [101]. Depending on the radius of the synchrotron, one needs to pre-accelerate the particles to specific energies before injecting them into a synchrotron for further acceleration. The accelerator complex at CERN enables functionality for larger synchrotrons with higher c.o.m. energy by using the older and smaller synchrotrons as feeders in an ascending chain. The maximum energy to which a synchrotron can accelerate

---

<sup>2</sup>The schedule is online and can be found at <http://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.htm>

a particle depends on the bending power of its dipole magnets and the energy loss of the beam while going through the synchrotron.

Ionising hydrogen atoms from a hydrogen gas cylinder creates protons. The protons are accelerated in the LINAC 4. About  $10^{11}$  protons are focused together spatially to form a proton bunch. Proton bunches are accelerated sequentially in the Booster, the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) before injecting them in opposite directions at a fixed timed interval of 25 ns to form two oppositely travelling proton beams which are accelerated out of phase in the LHC. The LHC was designed to accelerate the particles to energies of the order of TeV. When the two oppositely moving beams collide, the resulting c.o.m. of the collision,  $\sqrt{s}$ , is twice the energy of the individual beams. As of its current operational status in 2023, the LHC has achieved a c.o.m. energy of  $\sqrt{s} = 13.6$  TeV for  $p$ - $p$  collisions. The LHC is also designed to accelerate ions from heavy elements, e.g. Xe and Pb, which lie beyond the scope of this thesis.

The proton bunches are accelerated and compressed in radio-frequency (RF) electric resonance cavities situated in a linear sector of the LHC. As the energy of the proton beams increases by acceleration in the RF cavities, the strength of the bending magnets is gradually increased to contain the beams in the beam pipe. Quadrupole magnets are used to focus the beams and direct them to the collision points at the experimental sites. Liquid helium is used to bring down the temperature to 1.9 K to create super-conducting operation conditions.

The LHC provides reach to high-energy scales. This enables fundamental physics research in multiple areas including studying the properties of the Higgs boson to a high precision. The physics program of the LHC is diverse. CMS and ATLAS [111] are the general-purpose physics detectors situated diametrically opposite each other around the LHC. They have similar physics goals and they are meant to cross-check each others results. It includes studying the SM and searching for processes beyond the SM. A Large Ion Collider Experiment (ALICE) [112] studies the quark-gluon plasma and the physics of strongly interacting matter using heavy-ion collisions. Finally, the LHC beauty (LHCb) [113] experiment studies the CP violation and rare decays of the B hadrons. Besides these four main detectors, other smaller detectors are installed to study exotic physics, elastic scattering, neutrino physics and also to search for physics beyond the SM in a complementary phase space to that of CMS and ATLAS [114–119].

I illustrate the concept of the LHC design and introduce some essential variables that govern the dynamics of the physical processes at the LHC.

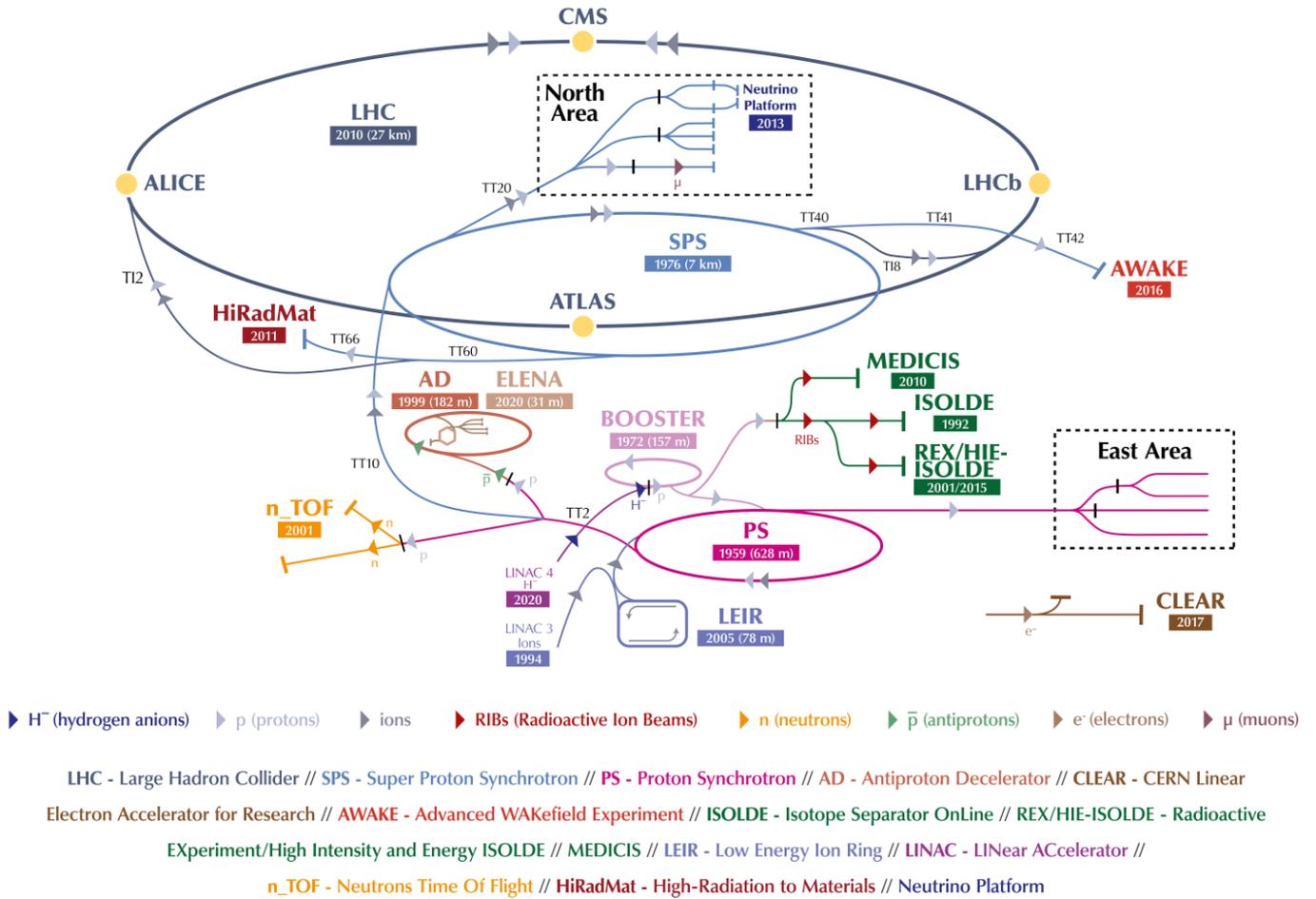


Figure 2.1: The CERN accelerator complex is shown. It contains some of the state-of-the-art facilities of particular interest to HEP physicists. The injection sequence to the LHC (dark blue) starts with the LINAC 4 (purple), followed by acceleration in the BOOSTER (light pink), the PS (dark pink) and the SPS (light blue). [110]

### 2.1.2 The LHC parameters

LHC was designed to study high-energy physical processes with goals to discover the BEH boson, dark matter and supersymmetry. Discovering a rare physics process or studying the corresponding physics process more precisely depends on two design factors - the  $\sqrt{s}$  and the luminosity.

The cross-section of a process quantifies the probability for an interaction to happen. It has the same dimension as area. However, the value for the cross-section of a process is too small to be described in SI units of area. The unit **barn** (b), see Eq. 2.1, and its multiples are chosen to measure the cross-section of an interaction.

$$1\text{b} = 10^{-28}\text{m}^2 \quad (2.1)$$

The cross-section of a physics process depends on the energy of the colliding partons which increases with  $\sqrt{s}$  [120]. Typically, the cross-section of a rare process increases with an increase in  $\sqrt{s}$  [121]. However, the diameter of the LHC tunnel and the strength of the dipole magnets constrain the maximum energy achievable for a beam of particles. If the protons are accelerated beyond a certain maximum energy, the bending strength of the magnets becomes too small to contain the protons in the ring. On the other hand, energy losses due to synchrotron radiation limit the maximum energy achievable for the electrons. The maximum energy of the particle beam limits the c.o.m and consequently the production cross-section of rare processes at the LHC.

The production of a rare physics process can be further enhanced by increasing the number of  $p$ - $p$  collisions per unit of time, quantified by the instantaneous luminosity ( $L$ ). It is the ratio of the total number of events to the production cross-section of a process per unit of time and depends on the design parameters of the colliding proton bunches. Assuming a Gaussian spatial distribution profile for the proton bunches in the proton beams,

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

where  $N_b$  is the number of protons per bunch,  $n_b$  is the number of bunches per beam,  $f_{rev}$  is the revolution frequency of the beam,  $\gamma_r$  is the relativistic Lorentz gamma factor,  $\epsilon_n$  is the normalised transverse beam emittance,  $\beta^*$  is the amplitude function of the beam at the interaction point and  $F$  is the luminosity reduction factor due to the crossing angle at the interaction point. The luminosity is expressed as number of events per unit area per unit time. In the CGS units, this is  $\text{cm}^{-2}\text{s}^{-1}$  and it can be expressed as  $\text{barn}^{-1}\text{s}^{-1}$ . The LHC was designed to operate at a peak luminosity of  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$  for  $p$ - $p$  collisions which has already been exceeded during the Run 2 and the Run 3 programme. The integrated luminosity is the integral of the instantaneous luminosity over time.

Each  $p$ - $p$  bunch crossing is called an event. An event consists of multiple proton pairs getting scattered due to the extended size of the bunch. The phenomenon of having multiple simultaneous scattered proton pairs is called pile-up. Most physics analysis choose the proton pair with the largest c.o.m. of parton scattering as the collision of interest. Pile-up encompasses all other elastically and inelastically scattered proton pairs in the same event. The LHC is designed to produce an event every at 25 ns to create an event rate of 40 MHz. Only a small fraction of the hard scattered events are stored by the trigger for physics analysis.

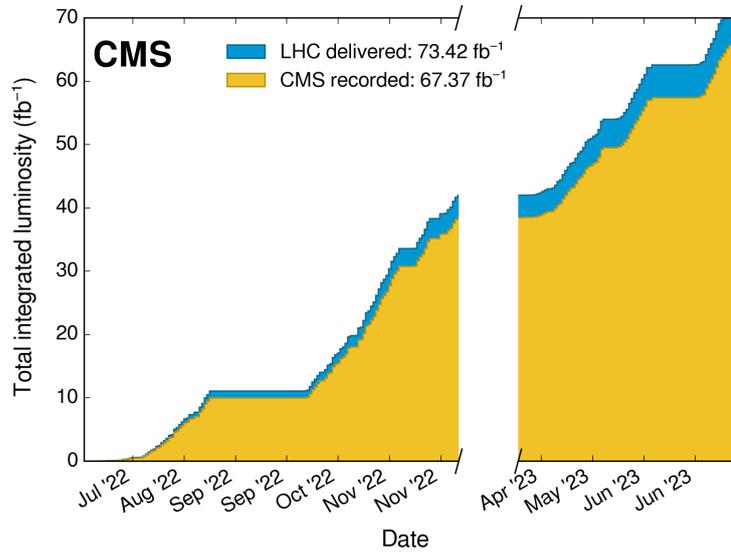


Figure 2.2: The CMS experiment recorded a total of  $67.37 \text{ fb}^{-1}$  of data in 2022 and 2023 at peak instantaneous luminosity of  $2.22 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1} \approx 22.2 \text{ nb}^{-1}\text{s}^{-1}$ , greater than twice the design luminosity of the LHC. [122]

## 2.2 Performance parameters of the LHC in Run 3

The LHC started the Run 3 phase of its program mid-July 2022. It delivered  $p$ - $p$  collisions at a  $\sqrt{s} = 13.6 \text{ TeV}$  until the first week of July 2023. Figure 2.2 shows the integrated luminosity delivered by the LHC and the luminosity recorded at the CMS experiment. There were periods of shutdown from August to October 2022 and December 2022 to March 2023. The LHC delivered an integrated luminosity of  $73.42 \text{ fb}^{-1}$  at  $\sqrt{s} = 13.6 \text{ TeV}$  with a peak instantaneous luminosity of  $2.22 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1} \approx 22.2 \text{ nb}^{-1}\text{s}^{-1}$ . A total luminosity of  $67.37 \text{ fb}^{-1}$  was certified by CMS as good data. An average PU of 46 and 52 simultaneous collisions were observed in 2022 and 2023 respectively, see Fig. 2.3.

## 2.3 The Compact Muon Solenoid experiment

The Compact Muon Solenoid [106], see Fig. 2.4, is a general-purpose physics detector with a cylindrical geometry at the Point 5 site of the LHC. It is 21.6 m in length, with a diameter of 14.6 m and weighs 14 kilo-tons.

A single collision event results in the emission of hundreds to thousands of sub-atomic particles from the collision vertex. The CMS detector has been designed to identify and measure the transverse momentum and the energy of each individual particle in an event in the detector fiducial region to a high precision and with high efficiency. It is designed to contain all of the particles radiating from the collision, a phenomenon known as hermetic sealing. Only muons and neutrinos escape the detector confinement due to a low probability of interaction and low loss of energy from multiple scattering respectively. The transverse measurement and the energy

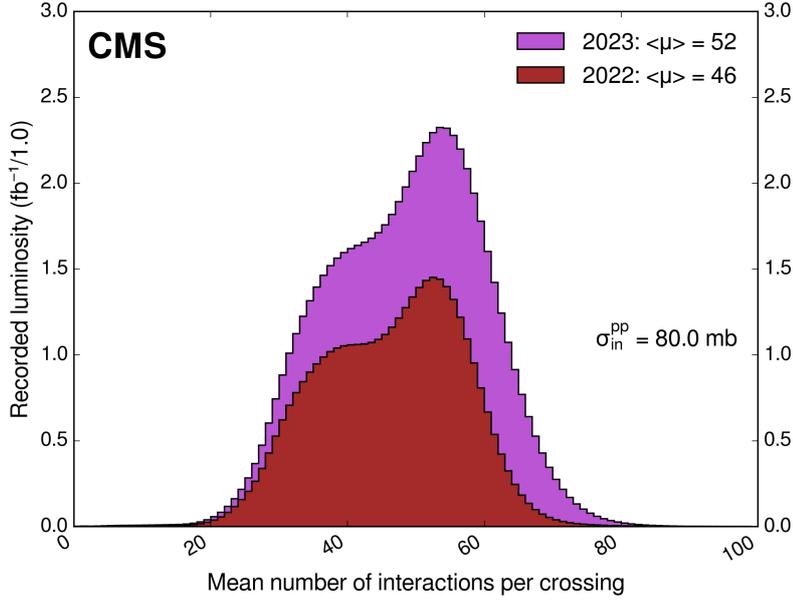


Figure 2.3: The pile-up profile of the 2022 and 2023  $p$ - $p$  collisions. [122]

of the neutrinos is obtained in terms of imbalance in the transverse momentum of the measured particles using momentum conservation in the transverse plane.

The particles have different mechanisms of interacting with the detector material. The electron and photon interact electromagnetically. Most mesons and baryons predominantly interact with the material via scattering with nuclei. The muons have a low interaction cross-section and lose little energy through multiple scattering and the neutrinos pass through matter with barely any interaction.

The CMS detector geometry uses cylindrical shaped layers of different materials which envelope the collision vertex. Going radially outward from the  $p$ - $p$  collision vertex and the beam pipe are the silicon tracker, the electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL) and the solenoid magnet. The solenoid magnet is a central feature of the CMS detector with a field strength of 3.8 T and results in a compact structure. Outside the solenoid magnet, the muon chambers detect the muons coming from the  $p$ - $p$  collisions. Finally, the CMS detector is sealed by the forward calorimeters close to the beam pipe.

The hermetic coverage and the excellent energy momentum resolution for particles enable the search for DM particles. If the dark sector couples to the SM particles, the production of the dark sector particles is possible through  $p$ - $p$  hard scattering processes. For the STHDM scenario, the electrically charged dark partner,  $\chi^+$ , could be produced at the hard scattering vertex. It decays to  $\chi^0$ , resulting in a charged lepton and a neutrino through the off-shell EW process. The  $\chi^0$  remains undetected in the detector. Due to the hermetic coverage of the detector, this results in a detectable imbalance in the transverse momentum of the leptons. With an efficient detection of the leptons, the kinematics of the produced DM particle can be accurately measured.

The CMS experiment uses a cylindrical right-handed coordinate system with the  $z$ -axis aligned with the beam pipe. The origin is at the interaction point midway between the beam

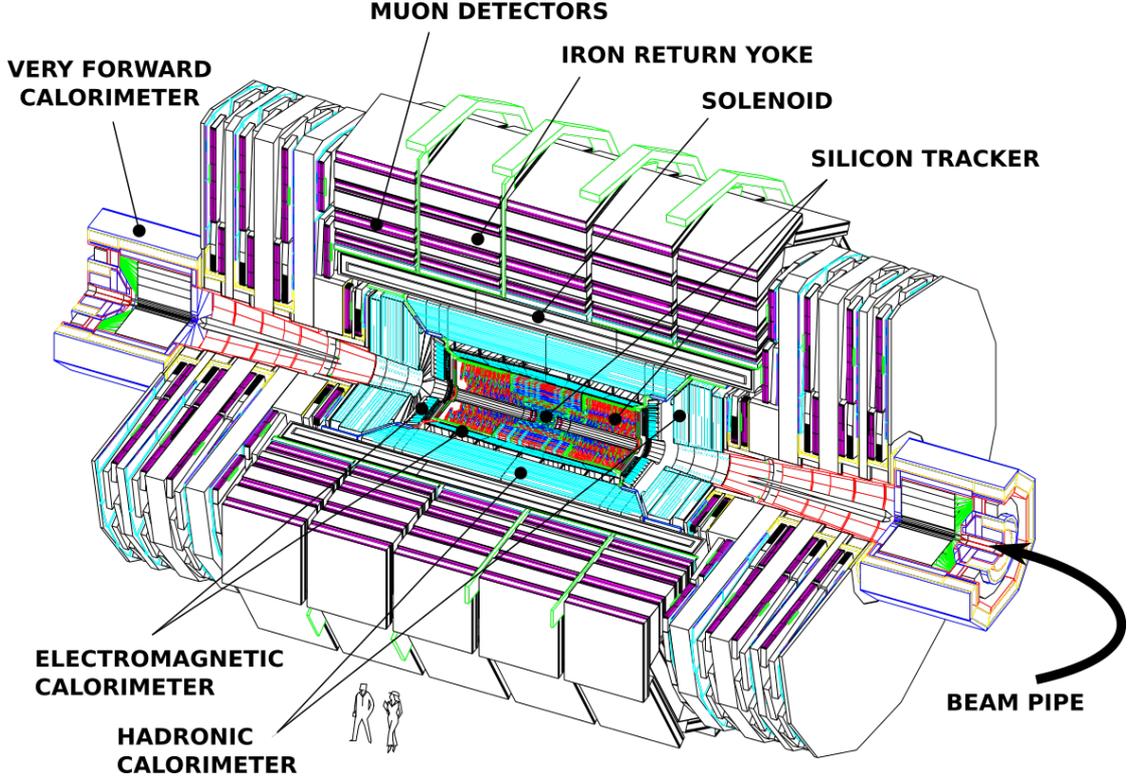


Figure 2.4: The Compact Muon Solenoid detector schematics with some critical sub components shown. The detector is hermetically sealed allowing for the entrance of the LHC beam pipe to carry in the proton bunches for collisions at the origin of the detector. Going radially outward from the origin are the silicon tracker, the calorimeters, the solenoid magnet and the muon detector system. A very-forward calorimeter seals the detector in the high- $\eta$  region. Figure is scaled and compared to the size of an average human. [123]

pipes. The  $y$ -axis points towards the sky and the  $x$ -axis points towards the centre of the LHC quasi-ring. The azimuthal angle  $\phi$  measures the inclination with respect to the  $x$ -axis. The radial vector is measured in the transverse  $x - y$  plane. The pseudo-rapidity,  $\eta$ , measures the inclination with respect to the  $z$ -axis, defined in Eq. 2.3 as a function of the polar angle.

$$\eta = -\ln \left[ \tan \left( \frac{\theta}{2} \right) \right] \quad (2.3)$$

The value of  $\eta$  is zero for a perpendicular  $\theta$  and goes to infinity in positive and negative values as  $\theta$  approaches 0 and 180° respectively. Low  $|\eta|$  values, usually  $|\eta| < 2$ , are called the central region, whereas high  $|\eta|$  values are called the forward region of the detector<sup>3</sup>. The differential count of particles emitted in a  $p$ - $p$  collision is constant as a function of the pseudo-rapidity. The differences in  $\eta$  are Lorentz invariant.

<sup>3</sup>The definition of the central and forward region is not strict. It is often used in relative scenarios when referring to a scenario of similar interest.

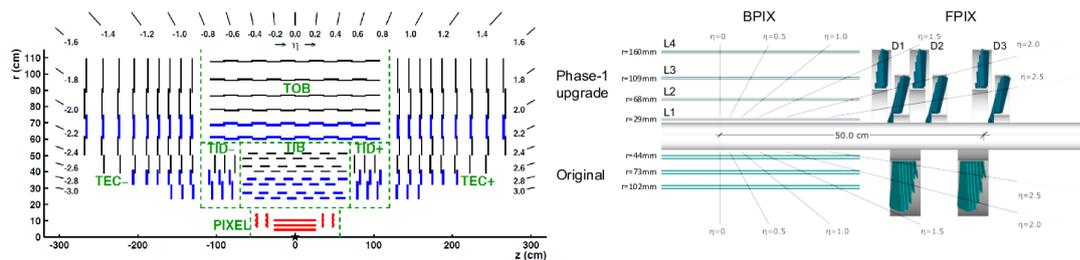


Figure 2.5: CMS tracker schema taken from [128]. The six major sections of the tracker are shown. The red section is built with pixel detectors. It was upgraded from three to four layers in BPix and from two to three layers in FPix. The blue modules are built with two layers of strip detectors and the rest have modules with a single strip layer.

### 2.3.1 Silicon tracker

Right around the beam pipe of the LHC, the CMS tracker detector [106, 124–126] is placed. The CMS tracker is built using doped silicon. The exact nature of the material used in the tracker varies depending on its radial distance from the collision vertex.

The CMS tracker is based on the principle that charged particles passing through a material interact electromagnetically and locally ionise the medium [127]. An electric field applied across the material separates the ions and a net current is observed in the material from the effect of ionising radiation passing through the material. The current is digitised to a numerical value. Above a certain a noise threshold, the net current is identified as a hit from the passage of ionising radiation or it is dismissed as random noise. Hits from multiple layers are combined, using dedicated algorithms described in the next chapter, to reconstruct the path taken by the charged particle in the detector. Neutral particles interact only through nuclear interaction. They pass through the tracker almost without any hits since the cross-section for a nuclear interaction is relatively low compared to the electromagnetic interaction.

The CMS tracker is the closest detector to the beam pipe. It faces the highest flux of particles out of all other CMS components. It is built to be radiation hard, cost-effective and with a high resolution to distinguish between the large number of charged particles that pass through it.

The Run 3 CMS tracker features cylindrical barrels in the central region and annular disks in the forward region. The inner region close to the beam pipe consists of pixelated silicon sensors. This includes the four barrel (BPix) cylinders of length 54 cm at mean radii of 2.9, 6.8, 10.9 and 16 cm and twelve end-cap disks (FPix) organised in three layers with two disks per layers on each side of BPix. The FPix layers are positioned at  $z = \pm 29.1, \pm 39.6$  and 51.6 cm. The inner disk extends radially between 45 and 110 cm and the outer disk from 96 to 161 cm. Strip-based silicon sensors occupy the region with radius between 20 cm to 116 cm. There are four distinct components - the Tracker Inner Barrel (TIB), the Tracker Inner Disks (TID), the Tracker Outer Barrel (TOB) and the Tracker End-Caps (TEC). The TIB is composed of four barrel layers extending to 55 cm in radius augmented by the three TID layers. The two layers closer to the beam pipe in the TIB and TID sections have two hits-sensitive silicon strip layers each. The

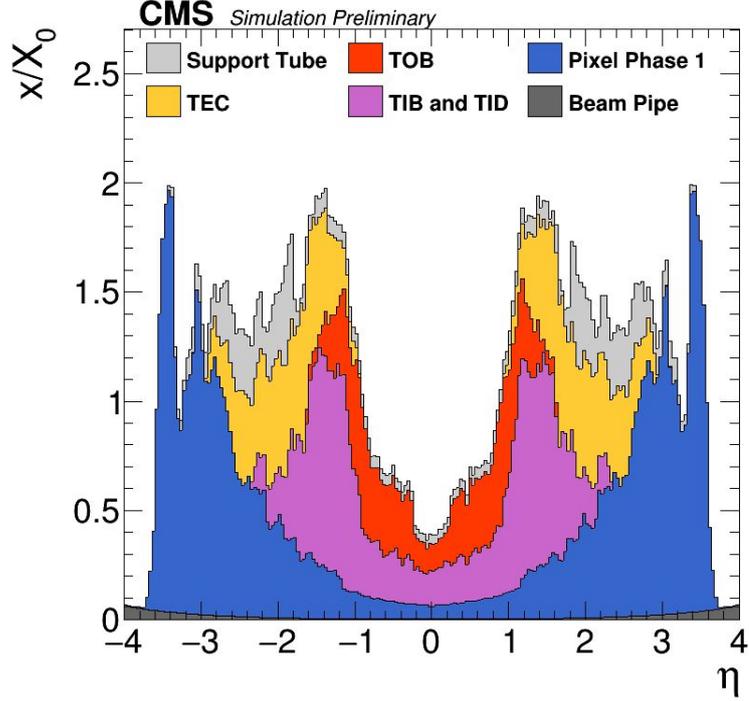


Figure 2.6: Material Budget of the CMS tracker showing contribution from individual sub-detectors in terms of effective radiation length as function of pseudorapidity. [129]

TOB has six more layers with a length of 236 cm. Beyond  $z = \pm 118$  cm, the TEC covers the region  $124 \text{ cm} < |z| < 282 \text{ cm}$  and  $22.5 \text{ cm} < |r| < 113.5 \text{ cm}$ . Each TEC is composed of 9 disks, carrying up to 7 rings of silicon strip detectors to provide high-resolution tracker coverage in the range  $|\eta| < 2.5$ . Under the extreme radiation environment of CMS, the tracker is operated at a temperature of  $-15^\circ\text{C}$  to reduce noise in electronics from thermal fluctuations. Figure 2.5 shows the CMS tracker layout in place since the beginning of the LHC Run 2 and a phase-1 upgrade between 2016 and 2017 that upgraded the pixel layer to take advantage of the reduction in beam pipe radius. The strip layers were made with a decade-long use and will be replaced for the upcoming high-luminosity LHC (HL-LHC) upgrade at the start of 2026.

The tracker is designed to provide a hit resolution of  $\mathcal{O}(0.01 \text{ mm})$  while the tracker alignment during installation is of  $\mathcal{O}(0.1 \text{ mm})$ , thus necessitating derivation of alignment corrections at the beginning of every LHC run. The resolution of tracking from the pixel and strip sensors is reduced due to multiple scattering in the support structure for the tracker. The support structures include the power supply cables, the readout mechanism, the electronics cooling mechanism and the mechanical frame. Hadronic interactions of pions and other hadrons in this material reduces the tracking efficiency for these hadrons. Photon conversions and bremsstrahlung radiation from electrons reduces the measurement accuracy for photons and electrons in the electromagnetic calorimeter. These irreducible effects have been quantified by measuring the effective radiation length of the material as a function of  $\eta$ , see Fig. 2.6, referred to as the material budget of the tracker.

### 2.3.2 Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) is a homogeneous calorimeter surrounding the tracker. Its primary goal is to absorb and provide an accurate energy estimation for the predominantly electromagnetically interacting particles, electrons/positrons ( $e$ ) and photons ( $\gamma$ ).

The CMS ECAL [106, 130, 131] is built using lead tungstate ( $\text{PbWO}_4$ ) crystals. An electron or a photon going through the  $\text{PbWO}_4$  crystals creates an electromagnetic (EM) shower through a cascade of pair production, bremsstrahlung and Compton scattering. [127, 132] The cascade is terminated at the critical energy to form an electron-positron pair, at which point the photon is absorbed by the crystal. The absorbed energy is released as scintillation light. The light yield from the scintillation is sent through digital photo multipliers which amplify and digitise the analogue signal. The digitised value is proportional to the energy of the particle that initiated the electromagnetic shower in the crystal.

The radiation length and Moliere radius characterise the longitudinal and the transverse shape of the EM shower for a given material medium [132]. The high density of the  $\text{PbWO}_4$  crystals result in a short radiation length and Moliere radius which allows for a compact and yet highly granular ECAL.

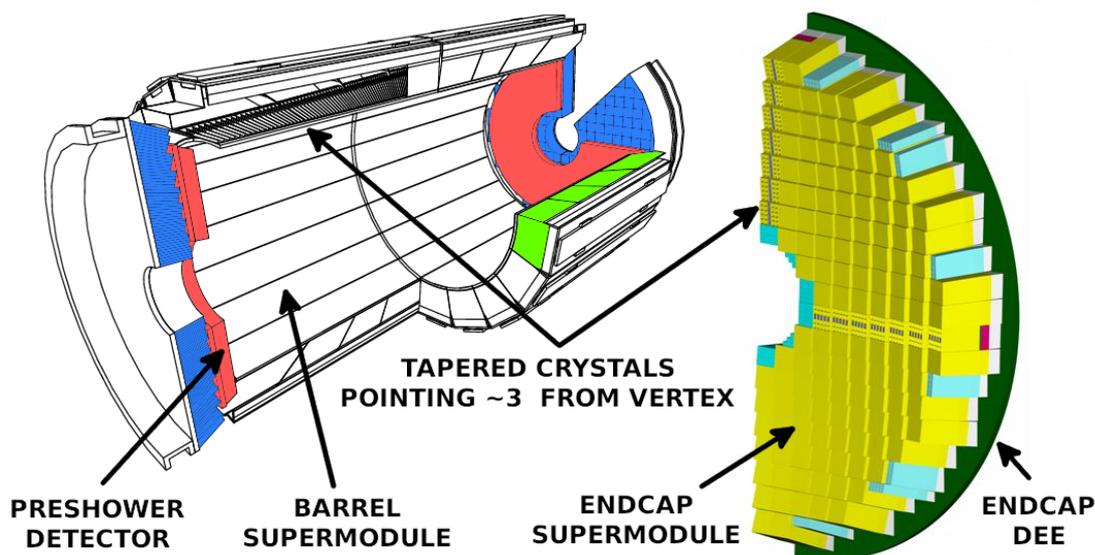


Figure 2.7: The schematics of the CMS electromagnetic calorimeter shows the placement of individual crystals. They point  $\sim 3^\circ$  off-centre to prevent the leakage of photons produced from  $p$ - $p$  collisions through the crystal gaps. [133]

The CMS ECAL, see Fig. 2.7, is structured in three sections - the barrel (EB) and two end-caps ( $EE^\pm$ ). The crystals are arranged in a cylindrical shape around the tracker closed by two disk-shaped crystal arrangements on both sides of the cylinder. The EB has 61 200 crystals placed in a grid of  $170 \times 360$  along  $\eta \times \phi$  in  $|\eta| < 1.479$  at a radial distance of 129 cm. To avoid particles leaking through gaps between the crystals, the central axis of the crystals points  $3^\circ$  off the origin. The  $EE^\pm$  covers the pseudo-rapidity range  $1.479 < |\eta| < 3.0$  and starts at  $z = \pm 315.4$  cm. Each EE has 7324 crystals in two halves arranged in the transverse plane with the shape of

an annular disk. The crystals point 130 cm beyond the origin to avoid  $e/\gamma$  leaking between the gaps.

The crystals in the CMS ECAL are trapezoidal in shape. Their length is between 22 and 23 cm. The face of the trapezoidal crystal with the shorter area pointing towards the collision vertex. Its size varies between  $22 \times 22 \text{ mm}^2$  and  $28.6 \times 28.6 \text{ mm}^2$ . The face with the larger surface area points radially outward with size varying between  $26 \times 26 \text{ mm}^2$  and  $30 \times 30 \text{ mm}^2$ .

Pre-shower (ES) detectors are annular disks placed in front of the EE disks to increase the effective radiation length in the region  $1.653 < |\eta| < 2.6$  and allow for separation of a single high energy photon from two angularly-close placed photons, *e.g.* from a high-energy  $\pi^0$  decay. Each ES is a 20 cm thick sampling calorimeter made from two layers of alternating lead absorbers and silicon sensors. The silicon sensors measure the energy deposit and transverse shower profile. The pre-shower adds about three radiation lengths of material in front of the EE.

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \quad (2.4)$$

The energy resolution of the CMS ECAL is given by Eq. 2.4, where  $S$  is a stochastic term due to fluctuations in the count of photo-electrons,  $N$  is due to noise in the ECAL electronics and  $C$  summarises any constant effect on the resolution of ECAL *e.g.* calibration effects. It is crucial to have low value for  $C$  as this will dominate the uncertainty at high energies. The energy resolution [134] has been measured by an electron test beam to be

$$\left(\frac{\sigma_E}{E[\text{GeV}]}\right)^2 = \left(\frac{0.3\%}{\sqrt{E[\text{GeV}]}}\right)^2 + \left(\frac{12\%}{E[\text{GeV}]}\right)^2 + (0.3\%)^2. \quad (2.5)$$

### 2.3.3 Hadronic calorimeter

A significant fraction of the particles coming from  $p$ - $p$  collisions consists of quark composites, *e.g.* baryons and mesons. They interact mainly via the strong force with the nuclei resulting in a complicated and highly stochastic chain of interactions. The interaction length,  $\lambda_{\text{int}}$ , parameterises the longitudinal nature of the hadronic shower shape.

The CMS hadronic calorimeter (HCAL) [106,135] encapsulates the ECAL and is placed inside the solenoid. It is designed to stop all particles in an event except for muons and neutrinos. It enables the measurement of the energy of all particles in an event. The radial extent of the solenoid (3.15 m) and the outer boundary (1.77 m) of the ECAL, combined with a minimum required  $\lambda_{\text{int}}$  to constrain the hadronic showers, pose geometrical design constraints. Given the high luminosity of the LHC and the high-energy particles that have to be stopped entirely in the HCAL, it needs to be radiation hard. A sampling calorimeter with fibre readout is the chosen design.

Figure 2.8 shows the components of the HCAL. The barrel (HB) covers the central region,  $|\eta| < 1.3$ . Alternating brass absorber layers are used with plastic scintillator tiles in the HB. At  $\eta = 0$ , the HB has the equivalent of  $5.82 \lambda_{\text{int}}$  on top of  $1.1 \lambda_{\text{int}}$  from materials in the ECAL. It increases to  $10 \lambda_{\text{int}}$  at  $|\eta| = 1.3$ . Further coverage in pseudo-rapidity is achieved using the HCAL end-cap (HE) and the forward calorimeter (HF). The material in the central region of the HB is not sufficient enough to contain hadronic showers, which might start later in the material. The HCAL is extended outside the solenoid in the HB angular region by adding scintillator layers to the first two layers of the muon chambers to ensure adequate sampling depth in the barrel region. The scintillator layers in the muon chambers form the outer hadronic calorimeter (HO).

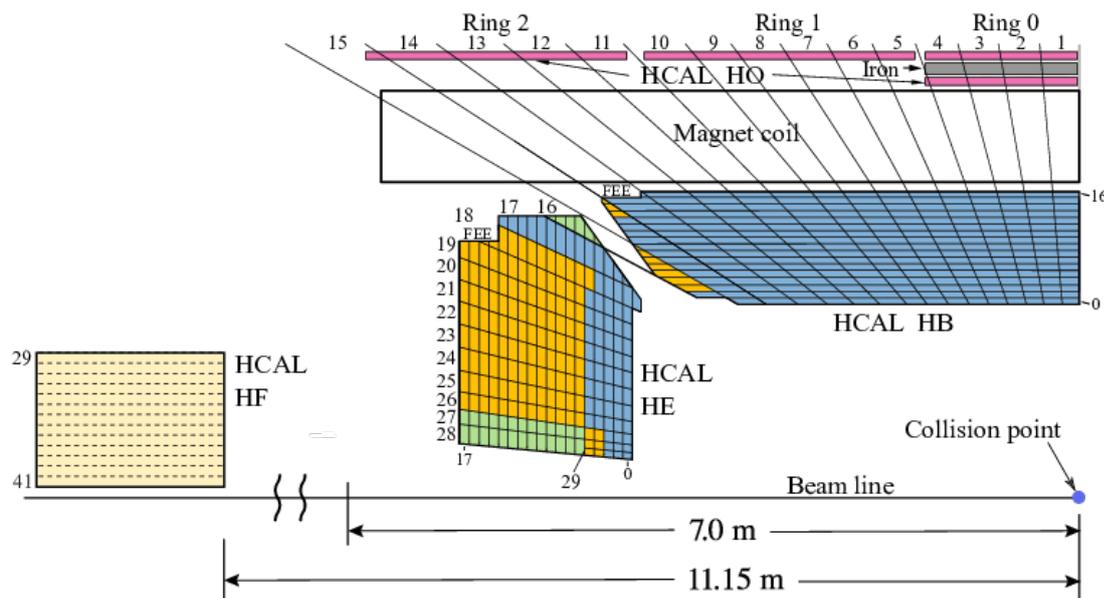


Figure 2.8: The components of HCAL in a quadrant of CMS show the relative placements of the HCAL components around the solenoid. [136]

In addition to the solenoid being used as an absorber material, the HO also benefits from an iron return yoke that contains the solenoid's magnetic field in the outer region.

The HE covers the region  $1.26 < |\eta| < 3$ . The HE absorbers are made from non-magnetic brass to prevent structural damage from the strong magnetic field. Moreover, the HE will also have a significantly high radiation flux, so quartz fibres are used for readout.

The HF extends the coverage of HCAL to  $|\eta| < 5.2$  and ensures that the HCAL is hermetic. The HF is located in a high radiation zone and, on average, receives more than seven times the energy than the rest of the detector combined per event. The active material is thus made entirely from quartz fibres interlaced into a steel absorber structure. HF is situated immediately next to the beam pipe at 11.2 m from the collision vertex. Its inner radius is 12.5 cm, outer radius is 1.4 m and depth is 1.65 m. The HF also monitors the luminosity of the  $p$ - $p$  collisions in the detector.

The HCAL is designed to contain all the particles produced in the  $p$ - $p$  collision. Combined with the ECAL, it provides energy measurements for the particles in an event. Combined with the position measurement from the tracker and the muon system, energy conservation can be applied in the transverse plane to estimate the missing transverse energy in an event. The missing transverse energy provides an estimate of the energy and the momentum of the SM neutrinos that might have been produced in the event. [137] Dark matter candidates, such as the  $\chi_0$  in STHDM can also provide a detector signature with missing transverse energy. [83] Chapter 3 discusses the various physics objects that can be reconstructed as a result of the energy deposit signature in the HCAL in combination with signatures from the other parts of the detector.

### 2.3.4 Solenoid magnet

Charged particles follow a helical path under the influence of a magnetic field [106]. The path of a charged particle can be obtained by combining hits in the tracker, as briefly conceptualised in Section 2.3.1 and described in more detail in the upcoming Section 3.2.1. A helical path enables us to measure the charge to momentum ratio of the charged particle.

Most of the particles of interest in the CMS detector have energies of  $\mathcal{O}(10 - 1000 \text{ GeV})$ . For such high-energy particles, a powerful magnetic field is required to achieve a sufficient sagitta in the path that can be measured under the resolution constraints of the detector. The strength of the magnet directly affects the size of the detector. If the detector has a larger fiducial region for the path of the charged particle, the sagitta is larger. Alternatively, you can reduce the size of the detector by using a stronger magnetic field that causes a larger bend in the path of the charged particle.

The solenoid magnet [106, 138], therefore, determines the size of the CMS detector. It is 12.9 m long along the  $z$ -axis with a diameter of 6.3 m and radial thickness of 31.2 cm. The CMS solenoid is cooled to 4 K to create superconducting operation conditions that generates a uniform field of 3.8 T inside the solenoid. A magnet return yoke made from iron, see Fig. 2.4, was placed around the solenoid to confine its magnetic field. As a result, the magnetic field outside the solenoid is inhomogeneous. The strength of the magnet enables a compact size for the CMS detector. A charged particle of momentum 1000 GeV obtains a sagitta of 0.15 mm in a path length of 1 m under the influence of the CMS solenoid.

### 2.3.5 Muon system

The muons are central to the physics program of the CMS collaboration. The leptons provide a natural window to study electroweak processes precisely. Dark sectors such as the STHDM that couple to SM particles through the EW sector provide easily verifiable signatures with leptons. In the lepton family, the muons provide the cleanest detector signature. They do not suffer from as strong energy loss as electrons from EM interactions and are stable enough to travel long interaction lengths unlike the taus. They can pass through calorimeters which stop almost all other SM particles and yet provide an easy to detect signature owing to EM interactions. The CMS has  $> 99\%$  efficiency in detecting muons with resolution of 1 – 2% in the barrel and  $< 10\%$  in the end-cap for muons with transverse momentum  $< 1 \text{ TeV}$  [139].

The CMS muon system [106, 141] is placed in the gaps of the iron return yoke of the CMS magnet. Fig. 2.9 shows a quadrant of the CMS muon detector along with the other components of the CMS detector. The muon chambers cover a large fiducial space in the region  $|\eta| < 2.4$  and  $\pm 5.6 \text{ m} < z < \pm 10 \text{ m}$ . They extend radially from 4 to 7.5 m. Gas-ionisation based particle detectors were chosen for muon detection to reduce the building cost. The barrel (MB) covers the region  $|\eta| < 1.2$  and is built using drift tubes (DT). Two end-caps (ME) on either side of the MB cover the region  $0.9 < |\eta| < 2.4$ . Since the ME is located in a highly irradiated environment with a higher flux of muons, cathode strip chambers (CSCs) are used. They provide a faster detection time, finer segmentation and higher radiation resistance [142]. A few unavoidable mechanical gaps exist to allow supply lines to the internal components of the CMS detector.

The DT and CSC systems are augmented with resistive plate chambers (RPCs) all over the muon system. The RPCs have lower uncertainties in measuring the beam crossing time of the LHC [143]. Correlations with the beam crossing time are essential to reduce background event rates at large luminosities. A single plane of RPCs was added to each multi-layered DTs and CSCs in the region  $\eta < 1.6$ . At the end of Run 2 of the LHC, gaseous electron multiplier

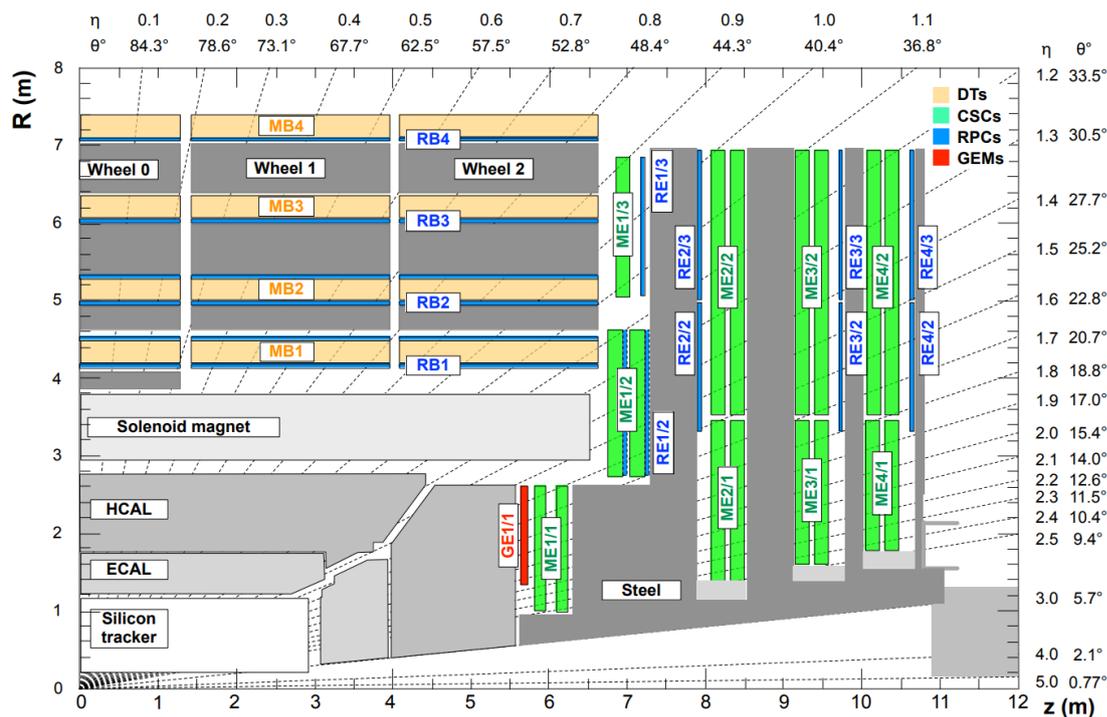


Figure 2.9: The position of the CMS Muon system between the iron yoke of the magnet is shown in comparison with the other components of the CMS detector. The red GEM-based muon detectors were added in the recent upgrade during LS2. [140]

(GEM) detectors were added in  $1.6 < \eta < 2.2$  [140]. The GEM detectors have a high radiation resistance [144]. Certain forward muon regions of the CMS will be upgraded with GEMs in preparation for the upcoming HL-LHC.

### 2.3.6 Triggering and data acquisition

All the information from the detectors is digitised and fed through electronic streams to the data acquisition systems where they are packed to a binary ROOT file [145]. The large amount of data generated at the CMS detector causes technical challenges to its output and storage. To achieve a high luminosity,  $p$ - $p$  collisions occur every 25 ns [2]. With a collision frequency of 40 MHz and an average event size of 1 MB/event [146], 4 TB/s of data is generated by the CMS detector. The data output rate exceeds practical limitations. Hundreds of exabytes of data are created in a year of LHC operation at the CMS experiment. Maintaining an infrastructure that stores and analyses such large amounts of data would be too expensive. The trigger system employs algorithms, developed by CMS physicists, that decide if an event is of relevance for analysis conducted by the CMS collaboration. It reduces the data output rate and keeps a high efficiency for the physics reach of the experiment.

The high luminosity reach of LHC is necessary to enable the production of rare physics processes. Figure 2.10 shows the production cross-sections of fundamental physics cross-sections at the LHC [121]. At the designed luminosity of the LHC, despite 40 million collisions every second, only about eight events of  $t\bar{t}$  and 20 events of  $Z \rightarrow e\bar{e}$  take place per second. The

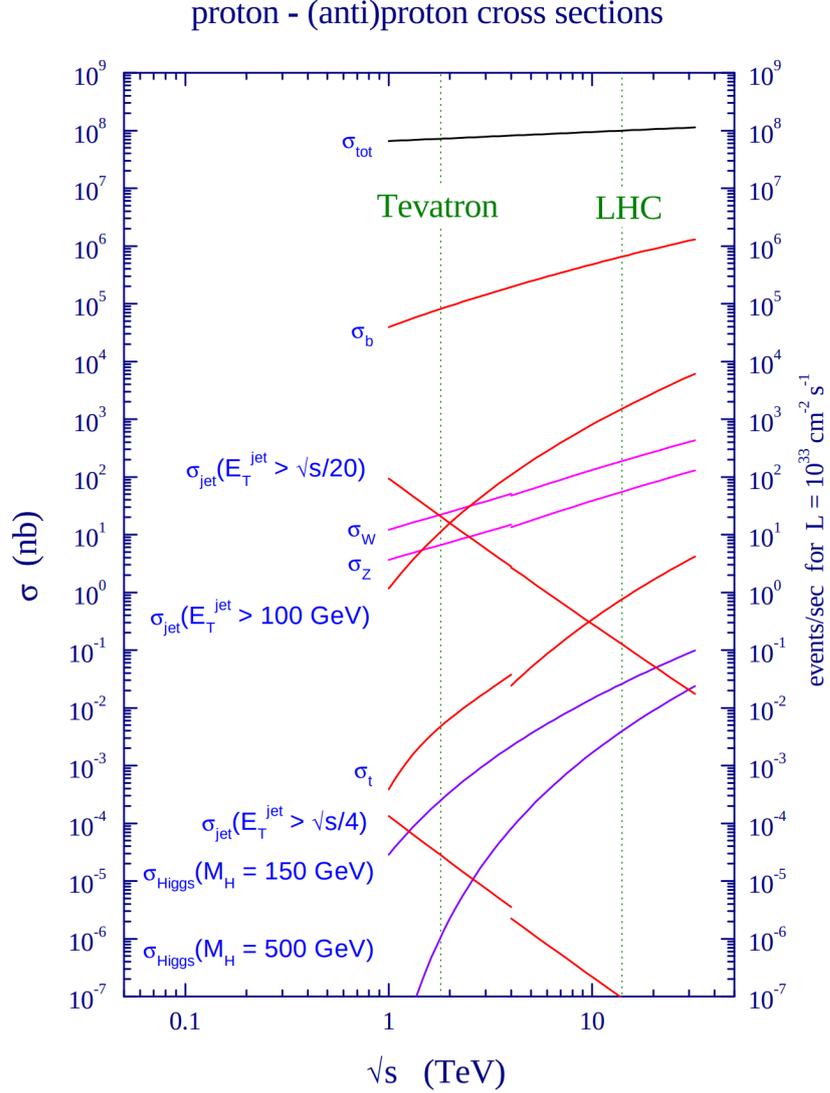


Figure 2.10: The cross-section of fundamental physics processes in a  $p$ - $p$  collision, typically increases with an increase in the centre-of-momentum energy of the collision. The plot is taken from the era before discovery of the BEH boson. [121]

BEH boson is produced only once every 2 seconds. The trigger system enables the selection of these rare events with maximum possible efficiency within the resource constraints of the CMS experiment. It is flexible to accommodate the changing interest of the experimental analysis groups. For the LHC Run 3, multiple new triggers were added to increase the physics reach of the CMS experiment to be sensitive to exotic physics and DM scenarios. I successfully included three triggers targeting the STHDM model.

The CMS trigger system [106, 147–150] reduces the event rate from 40 MHz to  $\sim 1$  kHz. It has been designed to retain events of physics interest with high efficiency. Events are selected

by imposing a certain reconstruction quality on a reconstructed physics object, e.g. a muon or electron. This is elaborated more in Chapter 3. In some cases, event selections, defined as correlations between the physics objects in the same event, are considered as well. Object selections increase the purity of the stored data and reduce the occurrence of events with mis-reconstructed objects in the data. In comparison, event selection will require a physics motivation that correlates objects in the detector phase space. The events selected with good object selection can be later used to study multiple physics cases compared to events with dedicated event selection which will be limited by the correlation requirement between the individual objects. Using object selections makes the trigger system inclusive such that the events retain significance to future physics analysis. The selections on physics objects are only made on attributes that are less affected by detector calibrations.

The trigger collects data to fulfil multiple objectives. In addition to physics goals, a part of the data is used for monitoring the data collected by the CMS experiment during operation. Some other objectives include offline trigger efficiency measurement, trigger studies in changed circumstances and physics object calibrations.

The CMS event triggering is done in two steps, the Level-1 (L1) trigger and the High-Level Trigger (HLT), respectively. The L1 trigger uses custom-designed electronic boards to reduce the event rate from 40 MHz to 100 kHz. Events that pass through the L1 trigger are sent from the detector to the Data Acquisition (DAQ) system. The HLT is part of the DAQ. It is a software algorithm implemented on commercial processing units that makes the final decision to keep the events. If the HLT accepts an event, it is sent to a temporary storage at the experimental site. The HLT further reduces the rate to  $\sim 1$  kHz.

Data is transferred from the experimental site to permanent storage in tapes at the main CERN computing site. A copy of the data is made available to the physicists through the Worldwide LHC Computing Grid (WLCG). The WLCG is structured to four tiers - Tier 0, 1, 2 and 3. Tier 0 stores the unprocessed detector file. This format of the data is called the RAW file format. Tier 0 performs a prompt reconstruction on the collected data to create the analysis-level (AOD) data format. Tier 1 copies the RAW and re-processes the data with new and improved reconstruction techniques on a periodical basis. Tier 2 computing resources are used by physicists to analyse the data. The Inter-University Institute for High Energies (IIHE) at Brussels, Belgium houses a Tier 2 facility for the CMS experiment. The Tier 3 resources are a relatively recent addition of local computing resources and cloud services that are accessed individually and are indirectly related to WLCG.

### The Level-1 Trigger

The L1 trigger [147, 149] is implemented using integrated circuits (ASICs), programmable lookup tables and Field Programmable Gate Arrays (FPGAs). The technological implementation of the L1 trigger is a balance between making fast decisions and necessary logical flexibility. Radiation-hard technology is used for the L1 trigger in the endcaps because of the high radiation flux. The FPGAs are used, where possible, to allow for flexibility.

The L1 trigger is split between the muon system and the calorimeters. The decision time for the L1 trigger is constrained by the requirement that it has to analyse all of the 40 million  $p$ - $p$  bunch crossings every second. The tracker does not contribute to the L1 trigger to reduce the time to make a decision. The L1 trigger system uses information from the detector sub-systems with a lower resolution to use a lesser number of readout channels. While the L1 trigger system makes a decision, the data is stored in memory pipelines in the front-end electronics detector. A total latency of  $3.2 \mu\text{s}$  is allowed for the L1 trigger to make a decision and convey it to the

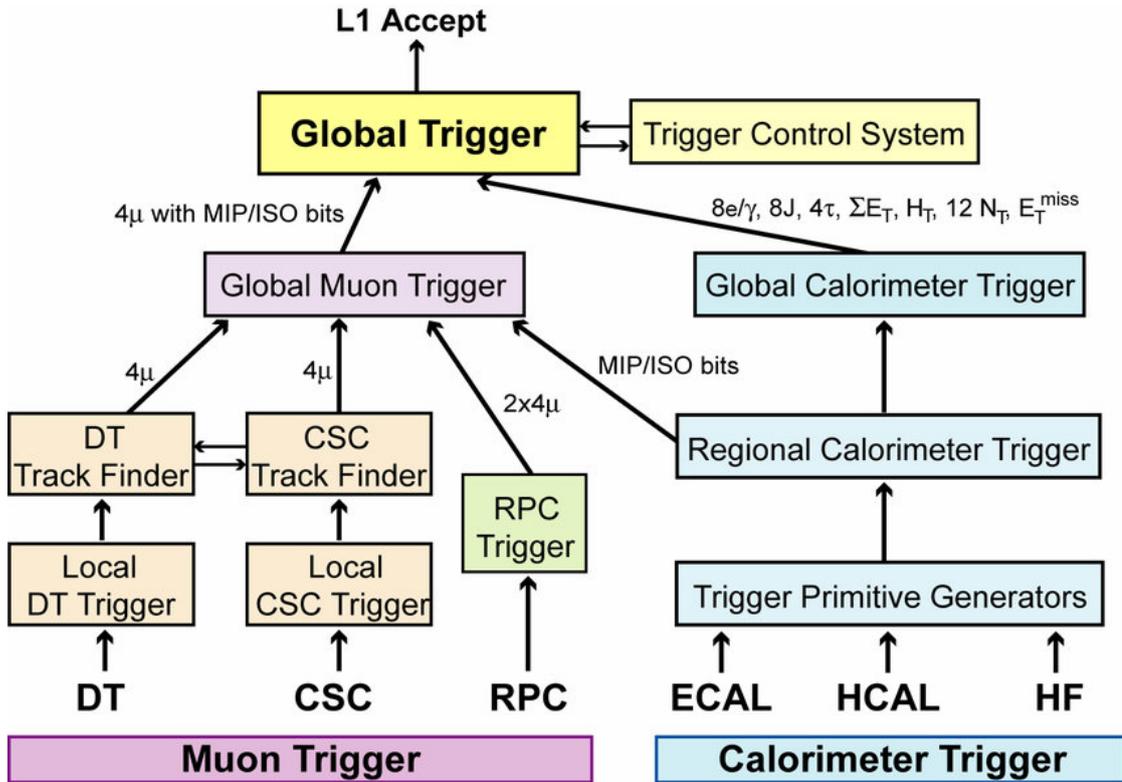


Figure 2.11: Architecture of the L1 trigger shows the information flow from local to regional to global trigger systems. The L1 Global trigger makes a decision based on the global trigger primitives. [149]

detector front-end electronics.

Figure 2.11 shows the two branches of the L1 trigger. Each of the muon and the calorimeter branches has three levels of data processing - the local, the regional and the global levels. The local component uses the digitised detector data to generate hits and patterns in the muon system and group the energy deposits in the calorimeters to create simple physics objects for a muon, an electron, a photon, a tau, a jet or missing energy. These are called trigger primitives. The regional component ranks and sorts the trigger primitives. The ranking is done based on their energy and a quality factor that quantifies the confidence in reconstructing the primitives. The global component groups data from multiple regions and selects a handful of highly ranked primitives. It passes them onto the L1 global trigger, which combines the muon and calorimeter objects to make a decision to accept or reject an event.

Based on its algorithm, the L1 trigger sends information about four muons, eight electrons or photons, four taus and other information about the missing energy in an event.

### Data Acquisition

If the L1 trigger decides to accept an event, the data stored in detector front-end memory pipelines is pushed by a read-out system to the DAQ [148]. The schematic for the DAQ is shown in Fig. 2.12. The event builder assembles all the information related to the same L1 trigger to form a complete event. There are eight nearly independent slices that form the DAQ. Each

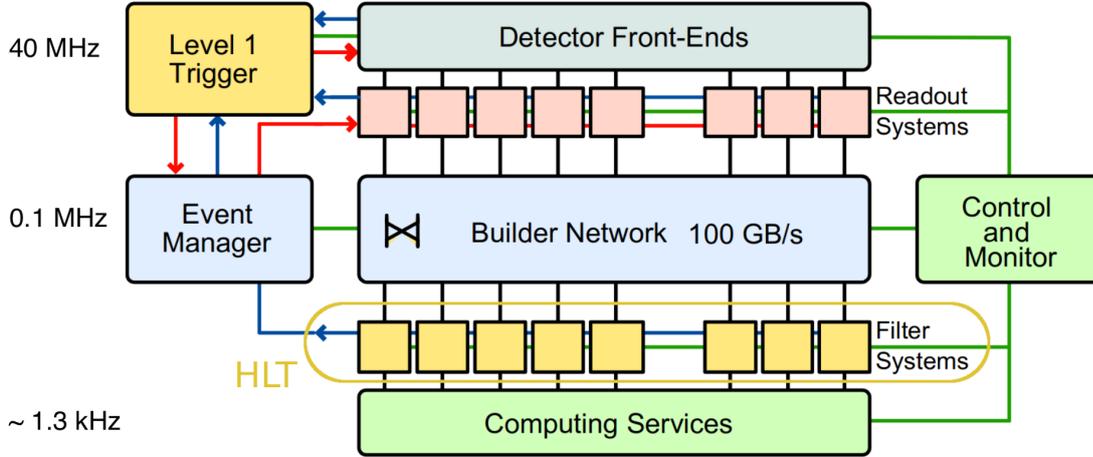


Figure 2.12: On acceptance by the L1 trigger, data is readout from the detector pipelines to the DAQ system which builds an event and sends required information to the HLT. The full data is sent to permanent storage for offline study when an HLT path accepts the event. [148]

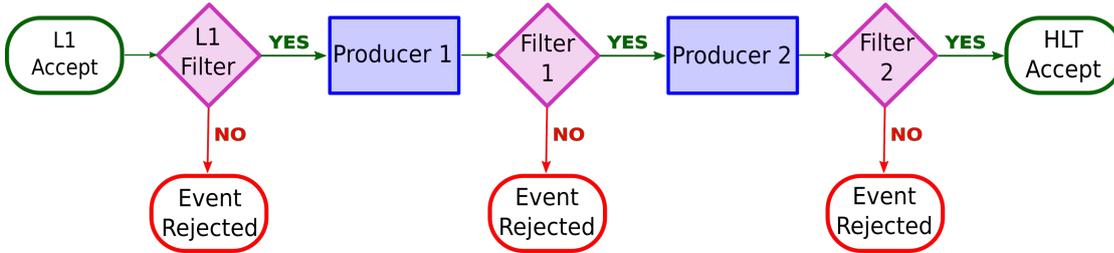


Figure 2.13: The basic building block of an HLT path consists of sequential producers and filters. The producers are shared between the HLT paths. Each HLT path is seeded by a combination of L1 paths.

of them is able to handle a 12.5 kHz event rate. The event builder sends the data from the underground detector to the surface. The trigger settings are not changed for a single luminosity (LS), where a LS is defined as the time period equivalent to  $2^{20}$  LHC orbits. If the data output rate is too high, the prescale setting for the L1 trigger is changed which reduces the event count by discarding events depending on the prescale factor. It accepts one event every  $P$  events where the prescale factor of  $P$  is defined as

$$P = \frac{\text{Event count allowed to DAQ}}{\text{Event count accepted by the L1 trigger}} \quad (2.6)$$

The full event information is passed onto the HLT for a decision to transfer it to permanent storage. The two-step trigger mechanism allows for a larger latency to make the decision on whether an event should be stored permanently for a physics analysis.

### The High-Level Trigger

The CMS HLT [150] runs a fast event analysis software. Full event information is available to be used for the HLT decision. The HLT runs a lightweight version of the software that is

used for offline physics analysis. The algorithm is run in steps of producers and filters. The producers use the full detector data to reconstruct physics objects such as an electron or a muon. Chapter 3 elaborates on some of the reconstruction methods used. The filters apply selections on the reconstructed physics object to either reject or pass an event to the next stage of the producer and filter.

Figure 2.13 shows the schematics for the HLT event rejection in stages. Objects that require data from a higher number of readout channels and hence require more combinatorics result in CPU-intensive algorithms. The reconstruction of such objects is kept for later stages. Traditionally, this follows the same principle as the L1 trigger. The data from the muon system and the calorimeters are used to reconstruct physics objects in the earlier stages, while the tracker data is only used for object reconstruction in the later stages. This approach results in a high CPU efficiency for the event rejection rate.

The HLT makes use of the full granularity of the detector. Specific L1 trigger algorithms seed all the HLT algorithms. Additional time reduction is obtained for the event selection at the HLT by performing object reconstruction in the geometrical vicinity of the L1 seed only. This performs partial reconstruction of the data instead of reconstructing data for the full detector.

The HLT runs on a hybrid computer farm of 200 nodes with each node equipped with two AMD Milan 64-core CPUs and two NVIDIA Tesla T4 GPUs<sup>4</sup>. The hybridisation of the computer farm was done in Run 3 in preparation for the future high luminosity LHC [151]. The event output rate at the HLT is  $\sim 1$  kHz. It takes  $\sim 200$  ms on average to make a decision on whether an event should be stored or not.

The events are split into multiple data files depending on the specific HLT algorithm that it was triggered by. The different triggers at the HLT can be grouped together on the basis of the physics object they filter. Some triggers filter the kinematical properties of multiple physics objects. These are called cross-triggers. The output from a certain group of triggers is connected to a specific output data stream. This results in different datasets. The name of these datasets such as `SingleMuon`, and `SingleElectron` is self-explanatory. Two specific modes of data output called Scouting [152, 153] and Parking [153] have been developed in addition to the regular data streams. Parking collects a large section of the data to be analysed on the empty cores available during long shut-down phases of the LHC. Scouting aims to collect events with reduced event size based on the event reconstruction done at the HLT. I shall discuss scouting in detail and the advantages it offers in this thesis.

---

<sup>4</sup><https://cms.cern/news/first-collisions-reconstructed-gpus-cms>

# Chapter 3

## Simulation and Physics Object Reconstruction

*The CMS detector is a complicated experiment. Many tools and techniques are required to extract meaningful information from the detector data. This chapter contains three broad sub-topics. I discuss the proton collision simulation, followed by reconstruction techniques that convert the detector data to meaningful physics information in the CMS detector. Lastly, a brief discussion about additional tools relevant to the research in this thesis is done.*

The high-energy  $p$ - $p$  collisions at the LHC are an extremely complicated environment that needs to be understood. Additionally, the CMS detector consists of multiple sub-components that must operate with extreme accuracy. The key for successful simulation and reconstruction of protons collisions is the precise measurement of the signature of the particles in the detector combined with the precise understanding of the physics processes. A detailed comprehension of the SM physics processes at the LHC is imperative to observe any new BSM physical processes. Consequently, theoretical understanding is driven to a higher accuracy from new experimental data and novel experimental techniques.

The CMS collaboration makes use of several numerical and analytical tools to ensure a detailed and accurate simulation of the proton collisions in the detector. The scattering matrix elements provide the cross-section of the physical interactions that probabilistically occur in the proton collisions. Usually, few partons participate in the hard-scattering event. The remnant partons of the proton undergo low-energy (soft) scattering, radiative decays and recombination, which are simulated with dedicated tools. The final result is passed through a detector simulation software to understand the signature produced in each CMS detector layer.

High-energy physicists pioneer the reconstruction of meaningful information from the signal observed in the detectors from proton collisions. Individual fundamental particles are reconstructed from the detector response. Electrons, muons, photons and hadrons are reconstructed as high-level physics objects in the CMS experiment.

Simulating the different components in the CMS experiment is covered in the first section of this chapter. The second section of this chapter describes an overview of the algorithms used to reconstruct the physics objects in the detector. Special attention has been paid to the detailed reconstruction of some physics objects at the HLT that are relevant for this thesis.

### 3.1 Event Simulation

The LHC is built for particle discovery. It collides protons which can achieve high collision energies because of low synchrotron radiation energy loss [154]. However, protons are composite particles. Simulating hard scattering processes in composite particles is a non-trivial task due to a fundamentally incomplete understanding of QCD at low-energy scales. Assumptions about asymptotic-freedom for running the strong coupling constant to high-energy scales is necessary to enable a perturbative evaluation for the QCD Lagrangian. Numerical methods relying on Monte-Carlo (MC) sampling techniques [155, 156] are extensively used in simulating proton collisions at the LHC.

Event simulations to understand the collision of proton beams at the LHC has undergone extensive theoretical studies and software development [157, 158]. The simulation of a  $p$ - $p$  collision comprises multiple stages. The probability distribution of the energy of partons as a fraction of the proton momentum is obtained from the parton distribution function (PDF) of the proton. The fraction of momentum of the proton carried by a parton is measured by the Bjorken- $x$ . Asymptotic-freedom based perturbative calculations provide a reliable estimate to next-to-leading-order (NLO) accuracy for the hard-scattered partons in a proton. The other remaining partons of the hard-scattered proton and other protons in the beam undergo radiative processes and soft scattering. The coloured partons from hard-scattering, soft-scattering and radiative processes undergo hadronisation to form colour-neutral stable or metastable particles. These particles are passed through a detector simulation to mimic the signals observed in the detector as an effect of radiation passing through matter. The resulting detector signals pass through identical reconstruction software as the data from observed proton collisions to study the signature of the physics processes in the detector. Due to the complexity of the whole process, the simulation is usually done in multiple stages with a combination of different software.

Physics processes from proton collisions essentially follow the form  $p p \rightarrow X$ . Under LHC conditions, most collisions are inelastic and the partons in the sub-structure of the proton participate in the interaction. Understanding the parton dynamics at all energy scales relies on an accurate theory of QCD, which has not been achieved yet. As a function of the Bjorken- $x$ , and the momentum exchange between two partons  $Q^2$ , the probability to find a parton  $i$  is given by the PDF,  $f_i(x, Q^2)$ . The PDF evaluation at a required energy scale is obtained by the using the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations on the PDF at another (usually lower) energy scale [159–161]. The input PDF to the DGLAP equations is an experimental input from deep inelastic scattering experiments. The proton PDF is maintained and updated by dedicated collaborations [162, 163].

$$\sigma_{pp \rightarrow X} = \sum_{i,j} \int dx_i \int dx_j f_i(x_i, Q^2) f_j(x_j, Q^2) \hat{\sigma}_{ij \rightarrow X} \quad (3.1)$$

The process rate at which  $p p \rightarrow X$  happens is given by the total cross-section of the process in Eq. 3.1. The total cross-section is the partonic cross-section  $\hat{\sigma}_{ij \rightarrow X}$  multiplied by the probability of obtaining partons  $i$  and  $j$  with momentum fraction  $x_i$  and  $x_j$  and exchange energy  $Q^2$  from the two protons integrated over all possible momentum fractions and summed over all possible initial state configurations that gives rise to the final state  $X$ . Multiple initial state partons can give rise to similar final state, e.g for the well-know DY process  $p p \rightarrow \mu^+ \mu^-$ , the initial state partons will include a quark anti-quark pair. Contributions from all but the top quark need to be summed over. Each initial quark anti-quark pair will also contribute a different partonic cross-section in Eq. 3.1. The parton combinations,  $u\bar{u}$  and  $d\bar{d}$ , dominate the DY production

process at the LHC.

The partonic cross-section  $\hat{\sigma}_{ij \rightarrow X}$  is calculated from computing the scattering matrix element factorised between the different initial states and integrated over the corresponding phase space. A clear distinction needs to be made between the initial states described here and in the total cross-section. In the partonic cross-section, the different initial states are separated by their helicity, nature of interaction ( $s$ -channel or  $t$ -channel) and not flavour. While the total cross-section considers the flavour of the leptons involved. Software such as `MadGraph` and `Mg5_aMC@NLO` used in this thesis provide automated computation for the matrix elements [164, 165]. They require an input physics model which specifies the particles, force couplings, and interactions to be used. The SM is available as an input in the software. New BSM models, e.g. STHDM, are generated using the `Feynrules` package, which computes the Lagrangian of the model and packs the relevant information of the model in a standard Universal Feynrules Output (UFO) file format [166, 167].

The partons involved in the hard scattering, as well as other partons in the proton, radiate particles and can decay as well. Either the quarks can emit gluons or the gluons can decay into a quark anti-quark pair. These processes happen in conjunction with the hard scattering. If they happen before the hard scattering, they are called initial state radiation (ISR); if after, they are called final state radiation (FSR). Together they form the parton shower and initiate the evolution of the proton collision. Calculating the cross-section for each radiation is a numerically expensive procedure and is substituted with alternative approaches. The software `Pythia` simulates the parton shower and is one of the multiple available tools for usage [168]. The parton shower is simulated up to the energy scale  $\Lambda_{\text{QCD}} = 300$  MeV where the perturbative description of QCD is still valid [169].

The evolution state of the simulated  $p$ - $p$  collision after the parton showering still consists of coloured particles. The collision must evolve further to combine the particles into colour-neutral states. Since the perturbative description of QCD no longer applies, the simulation for forming colour-neutral hadrons, called hadronisation, relies on phenomenological models. The two main types of phenomenological models used are the Lund string model and the cluster model [171–173], with `Pythia` using the former.

The hard scattering partons are not the only interaction possible in an LHC proton beam collision. The ISR and FSR can give rise to substantial activity in the event. Moreover, on average, at least one soft scattering is observed from other protons in the beam, along with the hard scattering. These result in additional event activity and, thus, additional signals in the CMS detector. The signatures from the additional parton interactions in the hard-scattered proton is called the underlying event while the signature from other protons in the event are collectively referred to as the pile-up (PU). The underlying event and PU consist of multiple components and are modelled from data [174].

A single collision event thus results in hundreds to thousands of particles being emitted from the collision vertex into the detector. The `Geant4` software models the detector effects by simulating the passage of radiation through matter [175–177]. A detailed geometry of the CMS detector is modelled with all the components, wires, and support structures. The modelling includes the details of the materials used to construct every component. The effect of the magnetic field is included in the simulation with numerically solved field values for each point in the detector. `Geant4` tracks the particles outward from the collision vertex through the CMS detector’s active modules and non-active supporting structures. The signal in the active components is digitised, considering the observed inefficiencies in the detector consistent with measured data. Figure 3.1 summarises the entire process of the  $p$ - $p$  collision simulation.

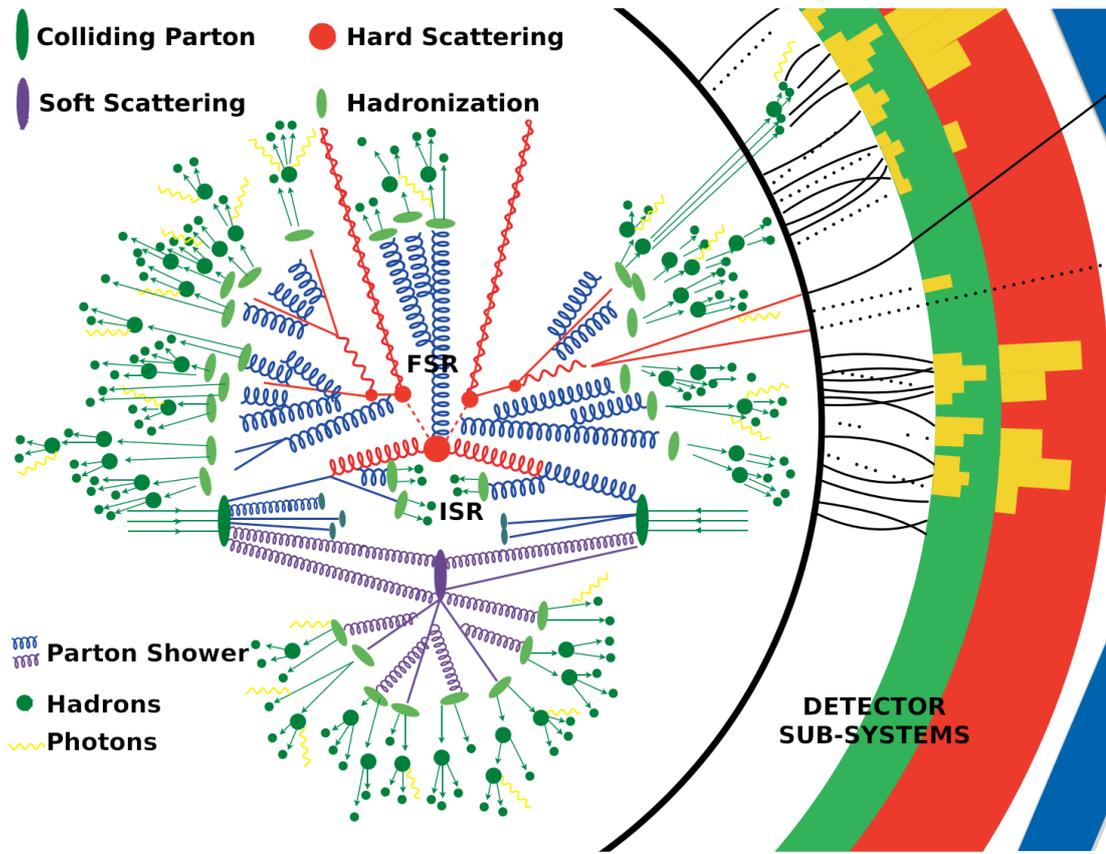


Figure 3.1: An overview of proton (green big ellipse) collision simulation at the LHC. The partons undergo radiative processes (ISR and FSR, shown in blue) alongside the hard scattering (shown in red). Soft scattering between other partons in the proton (violet) also occurs. The parton undergo hadronisation (light green small ellipse) after showering to form colour-less hadrons (green circle). The hadrons can decay or radiate photons (yellow). The final state particles are simulated through the detector layers and modules. Figure taken from [170].

### 3.2 Reconstruction of Physics Objects

The particles produced from the LHC proton collisions pass through the CMS detector and produce analogue signals, which are subsequently digitised by the detector electronics. The data output streams transfer the digitised data from the detector to permanent storage on an accept signal from the HLT. To study physics with this so-called detector raw data, the initial particles that give rise to the detector signals are reconstructed from the raw data to form physics objects. The physics object reconstruction has multiple facets. The reconstruction algorithm that is run on the detector data stored for analysis is called offline reconstruction and the resulting physics objects are called offline physics objects. A simplified and computationally light version of the offline reconstruction algorithm is run by the HLT called online reconstruction. The corresponding physics objects are called online physics objects.

There are multiple levels of reconstruction. The most basic level of reconstruction clusters

energy deposits in the calorimeters and builds tracks in the tracker and the muon system. The higher level of physics objects fit for analysis are the electrons, photons, muons, jets from partons, hadrons and taus and the missing transverse momentum ( $\cancel{p}_T$ ) in the detector. This thesis is mainly concerned with electrons, photons and muons in the HLT, so special emphasis has been paid to the online reconstruction.

The online reconstruction follows a different rule than the offline-based one with respect to the time complexity of a reconstruction algorithm. It reads information selectively from specific sub-detector modules based on the level-1 (L1) trigger input. The algorithm with the fastest reconstruction is prioritised. The time of running an algorithm is often related to the number of data readout channels input to the algorithm. The higher the number of readout channels used, the slower the algorithm because it usually involves multiple iteration over the readout channels.

### 3.2.1 Tracking

The innermost part of CMS is the silicon-based charged particle tracker [106]. Tracks identifying charged particles passing the CMS tracker can be reconstructed efficiently [178]. Charged particles traversing matter ionise the material. They produced charge clusters that can be spatially located with high precision in the CMS tracker through its highly granulated structure of pixels and strips. To segregate spurious noise from an actual charged particle passing through the tracker, a *hit* coming from a charged particle is defined only when it crosses a specific charge deposition threshold. The threshold for suppressing spurious charge deposition varies significantly depending on the local region of the tracker. On average, the tracker expects an  $\mathcal{O}(1000)$  of charged particles at the LHC luminosity. Clustering the hits to make tracks of all these particles in the detector is challenging and time-consuming and requires repeated iterations on the tracker hits.

A combinatorial track finder (CTF) approach is used. It consists of multiple iterations of four basic steps built with the Kalman Filter [179] as the base. The selections are loosened after every successive iteration. A track has to fit into the helical structure under the influence of the solenoid magnetic field. The four steps in every iteration are described in the following paragraphs.

The first step combines three hits from the collection of pixels and strips to fit the track parameters. Good quality tracks are iteratively used to identify the primary vertices. The reconstructed  $p$ - $p$  vertex is in return used by tracks to fit and update the track parameters. Track reconstruction in the tracker pixel is a looped process as the estimate of the reconstructed  $p$ - $p$  vertex, discussed in Section 3.2.2, is done using hit information in the inner pixel tracker and the first step in tracking uses the primary vertices as an input to track building. Usually, the approach starts from the innermost pixel region of the tracker. Despite being smaller and being in a region where tracks have lower spatial separation, the pixel has the highest granularity and, thus, the lowest number of particles per unit pixel. This step is called seed generation because the collection of hits with initial track estimates is referred to as a seed. In the case of online track reconstruction, generally the seed uses the pixel hits only to reduce the reconstruction time.

The second step employs a track-finding algorithm that finds hits in successive layers of the tracker moving radially outward from the beam pipe. It starts with a crude estimate from the track seed generated in the previous step. It uses the Kalman filter to find the hit and updates the initial set of track parameters with every propagation to the next layer. It also uses the information on the material budget at each layer to account for the energy loss from multiple scattering interactions. The third step fits the entire track with all the hits found using the

Kalman filter and smoother. Outlier hits are assigned to the track at this stage.

The fourth step selects or rejects the reconstructed track by applying selection on the number of hits in the track, the track parameters, the uncertainty in the parameters and the goodness of fit. The stringent nature of the selection depends on the iteration stage of the track reconstruction.

The iterative algorithm starts by forming tracks that are closer to the beam pipe and are highly energetic. A very high quality constraint is required on the parameters of this track such that the parameters have a low error margin and a high goodness of fit. Charged particles with lower momentum or tracks formed further from the beam pipe are reconstructed in the next iterations.

Tracking is implemented at the HLT to have a larger efficiency for signal events and higher background rejection. The offline tracking involves multiple iterations on the hits in the tracker modules, which entails reading data from approximately 50 million channels. The sheer bulk of data processed makes the tracking algorithm extremely time-consuming and computationally expensive. The full complexity of offline tracking is used at a limited capacity at the HLT, depending on the use case. A significant portion of tracking can be done for prompt particles by relying on the hits of the pixel detector. A significant amount of time is saved if the outer strip layers of the CMS detector are not considered. Other methods consider pixel and strip trackers but only perform regional tracking around other physics objects as described in this section, e.g. a muon. The number of iterations is reduced at the HLT to comply with the time constraints.

In preparation for HL-LHC, the CMS collaboration has begun integrating GPU farms into the HLT. A significant portion of tracking has been offloaded to GPUs. This results in 20% faster decision of events at the LHC [180].

### 3.2.2 Primary Vertex

Primary vertex reconstruction [181] is done to identify the exact position of the hard-scattered  $p$ - $p$  interaction in each event. It is essential to obtain event level information about the collision region in which the proton beams crossed each other and to estimate the PU in the event. Tracks close to the interaction region with a minimum number of hits in the pixel and strip detector are used for determining the primary vertex candidates in an event. The selected tracks are clustered together using the deterministic annealing (DA) method [182] in the  $z$ -axis of CMS. Vertex fitting is performed using the adaptive vertex fitter (AVF) [183] to obtain the 3-D position of the vertex, the covariance matrix and the results of the goodness of the fit. Figure 3.2 shows 20 primary vertices from an event in 2018 with a tau pair detector signature.

### 3.2.3 Particle Flow

The CMS collaboration has adopted a global event reconstruction strategy since the start of run 2, which combines information from multiple sub-detectors to create a set of physics objects. This strategy is called the Particle Flow (PF) algorithm [185]. It reconstructs muons, electrons, photons, and charged and neutral hadrons by combining information from multiple sub-modules in the entire detector as illustrated in Fig. 3.3. The reconstruction of each object is described in their respective sections.

The distinct PF objects are based on the philosophy that they leave a unique signature in the detector. Muons produce hits in the inner tracker, pass through the calorimeters with barely any interaction and then produce hits in the gaseous muon detectors. Charged hadrons form tracks in the tracker, deposit part of their energy in the ECAL and are stopped in the HCAL,

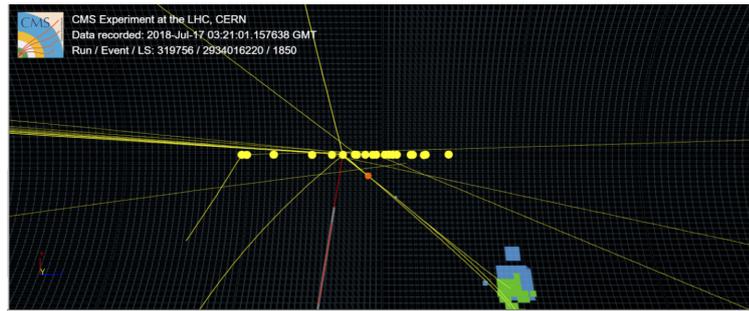


Figure 3.2: Around 20 primary vertices [184] observed alongside a hard-scattered event. Only high- $p_T$  tracks associated with the vertices are shown.

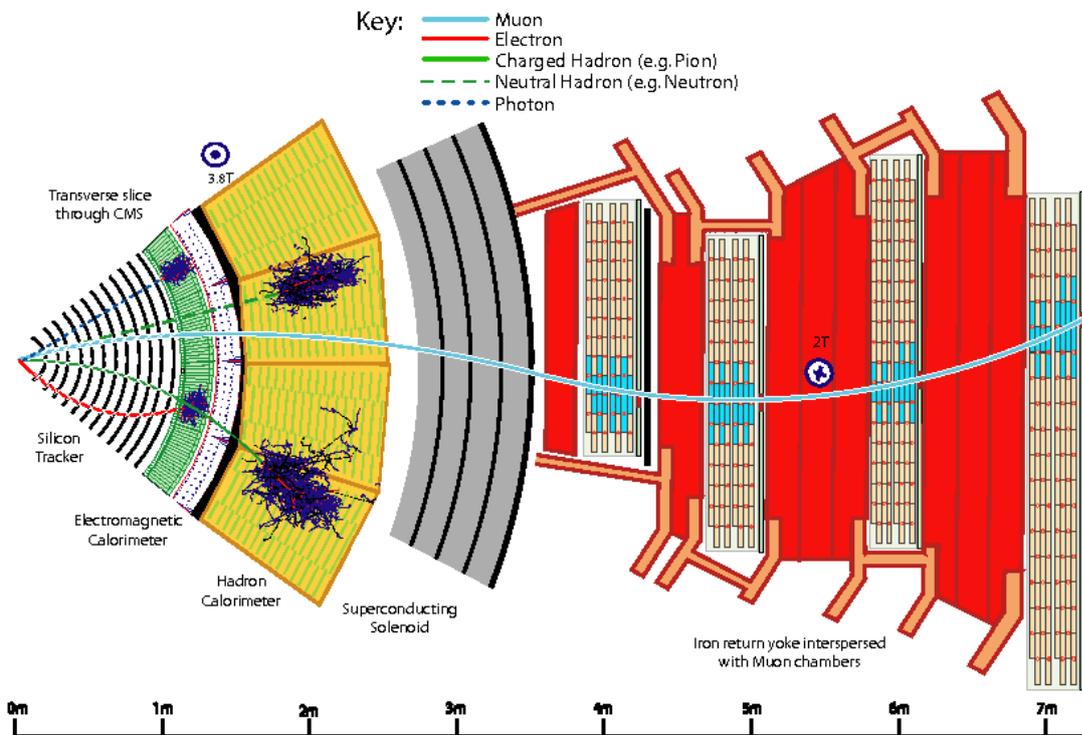


Figure 3.3: The Particle Flow algorithm combines information from the entire detector to identify the fundamental particles. A muon is identified as the only particle which traverses through the whole detector leaving only tracking signatures. An electron leaves a track in the tracker and is stopped at the ECAL, where its total energy is deposited. Like the electron, the photon deposits all its energy in the ECAL is not associated to a track. An analogous philosophy applies to the charged and neutral hadrons with the HCAL. The neutrino is the only particle that escapes the detector without any signature. However, its energy can be measured by the imbalance in the transverse momentum of the other detected particles.

where they deposit their remaining energy. The particle should be a neutral hadron if an energy deposition is observed in the ECAL and HCAL but no corresponding track is found. Electrons and photons deposit all their energy in the ECAL, with electrons leaving tracks in the tracker. The vector sum of the transverse momentum of all PF objects should add up to zero as the initial state partons have negligible transverse energy from the conservation of momentum in a head-on collision. If substantial energy in the transverse plane is unaccounted for, it is likely from a neutrino that passes through without any interaction in the detector. The missing transverse momentum ( $\cancel{p}_T$ ) characterises the neutrino energy.

The energy clusters from ECAL and HCAL, tracks from the tracker and the muon tracks from the muon detectors are inputs to the PF algorithm. Tracks reconstructed in the muon chambers are extrapolated to the tracker to identify muons. The PF process is initiated with muons because they have the cleanest signature of all particles in the CMS experiment. The electrons and photons are reconstructed next by associating the energy deposits in the ECAL to the tracks in the tracker. If the energy of the tracks can be related to the HCAL energy deposit, a charged hadron is identified. The un-associated event information is refined at every step by removing the signature associated with the reconstructed particles. This prevents repeated use of the same signature for multiple particles. Any excessive energy deposit in the HCAL and ECAL is associated with neutral hadrons and photons, respectively.

### 3.2.4 Electron and Photon

Electrons and photons ( $e/\gamma$ ) form electromagnetic showers in the ECAL. The ECAL absorbs almost all the  $e/\gamma$  produced in an event. The detection of  $e/\gamma$  is complicated by the presence of the tracker volume before the ECAL. Ideally, for an electron, an energy deposit in the ECAL can be associated with a helical track in the tracker. If the track is absent, it implies that the energy deposit in ECAL originates from a photon. As shown in Fig. 2.6, depending on the pseudorapidity, the tracker contains a sufficient amount of material in it to cause a photon to convert to electrons by pair production,  $\gamma \rightarrow e^+e^-$ , before reaching the ECAL. An electron can produce multiple photons through bremsstrahlung,  $e^{+/-} \rightarrow e^{+/-} + \gamma$ . Figure 3.4 shows the corresponding electromagnetic cascades that occur for an electron and photon and illustrates the complication of distinguishing between them. The diagrams are similar and differ by only one step in the shower chain. Both electron and photon will thus leave multiple tracks in the tracker followed by a distributed pattern of energy deposit spread in the azimuth of the ECAL. Furthermore, the tracks in the electromagnetic (EM) shower lose a significant amount of energy between the tracker layers via bremsstrahlung.

Each ECAL crystal is equipped with a clock along with its ADC, which allows a measurement of the time at which the energy deposit from a particle traversing through the ECAL peaks in the crystal [186]. The clocks are synchronised such that a prompt particle from the origin of the proton collision will arrive at the same stamp irrespective of its angular direction. The synchronisation results in an offset in the crystal clocks from the central region of the ECAL to the high  $\eta$  region, depending on its distance from the origin of the detector. The crystals with the same  $\phi$  will have the same time parameter. The ADC sends a group of ten amplitudes of the energy detected in the crystal spaced by a time difference of 25 ns into the data stream. The distribution of the ratio of these amplitudes as a function of the time difference has a shape as shown in Fig. 3.5 left, which is a universal property for a crystal. It does not depend on the particle's energy or the particle's type. It depends on the crystal and the time at which the maximum amplitude was observed. The shape of the distribution varies between EB and EE and is sensitive to a delayed energy deposit in the crystal. The resolution of the timing of the

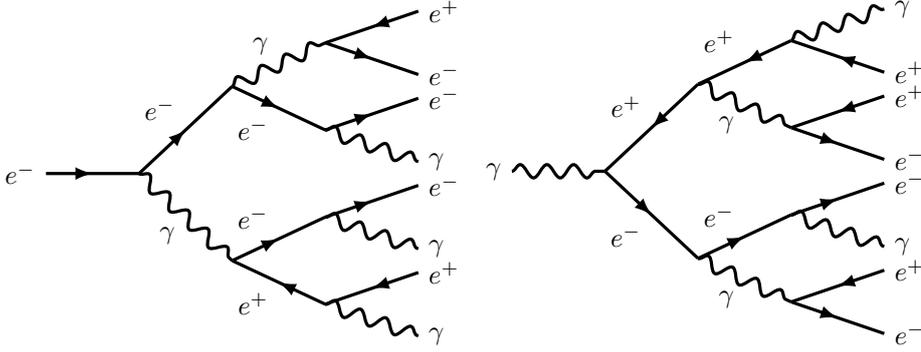


Figure 3.4: Illustration of bremsstrahlung (left) and pair production (right) event.

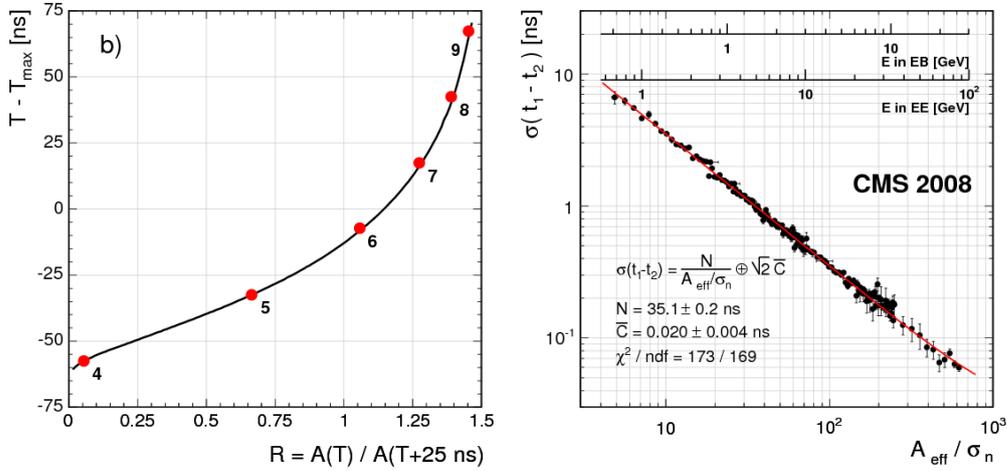


Figure 3.5: Left: The  $x$ -axis plots the ratio of the amplitude of energy deposit taken every 25 ns. The  $y$ -axis plots the time difference with the time of maximum energy deposit in the crystal. The ECAL timing of an energy deposit is obtained with a fit to this distribution. Right: The resolution of the time measurement shown as the time difference between energy deposit in two adjacent crystals as a function of the maximum amplitude of energy deposit in each crystal divided by the resolution of the energy deposit. The equivalent energy in EB and EE is plotted in two other  $x$ -axis to provide comparative energy values for easier understanding.

ECAL is dependent on the slope of the distribution of Fig. 3.5 left, the noise in the detector, the truncation during digitisation and the energy deposit in the ECAL. Figure 3.5 right shows the variation of time resolution with energy deposit in an individual crystal. The resolution of a 10 GeV energy deposit in a single crystal is less than 200 ps in EB and 500 ps in EE.

The detailed reconstruction of electrons and photons is given in Ref. [187]. The pulse shape from the ECAL is fitted with template functions to subtract the out-of-time PU contribution [188]. The out-of-time PU refers to events from the previous or the next bunch crossing that might have overlapping energy deposit with the current event. This happens because the dead-time of the ECAL crystal is 25 ns, similar to nominal bunch crossing period of the proton

collisions. Time-dependent crystal response corrections and calibrations correct the signal amplitudes to consider the degrading of crystals in the high radiation environment. The reconstruction begins with clustering the energy deposit in the neighbouring crystals in the ECAL above a certain noise threshold. ECAL clusters within a specific geometrical region are grouped together to form a supercluster (SC) to include losses from photon conversion and bremsstrahlung. A ‘mustache’ superclustering algorithm was used throughout run 2 and graph neural network based superclustering is developed to substitute it [189].

The reconstruction algorithm searches for tracks in the tracker. The Kalman Filter(KF) algorithm-based tracking is replaced with Gaussian sum filtered (GSF) tracking [190], which can correctly reconstruct tracks with radiative losses from bremsstrahlung. The track reconstruction for electrons can either be initiated starting from the ECAL SC or the tracker. In the ECAL SC driven tracking, ECAL SCs which have transverse energy  $> 4$  GeV and less than 15% energy in the HCAL crystals in a cone  $\Delta R < 0.15$  centred around the SC are selected and groups of two or three hits are searched for in the inner pixel tracker in the geometrical region of the SC compatible with a helical trajectory hypothesis. In the tracker driven tracking, the algorithm iterates over all generic tracks and identifies a KF track compatible with the ECAL SC. The identified hits in the inner pixel detector and the KF track initiate a GSF track reconstruction for the electron. A dedicated algorithm tries to find the track pairs likely to originate from photon conversions to  $e^+ e^-$  pairs.

The above information is input to the PF framework. The SC energy is further refined by identifying more ECAL clusters from tracks that were identified as part of the GSF track reconstruction. Identifying the ECAL SC and tracks as electrons and photons happens at this stage. If a refined SC with some loose selection requirements is associated with a GSF track, it is identified as an electron, otherwise it is identified as a photon. If the refined SC fails specific selection requirements, its energy is further considered to form neutral hadrons, charged hadrons or non-isolated photons.

Online reconstruction has many aspects similar to offline reconstruction. Depending on the energy and nature of the electron or photon that the trigger is built for they might have differences in the online reconstruction method. Most of these methods use the ECAL L1 candidates as a starting point. The L1 candidates are a group of neighbouring trigger towers defined by arrays of ECAL crystals and could be either an electron or a photon [130]. A group of trigger towers is used to account for losses from bremsstrahlung and pair production. Certain constraints on the shape of the energy deposit in the group and the energy deposit in the HCAL behind it are used to improve the ECAL signature identification for electrons and photons. The reconstruction for electrons and photons at the HLT is regionally centred around the L1 ECAL object. The regional reconstruction reduces the HLT processing time. Localised hits in the vicinity of the ECAL energy deposit are required in the pixel tracker to initiate GSF tracking. The integration into the PF algorithm is dropped to avoid reconstructing other objects in the detector that are irrelevant to the HLT. Full KF tracking is done with fewer iterations to calculate the isolation of the GSF track in the detector. Depending on the background from certain methods of tracking or energy of the object considered the selection parameters in the online electron and photon triggers at the HLT vary.

The online reconstruction proceeds with successive modules in the chain requiring an increased amount of time. Event filtering is done after every module to reject background and reduce the overall run time of the menu by not having to run the modules that require a larger time for processing background events. The ECAL superclustering is followed by GSF tracking, followed

by the isolation reconstruction of the track.

Other differences exist between the online and offline reconstruction concerning calorimetry. The calibrations for crystals are updated periodically for the ECAL crystals due to transparency loss in the high-radiation environment. The time selection applied to reject out-of-time hits in the crystal is removed because it does not significantly reduce the rate and risks losing rare signatures such as the ones from displaced electrons in STHDM decay.

A set of variables described below define the quality and kinematic properties of the reconstructed ECAL SC used to identify good electrons and photons.

#### $\sigma_{i\eta i\eta}$

$\sigma_{i\eta i\eta}$  is the second moment of the log-weighted energy distribution of an ECAL SC [187]. A  $5 \times 5$  matrix around the most energetic crystal of the SC is chosen. For every crystal  $i$  with pseudorapidity  $\eta_i$  and energy deposit  $E_i$ , a weight is assigned as

$$w_i = \max\left(0, 4.7 + \ln\left(\frac{E_i}{E_{5 \times 5}}\right)\right), \quad (3.2)$$

where  $E_{5 \times 5}$  is the summed energy deposit in the  $5 \times 5$  matrix. The added factor of 4.7 in the definition of  $w_i$  is equivalent to requiring an energy deposit in the crystal, which is at least 0.9% of the total energy in the  $5 \times 5$  matrix. This ensures proper ECAL noise mitigation. The mean pseudorapidity of the  $5 \times 5$  matrix is given by  $\bar{\eta}_{5 \times 5}$ . The mathematical expression of  $\sigma_{i\eta i\eta}$  is then given by

$$\sigma_{i\eta i\eta} = \sqrt{\frac{\sum_i^{5 \times 5} w_i (\eta_i - \bar{\eta}_{5 \times 5})^2}{\sum_i^{5 \times 5} w_i}}. \quad (3.3)$$

This is a measure of the  $\eta$  spread of the energy deposit in the ECAL. The energy deposit spread for single electron and photon showers is due to bremsstrahlung and photon conversions. This is expected to be narrower than two-photon showers in meson decays, which form the central component of the background in largely electromagnetic jet showers.

#### $H/E$

The  $H$  in  $H/E$  is the energy deposit in the HCAL in a cone  $\Delta R = 0.15$  around the ECAL SC energy deposit, denoted by  $E$ . The source of the energy deposit in the HCAL could be detector noise, energy deposit from PU or leakage of electrons and photons through the gaps in the ECAL. This variable shows the relative deposition of the energy deposit between the HCAL and ECAL and separates jets from electrons and photons.

#### $I_E/E$

The  $I_E$  measures the sum of the energy deposit in the ECAL around SC in a cone  $\Delta R = 0.3$ . Divided by the energy of the SC, it provides the relative isolation of the ECAL SC in the ECAL. This is the online definition of the variable. The offline definition uses physics objects from PF algorithm.

#### $I_H/E$

The  $I_H$  measures the sum of the energy deposit in the HCAL around SC in a cone  $\Delta R = 0.3$ . Divided by the energy of the SC, it provides the relative isolation of the ECAL SC in the HCAL. This is the online definition of the variable. The offline definition uses physics objects from PF algorithm.

#### $|\Delta\eta|_{in}^{seed}$

It is defined as the  $\eta$  difference between the central crystal of the SC and the track seed of the

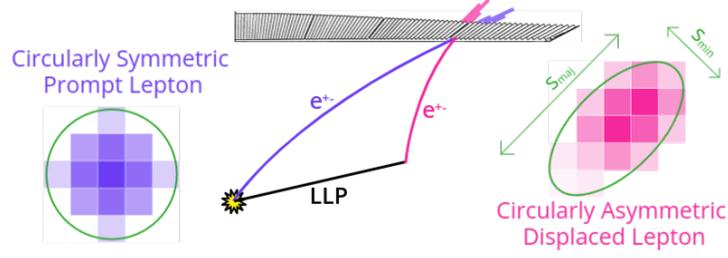


Figure 3.6: Illustration of the  $s_{min}$  and the  $s_{maj}$  variable.

tracker associated with it. The track seed reconstructed in the pixel region of the tracker is used as it has not gone through radiative processes yet and thus has the best estimate of the angular parameters of the electron. There would ideally be no difference for an electron with perfect reconstruction. Realistically, the value is non-zero and drops exponentially.

### $|\Delta\phi|_{in}$

This is the  $\phi$  difference between the SC and the track seed. The reason for using tracker seed is similar to  $|\Delta\eta|_{in}^{seed}$ .

### $1/E - 1/p$

The variable  $1/E - 1/p$  measures the energy difference between the ECAL SC and the track momentum. The reciprocal values for  $E$  and  $p$  were taken to make the variable sensitive to large differences for smaller energy deposits.

### $I_T/E$

$I_T$  defines the isolation of the ECAL SC with respect to the energy of the tracks in the tracker. This is the online definition of the variable. The offline definition uses physics objects from PF algorithm.

### $s_{min}$ and $s_{maj}$

The  $s_{min}$  and  $s_{maj}$  measure the circular symmetry of the energy deposit in the ECAL. The energy deposit in the ECAL can be envisioned as a  $2-d$  matrix in the  $\eta-\phi$  plane, with each matrix entry taking the energy deposit value in the corresponding crystal. The second moment of the energy deposit in a  $5 \times 5$  crystal is obtained along the  $\eta$  and  $\phi$  using equations analogous to Eq. 3.3. Three values of the second moment of the energy deposit are obtained -  $S_{\eta\eta}$ ,  $S_{\phi\phi}$ , and  $S_{\eta\phi}$ . The length of the minor and major axis of the energy deposit parameterized by  $s_{min}$  and  $s_{maj}$  respectively are given by,

$$\begin{aligned}
 s_{min} &= \frac{s_{\eta\eta} + s_{\phi\phi} - \sqrt{(s_{\eta\eta} - s_{\phi\phi})^2 + 4s_{\eta\phi}^2}}{2} \\
 s_{maj} &= \frac{s_{\eta\eta} + s_{\phi\phi} + \sqrt{(s_{\eta\eta} - s_{\phi\phi})^2 + 4s_{\eta\phi}^2}}{2}.
 \end{aligned} \tag{3.4}$$

They measure whether the shower shape of ECAL SC is circularly symmetric. For a prompt particle signature unaffected by a radiative process in the tracker, such as a photon from BEH decay that has not gone through pair production, the energy deposit will be symmetric and the values for  $s_{min}$  and  $s_{maj}$  will be equal and take values of about 0.8. This happens because the

ECAL has been placed to point towards the collision vertex. A prompt electron/photon will have a circularly symmetric energy deposit as illustrated in fig. 3.6. On the other hand, if the energy deposit is significantly skewed it can indicate the presence of signatures such as pair produced photon or meson decaying to a pair of electrons or an electron from the decay of a displaced long-lived particle (LLP), such as in STHDM.

### 3.2.5 Muon

Muons are the first PF object to be reconstructed because their signature is very clean [191]. The CMS detector is designed such that almost all the particles from the proton collisions deposit their energy and thus get absorbed entirely in the calorimeters. Any particles that may escape the calorimeters are stopped in the solenoid material or in the steel return yoke for the magnet. Only muons and neutrinos can escape the considerable interaction length of matter inside the solenoid. Of these, the neutrinos rarely interact with matter, while the muons leave a clean signature in the CMS muon chambers.

The muon reconstruction in the PF algorithm begins with KF tracking in the muon chambers. The tracks obtained from the muon chambers are matched with tracks in the tracker. If a match is found, these muons are labelled as global muons. A complementary algorithm matches every track in the tracker with a signature in the calorimeter and muon chamber. Such a signature is labelled as a tracker muon if a compatible signature is found.

A PF muon is identified if certain selection conditions described in Ref. [191] are met. All signatures with PF muons are removed from the collection before proceeding to identify other PF physics objects. If a PF muon is not identified and is a tracker muon with a high-quality track fit, it is retained in the PF collections.

The online reconstruction for muons takes place in three stages [192]. The L1 trigger system combines trigger primitives from three regional track-finders segmented based on the barrel, end-cap and the overlap region between the barrel and the end-cap to reconstruct muon tracks. The L1 Global Muon Trigger collects reconstructed muons from the three track finders and removes overlaps based on the  $p_T$  and track quality. Additional quality selections are imposed to form the L1 muons. The L1 muon reconstruction is more than 95% efficient in the barrel and gradually drops to 85% in the end-cap with extremely tight selection criteria.

The HLT muon reconstruction includes two levels. The first level, the second stage of muon reconstruction overall, is identical to the offline muon reconstruction in the muon chambers. The DT and CSC hits are combined to form a set of initial track states called L2 seeds. The L2 seeds must be in an angular cone  $\Delta R < 0.3$  of the L1 muons. The KF tracking method is used to build tracks from the L2 seeds. An L2 seed is extrapolated to the L2 seed in the innermost muon chamber. The L2 seed parameters are updated and track building is initiated. The track parameters are updated with every fit as the track finder propagates to the outer layers. In most triggers, the beam spot is used to constrain the track parameters to improve the momentum resolution of the track. The HLT paths for cosmic muons and paths targeting displaced muons are exempt from the beam spot requirement as it would reduce the efficiency of tracking for these paths.

The third stage of muon reconstruction in the HLT combines information from the muon chambers and the inner tracking detector. It uses a three stage approach. In the first stage, the L2 seeds obtained in the previous step are used to construct a helical track in the tracker. Only tracks satisfying specific quality requirements identical to the tracking are retained in this step. In the second step, regional tracking is performed based on an L2 seed or an L1 muon. A three

pixel hit seeded algorithm is followed by a two pixel hit seeded algorithm to form tracks with a helical match to either an L1 muon or an L2 seed. The track collection is then checked to remove duplicates. The muon tracks obtained are angularly matched to an L2 muon to obtain an L3 muon. The resulting efficiency is almost 100% compared to the L1 efficiency.

### 3.2.6 Jet

Once the reconstruction for electrons, isolated photons and muons is completed, the remaining observed data in the PF collections must be from hadrons and hadronic decays of mesons created from colour recombination of the beam remnants and hadronic decays of the  $\tau$  lepton. The hadrons move through the detector and radiate and decay along their trajectory. A single parent parton can thus give rise to a collimated spray of particles called a jet. They give rise to angularly close tracks if charged hadrons are part of the jet and energy deposit in the HCAL for both charged and neutral hadrons. Photons from radiative processes are also part of the jet and give rise to a significant energy deposit in the ECAL.

All PF tracking elements with energy deposit in the HCAL will have little energy deposit in the ECAL and are caused by charged hadrons. The fraction of energy deposit in the HCAL that cannot be ascribed to tracking elements comes from neutral hadrons. The ECAL energy component majorly arises from the photons. The jet is clustered using the anti- $k_T$  algorithm [193, 194] with a distance parameter 0.4 and contains multiple reconstructed hadrons and non-isolated photons. The raw jet energy is the summed energy of the calorimeter towers, and the raw jet momentum is the vector sum of the tower momenta. The raw jet energy is corrected to establish an uniform calorimeter response in  $\eta$  relative to other jets. The raw jet transverse momentum is corrected by a calibrated absolute response.

The jets originating from the  $b$  and  $c$  quark are displaced from the primary vertex and can be identified using dedicated algorithms [195]. Identifying heavy-flavour jets provides essential physics scope, especially with top quarks. The jet energy and transverse momentum corrections are also derived for the jets from  $b$  and  $c$  quark.

### 3.2.7 Missing Transverse Momentum

The vector sum of all transverse momentum in a head-on proton collision should be zero by momentum conservation. A momentum imbalance usually implies an undetected particle in the event, e.g. a neutrino. In terms of PF physics objects, the missing transverse momentum is defined as

$$\vec{p}_T^{\#} = - \sum_i^{\text{PF Objects}} \vec{p}_T^i. \quad (3.5)$$

The magnitude of the missing transverse momentum,  $p_T^{\#}$ , is generally used. Due to an imperfect reconstruction, this value can become artificially large. Mostly, this happens due to mis-reconstruction or mis-identification of a high- $p_T$  muon. Events with large  $p_T^{\#}$  go through a post-processing step, identifying some known cases that can give rise to a wrong significant value for  $p_T^{\#}$  [196]. The  $\vec{p}_T^{\#}$  is modified to account for the jet energy corrections in the event.

## 3.3 Delphes Fast-Simulation Framework

The Delphes framework is designed for fast emulation of detector effects on simulated samples to provide the scope to study the detector signature of the simulation under consideration [197].

No actual material radiation simulation, as done in `Geant4`, is included in `Delphes`. Instead, `Delphes` emulates the detector effects by artificially applying efficiency and energy smearing to the simulation particles through configuration cards. The experiments publish efficiency and resolution effects based on the kinematic properties and position of a physics object. This can be summarised in one configuration file in `Delphes`. The `Delphes` simulation framework reads the conditions and applies the effects on the particles simulated by the generator. Some objects, e.g. the  $\cancel{p}_T$  and isolation, are computed by `Delphes` after the detector effects have been applied to the particles.

## 3.4 Machine Learning

Machine Learning (ML) is the science of designing computer algorithms which can perform tasks without having to code explicit instructions [198]. It is beneficial for jobs where assigning a human algorithm developer becomes cost-prohibitive. ML has a history of usage in experimental particle physics. State-of-the-art jet taggers rely on advanced deep neural networks to identify jets from  $b$  or  $c$ -quark or the  $\tau$ -lepton [199–201]. Background simulation of events could benefit from a larger sample size and faster simulation using the generative models [202, 203]. ML methods can be used to identify BSM events in the LHC experiments [204]. The conventional iterative tracking can be replaced with a faster ML-based tracking implemented on FPGAs for faster decision at the HLT [205]. These are some of the regular contributions of computerised learning methods applied to experimental data. This section presents a brief discussion about a specific ML model which has been used in this thesis.

A significant part of ML involves solving problems where data is split into multiple distinct, mutually exclusive classes. Such problems are called classification problems. In experimental particle physics, classification problems include classifying a BSM signal from the SM background, identifying the jet flavours from the detector signatures, separating the ECAL signature of an isolated electron from a non-isolated one, etc. Mathematically, as given in Eq. 3.6, a solution to the classification problem is a function  $f$  that involves mapping  $m$  real numbers ( $\mathbb{R}^m$ ) input in data to  $k$  real numbers ( $\mathbb{R}^k$ ) representing the individual classes that the data is composed of, where  $\mathbb{R}^k$  is typically interpreted as the probability for the data to belong to a specific class.

$$f : \mathbb{R}^m \rightarrow \mathbb{R}^k : \vec{x} = \{x_1, x_2, \dots, x_m\} \mapsto \{d_1, d_2, \dots, d_k\} \quad (3.6)$$

The output  $d_i$  are called discriminant. The discriminant needs to satisfy the condition  $d_i > 0$  and  $\sum_i d_i = 1$  to facilitate a probabilistic interpretation. There are two classes in a typical problem involving signal and background.

Classification problems are approached using multiple ML models. The model of specific interest to us is the artificial neural network (ANN). The ANN was developed based on the ideas pioneered by Warren S. McCulloch and Walter Pitts [206]. The basic building block of a neural network was first pioneered in 1958 [207]. The basic building block has undergone a lot of evolution before the familiar form well-known to ML users today [198]. The concept of an ANN is illustrated using Fig. 3.7.

An ANN is made by propagating the input data through various layers of functions before computing the discriminator that decides the output class. For the model illustrated in Fig. 3.7, the input  $x_i$  are  $m$  numbers. The layers of functions that process the input data in an ANN are called hidden layers. Each hidden layer is composed of multiple nodes. Different hidden layers can have varying numbers of nodes. A node is a non-linear function of the values in the previous

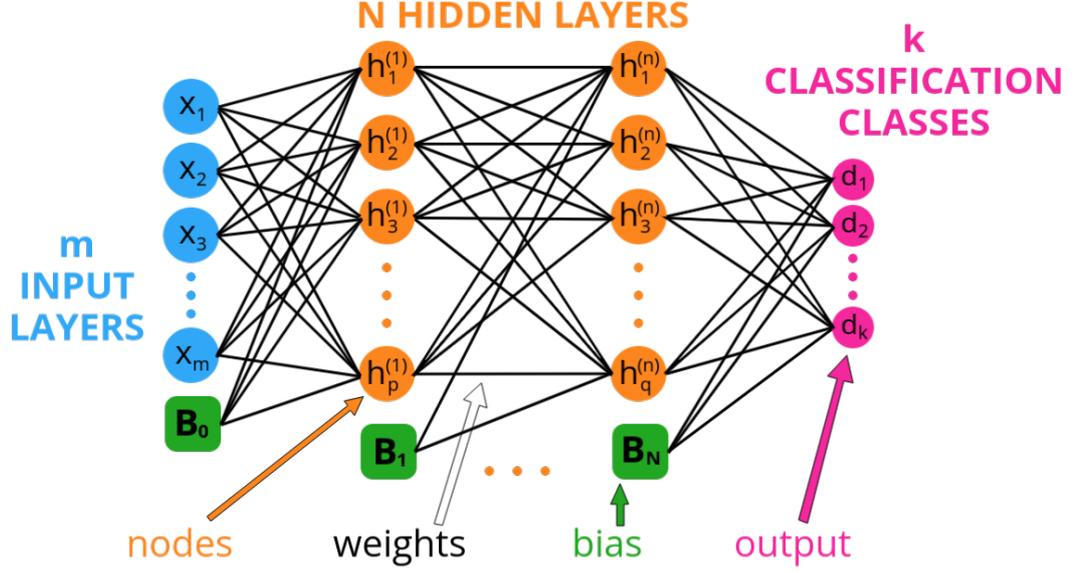


Figure 3.7: Illustration of Artificial Neural Network model. It shows the input in blue and the output in pink. The number of output numbers is the number of distinct mutually exclusive classes involved in the data. The ANN can have as many hidden layers (shown in orange) as required to classify input data properly. The hidden layers are composed of varying numbers of nodes. The black lines are weighted connections between the layers. Each layer includes a contribution from a constant factor.

node. The  $a$ -th node in the  $b$ -th hidden layer is defined mathematically as,

$$h_a^{(b)} = \mathcal{A} \left( B_{b-1} + \sum_i w_a^i h_i^{(b-1)} \right), \quad (3.7)$$

where,  $\mathcal{A}$  is a non-linear function of a weighted linear combination of nodes in the previous layer called the activation function, and  $B_i$  is a constant factor in every layer called the bias in a layer. There are several choices for the activation function. One of the most popular choices is the Rectified Linear Unit (ReLU) [208]. The ReLU activation is used for all the hidden layers nodes. It is defined as  $f(x) = \max(0, x)$ . The nodes in the output layer can be envisioned as having the same structure as the hidden layers but with the softmax activation function [198]. The softmax activation function ensures that the discriminant values are positive, less than one and sum up to one, thus enabling the probabilistic interpretation.

The numerous weights of the ANN, under ideal classification conditions, take a value such that the input data is categorised into the constituent classes perfectly. They are apriori unknown. Finding the correct value of these weights to obtain a high classification accuracy is called training the neural network. A neural network can be trained with simulations in what is known as the supervised training of neural networks. While training, the performance of the ANN is characterised by the cross-entropy loss function and the prediction accuracy [198]. For a 2-class

classifier, the loss function is defined as,

$$L = -( y \log(p) + (1 - y) \log(1 - p) ), \quad (3.8)$$

where  $y$  is the known class number from simulations and  $p$  is the prediction of the neural network based on the input data. The prediction accuracy is the ratio of the samples correctly predicted by the ANN to the total number of events in the sample. For training the ANN, the simulation sample is split into two parts. One is used to train the ANN, called the training sample, while the other is used to evaluate the performance independent of the bias the ANN may acquire from the training sample, called the testing sample. Training the ANN requires multiple iterations over the simulation sample. Each iteration is called an epoch. With increasing epochs, the weights are optimised such that the loss function decreases. The prediction accuracy is consequently observed to increase.

While training the ANN, the model may learn parameters specific to the training sample and not generic to the simulation sample. When this happens, the prediction accuracy on the training sample increases while the prediction accuracy on the testing sample decreases. Simultaneously, the loss function decreases on the training sample while the loss function increases on the testing sample with increasing epochs. When the performance of the ANN increases on the training sample and reduces on the testing, the mode is overtraining and is avoided by regularisation techniques. The dropout method of regularisation randomly drops a certain fraction of nodes in a training epoch [209].

The performance of the ANN in terms of signal classification from the background is done by the receiver operation characteristic (ROC) curve. The ROC curve is plotted with one signal selection efficiency on one axis and background rejection efficiency on the other. The exact definitions vary on the user and are defined in Sec. 4.4.3 where it has been used.



# Chapter 4

## Soft and Displaced Leptons - Phenomenology

*The phenomenology of the dark matter model introduced in the first chapter is studied in detail in the CMS experiment using a lightweight simulation framework generic to the particle physics community where the full complexity of an actual experiment simulation can be avoided. The aim is to establish a concept of proof for a more detailed study using the CMS experiment data. This study has been published.*

The Singlet Triplet Higgs portal Dark Matter (STHDM) model predicts a BSM signature with a low transverse momentum ( $p_T$ ) oppositely charged displaced lepton pair from the decay of two  $\chi^\pm$  in the CMS detector. This chapter studies the challenges to detecting the STHDM signature and makes predictions for a possible search with the CMS experiment. It has been shown that search sensitivity can be obtained for the STHDM model over four orders of lifetime of  $\chi^\pm$  with lepton  $p_T > 20$  GeV. The work in this chapter has been published in the Journal of High Energy Physics [83]. Significant overlap is expected between the ideas expressed in the paper and this chapter.

In order to search for events in the low energy displaced di-lepton final state, a reliable estimate of background events in the LHC experiments is needed. In the following sub-section, the existing searches with co-annihilating and co-scattering dark matter are described and an informed decision is made on the LHC analysis used to estimate the number of background events. The choice of analysis for background estimation motivates several factors in this study e.g. the final state comprises one displaced electron and one displaced muon ( $e\mu$ ) and the signal region has to be split with respect to different values for the impact parameter  $d_0$  of the leptons.

### 4.1 Existing LHC Searches for Displaced Leptons

Searches for prompt leptons with compressed mass spectra, searches for displaced leptons and searches for disappearing tracks are all sensitive to co-scattering or co-annihilation DM models. Prompt lepton refers to a lepton which is produced at the proton collision vertex either through scattering or radiative processes or from the decay of particles without significant lifetime, e.g. the Z-boson. Prompt lepton searches for DM models [210, 211] require leptons with transverse impact parameter  $d_0 \leq 0.1$  mm. They are, by selection, not sensitive to long-lived dark sectors.

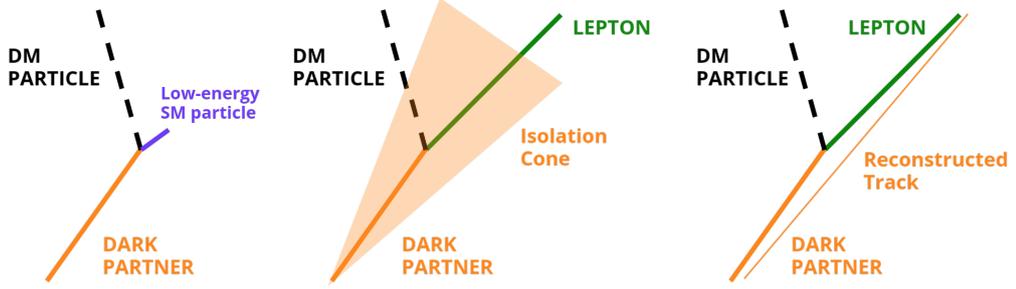


Figure 4.1: The orange track is the charged dark portal candidate whose track is visible. Depending on the dark portal, it decays to an unobserved dark matter candidate, represented in the figure by a dashed black track, and a low-energy SM particle (violet track in the left figure) or high-energy SM particle (green track in the centre and right figure). As shown in the left figure, a low-energy SM meson will not be reconstructed, leading to a disappearing track. However, a lepton will distort the reconstruction of the orange track by making it fail either the isolation or track goodness of fit. These scenarios are demonstrated in the orange isolation cone in the centre figure and the poorly reconstructed orange track in the right figure.

Searches for disappearing tracks [212–214] consider a signature where an electrically charged dark portal particle leaves a track in the CMS tracker before decaying to a neutral DM candidate and low energy lepton which is not reconstructed in the CMS detector, see Fig. 4.1 left. Hence, the track created from the proton collision disappears halfway. These searches probe lifetimes larger than a few centimetres. The experimental bounds from disappearing track searches apply only if the decay products of the original charged particle are invisible in the detector. The presence of an extra lepton in the final state may degrade the sensitivity, either by causing the original track to fail the isolation criteria demonstrated in Fig. 4.1 centre or because the kink from the lepton track modifies the reconstruction of the original track from the charged state leading it to fail track reconstruction goodness of fit shown in Fig. 4.1. Searches for prompt leptons and disappearing tracks are not sensitive to DM scenarios due to the presence of the additional track in the final state.

In the displaced region from a millimetre to several centimetres, the leptons produced in decays of dark states do not have prompt tracks. They are not displaced enough for the track of the dark partner to be a disappearing track. These tracks lie in the fiducial region of the CMS tracker. It should be possible to reconstruct tracks with these displacements allowing for an inclusive search. The proposed STHDM allows to search for DM signatures in this range.

Displaced leptons have been widely sought for at the LHC [93, 94, 215–217]. References [93, 215, 216] search for signatures where two leptons are produced by the decay of the same particle. This search differs from the STHDM leptons that stem from the decay of two different particles that travel in opposite directions of the collision vertex in the CMS detector. The vertex requirement for leptons in prompt searches will reduce the sensitivity to STHDM by a significant fraction or make it zero.

By designing the search for two individual displaced leptons without any vertex requirement, a broader class of event topologies is covered. References [94, 217, 218] search for similar event topologies, see Eq. (1.24). The searches at 8 TeV [217] and 13 TeV [94, 218] differ mainly in the

requirement of the lepton transverse momenta. Reference [218] is essentially an updated analysis of Ref. [94] with the entire run 2 data collected in the CMS experiment. At the time of this work in 2019-20, Ref. [218] was not yet published.

Both [94, 217] analyses were performed without imposing any model-dependent kinematic restrictions, e.g. on the  $\Delta R$ , on the leptons or requiring further activity in the event. The event selection required an oppositely charged pair of an electron and a muon with  $p_T(e) > 42$  GeV,  $p_T(\mu) > 40$  GeV, and  $|\eta(\ell)| < 2.4$ , for  $\ell = e, \mu$  at 13 TeV. At 8 TeV both leptons were required to have  $p_T(\ell) > 25$  GeV. The higher lepton  $p_T$  selection at 13 TeV was driven by the HLT paths used in the analysis. The 13 TeV analysis was chosen for obtaining background estimation, as this was closer to the expected collision energies in future programs at the LHC. Hence the 13 TeV analysis provides a more reliable background estimate.

The background estimate in the 13 TeV displaced non-vertexed lepton analysis has been calculated using a data-driven approach [94]. A significant proportion of the background was observed to be from events with displaced leptons from jets originating from the  $b$ -quark and the  $c$ -quark. In what comes, these jets will be referred to as the heavy-flavour (HF) jets. Displaced leptons are typically observed in the HF jets and serve to distinguish them from the jets of other quarks. Due to a large multiplicity of the HF jets from soft scattering in the proton collisions, the inefficiency in reconstruction can occasionally cause the displaced leptons in the HF jets to be misidentified as isolated. No dedicated simulation exists for such a background requiring a data-driven estimate. The simulation of misidentified isolated leptons from HF jets would anyway require too much modelling from the detector and large statistics to obtain a reliable result. A data-driven estimate was obtained from control regions (CRs) that were defined in the search phase-space where the signal was expected not to be statistically significant. The background estimated in the CRs was extrapolated to the region in the search phase-space where the signal was expected to be significant, called signal regions (SRs) by extrapolating based on relative number of events observed in simulation. The CRs and the SRs were defined based on the displacement of the leptons.

The unsigned impact parameter  $d_0$ , defined in Fig. 4.7 in Section 1.4.3, parameterises the displacement of the leptons from the primary vertex. The phase-space of the search, shown in Fig. 4.2, was parameterised on one axis by the muon  $d_0$  and on the other axis by the electron  $d_0$ . Four CRs were defined in the phase space for prompt leptons. A HF simulation gave the expected number of background events in each CR and SR based on the number of events in each lepton  $d_0$  bin in the simulation. CR1 was used to obtain data-driven background estimate which was used to scale the expected background observed in simulation by a numerical factor. The same numerical factor was used to scale the background simulation events observed in the other CRs and SRs. The other CRs were used to validate these scale factors. Three mutually exclusive signal regions, in Fig. 4.2, are defined.

$$\begin{aligned}
 \text{SR III} &: \text{both leptons satisfying } 1 \text{ mm} < d_0 < 10 \text{ cm}^1 \\
 \text{SR II} &: \text{one lepton failing SR III and both with } d_0 > 500 \mu\text{m} \\
 \text{SR I} &: \text{lepton failing SR II and both with } d_0 > 200 \mu\text{m}.
 \end{aligned}
 \tag{4.1}$$

Three possible leptonic signatures are possible for the STHDM model depending on the decay of the  $\chi^\pm$ : the  $\mu\mu$ ,  $e\mu$  and  $ee$ . The above method based on background estimation from the 13 TeV analysis limits the analysis case to  $e\mu$  only. The expected background contribution for the  $\mu\mu$  final state cannot be estimated by extrapolating the  $e\mu$  background contribution, as they stem from different sources. Examples are radiation from the experimental cavern,

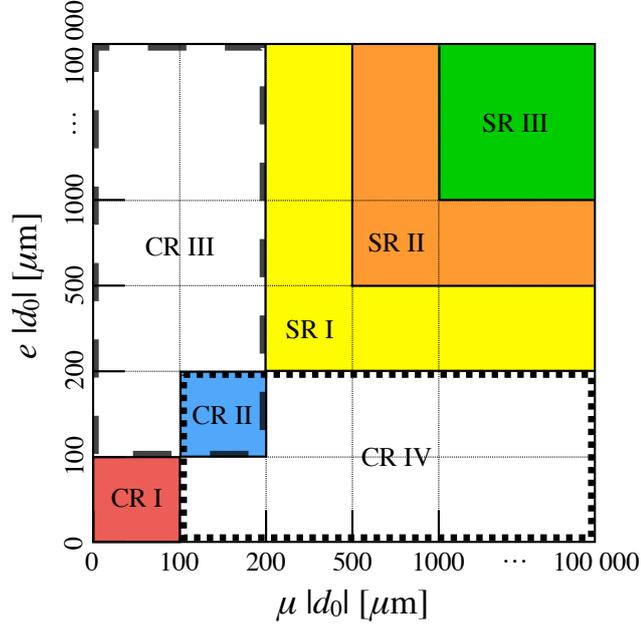


Figure 4.2: Definition of the signal regions based on the  $d_0$  of the leptons. The control regions (CRs) was for prompt particles where an excess in the number of signal events is not expected and has been used to make a data-driven estimate of the number of background events [94].

muons created by LHC beam interactions outside the experimental interaction region, cosmic rays, and particles from hadronic jets that escape the hadronic calorimeter and end up in the muon chambers. The sensitivity of the  $e e$  final state is expected to be lower than the  $e \mu$  owing to rapidly decreasing electron identification efficiency with increasing displacement. The background for  $e e$  also stems from different sources with a large background from jets with significant electromagnetic interactions and scattering from the PU vertices. Backgrounds for  $\mu \mu$  and  $e e$  are complicated and depend on the LHC and the CMS detector parameters which are not easy to model phenomenological. A data-driven background estimate is necessary for  $\mu \mu$  and  $e e$  final states as well. Due to the lack of a data-driven background estimate at the time of this analysis, it has been restricted to only the  $e \mu$  final state in specific SRs.

Both the 8 and 13 TeV searches had no excess with respect to the number of background events expected. 95% confidence level (CL) upper limits were quoted for the signal.

The results of this analysis will be compared to the existing analyses to numerically show the gain in ability to search for the STHDM model. The Poisson log-likelihood ratio has been used to define the signal significance as,

$$Q = \sum_{i=1,II,III} -2 \log \left( \frac{\mathcal{L}_{S_i+B_i}}{\mathcal{L}_{B_i}} \right), \quad \mathcal{L}_{S_i+B_i} = e^{-(S_i+B_i)} \frac{(S_i+B_i)^{N_i}}{N_i!}, \quad (4.2)$$

where  $N_i$  is the observed number of events in signal region  $i$ ,  $S_i$  and  $B_i$  are the expected signal

Table 4.1: Benchmarks of the singlet-triplet dark matter (DM) that will be used as a signal. Leptonic decays of the charged dark partner ( $\chi^+ \rightarrow \chi^0 \ell \nu$ ) are considered only. Benchmarks 1 and 2 are motivated by DM scenarios while benchmarks 3 to 6 extend the approach to a more model independent search, where the decay length is independently varied, keeping  $\Delta m$  fixed. Benchmark 7 has a larger mass splitting  $\Delta m$  that can be compared to the lower mass splitting. The last column shows the total branching ratio into lepton pairs. For benchmarks 1 and 2, see Eq. (1.25).

#	$m_c$ [GeV]	$\Delta m$ [GeV]	$c\tau_c$ [cm]	$\mathcal{B}(\ell^+ \ell^-)$
1	324	20	2	0.025
2	220	20	3	0.014
3	220	20	0.1	1
4	220	20	1	1
5	220	20	10	1
6	220	20	100	1
7	220	40	1	1

and background events in region  $i$ , and  $N = \sum_i N_i$  is the total number of observed events. The log-likelihood ratio for a measurement in the three SRs follows a  $\chi^2$  distribution with  $2(3-1)$  degrees of freedom. The 95% CL upper limit for such probability density is at  $\mathcal{Q} = 5.99$ . The ratio defined below defines the search significance with respect to the 95% CL upper limit,

$$R_{95} = \mathcal{Q}/5.99, \quad (4.3)$$

so that  $R_{95} = 1$  corresponds to the exclusion limit at the 95% CL. It was checked explicitly that the likelihood ratio  $\mathcal{Q}$  reproduces the exclusion limits from Ref. [217] within  $1\sigma$ .

## 4.2 STHDM Model Benchmarks

Section 1.4.3 discussed the STHDM model in detail and set out the independent parameters,  $(m_c, \Delta m, c\tau_c)$ , which determine the kinematics of the model. The limits of these parameters have also been discussed at LHC energies. The mass of  $\chi^\pm$ ,  $m_c$ , determines the rate of production of the STHDM from proton collisions. The mass range for  $m_c$  is between 100 – 500 GeV. If the relic abundance of DM was set by the co-scattering scenario of the STHDM, the limits on  $\Delta m$  vary between 10 – 40 GeV and the lifetime of  $\chi^\pm$  is set to  $\mathcal{O}(1 \text{ cm})$ .

Table 4.1 shows the values of the discussed model parameters that were chosen as benchmarks. Benchmarks 1 and 2 correspond to DM scenarios where the relic abundance can be explained by dark partner lifetimes with displacements that the LHC experiments can observe. The  $\chi^\pm$  decays via weak interactions. The branching ratio is determined by the gauge quantum numbers of the particles in the final state. Only the  $e \mu$  final state from leptonic decays of the charged partner are considered. As mentioned in the previous section, this choice is driven by the requirement of a data-driven background estimate from the CMS search of displaced leptons at 13 TeV in Ref. [94].

One can extend the scope of soft displaced leptons at the LHC by treating  $\Delta m$  and  $c\tau_c$  as

independent. In this case, the leptonic decay branching fraction is taken to be 1. Benchmarks 3 to 7 have a mass of  $m_c = 220$  GeV, with benchmarks 3 to 6 providing different displacements and benchmark 7 having a larger mass splitting. The lower  $m_c$  has been chosen such that it will have a larger production cross section from proton collisions. The mass splitting corresponding to  $m_c = 220$  GeV,  $\Delta m = 20$  GeV is close to the typical weak scale DM. The independent variation of the lifetime of  $\chi^\pm$  makes the study relatively independent of the model. If an excess is observed in the number of signal events compared to background in data for any lifetime in benchmarks 3 to 6, the result can be re-interpreted in other UV complete scenarios either by introducing an electroweak fermion doublet [84] or a scalar triplet [219].

Leptons in compressed dark sectors have low transverse momenta, as was illustrated in Fig. 1.7. A lepton  $p_T$  threshold of 40 GeV, as used in the analyses Ref. [94], would result in a vanishing sensitivity to all compressed sectors. Comparatively, the search at  $\sqrt{s} = 8$  TeV [217] has lower transverse momentum selection but the sensitivity for co-scattering DM in this search reduces with decreasing mass splitting values starting from  $\Delta m = 40$  GeV and vanishes for  $\Delta m = 20$  GeV. This is illustrated in Fig. 4.3.

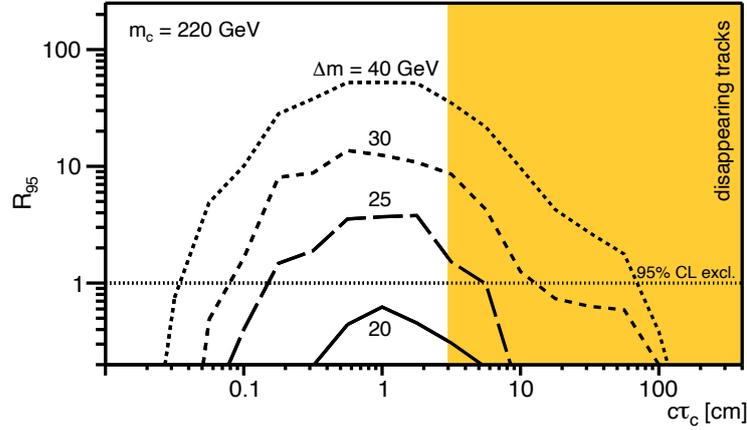


Figure 4.3: Existing limits on co-scattering dark matter as a function of the decay length of the dark partner,  $c\tau_c$ . The concentric curves show the values of  $R_{95}$  (defined in Eq. 4.3) from the 8 TeV CMS disappearing track search [217] for different mass splittings at  $\Delta m = 20, 25, 30$  and  $40$  GeV. The upper bound on the dark partner decay length from the most recent disappearing track search [214] is shown in yellow.

Figure 4.3 shows the log-likelihood ratio  $R_{95}$  defined in Eq. (4.3) used to estimate the search sensitivity as a function of the mediator decay length  $c\tau_c$ . For  $\Delta m = 20$  GeV, the search cannot exclude any range of  $c\tau_c$  at 95% CL, see Fig. 4.3. Thus, a search for the compressed dark sector can be targeted with soft *and* displaced leptons at  $\sqrt{s} = 13$  TeV.

### 4.3 Simulation of Signal and Background Events

The signal model is described using FeynRules [166]. The UFO [167] interface to MadGraph5\_aMC@NLO 2.6.6 [164] is used to generate  $2 \times 10^6$  leading order signal events of the process

$$pp \rightarrow Z^*/\gamma^* \rightarrow \chi^+\chi^- \rightarrow (\chi^0 e^\pm \nu)(\chi^0 \mu^\mp \nu) \quad (4.4)$$

for each of the various benchmarks from Table 4.1. Initial and final state showers as well as hadronisation are modelled with `Pythia v8.243` [168]. These have been discussed in section 3.1. The cross-section  $\sigma(m_c)$  for  $pp \rightarrow \chi^+ \chi^-$  is well estimated by calculations for pair production of supersymmetric wino-like charginos. For each benchmark, the event simulations are re-scaled by the corresponding LHC prediction for  $\sqrt{s} = 13$  TeV at NLO+NLL [220, 221].

$$\sigma(220 \text{ GeV}) = 903 \pm 54 \text{ fb}, \quad \sigma(324 \text{ GeV}) = 127.7 \pm 9.5 \text{ fb}. \quad (4.5)$$

As observed by the search for displaced leptons at CMS [94], HF events dominate the background by orders of magnitude. To simulate this background contribution,  $2 \times 10^7$  events of  $pp \rightarrow b\bar{b} \rightarrow \ell + X$  were generated. The events were simulated at NLO QCD with `MadGraph5_aMC@NLO 2.6.6` [164], followed by hadronisation using `Pythia v8.243` [168]. Pure hadronic decays for hadrons from  $b$  and  $c$  quarks were disallowed. This lepton enrichment was necessary to have sufficient number of background events after event pre-selections defined in the upcoming subsection. It was verified that the  $p_T(\ell)$  and  $d_0$  distributions after lepton enrichment were similar to those of the leptons from a sample of inclusive  $b\bar{b}$  events in our regions of interest. This method improved the available number of simulated events by at least three orders of magnitude.

The detector effects for both signal and background samples were emulated by passing the events through `Delphes 3.4.1` [197], using the CMS detector response model.

### 4.3.1 Event pre-selections

As in the analysis of the CMS collaboration performed at  $\sqrt{s} = 13$  TeV [94] in 2015, the pre-selection requires one electron and one muon of opposite charge in the final state. Electrons and muons are required to fulfil  $|\eta(\ell)| < 2.4$ . A pre-selection criterion of  $p_T(\ell) > 15$  GeV is required for all leptons. The lepton isolation criteria are  $\text{Iso}(e) < 0.12$  and  $\text{Iso}(\mu) < 0.15$ , with the isolation defined as the sum of the  $p_T$  of all reconstructed particles within a cone of  $\Delta R = 0.2$  around the lepton, divided by the  $p_T$  of the lepton. The pre-selection restricts the events studied, based on an educated estimate dependent on the phenomenology of the STHDM, where is expected the background to dominate completely. At the analysis level, the kinematic properties of the events have been studied to facilitate separation of the signal events from the background events. A selection of  $p_T(\ell) > 20$  GeV is required at the analysis level. `Delphes` also provides information about the lepton transverse impact parameter  $d_0$  and includes a veto for leptons that are created outside the CMS tracking volume.

### 4.3.2 Background extrapolation to low momenta

The 13 TeV 2015 search [94] provides an estimate for background, which will be a reference for the background estimation in this analysis. The  $B$  hadron in an HF jet decays semi-leptonically with a mean free path in the order of a few millimetres. Due to statistical fluctuations in the hadronisation process, as well as imperfect reconstruction, the lepton from the  $B$  hadron decay can be mis-identified to be isolated. This happens with low probability in CMS. Due to the large production cross section of HF jets at hadron colliders in order of microbarns, a sizeable number of background events is still obtained. The estimate is derived in CR1 and cross-checked for closure in CR2. This is extrapolated to obtain the estimate in SRs in Eq. (4.1). The 95% CL upper limits on the number of background events in each SR is obtained from the 13 TeV 2015 CMS analysis [94]. The upper limits on the background, obtained with  $2.6 \text{ fb}^{-1}$  of data, are given below.

$$B_{\text{I}} = 3.2; \quad B_{\text{II}} = 0.50; \quad B_{\text{III}} = 0.019. \quad (4.6)$$

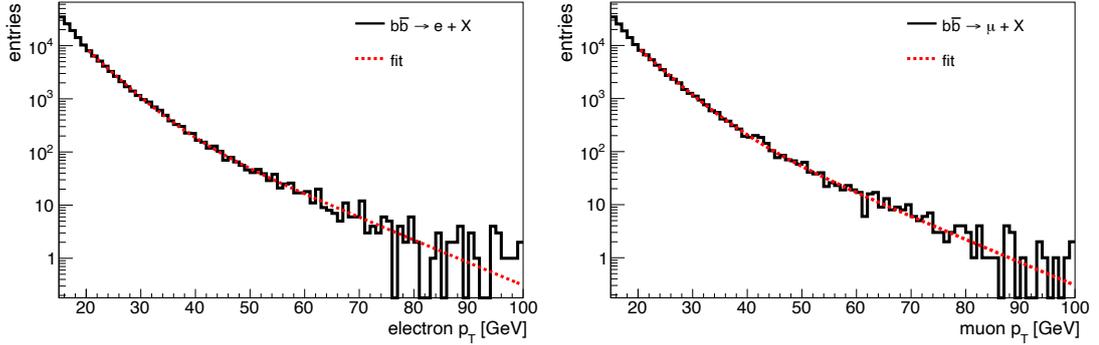


Figure 4.4: The transverse momentum distributions of electrons (left) and muons (right) in  $pp \rightarrow b\bar{b} \rightarrow \ell + X$  at  $\sqrt{s} = 13$  TeV. The distribution is well estimated by the double-exponential fit with Eq. (4.7) (red dotted curve).

In each region, multijet events stemming from HF production are the dominant background. All other SM backgrounds, for instance from top quark-antitop quark production or from production and decay of  $W$  and  $Z$ -boson, are orders of magnitude smaller.

For the simulated HF background process used for background estimation in this study, the opposite charge requirement is not imposed on the lepton pair to retain a sufficient amount of events. The HF background simulation is only relevant to scale the data-driven estimate of 13 TeV CMS analysis to lower lepton  $p_T$ . The leptons come from independent hadronic decays and should have independent kinematics, so can be factorised if no correlation exists. The correlation was verified to be negligible compared to possible statistical fluctuation. For studying the lepton  $p_T$  correlation in background, the opposite sign selection was re-instated.

The  $p_T(\ell)$  distribution of the background leptons is obtained from the simulated lepton-enriched HF background. Leptons stemming from this background show no correlation between  $p_T(\ell)$  and  $d_0$  distributions. A similar feature has been observed in Ref. [217]. Background events that impose a lower  $p_T$  threshold on the leptons can thus be estimated independent of  $d_0$ . The number of background events in the different SRs can be estimated from Eq. (4.6) in Ref. [94] using an overall multiplicative factor that is only a function of  $p_T$ .

Fig. 4.4 shows the transverse momentum distribution of all isolated electrons (left) and muons (right) in the simulated background processes in log  $y$ -scale. An exponential decrease can be observed. The electrons in low- $p_T$  region have a much higher decrease rate compared to the electrons with high  $p_T$  values due to a change in the nature of the leading order physics process that gives rise to the electrons. The exact nature of the leading order physics process is not know. The muons have behave similarly.

Given the shape, the distributions are parameterised with a double-exponential fit to allow extrapolation from high to low transverse momenta. The scaling constant used to define

Table 4.2: Transfer factors  $\kappa_\ell(p_T)$  to extrapolate the number of expected background events from Ref. [94] to regions with a lower threshold on the lepton transverse momentum  $p_T(\ell)$ . The selection used in this analysis is indicated in bold font.

$p_T(\ell)$	$\kappa_e(p_T)$	$\kappa_\mu(p_T)$	$\kappa_{e\mu}(p_T) \pm \kappa_{e\mu}$
35	2.67±0.14	2.06±0.12	5.51±0.42
30	5.94±0.28	4.57±0.24	27.12±1.92
25	14.55±0.66	11.04±0.55	160.65±10.83
<b>20</b>	<b>39.19±1.73</b>	<b>29.08±1.45</b>	<b>1139.46±75.97</b>

extrapolation for electrons and muons called transfer factors are, determined as

$$\begin{aligned} \kappa_e(p_T) &= \frac{\int_{p_T}^{70} d\tilde{p}_T f_e(\tilde{p}_T)}{\int_{42}^{70} d\tilde{p}_T f_e(\tilde{p}_T)}, & f_e(\tilde{p}_T) &= e^{13.65-0.23\cdot\tilde{p}_T} + e^{8.62-0.10\cdot\tilde{p}_T}, \\ \kappa_\mu(p_T) &= \frac{\int_{p_T}^{70} d\tilde{p}_T f_\mu(\tilde{p}_T)}{\int_{40}^{70} d\tilde{p}_T f_\mu(\tilde{p}_T)}, & f_\mu(\tilde{p}_T) &= e^{13.50-0.22\cdot\tilde{p}_T} + e^{8.73-0.10\cdot\tilde{p}_T}, \end{aligned} \quad (4.7)$$

where  $p_T$  is given in units of GeV. The lower limit of the integral over transverse momentum in the denominator at 40 GeV for electrons and 42 GeV for the muons is taken for the CMS 13 TeV 2015 search [94]. The upper limit is taken at 70 GeV. Taking a higher upper limit increases the error margin in the estimation due to the lower number of events in the high  $p_T$  bins and gives only second order contributions to the integral value. This was verified. The transfer factors are listed in Table 4.2 for different thresholds of the lepton momentum  $p_T(\ell)$ .

The total transfer factor for the event rate can be split into independent contributions as

$$\kappa_{e\mu}(p_T) = \kappa_e(p_T) \times \kappa_\mu(p_T). \quad (4.8)$$

owing to minimum correlation between the electrons and muons in the HF background as discussed earlier.

### 4.3.3 Event yields for signal and background

Using the transfer factor from Eq. 4.8 and the upper limits derived at high lepton  $p_T$  from Eq. 4.6, the upper limits on the expected background yield  $B_i$  in each SR  $i$  is obtained via

$$B_i(p_T(e), p_T(\mu)) = (\kappa_{e\mu}(p_T) + 2\Delta\kappa_{e\mu}) \times N_i, \quad (4.9)$$

where twice the uncertainty of the value of the total transfer factor,  $\Delta\kappa_{e\mu}$ , is added to it have a 95% C.L. upper limit on the scaled background event count. The background event count is listed in Table 4.3.

As expected, lowering the lepton  $p_T$  threshold increases the number of expected background events coming from the  $b\bar{b}$  process. At low momentum, the background from the electroweak processes e.g decay of top quark, is at least three orders of magnitude smaller. The overall fraction

Table 4.3: 95% CL upper limits on the expected number of background events, obtained using Eq. 4.9, in the individual signal regions from Eq. (4.1), at  $\mathcal{L} = 2.6 \text{ fb}^{-1}$ ,  $\sqrt{s} = 13 \text{ TeV}$ . The upper limits for (42,40) are adopted from Ref. [94]. The selection used in our analysis is indicated in bold font. The background event count is scaled to an integrated luminosity of  $140 \text{ fb}^{-1}$  in the last row.

$(p_T(e), p_T(\mu))$	$B_I$	$B_{II}$	$B_{III}$
(42,40)	3.2	0.5	0.019
(35,35)	20	3.2	0.120
(30,30)	99	15	0.587
(25,25)	582	91	3.5
<b>(20,20)</b>	<b>4123</b>	<b>644</b>	<b>25</b>
<b>(20,20) [at <math>140 \text{ fb}^{-1}</math>]</b>	<b>221997</b>	<b>34688</b>	<b>1318</b>

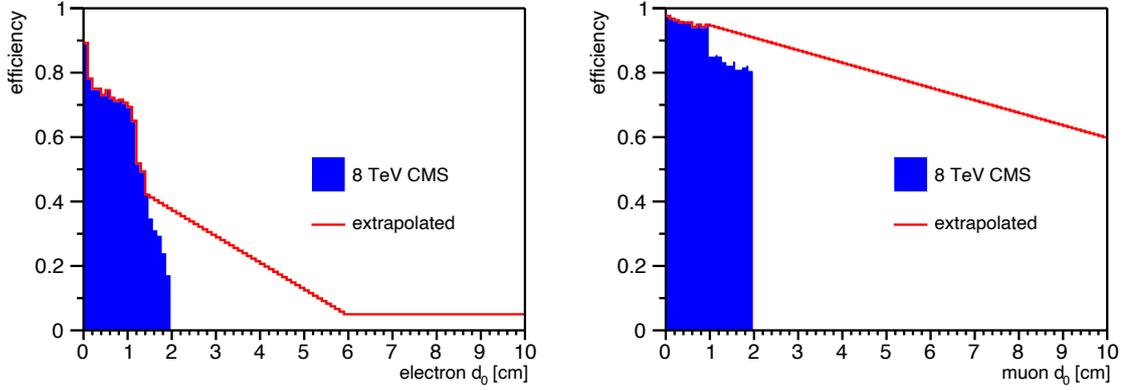


Figure 4.5: The efficiency factors due to electron (left) and muon (right) identification obtained by linearly extrapolating numbers from the 8 TeV search [93]. The electron reconstruction is limited by the extent of the range of the inner tracker while for the muons this range is higher because of the way the detector is constructed with the muon system outside the solenoid. The 8 TeV efficiency in the last bin has been extrapolated to a hypothetical efficiency at 10 cm. A wide range of efficiency factors were tested. The results between these trials were fairly consistent among each other.

from HF events for the total background is thus larger than at high lepton  $p_T$  thresholds.

The signal cross sections from Section 4.2 in addition with the event pre-selection requirements from Section 4.3.1 and lepton identification efficiency are applied to obtain these yields.

Lepton reconstruction deteriorates with increasing  $d_0$ . It is caused by requirements in the reconstruction algorithm that target prompt leptons and consider displaced leptons as fake signature at the trigger level. This effect is modelled using efficiency factors from the 8 TeV search [93]. The 8 TeV search provides these numbers up to the limit  $d_0 = 2 \text{ cm}$ . It has been extrapolated linearly to 10 cm, as shown in Fig. 4.5, the full range of the 13 TeV search [94].

Table 4.4: Expected number of events for the signal benchmarks and 95% CL upper limit on the HF background events in the three signal regions from Eq. (4.1) [94], for a threshold of  $p_T(\ell) > 20$  GeV on the lepton transverse momentum, assuming an integrated luminosity of  $140 \text{ fb}^{-1}$  at  $\sqrt{s} = 13$  TeV.

#	$(m_c [\text{GeV}], \Delta m [\text{GeV}], c\tau_c [\text{cm}])$	$S_I$	$S_{II}$	$S_{III}$
1	(324, 20, 2)	0.38	0.43	1.18
2	(220, 20, 3)	1.18	1.40	5.55
3	(220, 20, 0.1)	139	37	5.98
4	(220, 20, 1)	174	157	283
5	(220, 20, 10)	32	93	318
6	(220, 20, 100)	1.35	2.15	31
7	(220, 40, 1)	1067	980	1826
HF background		221997	34688	1318

The expected signal and background events for  $\mathcal{L} = 140 \text{ fb}^{-1}$  are listed in Tab. 4.4. There is an important caveat in the estimation of the signal and background yield. Owing to the phenomenological nature of the analysis, an estimation of the systematic uncertainty in the yield could not be obtained. For now the analysis assumes that in a search with the data, the systematic uncertainty will be small compared to the yield. The same is also true for the background estimate where the assumption is that the systematic uncertainty causes small changes to the 95% CL upper limit to the background yield. The yield increases from small to moderate displacements and maximises at around an order of displacement in centimetres. The yield for the compressed mass splitting has a low signal-to-background ratio. Only benchmark 7 has a decent signal yield.

To improve the signal sensitivity, it is therefore essential to further reject the  $b\bar{b}$  background using additional techniques. Lowering the  $p_T$  accepts larger number of signal events. Multi-variate analysis can help improve the signal sensitivity and increase the phase-space of the search.

## 4.4 Multi-variate analysis

In what follows, I motivate the event kinematics that are used in the multi-variate analysis and go on to show that neural networks are essential for successfully extracting signal events in LHC data. A basic approach employing only one single kinematic variable was also done to compare its performance to the neural network and it was found the neural network outperforms the cut based criteria to separate signal type events from background.

### 4.4.1 Kinematic distributions

The kinematic properties of the leptons from  $\chi^\pm \rightarrow \chi^0 \ell \nu$  decays, except for the impact parameter, are dependent on  $\Delta m$  only. Fig. 4.6 shows that for  $\Delta m = 20$  GeV, the  $p_T$  distribution is similar to that from the HF background. For higher  $\Delta m$ , e.g 40 GeV, one could consider using the  $p_T$

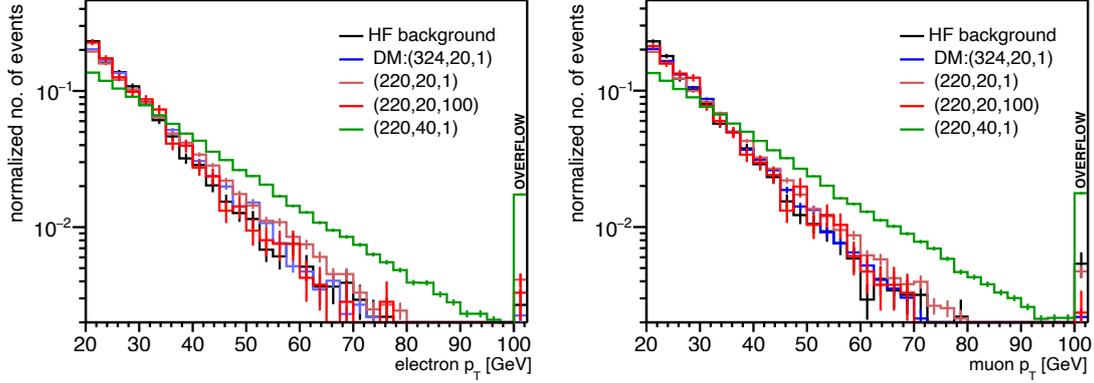


Figure 4.6: The transverse momentum of muons (left) and electrons (right) compared between the background (black) and different signal benchmarks characterised by  $(m_c, \Delta m, c\tau_c)$  shows that for  $\Delta m = 20$  GeV (red and blue), the  $p_T$  distributions are similar for signal and background. Unlike the scenario for  $\Delta m = 40$  GeV (green),  $p_T$  cannot be used to discriminate signal from background for the compressed co-scattering dark matter scenario. The  $p_T$  of the lepton is independent of its displacement, as can be seen by comparing the two red curves. Distribution normalised to unity.

distribution as a discriminant. For the soft lepton signals, using only the  $p_T$  distribution as a discriminant is no longer an option.

The transverse impact parameter is a good discriminator as expected because it was used in Ref. [94]. Fig. 4.7 shows normalised distributions of  $d_0$  for the benchmarks defined in Tab. 4.1 and for the HF background. The background peaks at  $\mathcal{O}(0.1\text{mm})$ , which is consistent with the decay length of a  $B$  meson from background,  $c\tau_B \approx 0.5$  mm. Benchmark 3 ( $m_c = 220$  GeV,  $\Delta m = 20$  GeV,  $c\tau_c = 0.1\text{cm}$ ) offers a good comparison, as its displacement  $c\tau_c = 1$  mm is approximately of the same order as the  $B$  meson decay length. All other signal distributions are shifted to larger displacements. Depending on their displacement, they achieve higher sensitivity in different signal regions. SR I is sensitive to decay lengths up to  $c\tau_c \approx 2$  mm and SR II and III probe larger mean decay lengths up to 1 m. Sensitivity beyond  $d_0 \approx \mathcal{O}(10$  cm) is limited by the low lepton reconstruction efficiency in the tracker of the CMS detector.<sup>2</sup>

The proton beams at LHC, discussed in section 2.1, have equal and opposite momentum with  $\sqrt{s} = 13$  TeV. The net momentum along the beam axis is negligible compared to the transverse component. The charged mediators,  $\chi^+$  and  $\chi^-$ , are thus produced in opposite angular direction to each other with a moderate effective boost in the transverse direction. Due to this overall boost, daughter products from their decay, the DM candidate ( $\chi^0$ ),  $\nu_\ell$  and  $\ell$ , have low angular separation between them. This results in a signature of oppositely charged leptons produced with in angular opposite direction, with a significant fraction in the central region (low  $|\eta|$ ) of the detector, and a moderate missing transverse momentum ( $\cancel{p}_T$ ).

The background leptons come from a distinctly different physics process. HF hadrons e.g  $B$  hadrons are copiously produced via gluon-gluon fusion in a hadronic collision, like the pp collisions at the LHC, and are highly boosted. In the transverse region of the detector, the HF hadrons have a uniform distribution in pseudorapidity. Leptons are frequent decay products

<sup>2</sup>Recent developments allow for reconstruction of muons with decay length of  $\mathcal{O}(1$  m).

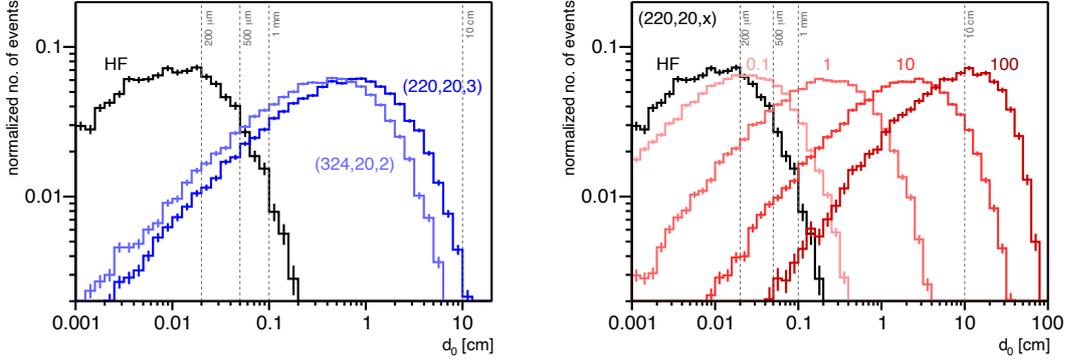


Figure 4.7: The transverse impact parameters  $d_0$  of the selected electron-muon pair for dark matter benchmarks 1 and 2 (left) and benchmarks 3 to 6 with decay length of the charged dark state  $c\tau_c$  varying over orders of magnitude (right), compared to that of the heavy flavor (HF) background. To allow a comparison of the shapes, the distributions are normalised to unity. The  $d_0$ -based identification efficiencies are not included in this figure so that the displacement of the leptons with increasing decay length can be observed without any bias. They are included in the calculation for the final yields.

from leptonic  $b$  to  $c$  decays. The leptons coming from the decay of HF jets are as highly boosted as the mother hadron. This can result in the leptons from the background being collimated or going in the opposite direction. Collimated leptons in the background arise from the decay of the same HF hadron. Leptons going in opposite directions stem from separate mother hadrons.

The leptons, reconstructed jets<sup>3</sup>, and missing transverse momentum in the event are used to define variables to understand and separate signal from background. Based on the event physics properties, nine kinematic variables are identified to distinguish the signal-like events from background events.

### $\Delta R(e, \mu)$

The angular correlation between the leptons serves as an important distinction feature between signal and background. The angle between the two leptons is quantified in terms of the angular separation in  $\eta$  and  $\phi$  as

$$\Delta R(e, \mu) = \sqrt{\Delta\eta^2(e, \mu) + \Delta\phi^2(e, \mu)}. \quad (4.10)$$

Fig. 4.8, left, shows the  $\Delta R(e, \mu)$  distributions for the signal benchmarks and for the HF background. Soft leptons of the signal processes tend to be in the opposite direction, with a maximum around  $\Delta R(e, \mu) = \pi$ . In comparison, background leptons feature a different distribution with two peaks at 0 and  $\pi$ , corresponding to collimated and back-to-back leptons, respectively.

### $\Delta\phi(\ell_1, \mathbf{p}_{T\ell_j})$

This variable is the azimuthal angular difference between the leading lepton and the transverse component of the the vector sum of the reconstructed leptons and jets in the event. In the

<sup>3</sup>For the jets the default anti- $k_T$  definition from Delphes has been used and the selections  $p_T(\text{jet}) > 20 \text{ GeV}$ ,  $|\eta(\text{jet})| < 2.5$  [193] are required

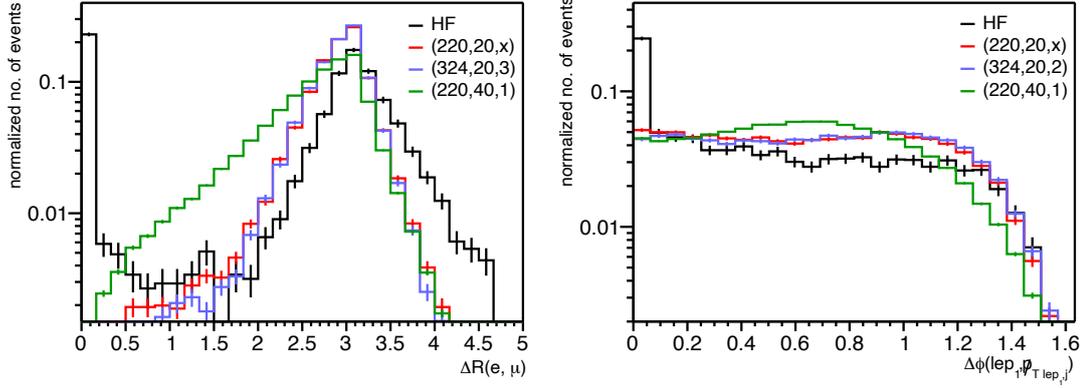


Figure 4.8: Angular lepton separation  $\Delta R(e, \mu)$  (left) and azimuthal angular separation between leading lepton and vector sum of reconstructed objects  $\Delta\phi(\ell_1, \vec{p}_{T\ell_j})$ . The distributions are normalised to unity. The mean decay of the leptons in the DM model benchmarks do not affect the kinematics, hence all benchmarks with  $\Delta m = 20$  GeV are represented by distribution with deep blue color.

background process, events with collimated leptons tend to peak at zero. For the background events with leptons emitted angularly opposite to each other, the neural network should find correlations with additional information from a reconstructed jet. The signal distribution is remarkably different in nature from the background as can be seen from Fig. 4.8 right.

### $\cancel{p}_T/\sqrt{H_T}$

$\cancel{p}_T$ , the missing transverse momentum of the event is calculated by taking the vector sum of all available particle flow candidates in the event. This value is divided by the  $\sqrt{H_T}$ , with  $H_T$  being the scalar sum of the transverse energy of all the particle flow candidates, which is a measure of the significance of the missing transverse momentum in an event. This is different from  $H_T(j)$  which is the scalar sum of transverse momenta of reconstructed jets. For the background heavy flavour processes, there is very little imbalance in the net transverse momentum of the event arising from the systematic errors and the incomplete coverage of the detector in the forward region. For the signal processes, however, there is a net imbalance in the vector  $p_T$  sum of all particles in the detector. This comes from the neutrino pair and the DM candidates that pass through the detector without being detected. From Fig. 4.9 left, it can be seen that there is a missing transverse momentum significance for the signal benchmarks higher than the HF background. Increasing the mass of the candidate increases the net loss in transverse momentum while increasing the mass splitting leads to centrally emitted leptons decreasing the net transverse momentum.

### $(\cancel{p}_T/\sqrt{H_T})_{\ell_j}$

The missing transverse momentum significance was also computed for the objects that passed our object identification and event selection requirements. Comparing Fig. 4.9 left and right, this variable has a distinctly different shape. Since the leptons in the signal are emitted angularly opposite to each other, their vector transverse momentum sum is much smaller than the total energy deposit. In the background, there is a large number of events where the leptons that are

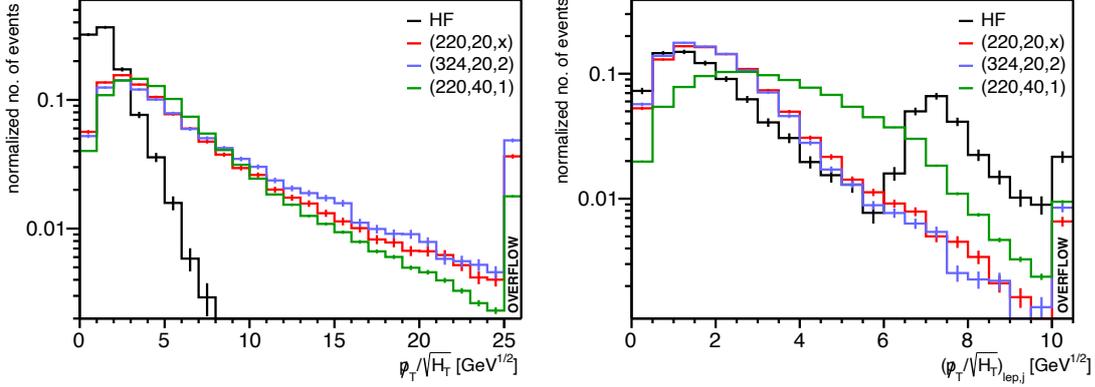


Figure 4.9: Missing transverse momentum significance  $\not{p}_T/\sqrt{H_T}$  (left) for representative signal benchmarks and for the heavy flavor (HF) background and object based missing transverse momentum significance (right) from objects that pass selection criteria. The distributions are normalised to unity. The mean decay of the leptons in the DM model benchmarks do not affect the kinematics, hence all benchmarks with  $\Delta m = 20$  GeV are represented by distribution with deep blue color.

produced in the same direction pass the event selection and have a comparable vector  $p_T$  sum to the total scalar energy deposit in the event. This causes the peak at  $7.5 \sqrt{\text{GeV}}$ . The events with value for this variable below  $6\sqrt{\text{GeV}}$  are correlated to the angular distance  $\Delta R(e, \mu) \approx 3$ . For  $(\not{p}_T/\sqrt{H_T})_{\ell,j} > 6$ , this variable is expected to be a function of the kinematics of leptons and jets in the event.

### $m_T(\ell_1, \not{p}_T)$

Transverse mass of the lepton with the missing transverse momentum in the event. The leading lepton is projected to the transverse plane and its invariant mass with  $\not{p}_T$  is taken. From Fig. 4.10 left, it can be seen that the background for this variable peaks at values close to zero whereas the signal has a non-negligible peak at  $\sim 50$  GeV that fluctuates slightly with the mass splitting value. For different mass splitting benchmark and the same mass of the charged mediator, the peak shifts by a little, showing that this variable is robust to mass splitting. The variable could be largely dependent on the missing transverse momentum from the neutrino that could explain this robustness.

### $\alpha_T$

$\alpha_T$  has been traditionally used to understand the imbalance in transverse momentum [222] among various objects in the event. It is a well-established variable and would provide easy comparison to existing results.

$$\alpha_T = \frac{p_T(\ell_2)}{m_T} \quad (4.11)$$

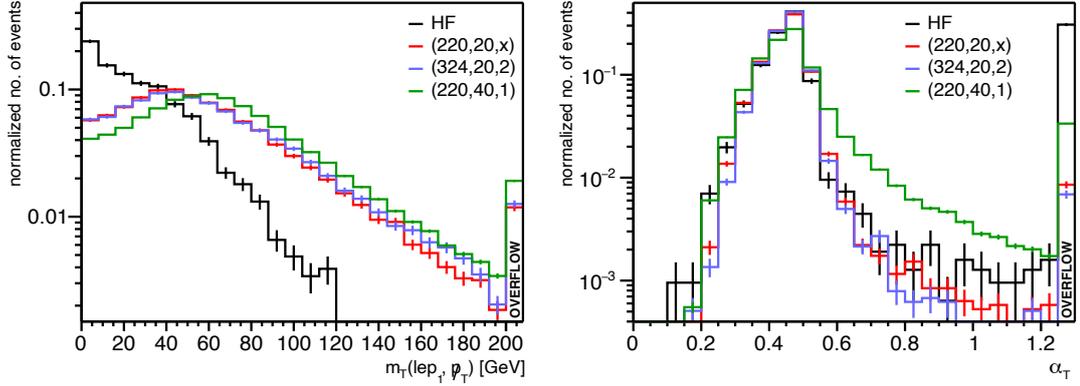


Figure 4.10: Invariant mass of the transverse component of leading lepton and missing transverse energy  $m_T(\ell_1, \cancel{p}_T)$  (left) and  $\alpha_T$  (right), a variable which characterises the imbalance in lepton transverse momentum. The distributions are normalised to unity. The mean decay of the leptons in the DM model benchmarks do not affect the kinematics, hence all benchmarks with  $\Delta m = 20$  GeV are represented by distribution with deep blue color.

with  $m_T$  defined as

$$m_T = \sqrt{\left(\sum_{\ell} |\vec{p}_T|\right)^2 - \left(\sum_{\ell} \vec{p}_T\right)^2} \quad (4.12)$$

$m_T$  is the transverse mass between the leading lepton  $\ell_1$  and the subleading lepton  $\ell_2$ , calculated from the Lorentz vector sum of the leptons in the transverse plane. For a symmetric process as the singlet-triplet model, a smaller imbalance in the lepton  $p_T$  compared to the background where the kinematics of the two leptons are independent, Fig. 4.10, is observed.

### sphericity and transverse sphericity

Sphericity and sphericity [223, 224] measure the circular symmetry of an event. Only leptons in an event have been used to calculate the symmetry in an event. Events with circular symmetry peak at 1 while events without circular symmetry peak at zero. Background events are less balanced in their lepton  $p_T$  compared to the signal events. The signal events are not perfectly spherical due to the  $\cancel{p}_T$ , but have higher lepton  $p_T$  balance. This is reflected in Fig. 4.11 where the background process peaks at zero.

### $H_T(j)$

The scalar sum of transverse momenta of all reconstructed jets is motivated by the expected presence of jets in background events. A significant difference between signal and background is not seen, however, after the event pre-selections. In background events these jets will be kinetically correlated to the  $\cancel{p}_T$  of the event. In signal, jets arise from the ISR and FSR and do not have any association to the kinematic properties of the signal. Reconstructed jets in an event can be used to identify background, hence it is provided to the discriminator which might succeed in finding correlation among the background events.

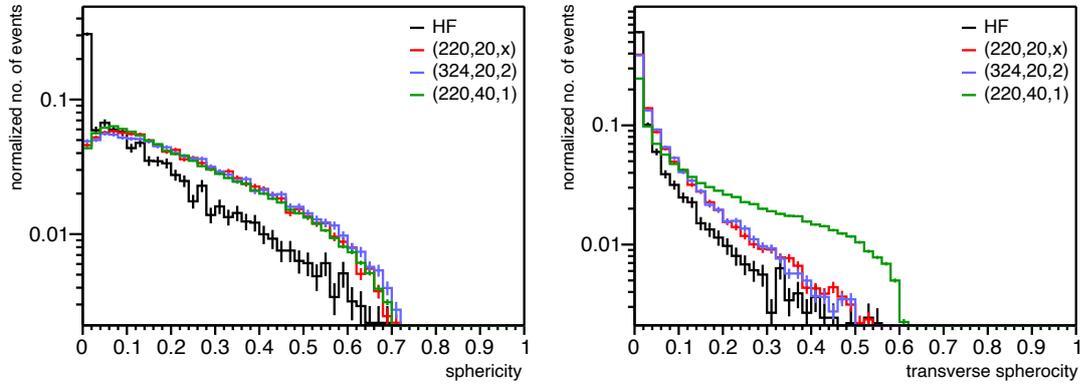


Figure 4.11: Sphericity (left) and transverse sphericity (right) capture the spherical symmetry of the final state of leptons in an event. The distributions are normalised to unity. The mean decay of the leptons in the DM model benchmarks do not affect the kinematics, hence all benchmarks with  $\Delta m = 20$  GeV are represented by distribution with deep blue color.

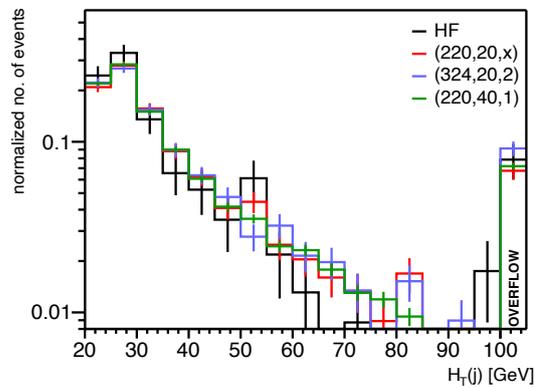


Figure 4.12: The scalar sum of energy of reconstructed jets in the event  $H_T(j)$  in the case that the final state of an event comprises jets. The distributions are normalised to unity. The mean decay of the leptons in the DM model benchmarks do not affect the kinematics, hence all benchmarks with  $\Delta m = 20$  GeV are represented by distribution with deep blue color.

These variables are well known for their robustness in hadron collider environment and have been verified to be almost completely independent of  $d_0$ . This is also observed in the signal kinematics which is dependent only on  $\Delta m$  and shows only modest kinematic differences due to statistical effects during MC sample generation among the benchmarks.

The variables were used in a simple multi-layer neural network that outputs the probability for an event to be of signal-like type. The neural network based approach establishes correlations between the variables and finds the optimal phase space of the variables for signal-like events. It is optimised to obtain the maximum sensitivity for discrimination between signal and background type events.

#### 4.4.2 Neural network structure

The nine kinematic variables described in Section 4.4.1 are combined in a neural network. Ranked by performance, these are

$$\{\not{p}_T/\sqrt{H_T}, \Delta R(e, \mu), m_T(\ell_1, \not{p}_T), (\not{p}_T/\sqrt{H_T})_{\ell,j}, \Delta\phi(\ell_1, \not{p}_T), \text{sphericity}, \alpha_T, \text{sphericity}, H_T(j)\}. \quad (4.13)$$

The transverse impact parameter of the leptons have not been used as a parameter to distinguish signal from background. This avoids the possibility of any unrealistic modelling of the transverse impact parameter  $d_0$ . This also makes the search inclusive to all possible coupling scenarios between the mediator and the dark partner.

The neural network is implemented in SCI-KIT LEARN and TENSORFLOW, including KERAS [225, 226]. It has two fully connected layers, each with 18 nodes and a constant offset factor. The fully connected layers use activation with a rectified linear unit [227]. The neural network output is a number between 0 and 1, with 1 being maximally signal, and 0 being maximally background-like according to the neural network classification. Only the DM model with  $m_c = 324$  GeV,  $\Delta m = 20$  GeV and  $c\tau_c = 2$  cm, listed as benchmark 1 in Tab. 4.1 has been used as the signal, together with the HF background sample to train the neural network. The performance of the trained NN model has been evaluated on all others but benchmark 7 and found to work equally well implying a benchmark agnostic performance. The performance on benchmark 7 is also good as discussed in detail in the following section.

Equal number of simulations in the background and signal sample are used for training the model. This is limited by the number of background simulations at 4108. A dropout regularisation of 20% has been used between the hidden layers [228]. The training was performed on 80% of the available data. The rest of the data was used to independently evaluate the performance of the model. This enabled a crosscheck to prevent over-training.

The model was trained for a 1000 epochs. Figure 4.13 shows the progress of the loss function (left) and the prediction accuracy of the model (right) with the epochs. The loss function reduces rapidly for the model and converges and stabilises by the higher epoch counts. Consequently, the prediction accuracy of the model increases and plateaus. This resulted in an overall accuracy of  $\sim 85\%$ . Training further does not result in an improvement of accuracy. The training and testing sample converge in the higher epochs. This quantifies the optimal performance of a model on an independent data-set drawn from the same sample space. It is able to classify effectively between signal and background.

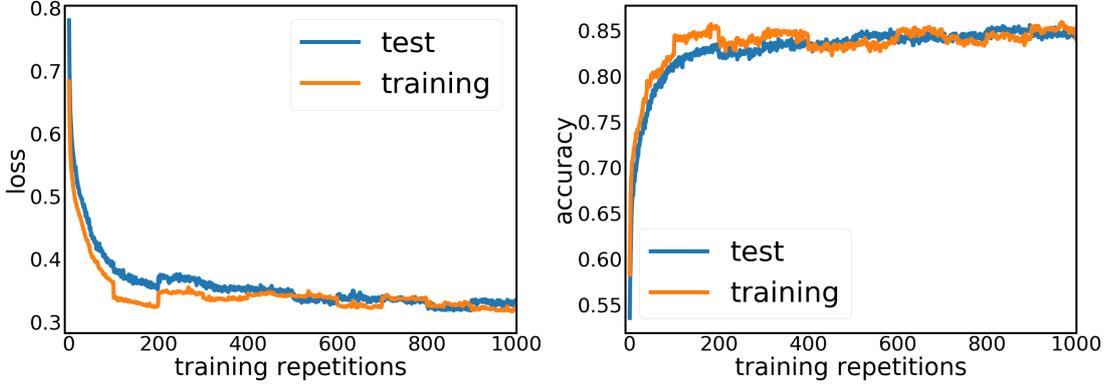


Figure 4.13: Loss (left) and accuracy (right) of the neural network at the end of each training repetitions (epoch). The loss decreases with each epoch and consequentially the accuracy of the model to predict signal and background increases. After sufficient number of epochs both loss and accuracy flatten signifying that optimisation has been achieved. For a well-trained model, the training and testing values for loss and accuracy converge. This also implies that the model was neither over-trained nor under-trained.

### 4.4.3 Classification performance

The neural network outputs two numbers that, by construction, sum to unity. These can be interpreted as the probability for an event to be of signal type or background type. The neural network output discriminant,  $NN$ , is chosen to quantify the probability to be signal type. Figure 4.14, left, shows the distribution for this discriminant. Based on the MC information, the simulated events can be tagged as signal or background. As observed from the MC truth information of the simulations, signal type simulations peak at 1, while the background type simulations peak at 0. The same model, that was trained on one of the benchmarks, was used to discriminate the other signal benchmarks as well. It was found to be effective for all benchmarks including benchmark 7 which had a 40 GeV mass splitting and hence significantly different kinematics. The performance deteriorated mildly for benchmark 7 as can be seen from the ROC curve in Fig. 4.14 right.

Fig. 4.14, right, is the ROC curve, that quantifies the performance of the model. The x-axis of this curve is the fraction of rejected background events and the y-axis of the curve is the fraction of accepted signal events. To decide whether an event is signal or background a *cut* value is chosen such that, for all  $NN > cut$ , the event is considered to be signal and  $NN \leq cut$ , the event is considered as background. The higher the required background rejection, the larger is the *cut* value, the lower is the signal efficiency.

$$\text{background rejection} = \frac{\text{HF events with } NN \leq cut}{\text{total no. of HF events}} \quad (4.14)$$

$$\text{signal efficiency} = \frac{\text{MC truth signal events with } NN > cut}{\text{total no. of MC truth signal events}} \quad (4.15)$$

The ROC curve, Fig. 4.14, right, shows that  $NN$  is relatively independent of the benchmark model, even though it was trained on one specific model. This was expected, because the input

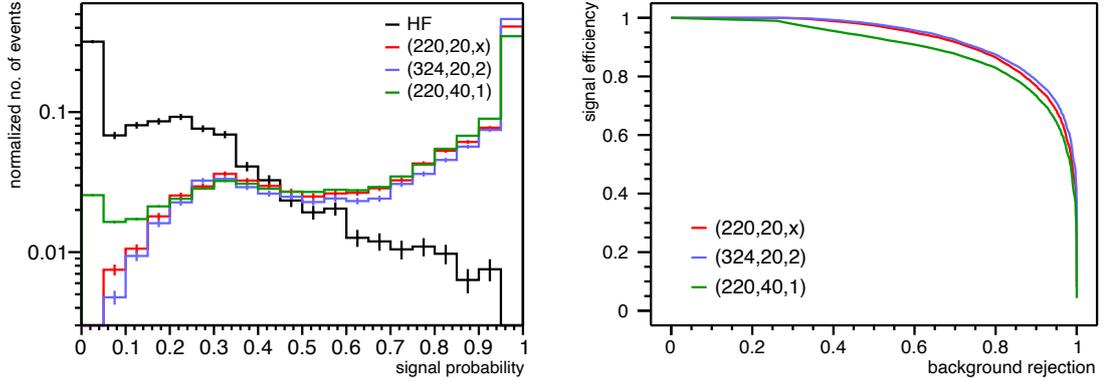


Figure 4.14: The neural network discriminant (left) and ROC curve (right) for selected signal benchmarks and heavy flavor (HF) background. The neural network discriminant, by construction, is the probability for an event to be signal type and hence is termed as signal probability. The distributions are normalised to unity. The classifier model has been trained using only the benchmark 1 - (324,20,2). It has been evaluated on the other signal benchmarks.

variables in Section 4.4.1 show only minor kinematic differences between the different signal benchmarks. Hence, the NN training based on benchmark model 1 can be used with good sensitivity to other benchmarks, without the need to retrain the network.

Fig. 4.15 and Fig. 4.16 show the neural network discriminant in the different signal regions for MC simulation scaled to an integrated luminosity of  $140 \text{ fb}^{-1}$ . SR III has the highest discrimination power between signal and background for most benchmarks. There is a general trend of an increase in the fraction of signal events with increasing displacement in the signal regions. For moderate decay lengths of the lepton, additional sensitivity is obtained by combining all three signal regions.

For the current neural network, background is highly reduced at high NN values. For the benchmarks 3 to 6, with leptonically decaying dark states, there is a clear excess at high NN values (Fig. 4.16 left). Applying a cut at  $\text{NN} \geq 0.9$  reduces the background to less than 1% of its initial value while retaining a large fraction of the signal. These yields are listed in Tab. 4.5 for all the signal regions. The DM motivated benchmarks, have a very low yield due to suppressed mediator decays into leptons, see Tab. 4.1.

Benchmark 7 with  $\Delta m = 40 \text{ GeV}$  (Fig. 4.16 left) has a higher signal-to-background ratio than for  $\Delta m = 20 \text{ GeV}$  at all NN values. The signal leptons generally carry more transverse momentum than those from HF decays for the earlier resulting in an easier signal like classification.

The sensitivity can be improved by using a binned likelihood fit of the NN output instead of a cut at  $\text{NN} \geq 0.9$ . This approach would especially matter for the second DM benchmark scenario, where over five events are expected in the SR III before applying the NN cut.

## 4.5 Predictions for the LHC

From the expected yields and the upper limit on the number of background events in Tab. 4.5 with the neural network cut  $\text{NN} \geq 0.9$  and Tab. 4.4 without employing the multi-variate analysis, the

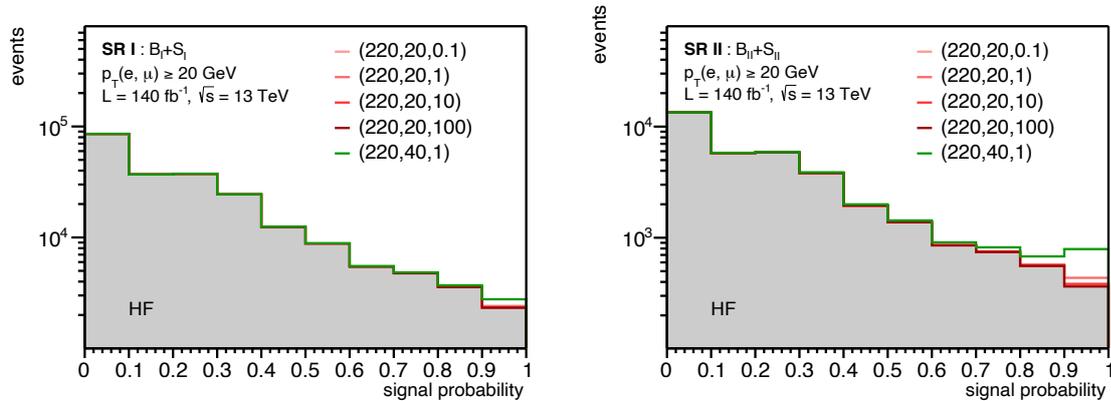


Figure 4.15: The total signal yield in SR I (left) and SR II (right) in log scale as function of the neural network discriminant for MC simulation scaled to  $\mathcal{L} = 140\text{fb}^{-1}$ . From 0 to 1, the background decreases. Signal is expected in bins close to 1. The overall number of events is dominated by background (grey shade). Barely any signal is seen for the larger mass splitting  $\Delta m = 40\text{ GeV}$  (pink) in SR I in the final bin. There is larger proportion of signal events in the more displaced SR II with the moderate displacement 1 cm,  $\Delta m = 20\text{ GeV}$  benchmark showing excess of signal events compared to the background in the final bin.

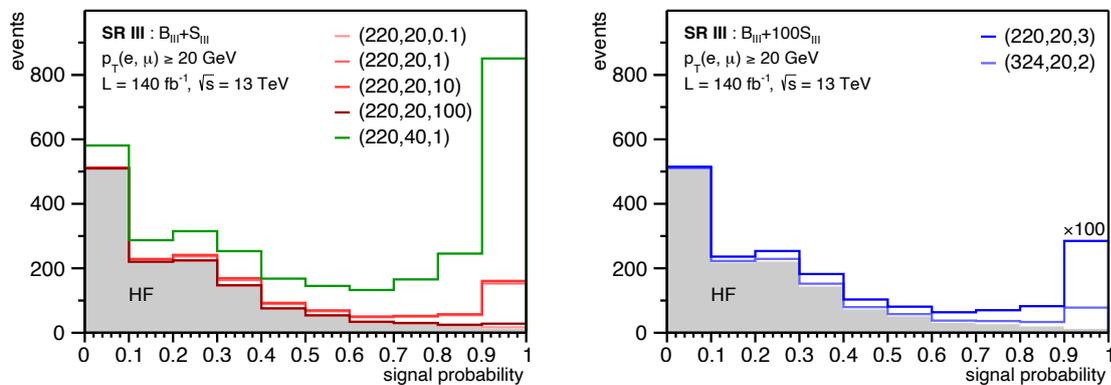


Figure 4.16: Distributions of the neural network discriminant for the signal benchmarks 3 to 7 (left) and the dark matter benchmarks 1 and 2 (right), and the HF background in the signal region SR III, determined for MC simulation scaled to an integrated luminosity of  $140\text{fb}^{-1}$  at  $\sqrt{s} = 13\text{ TeV}$ . The dark matter benchmarks (blue), right, have relatively low yields. The signal distributions have been scaled up by a factor 100 to make them visible to the reader.

Table 4.5: Expected number of events for the signal benchmarks and 95% CL upper limits on the HF background events in the three signal regions from Eq. (4.1), for  $p_T(\ell) > 20$  GeV and  $NN \geq 0.9$  (see text). Assumed is an integrated luminosity of  $140 \text{ fb}^{-1}$  at  $\sqrt{s} = 13$  TeV.

#	$(m_c [\text{GeV}], \Delta m [\text{GeV}], c\tau_c [\text{cm}])$	$S_I$	$S_{II}$	$S_{III}$
1	(324, 20, 2)	0.21	0.23	0.64
2	(220, 20, 3)	0.57	0.67	2.71
3	(220, 20, 0.1)	68	19	3.06
4	(220, 20, 1)	84	72	139
5	(220, 20, 10)	15	20	147
6	(220, 20, 100)	0.79	0.70	14
7	(220, 40, 1)	449	427	837
HF background		2323	363	14

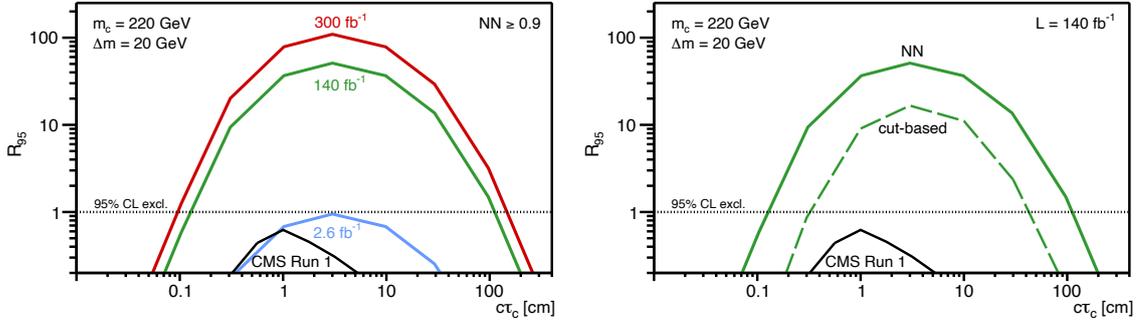


Figure 4.17: Projected signal exclusion limits at the LHC as a function of the decay length of the charged dark state,  $c\tau_c$ . The expected sensitivity obtained from a neural network analysis is shown for different integrated luminosities (left) and in comparison with a simple kinematic cut-based analysis for  $140 \text{ fb}^{-1}$  (right) at  $\sqrt{s} = 13$  TeV. A likelihood ratio  $R_{95} > 1$  indicates that a certain decay length is excluded at more than 95% CL. For comparison, existing upper bounds from a displaced lepton search at  $\sqrt{s} = 8$  TeV (plain black curve) [217] is shown with a similar threshold at  $p_T > 20$  GeV.

significance of rejecting a background only hypothesis for both scenarios using  $R_{95}$ , see Eq. 4.3, can be calculated. This expected significance is determined assuming an integrated luminosity of  $\mathcal{L} = 140 \text{ fb}^{-1}$  and is the combined sensitivity from all three signal regions. The sensitivity is shown in Fig. 4.17 as a function of the decay length of the charged dark partner, illustrating the sensitivity if full Run 2 data with lowered lepton  $p_T$  would be available.

With  $140 \text{ fb}^{-1}$  of data, compressed dark sectors with decay lengths over 3 orders of magnitude from 1 mm to 1 m can be excluded. Maximum sensitivity is obtained for 3 cm. For benchmarks with larger or smaller displacement, The available limitation of the detector or of the reconstruction algorithm is reached and thus cannot increase sensitivity.

Fig. 4.17 right compares the significance with and without the neural network classification.

Just by reducing the transverse momentum cut to 20 GeV, sensitivity to compressed sector is observed, NN being essential to have wider coverage of decay length.

One of the basic problems, that can be addressed with the upcoming Run 3 program of LHC, is increased sensitivity to displaced objects. The CMS program was highly focused on prompt objects in the scope of new physics beyond DM models. However, given the lack of any evidence in prompt searches and increasing models with displaced objects, it is important to open up the CMS searches to displaced objects.

The first technical challenge would be to improve the triggering sensitivity for these models. Displaced particles require dedicated algorithms at the trigger level, for events with displaced objects to be selected. This is a major challenge. The following chapters of this thesis will try to address this for a small variety of the CMS physics program.

Reducing the transverse momentum threshold increases the background exponentially. This is a major drawback especially at the trigger level. The increase in rate can be mitigated by use of event variable requirements of tighter restrictions on the object to reduce the background like events by as much as possible. This study already suggests some of the kinematic features that can be exploited to this end.



# Chapter 5

## New HLT Paths for Soft Displaced Leptons

*The work related to the standard trigger menu of creating new triggers for the STHDM in the CMS experiment is discussed in this chapter.*

It was concluded from the phenomenology study of the Singlet-Triplet Higgs Portal Dark Matter (STHDM) model in the previous chapter that it is possible to search for dark matter in CMS as a function of the lifetime of the  $\chi^\pm$ -particle. The gain in search sensitivity was achieved by lowering the transverse momentum ( $p_T$ ) threshold on the leptons to 20 GeV from 40 GeV for muons and 42 GeV for electrons. The high  $p_T$  threshold on the leptons arises from the limited event storage rate at the HLT. Lowering the threshold in the trigger increases the acceptance for the STHDM model while causing the event rate to drastically surge by  $\mathcal{O}(10 - 100)$ , well beyond the resource limitations of the CMS experiment.

The run 3 program of the LHC provides an opportunity to build triggers to increase acceptance for the soft displaced leptons predicted by the STHDM. The expected increase in the background rate can be mitigated by placing additional constraints on the leptons. Leptons from the decay of  $\chi^\pm$  in the CMS detector have a characteristic signature which can be used to build object-level selections on the leptons. Such selections will reduce the background rate with a low impact on the signal events. This results in a low event rate at the HLT. Special attention has to be paid to the time complexity of the implemented algorithm for computing the new identification variables. In addition, the rate constraints of the HLT computer farm have a limited processing capacity. If the timing of the HLT menu is so high that the HLT farm cannot process events at the level-1 (L1) trigger throughput rate, it could lead to disastrous consequences, including a crash of the HLT farm during proton collisions.

As was observed from the STHDM phenomenology, the properties of particular interest are the displacement of the leptons and the angular correlation of the leptons in the event. The missing transverse energy of events with  $\chi^\pm$  decay is typically less than 60 GeV. Comparatively, the lowest energy threshold for unscaled missing transverse energy (MET) based triggers was 200 GeV, and the lowest MET in cross-triggers with other objects was 50 GeV in the run 2 CMS HLT menu. The missing transverse energy of events in the STHDM is too soft by CMS standards at the trigger level and hence is not used in the triggers.

The LHC run 3 is designed to provide a similar luminosity as run 2. A large effort was

dedicated to taking advantage of the long shutdown of the LHC before run 3 to design triggers targeting exotic signatures in the search for physics beyond the SM. An additional rate of 100 Hz was allocated for searches of exotic particles. The new triggers could take advantage of this additional rate allowance for exotics at the CMS HLT.

The CMS physicists have performed many searches for high-energy displaced particles previously. In particular, a search exists for displaced leptons in the CMS detector [218]. It targets the decay of the supersymmetric top squark. Events for this search are triggered by a set of five triggers, one for  $\mu e$  and two each for  $\mu\mu$  and  $ee$  detector signatures. The HLT paths for  $\mu\mu$  require two muons with either  $p_T > 33$  GeV and  $d_0 > 0.01$  cm or  $p_T > 43$  GeV. The HLT path for  $\mu e$  requires a muon and a photon with  $p_T > 43$  GeV. The HLT paths for  $ee$  require two photons with either transverse energy above 22 GeV and invariant mass above 90 GeV or transverse energy above 70 GeV. The muons in the triggers are exempt from being produced at the primary vertex. Photon triggers are used for electrons because the standard electrons are not designed to recognise displaced tracks. This will be elaborated further in the section on electron triggers. The search requires two isolated offline leptons that are produced from the decay of two different particles. The isolation is required in an angular cone  $\Delta R < 0.3(0.4)$  for electrons(muons). The offline  $p_T$  threshold is required to be 45 GeV for the  $\mu\mu$ , 75 GeV for the  $ee$  and 45 GeV for the  $\mu e$  detector signature. The offline  $p_T$  threshold was chosen based on the region where the trigger efficiency plateaued. No excess was observed in the displaced lepton search.

The triggers used in the above search can potentially be used to search for the STHDM  $\chi^\pm$ . One of the major limitations in using these triggers comes from their high  $p_T$  thresholds. As was observed from the phenomenological study in the previous chapter, the CMS experiment has the potential to search for STHDM in over four orders of the lifetime of  $\chi^\pm$  from  $\mathcal{O}(0.1)$  cm to  $\mathcal{O}(100)$  cm if the offline  $p_T$  threshold is reduced to 20 GeV. In the following study, the existing triggers used in the displaced lepton search are modified to be sensitive to low energies using properties unique to displaced leptons. Especially the transverse impact parameter,  $d_0$ , is useful for both electrons and muons. The displaced electrons provide a time delay in the CMS ECAL that has also been used in the triggers.

Three different signatures are targeted:  $\mu\mu$ ,  $\mu e$  and  $ee$ . The methodology followed to create the new soft displaced di-lepton triggers is similar among all the final states. I start by choosing a reference run 2 trigger, which gives me an estimate of the trigger rate. The run 2 trigger with the lowest  $p_T$  threshold in the latest CMS search for displaced leptons, mentioned above, is chosen for each di-lepton final state [218]. This strategy also helps maintain consistency by using a similar reconstruction algorithm for displaced leptons at the HLT as other CMS triggers. The selections in the run 2 trigger can be removed completely to obtain the maximum signal acceptance, which consequentially results in an extremely high rate. A new set of selections is imposed on the physics objects in the above trigger to reduce the rate to fit within HLT rate constraints. The selections are optimised for maximum signal acceptance for the allowed trigger rate. The  $p_T$  threshold is the last variable on which the selections are made for the final trigger rate of the new HLT. The signal MC can be used to assess the change in signal acceptance with each selection.

As explained in Section 2.3.6, each HLT path is an aggregation of physics objects and trigger filters that implement selections and hence filter events based on the attributes of the physics objects. Conventionally, only the four-momentum information of the trigger filters is stored. For trigger studies, I also stored the physics objects and their relevant kinematics and identification variables which involved re-running the HLT offline. It can happen while re-running an HLT configuration with the updated CMS software that the objects and filters are not created with

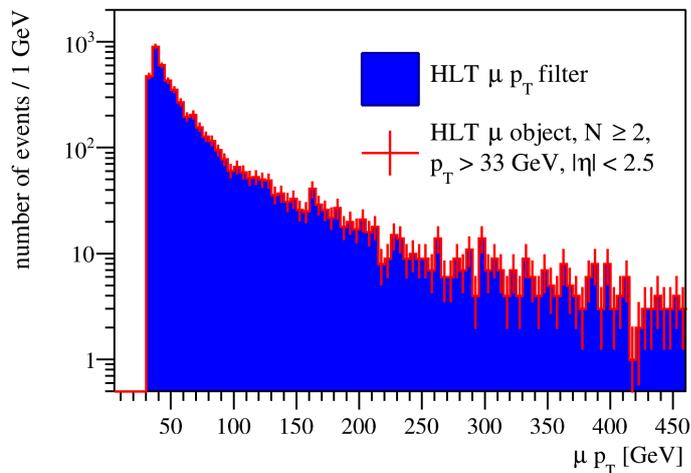


Figure 5.1: The plot shows the  $p_T$  distribution for the online muon  $p_T$  filter output (blue filled) and the online reconstructed muon objects (red hollow) for the trigger `HLT_DoubleMu33NoFiltersNoVtx_v`. An identical set of selections used in the online filter have been applied to the reconstructed objects. The  $pt$  distribution starts at 33 GeV, which is one of the selections. The distributions are identical, implying a successful cross-check of the HLT physics objects with the HLT filters.

identical modules. Using the same selection as the HLT filters on the physics objects, I performed successful cross-checks with the filter outputs. Figure 5.1 shows the comparison between the muon filter output and the muon objects remaining after a selection,  $p_T > 33$  GeV, identical to the filter in the HLT path `HLT_DoubleMu33NoFiltersNoVtx_v`. The names of triggers usually end with a version which refers to minor technical changes in the trigger during data taking in a year and is usually neglected when using the trigger name. The two outputs have an identical distribution signifying a successful reproduction of the run 2 trigger.

This chapter is divided to four sections depending on the detector signature they target: Section 5.2 for  $\mu\mu$ , Section 5.3 for  $\mu e$  and Section 5.4 for  $ee$  and one section in the beginning which discusses the requirements for adding a new trigger.

## 5.1 Factors Affecting New Trigger Inclusion

The HLT system is a critical component of the CMS experiment. It needs to be fast and faultless. If for some reason the HLT system encounters an error, it could lead to a potential system fault. The experiment will not be able to collect data if the HLT encounters an error, which could lead to the loss of data from valuable LHC run-time. A new HLT system is thus required to pass a series of tests and cross-checks before being deployed for operation at the CMS detector. A dedicated community operates and maintains the HLT system in the CMS collaboration.

For adding a new HLT path to the HLT menu for an LHC run, two tests are absolutely necessary. They are the rate and the timing measurement. The rate test checks the amount of additional number of events added to the total HLT rate. The timing test checks the additional time required to process events with the new HLT path included in the HLT menu. The addition

of the new HLT path to the HLT menu should be justified by an improved efficiency for the physics search targeted by the new trigger.

### 5.1.1 Dataset Used for Trigger Studies

Two datasets are collected with minimal constraints at the trigger level. The `ZeroBias` dataset is collected at a 27 MHz event rate without any constraint and with a very high prescale. The `HLTPhysics` dataset is collected after any of the L1 trigger conditions pass. It is collected with a heavily prescaled path that stores an event randomly if any of the L1 trigger conditions are satisfied without any selection at the HLT level. The `ZeroBias` dataset is used for multiple purposes because it provides an estimate of the event structure without any filters. It also has its use for triggers studies, especially if the L1 trigger conditions change. For most HLT studies, the `HLTPhysics` dataset is used because it is collected after the L1 trigger conditions and has higher statistics for HLT studies compared to the `ZeroBias` dataset. Most events in the `ZeroBias` dataset fail the L1 trigger conditions, hence have low statistics for the HLT studies.

The event rate at the HLT is measured with the `HLTPhysics` dataset. The detector information for events in the `HLTPhysics` dataset is emulated with the most up-to-date configuration of the CMS detector in preparation for the new LHC run. The events in the `HLTPhysics` dataset can be used to understand the rate variation with kinematic selections on the physics objects.

### 5.1.2 Rate Measurement

The event rate of a new trigger can be estimated using either the data collected by the CMS experiment in the previous LHC runs or with simulation samples. The simulation-based approach is generally only used for estimating the event rate for SM physics processes after major detector changes. The data-based approach provides more reliable measurements for almost all other use cases. The simulation sample based approach is only used when the collected data cannot be reconfigured with the changes made to the CMS detector. The trigger rate measurement method is maintained as a collaboration-wide standard procedure.

In the data-based approach, the rate of the events for a particular trigger is measured using either the `ZeroBias` or the `HLTPhysics` dataset. The `HLTPhysics` dataset is used in most cases since it provides larger statistics for event rate estimation for any given trigger after the trigger selections. This is due to the L1 pre-selection applied before filling the dataset. The statistics after a trigger selection on `ZeroBias` is low since most of the events collected in this dataset are stemming from low-energy QCD interaction. The `ZeroBias` dataset is helpful for estimating the rate of new L1 triggers, HLT paths that use new L1 conditions or for understanding the effect of mandatory changes, such as the ECAL energy calibration on the event rate.

Assuming the rate estimation is run on events corresponding to  $N_{LS}$  lumi-section of collisions where the data was collected at an instantaneous luminosity of  $L_{\text{reference}}$  with a prescale  $P_{\text{HLT}}$ , the trigger rate for the upcoming target phase of data collection is defined as

$$\text{Target HLT Path Rate} = \frac{N_{\text{path}}}{N_{LS} \times 23.31 \text{ s}} \times \frac{L_{\text{target}}}{L_{\text{reference}}} \times P_{\text{HLT}}, \quad (5.1)$$

where  $N_{\text{path}}$  is the number of events passing the HLT path of interest, 23.31 s corresponds to the nominal LHC lumi-section time period and  $L_{\text{target}}$  is the target luminosity for collection of new data. The added rate to the HLT menu by a new path is disentangled by calculating the rate using two definitions. The total rate of an HLT path is defined as the total number of events

accepted by the path, while the pure rate is defined as the number of events accepted by the HLT path only.

### 5.1.3 Time Measurement

The HLT has a limited time to process a decision on an event input from L1. This is dependent on the input rate from L1 and the output event rate at the HLT to storage. The software-based HLT system should have sufficient throughput to achieve the desired event rate reduction within limitations of its processing power. A new path should be able to perform event selection under these constraints and should not cause significant or accidental delays.

The current HLT farm is mounted on 200 nodes containing two AMD Milan 64-core CPUs and two NVIDIA Tesla T4 GPUs [180]. The timing studies should ideally be done with a smaller version of an identical setup. However, currently, a system with 32 cores mounted on an Intel Xeon Silver 4216 CPU and one NVIDIA Tesla GPU is being used for time measurements offline. The software conditions are maintained identical to the online conditions, allowing flexibility to include customised changes that a user wants to test.

A single event is processed on a single thread in the CPU. The timing distribution computes the time taken to process a single event. The distribution bins the number of events with the time taken by each event. On average, the timing of the menu should be less than 500 ms.

## 5.2 The $\mu\mu$ trigger

This HLT path targets two displaced muons from the decay of two long-lived particles. The muons will be observed at different locations in the detector. In the STHDM framework, this is the scenario where both the leptons from  $\chi^\pm \rightarrow \chi^0 \ell \nu$  are muons. Hence the muons are oppositely charged as well.

When reducing the muon  $p_T$  threshold, the change in event rate is shown as a function of the muon with the second highest  $p_T$  in the event. In an event, the muon with the highest  $p_T$  is called the leading muon, denoted by  $\mu_1$ , and the muon with the second highest  $p_T$  is called the sub-leading muon, denoted by  $\mu_2$ . For the requirement of at least two muons passing a minimum  $p_T$  selection, the change of rate with the  $p_T$  selection is given by the  $\mu_2 p_T$ .

The `HLT_DoubleMu33NoFiltersNoVtx_v` was chosen as the reference trigger for the di-muon final state. This trigger required two global level-3 muons with  $p_T > 33$  GeV,  $|\eta| < 2.5$  and  $d_0 > 0.01$  cm. The non-standard annotation `NoFiltersNoVtx` in naming this trigger implies that the reconstructed muons did not have any primary vertex constraints and did not need to come from the same position in the detector. The former is used in standard triggers to reduce the number of muons from pile-up(PU) vertices and the latter is used to study particles that decay to a pair of muons, such as the  $Z$ -boson. The trigger rate was  $< 0.5$  Hz throughout the proton collisions in 2018 and varied between runs due to changes to the beam profile of the colliding proton beams. I chose a reference of 0.5 Hz in these studies to benchmark event rate change with different kinematic selections. The final rate is obtained by running the new HLT paths on the `HLTPHysics` dataset using the method mentioned in section 5.1.2.

Figure 5.2 right shows the  $\mu_2 p_T$  distribution for the data and the signal simulation. The individual distributions are scaled down by their integral to obtain the normalised distribution in Figure 5.2 left. Normalisation is the process where the number of events in each bin in the histogram of a distribution are divided by the total number of events so that the integral of resulting distribution is one. The normalised plots for the sub-leading muon  $p_T$  compare the

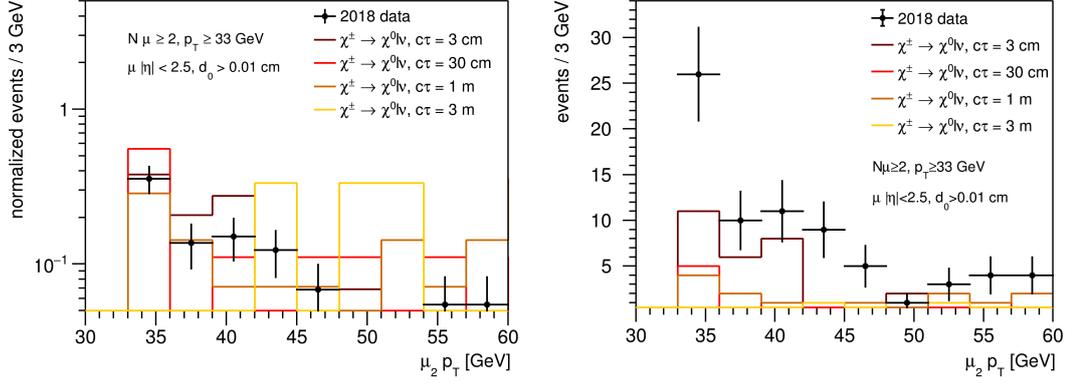


Figure 5.2: Left: Sub-leading muon  $p_T$  distribution for `HLT_DoubleMu33NoFiltersNoVtx_v` normalised to show the comparison between signal MC simulation and data. Right: Same distribution as the left but not-normalised to assist in the explanation of the rate and signal acceptance distribution in Fig. 5.3.

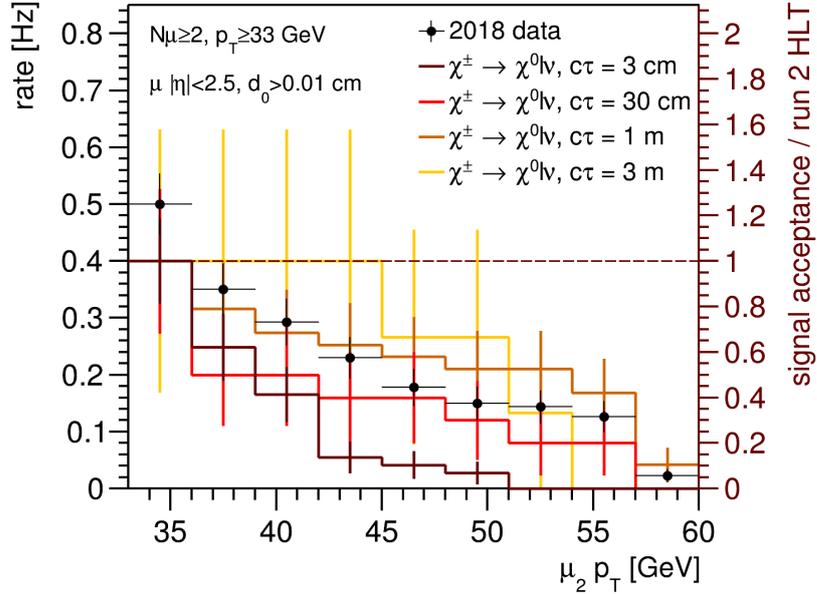


Figure 5.3: Rate (left ordinate) and signal acceptance (right ordinate) obtained by integrating and scaling Fig. 5.2 left from high to low  $p_T$  values.

distribution between data and the signal simulation. An exponential decay can be observed in the number of events for data. No definite conclusion can be made for the signal MC due to statistical limitation on the number of events accepted by the reference trigger. Figure 5.2 left is also useful to illustrate the definition by which change in rate and signal acceptance, shown in the Fig. 5.3, can be measured.

Figure 5.3 shows the variation of rate in data (black dots) as well as the variation of signal acceptance for different benchmarks in the signal MC (coloured histograms) with  $\mu_2 p_T$  after

selections identical to the `HLT_DoubleMu33NoFiltersNoVtx_v`. The plot shows the gain in signal acceptance and the increase in event rate as the selection on the  $\mu_2 p_T$  is lowered. Figure 5.3 is obtained from Fig. 5.2 left by the following equation.

$$g(i) = \left( \sum_{j \geq i} f(j) \Delta_j \right) \times \left( \frac{\text{Expected rate}}{\text{Reference HLT event count}} \right) \quad (5.2)$$

where  $g(i)$  is the event count in bin  $i$  of the rate and signal acceptance distribution in Fig. 5.3,  $f(j)$  is the event count in bin  $j$  of the distribution in Fig. 5.2 left and  $\Delta_j$  is the bin width of bin  $j$ . Each event was weighted by a scaling factor that differed between the data and the signal simulations to enable easy interpretation of the plot. In data, Fig. 5.3 shows the rate variation with a kinematic selection. The scaling of the signal simulation samples is such that the variation of signal acceptance with kinematic selections could be plotted on the same plot as the data. The scaling factor was set to use the reference trigger as a comparison benchmark. For data, the scaling factor was the ratio of the expected rate of the reference trigger to the number of events that pass the reference trigger from the `HLTPhysics` dataset. The values for the `HLT_DoubleMu33NoFiltersNoVtx_v` were 0.5 Hz and 87, respectively. The data has been plotted with the left (black) ordinate of Fig. 5.3. It can also be observed that the rate is integrated to lower  $p_T$  values and reaches 0.5 Hz, by construction, at  $p_T = 33$  GeV.

The scaling factor in signal MC was taken to be the ratio of the passing count of events to the reference event count that passed the run 2 reference trigger. The signal simulations were plotted with the coloured ordinate on the right side of the plot. The gain in signal acceptance with different selection criteria is essentially represented in multiples of the acceptance with the reference run 2 trigger. In Fig. 5.3, the gain in acceptance integrates to one from higher  $p_T$  values to  $p_T = 33$  GeV, which is expected for the reference trigger. Additionally, the gain in acceptance for the 3 cm signal MC lifetime was scaled to 80% that of the data rate to enable plotting in the same figure. The other lifetimes were scaled accordingly, which for the case of the reference trigger plot in Fig. 5.3 is 80%. Plotting the data and the signal simulations in the same figure enable study the change in rate with signal acceptance in a single figure as a function of our chosen kinematic,  $\mu_2 p_T$  in this case. The bars in Fig. 5.3 show the statistical uncertainty of the measurement. The plotting technique is used for all the rate and signal acceptance curves for all the triggers in this and upcoming sections in this chapter.

Figure 5.4 left shows the change in rate and signal acceptance by reducing the requirement from  $p_T > 33$  GeV to  $p_T > 16$  GeV. The rate increases to 3.9 Hz, about eight times that of the run 2 reference trigger while providing a signal acceptance of  $> 6 - 8$  times depending on the  $\chi^\pm$  lifetime. A similar increase in the event multiplicity between data and signal simulation is observed because the signal samples behave similarly to the displaced background from  $B$ -mesons and have a large number of events with low-energy muons. Signal acceptance can be gained by keeping the  $p_T$  threshold low while selecting other properties which reduce the event rate. The lower threshold of  $p_T > 16$  GeV came from the muon reconstruction algorithm in the muon chambers. Removing the displaced requirement of  $d_0 > 0.01$  cm gives an idea of the maximum signal acceptance possible with a new trigger without changes to the reconstruction algorithm for the displaced muon. Figure 5.4 right shows an increase in rate to approximately 30 Hz without a significant increase in acceptance for any of the signal lifetime benchmarks compared to Fig. 5.4 left.

The lower increase in signal acceptance on the removal of the  $d_0$  selection compared to the  $p_T$  selection is expected as most of the muons coming from the  $\chi^\pm$  decay are displaced. Figure 5.5 left makes a comparison in the distribution of base 10 logarithm ( $\log_{10}$ ) of the muon  $d_0$  ( $\log_{10} d_0$ )

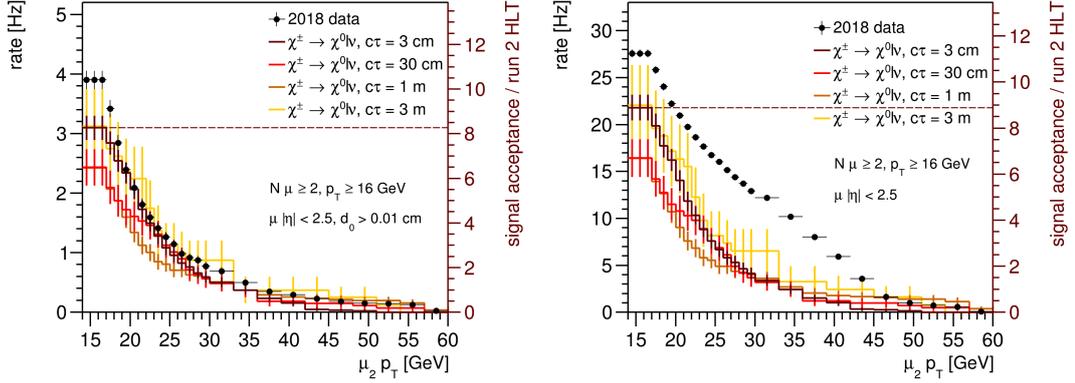


Figure 5.4: The rate and signal acceptance distribution after removing selections in the `HLT_DoubleMu33NoFiltersNoVtx_v`. Left: The muon  $p_T$  selection loosened from 33 GeV to 16 GeV. Right: Additionally the  $\mu d_0$  selection removed. The signal acceptance does not change much with the removal of the  $d_0$  selection for displaced muons while the rate increases by  $> 6$  times due to background events.

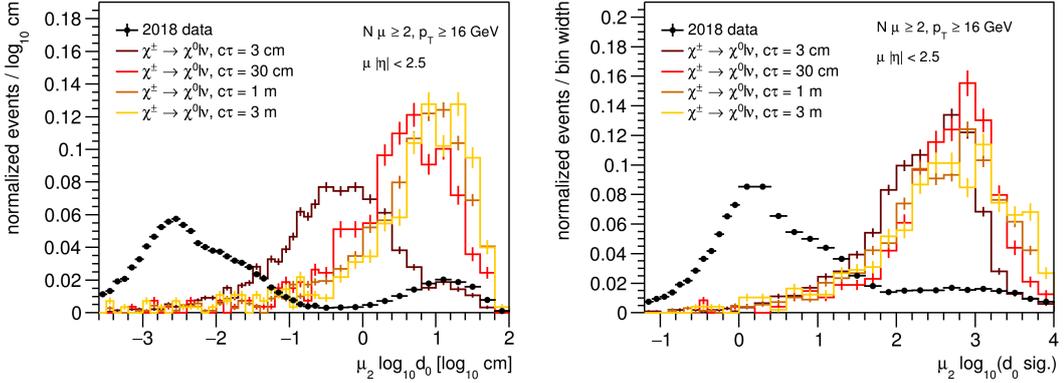


Figure 5.5: The base 10 logarithm of the  $d_0$  (left) and  $d_0$  significance (right), defined in Eq. 5.3, after a basic selection for the  $\mu\mu$  signature. Distribution trends quantify the displacement of the muons and help reduce data rate.

between data and the signal MC. The nature of the  $\log_{10} d_0$  distribution is detailed in appendix A for an exponentially decaying and a Gaussian  $d_0$  distribution. All the muons in the signal MC will have an exponentially decaying  $d_0$  distribution with different mean free paths. The peak in the  $\log_{10} d_0$  distribution corresponds to  $\log_{10} d'_0$ , where  $d'_0$  is the transverse impact parameter of the muon which is defined by the lifetime,  $c\tau$ , of the parent  $\chi^\pm$ .

The distribution of  $\log_{10} d_0$  in data is more complicated. A peak is observed at extremely low displacements with an immediate shoulder and another peak is observed at high displacements. The peak at extremely low values of  $\mathcal{O}(10 \mu\text{m})$  is the base-10 logarithm of the  $d_0$  resolution of the detector for prompt muon tracking. The shoulder corresponds to displaced muons from decay of  $B$  mesons and  $\tau$  lepton with mean free path  $\sim \mathcal{O}(0.1 \text{ mm})$ . The highly displaced muons at  $\mathcal{O}(10 \text{ cm})$  are a result of the decay of charged pions and kaons. Highly displaced muons can also be produced from decays involving BSM scenarios, such as the STHDM, and form an attractive

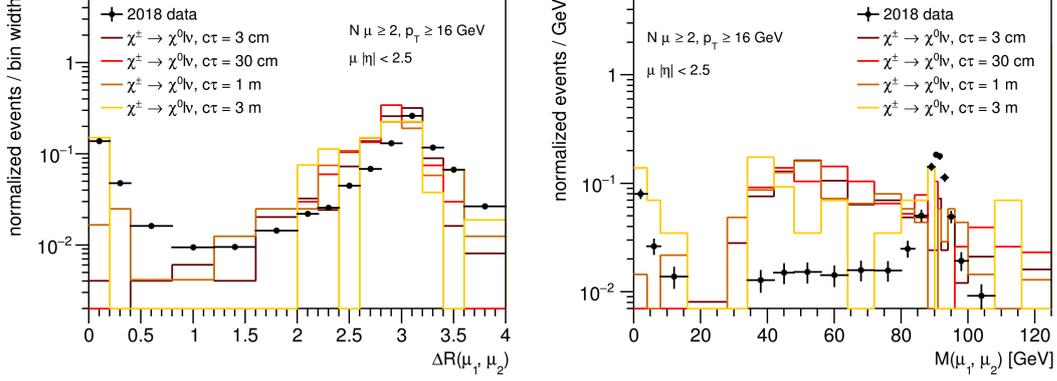


Figure 5.6: The  $\Delta R$  (left) and the invariant mass (right) of the highest energy muon pair in an event normalised to show characteristic difference between signal MC and data. This is used to select a phase space for triggers with low event rate and high signal efficiency.

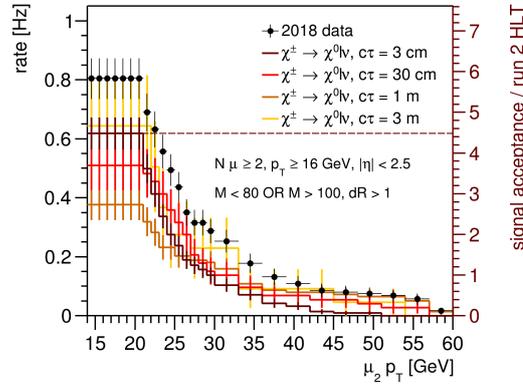


Figure 5.7: Event rate (left ordinate) and signal acceptance (right ordinate) with  $\mu\mu$  based event selections show low rate and upto 4.5 times larger no. of events in signal compared to the run 2 trigger.

prospect for DM research. Removing the lower cut  $d_0 > 0.01$  cm corresponds to removing the selection  $\log_{10} d_0 > -2$  on the x-axis of fig 5.5 left. Very few signal events are kept while many prompt background events are gained in data.

Often the transverse impact parameter significance, defined in Eq. 5.3 is used in lieu of  $d_0$  itself.

$$d_0 \text{ sig.} = d_0 / \sigma(d_0) \quad (5.3)$$

The  $\sigma(d_0)$  is the uncertainty in the  $d_0$  value and takes into account the resolution and uncertainty of the vertex measurement and the momentum reconstruction of the track. A well-reconstructed track will have a low  $\sigma(d_0)$  and hence high  $d_0$  significance. The base-10 logarithm of the  $d_0$  significance distribution, see Fig. 5.5 right, creates a clear separation between the data and signal MC. The  $d_0$  significance of muons in signal MC have values in the same order irrespective of their displacement. For highly displaced muons in data, it has spread to lower values signifying a considerable proportion of mis-reconstructed tracks.

The rate was reduced from  $\sim 27.5$  Hz in Fig. 5.4 right by using two different strategies. From the phenomenological study, it was observed that the STHDM leptons have a significant invariant mass and go angularly opposite to each other. Figure 5.6 shows the normalised angular distribution for the angular difference between the muons (left) and the invariant mass of the muons (right), comparing the signal MC to the data after a primary event selection with at least two muons with  $p_T > 16$  GeV,  $|\eta| < 2.5$ . The muon pair with the highest  $p_T$  are chosen to fill the event. The angular difference distribution, Fig. 5.6 left, has two distinct peaks in data which correspond to muon pairs in the same direction with angular difference close to zero or muons going opposite to each other with an angular difference of  $\pi$ . These peaks correspond to the background along with the low-mass SM particles and the Z-boson, respectively. The signal events, however, have a significant fraction of events with muon pairs angularly opposite to each other. The signal MC with  $c\tau = 3$  m has significant background contamination due to the low number of events that pass the event selections with multiple background events in the first bin. The events with low  $\Delta R$  could be from additional event activity where a low mass SM meson decayed to a muon pair. Due to a lower acceptance for highly displaced muons, the fraction of muons from background activity is more prominent in the longer signal lifetime simulations. This is evident in the di-muon invariant mass distribution in Fig. 5.6 right. Using a selection  $\Delta R(\mu, \mu) > 1$  removes a significant number of events in data with high efficiency for all the STHDM lifetimes.

The di-muon invariant mass distribution, see Fig. 5.6 right, is a well-understood distribution in data. The peak at low invariant mass values corresponds to low-mass SM resonances and background. The peak at values  $\sim 90$  GeV corresponds to the Z-boson. The plateau in the region 40 GeV to 80 GeV contains events from Z-boson and W-boson produced with a top quark or with each other as well events with hard scattered partons. The events from the signal MC have no distinct peak. They are spread over a plateau, starting around 40 GeV and reducing gradually by mass  $> 100$  GeV. The events in the Z-boson peak in the invariant mass distribution are directly correlated with the events with muons angularly opposite to each other in data. In addition to the angular difference selection, removing events from Z-boson mass peak significantly reduces the number of events in data and hence lowers the trigger rate. The selection  $M(\mu, \mu) < 80$  GeV or  $M(\mu, \mu) > 100$  GeV is used to obtain the estimate of rate and gain in signal acceptance for the new trigger shown as a function of the  $\mu_2 p_T$  in Fig. 5.7.

For a rate  $< 1$  Hz, an increase in signal acceptance from 2.7 – 4.5 times is obtained. The community of the CMS experiment, which works with exotic detector signatures, suggested alternative approaches to design the trigger in place of the event level selections on  $M(\mu_1, \mu_2)$  and  $\Delta R(\mu_1, \mu_2)$ .

Figure 5.5 showed the distributions of  $\log_{10} d_0$  and  $\log_{10} (d_0 \text{ sig.})$  in events with two muons with  $p_T > 16$  GeV and  $|\eta| < 2.5$  selection. I reused the  $d_0 > 0.01$  cm selection. The event rate was further reduced by requiring a selection on the  $d_0$  significance, see Fig. 5.8 left. It shows the  $\log_{10} (d_0 \text{ sig.})$  distribution after requiring the two muons to have a displacement,  $d_0 > 0.01$  cm. A selection  $\log_{10} (d_0 \text{ sig.}) > 1.4 \approx d_0 \text{ sig.} > 25$  resulted in a rate 1.23 Hz with a signal acceptance varying between 6 – 7.5 times that of the run 2 acceptance.

Table 5.1 summarises the gain in the signal acceptance with the rate of the triggers for the two triggers developed for the  $\mu \mu$  signature. The trigger proposals were superseded by a parallel effort who created a 20 Hz trigger for displaced di-muon signature without any additional constraints by utilising the additional rate made available for exotic signatures in run 3.

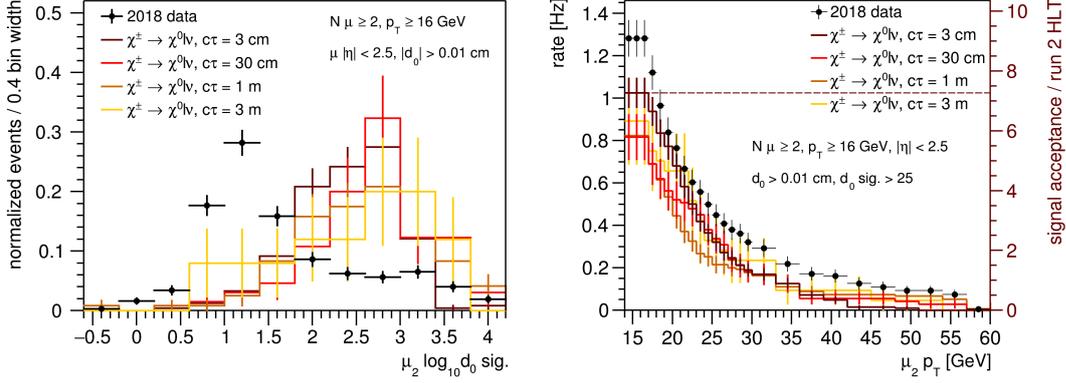


Figure 5.8: Left: The normalised distribution for  $\log_{10}(d_0 \text{ sig.})$  shows a large no. of events in data with a peak at 1.2. Right: Rate (left ordinate) and signal acceptance (right ordinate) with at least two muons in an event passing  $p_T > 16 \text{ GeV}$ ,  $|\eta| < 2.5$ ,  $d_0 > 0.01 \text{ cm}$  and  $d_0 \text{ sig.} > 25$ .

Table 5.1: Rate and signal acceptance gain with the new triggers.

Name	Rate [Hz]	Signal Acceptance [ $\times$ Run 2 Acceptance]
HLT_DoubleMu33 NoFiltersNoVtx_v	0.5	1
HLT_DiMu16NoFilters NoVtx_Dxy0p01DxySig1p4_v	$1.28 \pm 0.09$	$5.8 \pm 0.8 - 7.3 \pm 0.5$
HLT_DiMu20NoFilters NoVtx_dRgt1_Mlt80ORMgt100_v	$0.8 \pm 0.07$	$2.7 \pm 0.4 - 4.7 \pm 1.2$

### 5.3 The $\mu e/\gamma$ trigger

The  $\mu e$  signature arises when either of the lepton in  $\chi^\pm \rightarrow \chi^0 \ell \nu$  decay is a muon while the other is an electron. Such a signature relies on both the muon reconstruction and the ECAL energy deposit. The electron track relies on primary vertex constraints and hence is not used. The ECAL energy-deposit based trigger selection is also applicable for signature with photons. Hence this is termed the  $\mu e/\gamma$  trigger.

The HLT path `HLT_Mu38NoFiltersNoVtxDisplaced_Photon38_CaloIdL_v` was chosen as the reference trigger. It required one muon and one ECAL supercluster (SC) in an event. The muon is reconstructed using identical producers as the `HLT_DoubleMu33NoFiltersNoVtx_v`. At least one muon with  $p_T > 38 \text{ GeV}$ ,  $|\eta| < 2.5$  and  $d_0 > 0.01 \text{ cm}$  is required in the event before running the ECAL clustering producer. The photon is reconstructed starting from an L1  $e/\gamma$  trigger and was required to have  $p_T > 38 \text{ GeV}$ . A loose identification criterion, discussed in section 3.2.4, of  $\sigma_{i\eta i\eta} < 0.014(0.035)$  and  $H/E < 0.15(0.1)$  was required for the SC in the ECAL barrel (end-cap). The identification rejected background events from jets, soft scattering, PU and spurious noise

at the trigger level.

The rate distribution for the  $\mu e/\gamma$  HLT can be expressed as a function of the leading  $\mu$  or the leading ECAL SC ( $e/\gamma_1$ ) in the event. The results should be similar as long as both plots' selections are identical.

Figure 5.9 left is the normalised distribution of an event's leading muon  $p_T$ . Figure 5.9 right is the integral of the event count from high to low  $p_T$  values scaled to the reference trigger rate and signal acceptance using the non-normalised version of the plot in Fig. 5.9 left following an analogous procedure as discussed using Fig. 5.3 and Eq. 5.2 in section 5.2. The data in Fig. 5.9 right, to be read with the left (black) ordinate, shows the variation of the event rate with  $p_T$  selection on the leading muon in the event. It reaches a value of 2 Hz for  $p_T = 38$  GeV, which is the value chosen to round the rate of the reference trigger for comparison. The upper bound is chosen based on observing the rate for multiple runs for this trigger in 2018. The signal MC lifetimes, to be read with the right (coloured) ordinate, are scaled such that the number of events accepted by the reference trigger is taken to be one. The new trigger has been designed to have a rate in the ballpark of HLT\_Mu38NoFiltersNoVtxDisplaced\_Photon38\_CaloIdL\_v with a larger signal acceptance for the STDHM in as lifetime agnostic way as possible.

Table 5.2 lists the L1 triggers used for the new HLT path with  $p_T$  threshold reduced to 20 GeV at the HLT. These triggers were directly copied from the reference trigger.

It is a choice to represent the rate and signal acceptance in terms of  $\mu_1 p_T$ . Relaxing the constraints on the  $\mu_1$  and  $e/\gamma_1$ , I obtain the maximum rate and signal acceptance for  $\mu_1 p_T > 16$  GeV and  $e/\gamma_1 p_T > 20$  GeV. The lower limit on the muon transverse momentum comes from the filter on the muon reconstruction in the muon chambers. The lower limit on the  $e/\gamma_1$  resulted from the constraint on the corresponding L1 triggers that initiate this HLT path. Figure 5.10 shows that for a maximum gain in signal acceptance by a factor of  $\mathcal{O}(10)$ , the rate will increase to 130 Hz, which is inadmissible for a single DM search in the CMS detector. The rate needs to be constrained while retaining significant gain in signal acceptance.

The standard muon selection was kept unchanged. Figure 5.11 shows the change in rate and signal acceptance with  $p_T$  of the leading muon after a  $d_0$  selection,  $d_0 > 0.01$  cm, with  $\mu_1 p_T > 20$  GeV,  $e/\gamma_1 p_T > 20$  GeV and both  $\mu_1$  and  $e/\gamma_1$  in the region  $|\eta| < 2.5$ . The event

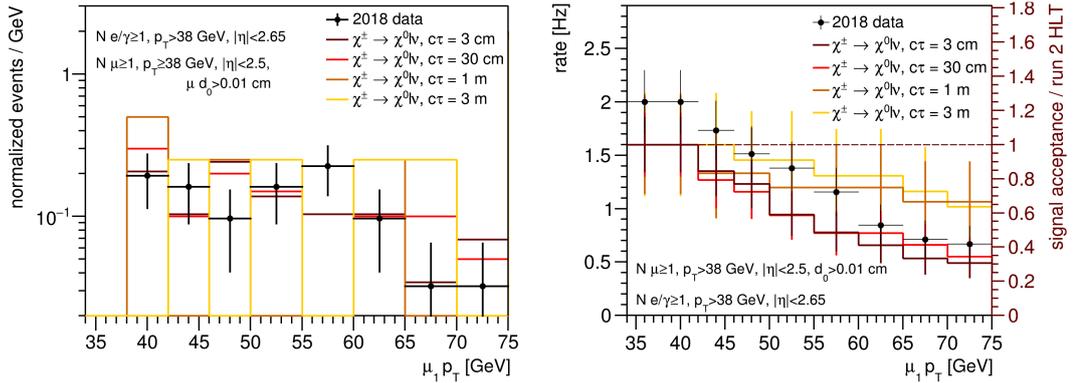


Figure 5.9: Left: The  $p_T$  distribution of the leading muon in the event, and Right: The rate and signal acceptance as a function of leading muon  $p_T$  using a definition identical to Fig. 5.3 for HLT\_Mu38NoFiltersNoVtxDisplaced\_Photon38\_CaloIdL\_v.

Table 5.2: Level-1 triggers that have been used to initiate the HLT\_Mu38NoFiltersNoVtxDisplacedPhoton38\_CaloIdL\_v HLT path. The unprescaled paths accept all events that pass the trigger while the prescaled paths choose an event randomly out of a number of events equal to the prescale factor. An L1 trigger with prescale factor 0 implies it accepts no events.

L1 trigger name
UNPRESCALED (Prescale = 1)
L1_Mu7_EG20er2p5
L1_Mu7_EG23er2p5
L1_Mu20_EG10er2p5
L1_SingleMu22
PRESCALED (Prescale factor)
L1_Mu5_EG23er2p5 (0)

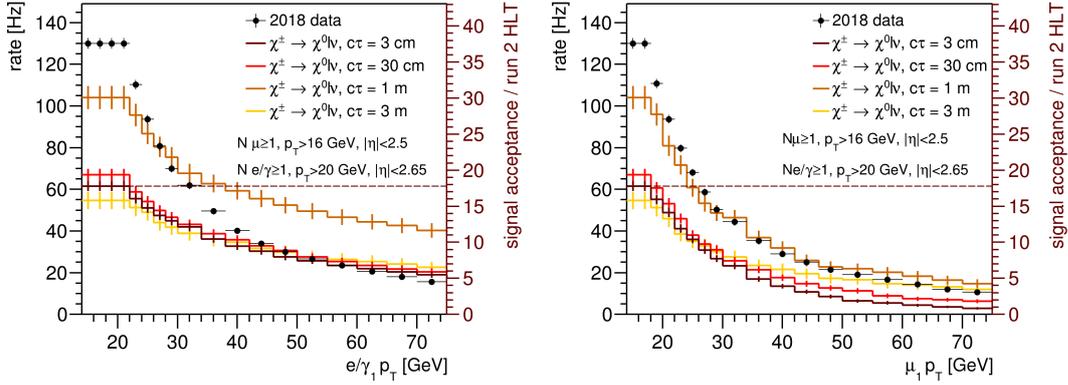


Figure 5.10: The rate and signal acceptance distribution with  $p_T$  threshold lowered and  $\mu d_0$  constraint removed shown as a function of the leading  $\mu p_T$  (left) and leading  $e/\gamma$  (right). At the respective low  $p_T$  threshold the distribution has identical rate and signal acceptance as expected. The difference in the variation of the event rate and signal acceptance with increasing  $p_T$  threshold is expected.

rate reduces by 67% to 50 Hz compared to the signal acceptance, which reduces by 20 – 25% only. Further rate reduction can be achieved by using the electron identification variables.

Electrons and photons are copiously produced from multiple scattering and decay processes from the product of proton collisions in the CMS detector. The background sources comprise SM electrons from hard scattered partons that are mis-constructed to be isolated, electrons from PU vertices, fakes from an imperfect reconstruction algorithm and other SM physics processes that have or can fake a similar detector signature as the signal model. The properties of displaced electrons produced from the decay of  $\chi^\pm$  are similar to those from isolated electrons and photons produced at the primary vertex. The identification variables identify and obtain

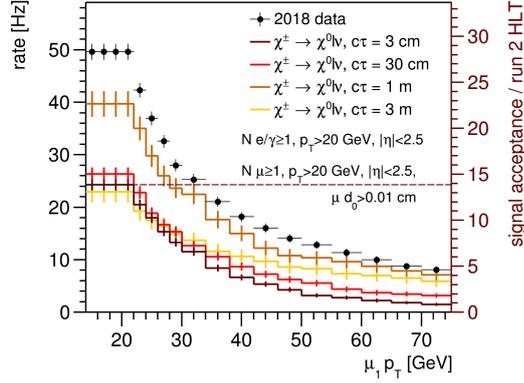


Figure 5.11: Rate decreases by  $> 50\%$  to 50 Hz with  $< 25\%$  reduction in signal acceptance compared to Fig. 5.10 with the selection  $\mu_1 d_0 > 0.01$  cm. The displaced muon selection reduces rate significantly with less effect on the signal acceptance.

well-reconstructed electrons.

The identification properties of electrons from  $\chi^\pm$  decay in the STHDM simulation must be studied separately from the background contamination in signal simulation. This is necessary to identify good displaced electrons. Good displaced electrons are separated from the background by matching electrons at the HLT to the electrons from the generator using angular constraints. This procedure is referred to as generator matching or gen-matching in short.

### 5.3.1 Electron gen-matching

Electrons and photons from various radiative processes dominate low-energy isolated ECAL signatures. To properly study the detector signature of displaced signal electrons from LLP decay, I associate a generator-level electron with an ECAL-based SC. The truth information of the electron from  $\chi^\pm$  decay is available from the MC simulation. The association of the SC to the MC-truth electron is done via angular matching with the prompt equivalent ECAL signature of the displaced electron. The properties of the  $\chi^\pm$  decay electrons are compared to electrons from  $Z$ -boson decay which serve as a standard for isolated electron identification variables. They are also compared to the detector signature of electrons in data with a predominant background component to understand the difference in signature with background  $e/\gamma$  and use the difference to reduce the event rate at HLT.

The geometric identification of an  $e/\gamma$  signature caused by an electron from  $\chi^\pm$  decay is made using the  $\eta$  and  $\phi$  of the respective particles. The angular parameters  $\eta$  and  $\phi$  of an  $e/\gamma$  signature assume the production of the corresponding  $e/\gamma$  at the primary vertex. For electrons produced at a displaced vertex, as in the case of  $\chi^- \rightarrow \chi^0 \nu_e e^-$  decay, the angular parameters of the corresponding SC can be significantly different from the actual  $\eta \sim \phi$  of the electrons as schematically illustrated in Fig. 5.12. The generator-level information provides the vertex and momentum of the electron. The generator vertex and momentum can be used to calculate the position at the ECAL where the displaced electron will arrive. The expected  $\eta$  and  $\phi$  of the ECAL signature can thus be computed from the generator-level information. I will refer to the changed value of  $\eta$  and  $\phi$  of the displaced electron as prompt equivalent angular parameters denoted by  $\eta_{prompt}$  and  $\phi_{prompt}$ .

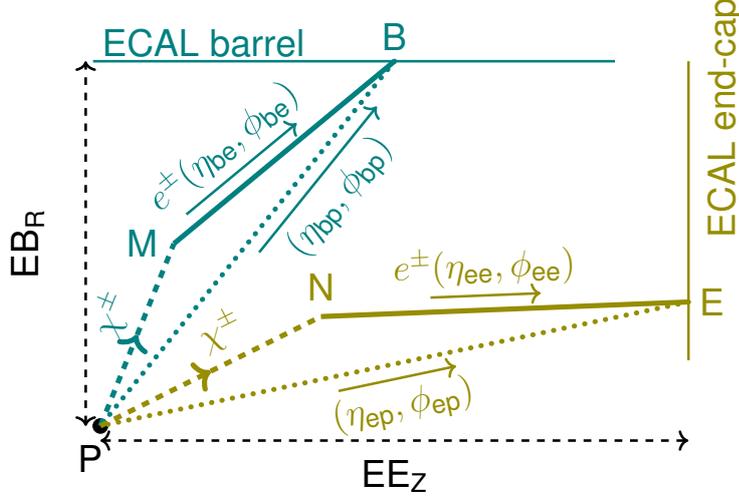


Figure 5.12: Electrons produced from the decay of  $\chi^+$  from a displaced vertex will have a longer flight path before reaching the ECAL. The angular parameters for the resulting ECAL signature will differ from their true value due to the algorithm of ECAL which assumes prompt production at the  $p$ - $p$  collision vertex.

To estimate the prompt equivalent for the angular kinematic parameters of a displaced electron, I consider the cases where the displaced electron is detected in the ECAL barrel (EB) separate from when the displaced electron is detected in the ECAL end-cap (EE). In Fig. 5.12, the EB case is coloured blue while the case of EE is coloured yellow. The  $\chi^\pm$  is produced at the  $p$ - $p$  collision vertex denoted by P, followed by its decay at vertex M(N) for the EB(EE) scenario. The displaced electron moves in the angular direction  $\eta_{be}$  ( $\eta_{ee}$ ) and  $\phi_{be}$  ( $\phi_{ee}$ ) before reaching the ECAL at B(E) for the EB(EE) scenario. This results in a signature at the ECAL in an angular direction  $\eta_{bp}$  ( $\eta_{ep}$ ) and  $\phi_{bp}$  ( $\phi_{ep}$ ) for EB(EE).

For an electron with  $p_T > 10$  GeV, the maximum sagitta of the track due to the CMS magnetic field is 2.5 cm, assuming no significant bremsstrahlung losses before reaching the ECAL barrel. The bend is small enough for the curve along the azimuth due to the magnetic field to be neglected. The electron path is thus assumed to be straight in Fig. 5.12.

An electron is only detected in the CMS experiment if it is produced inside the fiducial region of the ECAL detector. The vertex where the  $\chi^\pm$  decays to make the charged electrons, marked in the Fig. 5.12 as  $M(m_x, m_y, m_z)$  for EB and  $N(n_x, n_y, n_z)$  for EE, has the following constraints.

$$\begin{aligned} i_x^2 + i_y^2 &< EB_R^2, \text{ and} \\ i_z &< EE_z \text{ for all, } i = m, n \end{aligned} \quad (5.4)$$

where  $EB_R$ , the inner radius of the EB is taken to be 130 cm and  $EE_z$ , the inner  $z$ -coordinate where EE starts is  $|EE_z| = 130$  cm.

The EB scenario results in a more involved calculation. The vertex  $M(m_x, m_y, m_z)$  as well as the four-momentum of the displaced electron are known from the generator truth value. The four momentum is known in terms of  $E_e$ ,  $p_{Tbe}$ ,  $\eta_{be}$  and  $\phi_{be}$  and thus equivalently in terms of,  $(E_e, p_{ex}, p_{ey}, p_{ez})$ . Assuming an effectively straight path for the displaced electron coordinates for the point at which the electron deposits its energy in the EB, represented by  $B(b_x, b_y, b_z)$ ,

can be obtained for all  $i = \{x, y, z\}$  as,

$$b_i = m_i + v_i t_{be}, \quad (5.5)$$

where  $t_{be}$  is the time taken by the electron to reach EB.

$$\Rightarrow b_i = m_i + \frac{p_{ei}c}{E_e} t_{be}. \quad (5.6)$$

The constraint on the time of flight of the electron is given by,

$$b_x^2 + b_y^2 = EB_R^2 \quad (5.7)$$

$$\Rightarrow \left( m_x + \frac{p_{ex}c}{E_e} t_{be} \right)^2 + \left( m_y + \frac{p_{ey}c}{E_e} t_{be} \right)^2 = EB_R^2 \quad (5.8)$$

$$\Rightarrow \frac{(p_{ex}^2 + p_{ey}^2) c^2}{E_e^2} t_{be}^2 + \frac{2(p_{ex}m_x + p_{ey}m_y)c}{E_e} t_{be} + m_x^2 + m_y^2 - EB_R^2 = 0. \quad (5.9)$$

Using the only positive solution to the quadratic equation,

$$t_{be} = \frac{-B + \sqrt{B^2 - 4AC}}{2A}, \text{ where,} \quad (5.10)$$

$$A = \frac{(p_{ex}^2 + p_{ey}^2) c^2}{E_e^2}$$

$$B = \frac{2(p_{ex}m_x + p_{ey}m_y)c}{E_e}$$

$$C = m_x^2 + m_y^2 - EB_R^2.$$

Using the expression from Eq. 5.10, a value for  $t_{be}$  can be obtained. The coordinates for  $B(b_x, b_y, b_z)$  can be calculated using 5.6. The equivalent prompt angular coordinates for the barrel, given by  $(\eta_{bp}, \phi_{bp})$ , can thus be obtained by using the cartesian to spherical coordinate conversion given below.

$$\theta_{bp} = \cos^{-1} \left( \frac{b_z}{\sqrt{b_z^2 + EB_R^2}} \right)$$

$$\eta_{bp} = -\ln \tan \frac{\theta_{bp}}{2} \quad (5.11)$$

$$\phi_{bp} = \text{sgn}(b_z) \cos^{-1} \left( \frac{b_x}{EB_R} \right)$$

where  $\text{sgn}(b_z)$  is +1 if  $b_z > 0$  and -1 if  $b_z < 0$ .

Similarly, the prompt equivalent of angular coordinates for the displaced electrons in EE can be obtained. The EE is assumed to be a disk starting from  $|EE_z| = 310$  cm in the  $z$ -coordinate. The vertex where the displaced electron is produced,  $N(n_x, n_y, n_z)$ , should be in the ECAL fiducial acceptance region so it should satisfy Eq. 5.4.

The four-momentum of the displaced electron is known from the generator truth value in terms of  $E_e, p_{Te}, \eta_{ee}$  and  $\phi_{ee}$  or equivalently  $(E_e, p_{ex}, p_{ey}, p_{ez})$ . The sign of the  $z$ -momentum of the displaced electron decides if the electron will be detected at the EE disk at the EE- $z+$  or the EE- $z-$ . The point of EE where the displaced electron is detected is denoted by  $E(e_x, e_y, e_z)$ .

Assuming the time of flight of the electron before it is detected at the EE to be  $t_{ee}$ , E can be obtained in terms of the N and the electron momentum. For all  $i = x, y, z$ ,

$$\begin{aligned} e_i &= n_i + v_i t_{ee}, \\ e_i &= n_i + \frac{p_{ei}c}{E_e} t_{ee}. \end{aligned} \quad (5.12)$$

The z-coordinate of E is on the EE boundary. Thus,

$$\begin{aligned} n_z + \frac{p_{ez}c}{E_e} t_{ee} &= \pm EE_z \\ \Rightarrow t_{ee} &= (\pm EE_z - n_z) \frac{E_e}{p_{ez}c} \end{aligned} \quad (5.13)$$

The solution for  $t_{ee}$  leads to the solution for the prompt equivalent for the angular momentum in the scenario the EE detects the displaced electron. The method is analogous to the EB once the coordinates of E are known.

The value of  $t_{ee}$  obtained from Eq. 5.13 was compared with the value of  $t_{be}$  from Eq. 5.10. The EB and EE scenario calculation is done for every displaced electron as it is unknown if it will reach the EB or EE. The shorter time of flight decides the region of ECAL where the displaced electron is detected.

To perform angular matching, the difference in the pseudo-rapidity,  $\Delta\eta$ , and the difference in the azimuth,  $\Delta\phi$  are considered. The angular difference is expected to be symmetric around zero in  $\Delta\eta$ . However, due to the bending in the azimuthal plane under the influence of the CMS magnetic field, the azimuthal angular difference is expected to differ depending on the lepton's charge. I multiplied the azimuthal angular difference with the charge to plot electrons and positrons on the same plot. The resulting distributions for  $\Delta\eta$  and  $q \times \Delta\phi$  are shown in Fig. 5.13 and Fig. 5.15 for the different benchmark points.

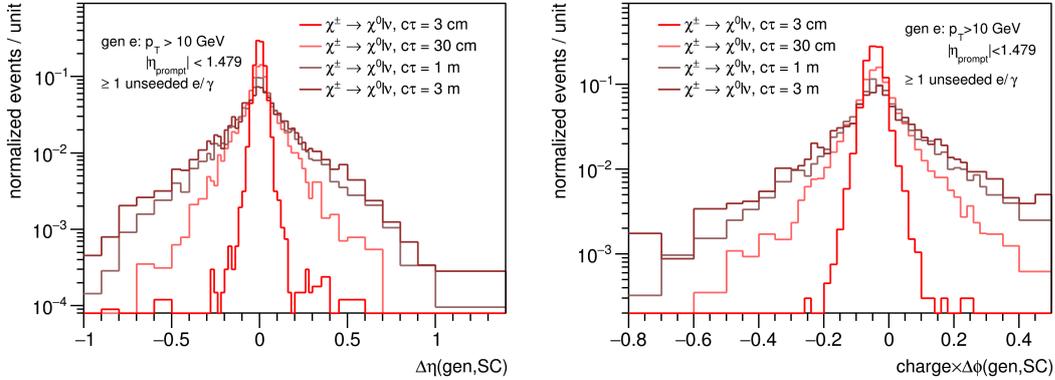


Figure 5.13: Lowest angular difference between a generator-level electron and an  $e/\gamma$  signature with normalised event count shows that there is a larger angular mismatch with increasing displacement. The angular pseudo-rapidity difference is symmetric. The azimuthal angular difference multiplied by the generator electron charge is asymmetric due to the bending of the charged particle in the transverse magnetic field of the CMS solenoid.

We observe a larger spread in the angular differences with increasing electron displacement. The spread is because the  $e/\gamma$  angular parameters are calculated, assuming the electron will come

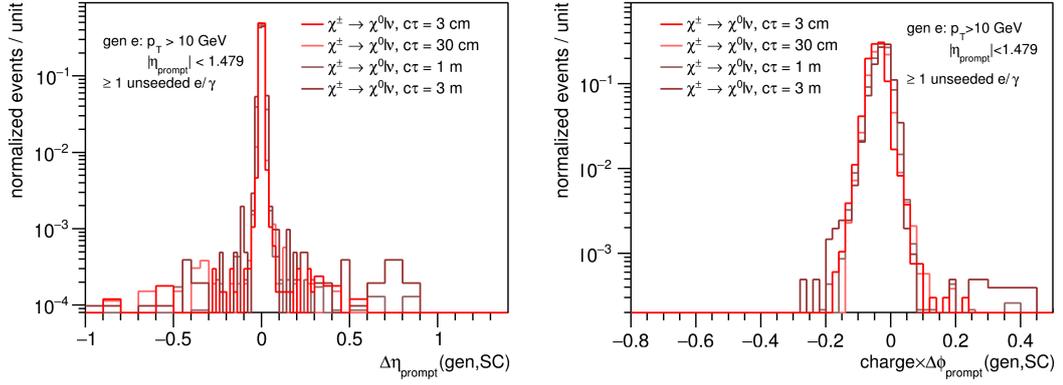


Figure 5.14: Angular difference plots with prompt equivalent angular parameters for the displaced electrons show uniform behaviour for generator matching for all the signal MC lifetimes.

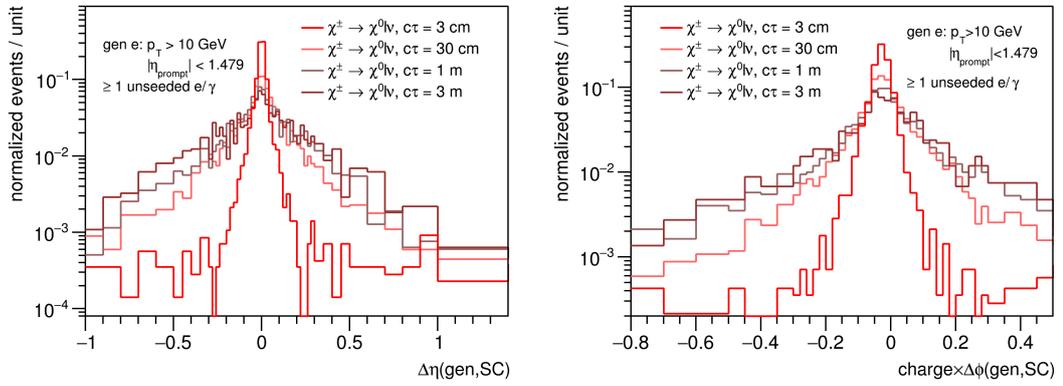


Figure 5.15: Similar to Fig. 5.13 in EE.

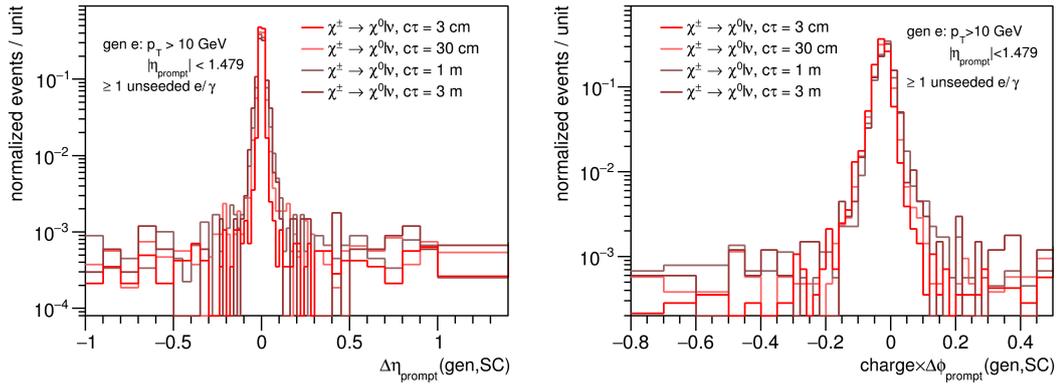


Figure 5.16: Similar to Fig. 5.14 in EE.

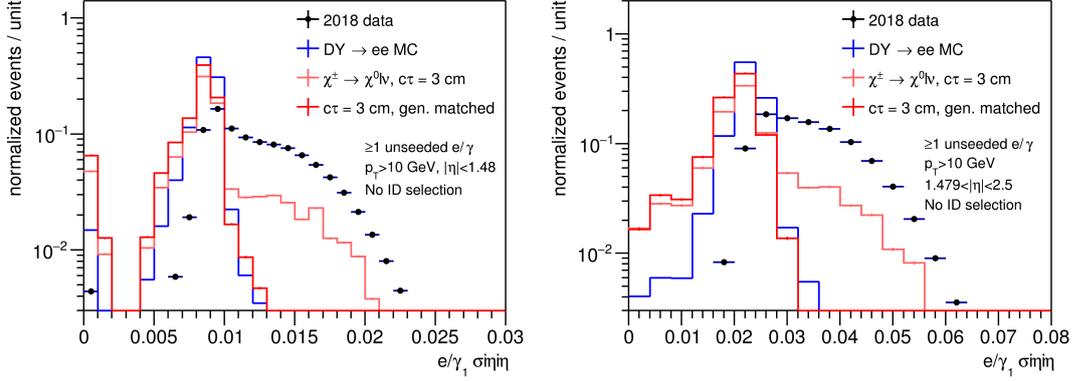


Figure 5.17: The plot shows that  $\sigma_{i\eta i\eta}$  variable for the electrons from gen-matching have a similar behaviour as the prompt electron in MC and data. The distribution of the variable without gen matching is significantly different and has a major component similar to background electrons in data. Left: EB and Right: EE.

from a prompt vertex. Knowing the generator-level information of the electron, I have computed the difference with the equivalent prompt angular parameters shown in Fig. 5.14 and Fig. 5.16 using Eq. 5.11 in EB and corresponding angles obtained in EE from solution using Eq. 5.13.

Based on Fig. 5.14 and Fig. 5.16 the angular matching criteria for the displaced electrons with the ECAL signature has been set to the following.

$$\begin{aligned} |\Delta\eta(\text{gen}, \text{SC})| &< 0.1(0.05), \text{ and} \\ -0.25(-0.15) &< |q_{\text{gen}}\Delta\phi(\text{gen}, \text{SC})| < 0.15(0.1) \quad \text{for EB(EE)}. \end{aligned} \quad (5.14)$$

### 5.3.2 Electron identification variables

The electron identification variables, introduced in Section 3.2, are essential to distinguish the signal electrons from the electrons due to background processes as discussed earlier. A loose selection based on these identification variables improves the purity of electrons in the selected events and reduces events with background-like behaviour.

For studying the identification variables, the  $e/\gamma_1$  was chosen in the data and the gen-matched  $e/\gamma$  was chosen in MC simulation. A comparison with electrons from  $Z$ -boson simulation was also included. A good generator-level electron is chosen by taking the final copy of an electron with mother  $Z$ -boson (in DY) or  $\chi^\pm$  (in signal STHDM MC) after the electron has gone through all final-state radiations. Additionally, the generator electron is required to have  $p_T > 10$  GeV and  $|\eta| < 2.5$ , and is required to be produced in the ECAL fiducial region defined by Eq. 5.4.

Fig. 5.17 shows the normalised distribution of the  $\sigma_{i\eta i\eta}$  variable in EB (left) and EE (right) for the gen-matched displaced electrons compared to all of the electrons in the signal MC for the lifetime  $c\tau = 3$  cm. Additionally, the distribution of the DY electrons and electrons in data allows an understanding of the gen-matching's effectiveness.  $e/\gamma$ s angularly matched to generator truth information have a significantly different distribution similar to the electrons in the DY MC from the primary vertex. The minor mismatch with the DY MC at lower values comes likely from an improper generator matching and could also point to differences for displaced electrons. In contrast, the distribution for the non-generator matched electrons shows a significant number of electrons with high  $\sigma_{i\eta i\eta}$  value as expected from background processes. The gen-matching is

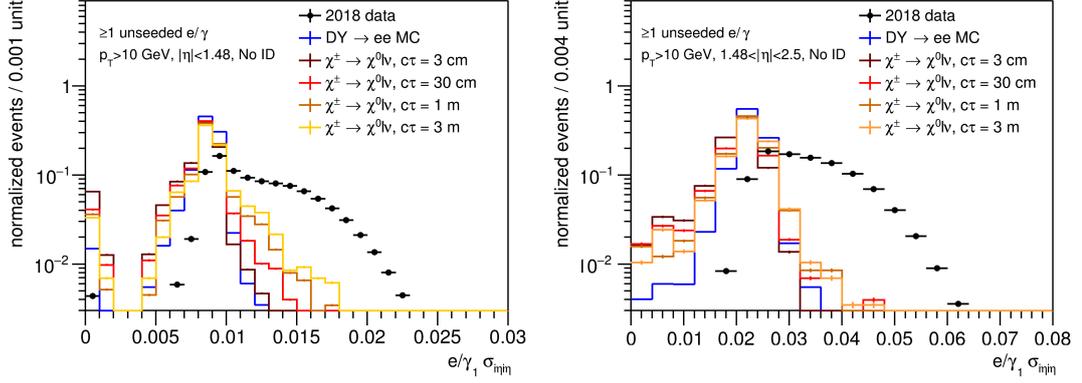


Figure 5.18: The plot shows the  $\sigma_{i\eta i\eta}$  variable has similar peaked distribution for gen-matched signal electrons as the  $Z \rightarrow ee$  simulation. At least one unseeded  $e/\gamma$  is required in the event with  $p_T > 10$  GeV and in the respective  $\eta$  region. The STHDM signal MC has additional gen-matching and the  $DY \rightarrow ee$  has the  $Z$ -boson mass window applied. Left: EB and Right: EE.

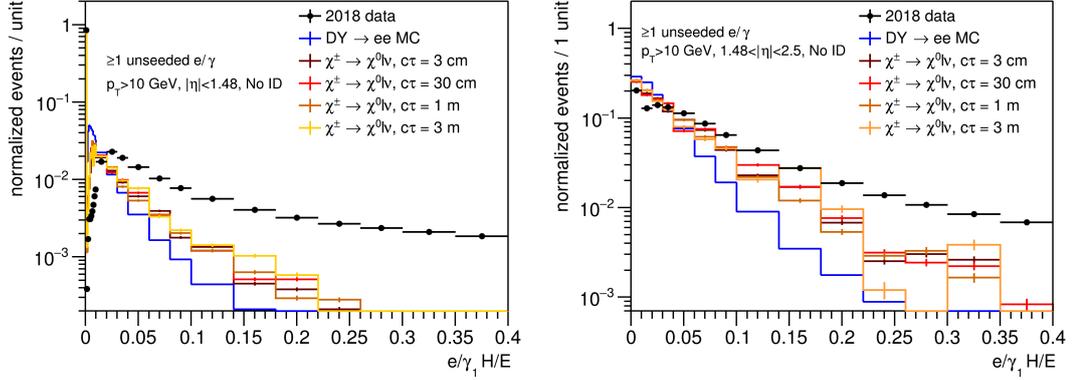


Figure 5.19: The H/E plot for the leading  $e/\gamma_1$  in the event. At least one unseeded  $e/\gamma$  is required in the event with  $p_T > 10$  GeV and in the respective  $\eta$  region. The STHDM signal MC has additional gen-matching and the  $DY \rightarrow ee$  has the  $Z$ -boson mass window applied. Left: EB and Right: EE.

necessary and effectual to properly study the soft displaced electrons from  $\chi^\pm$  decay to build the triggers.

Figure 5.18 compares the  $\sigma_{i\eta i\eta}$  variable in gen-matched displaced electrons among the different lifetimes and with the electrons from  $Z$ -boson decay in simulation in EB (left) and EE (right). With increasing lifetime, the distribution of  $\sigma_{i\eta i\eta}$  spreads by a small factor and retains the sharp peak with exponentially scaling down similar to  $Z \rightarrow ee$  MC. The shower shape variable selection  $\sigma_{i\eta i\eta} < 0.012(0.03)$  is used in EB(EE). It is a commonly used selection for electrons and photons at the HLT and results in a good signal acceptance while reducing the data rate by losing the background electrons.

Figure 5.19 compares the gen-matched displaced electrons between the different lifetimes and with the  $Z \rightarrow ee$  electrons for the  $H/E$  variable in EB (left) and EE (right). A sharp decrease is

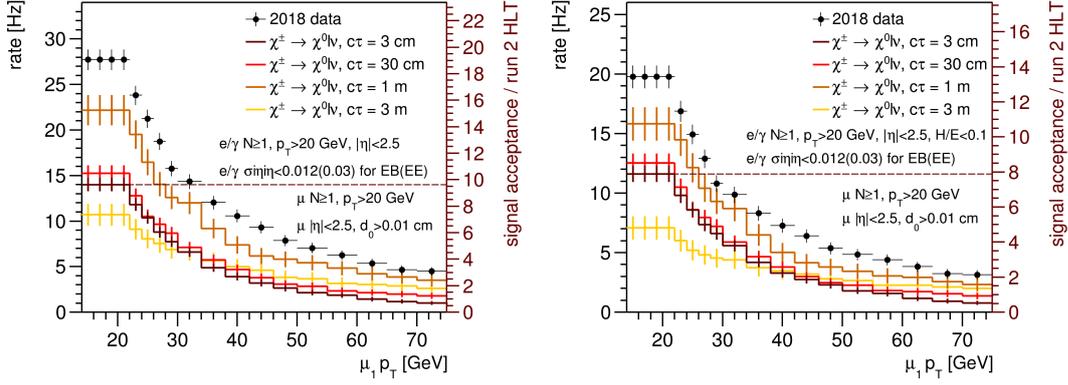


Figure 5.20: The rate and signal acceptance with additional selections on  $e/\gamma_1 \sigma_{in\eta\eta}$  only (left) and both  $\sigma_{in\eta\eta}$  and  $H/E$  (right). The event rate decreases considerably to 20 Hz from 50 Hz compared to selections in Fig. 5.11.

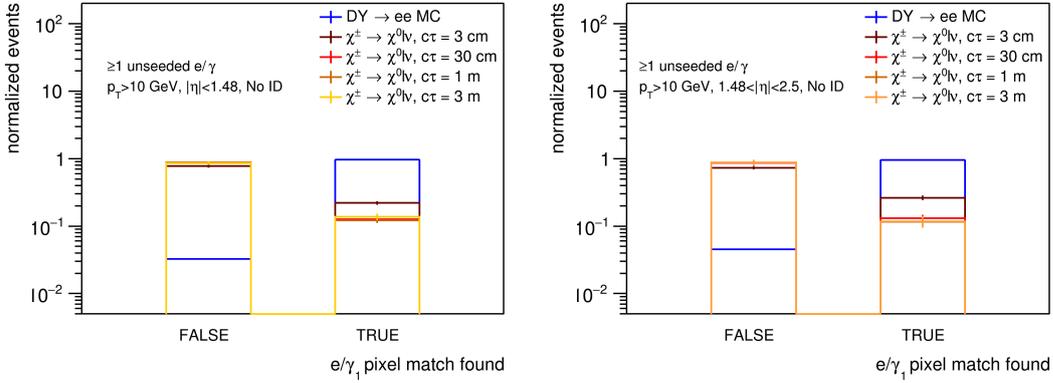


Figure 5.21: Plot shows if a pixel hit has been found for a selected  $e/\gamma$  after loose selections. This determines if the GSF tracking algorithm for electrons is run. Hits are found in the pixel region for most of the events in the  $Z \rightarrow ee$  MC while a significant fraction of events in the STHDM MC are unable to find a hit in the tracker pixels owing to the moderate displacement of these tracks. It is also observed that for higher displacements, a larger fraction of events are not able to run the GSF electron tracking. Left: EB and Right: EE.

observed for the signal electrons compared to the electrons in the data. The standard selection of  $H/E < 0.15(0.1)$  in EB(EE) is retained for this variable as well.

Figure 5.20 shows the change in rate and signal acceptance with the electron identification selection. The rate drops by  $\sim 60\%$ . The drop in signal acceptance depends on the lifetime and is  $\sim 30 - 40\%$  for the low displacement and  $\sim 40 - 50\%$  for the high displacements. The higher drop in high displacements could result from a larger proportion of background  $e/\gamma$ s in events where the displaced electron lies outside the ECAL fiducial volume. Electron identification using the tracks could not be done due to a lack of hits in the tracker pixel region.

### Missing inner tracker hits

Standard selections that rely on tracks for electrons cannot be used for displaced electrons. The

standard HLT signature for an electron uses GSF track reconstruction in an angular region compatible with an  $e/\gamma$ . However, this approach has two main limitations that keep the timing of the reconstruction algorithm within HLT timing constraints. Firstly, it requires a compatible hit in the first three pixel layers. Secondly, it requires matching the track to the primary vertex. These limitations decrease its sensitivity for displaced electrons.

Figure 5.21 shows if an initial hit has been found for an electron to initiate the GSF track reconstruction. Most of the electrons from  $Z \rightarrow ee$  MC,  $> 95\%$ , have a hit in the inner layer of the tracker while  $> 75\%$  of the electrons in the signal MC do not have a hit in the pixel layer. Larger displacements in the signal MC result in lesser electrons that initiate the HLT GSF tracking. Any reliance on the current tracking for electrons at the HLT will drastically reduce signal acceptance.

The rate after standard ECAL shower shape variables is still too high and needs to be reduced. A loose ECAL isolation is applied for this. The ECAL isolation is defined as the energy deposit in the ECAL in an angular cone of  $\Delta R < 0.3$  around the  $e/\gamma$  with a possible energy contribution from PU and the energy deposit from the SC itself reduced. The  $\Delta R$  is defined as follows.

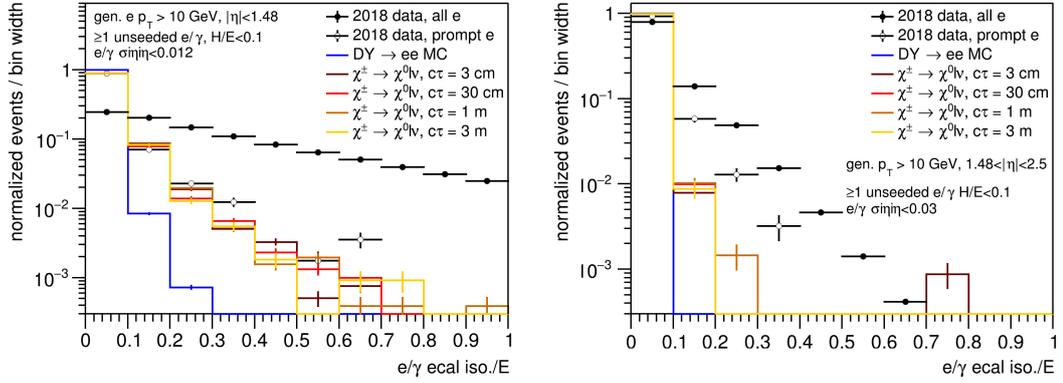


Figure 5.22: The isolation in ECAL divided by the energy of the  $e/\gamma$ . The electrons from  $Z \rightarrow ee$  have higher energy so sharply falling curve in isolation compared to the STHDM MC. Left: EB and Right: EE.

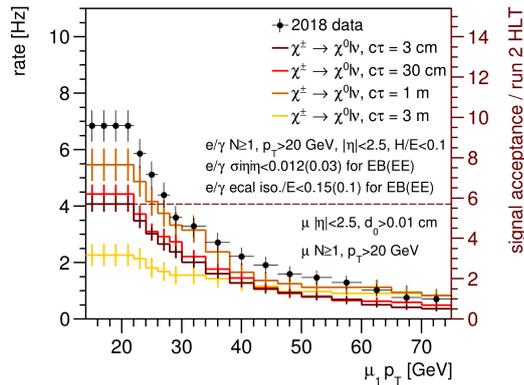


Figure 5.23: Rate and signal acceptance for the new trigger selection in the  $\mu e/\gamma$  channel.

Table 5.3: Rate and signal acceptance gain with the low-energy displaced  $\mu$   $e/\gamma$  HLT path, see Fig. 5.23. The rate is obtained from data to be read with the left black axis and the signal acceptance is to be read from the coloured signal MC simulations to be read with the right coloured axis. The minimum signal gain is observed for the 3 m lifetime simulation while the maximum gain is obtained for the 1 m lifetime simulation.

Name	Rate [Hz]	Signal Acceptance [ $\times$ Run 2 Acceptance]
HLT_Mu38NoFiltersNoVtxDisplaced_ Photon38_CaloIdL_v	2	1
HLT_Mu20NoFiltersNoVtxDisplaced_ Photon20_CaloCustomId_v	$6.8 \pm 0.6$	$3.2 \pm 0.4 - 7.7 \pm 0.6$

Figure 5.22 shows the ECAL isolation of an  $e/\gamma$  divided by its energy in EB (left) and EE (right). The isolation is divided by the energy of the SC to select the relative isolation of an electron. Relative isolation is effective for low-energy electrons. The relative isolation cuts are chosen to be less than 15%(10%) of the energy of the  $e/\gamma$  in EB(EE).

Figure 5.23 shows the final rate and signal acceptance for the  $\mu$   $e/\gamma$  final state. The trigger was named as

HLT\_Mu20NoFiltersNoVtxDisplaced\_Photon20\_CaloCustomId required one muon in  $p_T > 20$ ,  $|\eta| < 2.5$  with  $d_0 > 0.01$  cm and one electron in  $p_T > 20$ ,  $|\eta| < 2.5$  with  $\sigma_{in\eta} < 0.012(0.03)$ ,  $H/E < 0.1$ , ECAL iso./ $E < 0.15(0.1)$  in EB(EE). It is part of the CMS run 3 menu and has been collecting data throughout 2022 and 2023. Table 5.3 summarises the gain in the signal acceptance and expected rate compared with the run 2 reference.

## 5.4 The $e/\gamma$ $e/\gamma$ trigger

The  $e/\gamma$   $e/\gamma$  trigger targets the scenario when both the  $\chi^\pm$  decay to electrons. The methodology for studying the HLT electrons in the  $\mu$   $e/\gamma$  trigger applies to both the electrons in the  $e/\gamma$   $e/\gamma$  trigger. A generator matching was performed to separate the signature of signal electrons from electrons and photons that arise from the background. Identical angular matching conditions, as the  $\mu$   $e/\gamma$  HLT, have been used to associate a generator electron from  $\chi^\pm$  decay, given in Eq. 5.14. The kinematic distributions and variables of both the electrons were checked to match the electron distributions in the  $\mu$   $e/\gamma$  trigger.

As was observed in the previous case, the track reconstruction efficiency for displaced electrons is low at the HLT. Hence, no track-related kinematic variable selection could be applied for HLT paths targeting physics with displaced leptons. The ECAL signature is used to identify the electrons from  $\chi^\pm$  decay in the STHDM model. Because the track parameters are not used, the corresponding trigger is named  $e/\gamma$   $e/\gamma$  to clarify that the selections are ECAL-based only. In what follows, the  $e/\gamma$  will denote an energy deposit in the ECAL without a track requirement.

The  $e/\gamma$  objects with unseeded reconstruction have been used to provide the maximum possible efficiency for  $e/\gamma$   $e/\gamma$  signature. Realistically, in an HLT, the unseeded reconstruction is only run after at least one good  $e/\gamma$  has been observed using the seeded reconstruction to keep the computation time of the trigger within the constraints of the HLT menu.

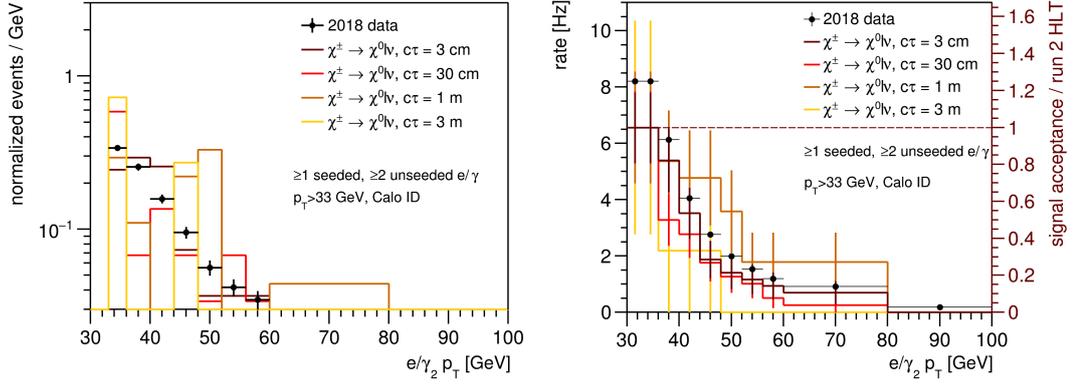


Figure 5.24: Left: The  $p_T$  distribution of data and signal simulations normalised to unity with selections identical to the HLT\_DoublePhoton33\_CaloIdL\_v. Right: Definition of the rate and signal efficiency with the reference trigger.

The HLT\_DoublePhoton33\_CaloIdL\_v was used as a reference to build the  $e/\gamma e/\gamma$  trigger because it was the path with the lowest energy threshold in the 2018 HLT menu. This HLT path required at least one L1 seeded  $e/\gamma$  with  $p_T > 33$  GeV,  $H/E < 0.15(0.1)$  and  $\sigma_{i\eta i\eta} < 0.014(0.035)$  in EB(EE). Following the L1 seeded  $e/\gamma$ , it required at least two unseeded  $e/\gamma$  with the same conditions as the L1 seeded  $e/\gamma$ . The path had a rate of about 1.28 Hz in Run 325022, one of the recommended runs for HLT studies in the 2018 HLTPhysics dataset. The path was also prescaled by a factor of 6.4. A reference rate of  $6.4 \times 1.28 = 8.19$  Hz was chosen for this trigger. A definition similar to Eq. 5.2 was chosen to express the rate and signal acceptance as explained earlier in Section 5.2. The rate and signal acceptance with the reference trigger are shown in Fig. 5.24 right.

For the  $e/\gamma e/\gamma$  HLT path, a  $p_T$  lowering to a threshold of 10 GeV was targeted. The low  $p_T$  choice was because the parallel effort for muons had a target to lower their  $p_T$  to 10 GeV. Keeping all the selections of the reference trigger unchanged and lowering the  $p_T$  threshold to 10 GeV increased the rate to 100 Hz. Simultaneously, a 15 to 25 times signal acceptance is obtained. This is shown in Fig. 5.25.

The list of L1 triggers in Table 5.4 was chosen to initialise the  $e/\gamma e/\gamma$  HLT paths. The choice includes all the unrescaled and prescaled L1 triggers that select an L1  $e/\gamma$  object with energy 10 GeV. The rate at HLT is lowered using two strategies as explained in the subsequent sections.

#### 5.4.1 Time delay in ECAL signature

Figure 5.26 shows the path taken by an electron coming from the decay of  $\chi^\pm$  and for a prompt electron. Analogous scenarios have been shown for EB in blue and EE in yellow. Similar assumptions have been made as for Fig. 5.12. The path has been assumed to be straight and bending due to the magnetic field has been neglected. The electron from the decay of  $\chi^\pm$  will be delayed compared to the prompt electron by,

$$\begin{aligned} t_{\text{delay}} &= t_\chi + t_{be} - t_{bp} \text{ in EB and} \\ t_{\text{delay}} &= t_\chi + t_{ee} - t_{ep} \text{ in EE.} \end{aligned} \quad (5.15)$$

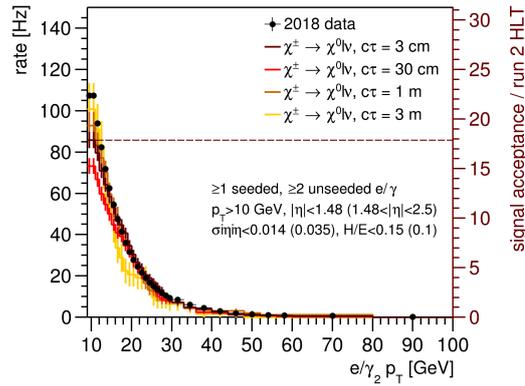


Figure 5.25: The rate (data to be read with the left black ordinate) increases to  $> 100$  Hz with identical selection as `HLT_DoublePhoton33_CaloIdL_v` with lower  $p_T$  threshold at 10 GeV. An increase in signal acceptance (coloured histogram to be read with the right ordinate) is observed by a factor between 15-25 depending on the lifetime of  $\chi^\pm$  in the sample.

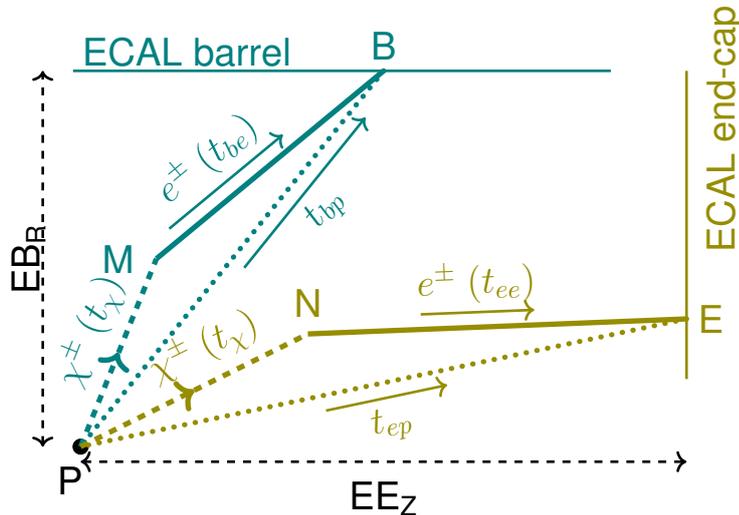


Figure 5.26: The figure shows the path and time an electron takes to reach the ECAL in cases it is produced from the decay of  $\chi^\pm$  or the  $p$ - $p$  collision vertex. Electrons produced from the decay of  $\chi^\pm$  from a displaced vertex will take longer to reach the ECAL than a prompt electron owing to the extra step. The delay in the ECAL signature is measured compared to an equivalent prompt signature. The blue shows the EB scenario, whereas the yellow shows the EE scenario.

This gives a conservative estimate of the time delay in ECAL signature from an electron from  $\chi^\pm$  decay. The actual delay will be larger than what is obtained by the calculation here.

The time delay can be estimated from the generator information of the particles to measure if the delay is significant compared to the standard deviation of the ECAL time measurement. Time taken by an individual particle is the distance travelled by the particle divided by its speed.

Table 5.4: Level-1 triggers that have been used to initiate the  $e/\gamma$   $e/\gamma$  HLT paths. The un-prescaled paths accept all events that pass the trigger while the prescaled paths choose an event randomly out of a number of events equal to the prescale factor. An L1 trigger with prescale factor 0 implies it accepts no events.

L1 trigger name
UNPRESCALED (Prescale = 1)
L1_Mu6_DoubleEG12er2p5
L1_SingleEG36er2p5
L1_SingleEG38er2p5
L1_SingleEG40er2p5
L1_SingleEG42er2p5
L1_SingleEG45er2p5
L1_SingleEG60
L1_DoubleEG_25_12_er2p5
L1_DoubleEG_25_14_er2p5
L1_DoubleEG_27_14_er2p5
L1_DoubleEG_LooseIso22_12_er2p5
L1_DoubleEG_LooseIso25_12_er2p5
L1_TripleEG_18_17_8_er2p5
L1_TripleEG_18_18_12_er2p5
L1_DoubleEG8er2p5_HTT300er
L1_DoubleEG8er2p5_HTT320er
PRESCALED (Prescale factor)
L1_SingleEG10er2p5 (9000)
L1_SingleEG15er2p5 (1550)
L1_SingleEG34er2p5 (0)
L1_DoubleEG_15_10_er2p5 (0)
L1_DoubleEG_20_10_er2p5 (0)
L1_DoubleEG_22_10_er2p5 (0)

The speed of the particle is obtained from the generator four-momentum of the particle. The distance travelled by  $\chi^\pm$  will be the difference between the proton collision vertex and the vertex at which the daughter electron is produced (shown in Fig. 5.26 as  $M$  in EB and  $N$  in EE). Equations 5.10 and 5.13 are used to obtain the time values for  $t_{be}$  and  $t_{ee}$ . The shorter flight time between  $t_{be}$  and  $t_{ee}$  determines if the electron is detected at EB or EE. The vertices  $M$  and  $N$  are obtained from the generator electron vertex by propagating a linear track along the generator electron momentum for the shorter time between  $t_{be}$  and  $t_{ee}$  using Eq. 5.5 or Eq. 5.12. The corresponding prompt angular parameters obtained in Eq. 5.11 can be used to calculate the momentum of the equivalent prompt particle. A similar energy as the displaced electron is assumed for its prompt equivalent. The speed from the prompt equivalent momentum was used

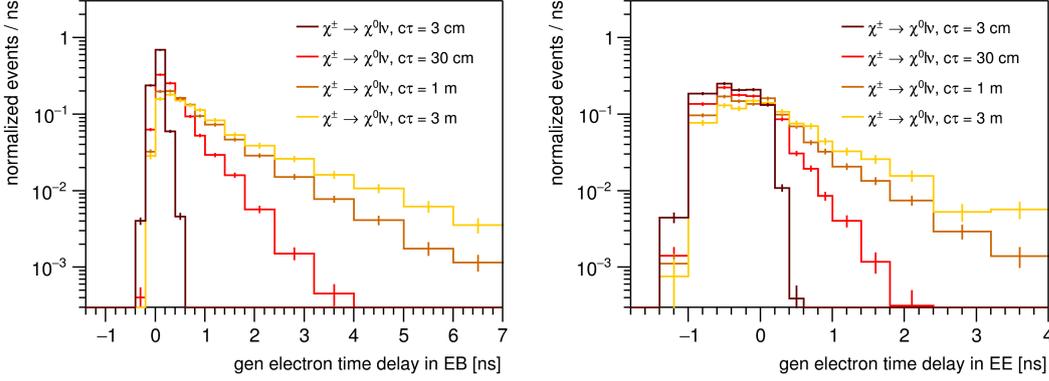


Figure 5.27: The time delay in the ECAL signature, defined in Eq. 5.15, from the generator parameters for the electron from decay of the  $\chi^\pm$  is shown for the different benchmarks. The distributions have been normalised to one to enable a comparison between the simulations with different  $\chi^\pm$  lifetime. Except for a mean decay path length of 3 cm, the others have a substantial number of events with delay above 1 ns in the ECAL signature. Left - EB, Right - EE.

to calculate  $t_{bp}$  and  $t_{ep}$  for a prompt equivalent of an electron that was detected at  $M$  or  $N$ . The time delay in the ECAL signature measured by the ECAL clock is obtained using the Eq. 5.15 and is shown in Fig. 5.27 left(right) for EB(EE) using the generator parameters for electrons produced from  $\chi^\pm$  decay in the ECAL fiducial region for all electrons with  $p_T > 10$  GeV.

The time delay in the ECAL signature from the generator information for the electron is shown in Fig. 5.27 for all mean free path length simulations of  $\chi^\pm$ . As expected, a longer delay in the ECAL signature is observed for the larger mean free path of the  $\chi^\pm$ . The number of events falls exponentially with time delay due to the probabilistic nature of the decay process of  $\chi^\pm$ . Some events are observed with a negative time delay as the time taken by the prompt equivalent particle was calculated from the origin instead of the actual proton collision vertex. However, this is a minor correction and is only visible as a significant effect for the 3 cm mean free path. It can be dealt with by shifting the bins so that all events begin from time delay zero. It only affects the conclusions made in EE to a small extent. The overall effect is negligible because the number of STHDM events is low in EE. The time delay distribution is more spread in EE compared to EB owing to the longer path that the electron takes. A significant number of events is observed with a delay longer than 1 ns compared to the ECAL time resolution of 0.4 ns for electrons with  $p_T > 10$  GeV [187].

The corresponding time delay in EB and EE from detector simulation with generator-matched electrons for the different lifetimes of  $\chi^\pm$  is shown in Fig. 5.28. It shows the time delay for unseeded  $e/\gamma$  with  $p_T > 10$  GeV in the STHDM MC simulation samples that are angularly matched to a generator electron produced in the ECAL fiducial region in the coloured plots. These values are comparable to Fig. 5.27. Additionally, Fig. 5.28 also includes values from the  $Z \rightarrow ee$  MC in the blue histogram to show the resolution for prompt ECAL signature in the simulation. A significant delay longer than 1 ns is observed for all the STHDM lifetimes. The ECAL timing resolution for prompt  $e/\gamma$  is represented in hollow dots and is observed to be slightly larger than in the simulation for prompt electrons; however, it is still well below 1 ns in EB. The ECAL time resolution for prompt electrons is considerably larger in EE owing to

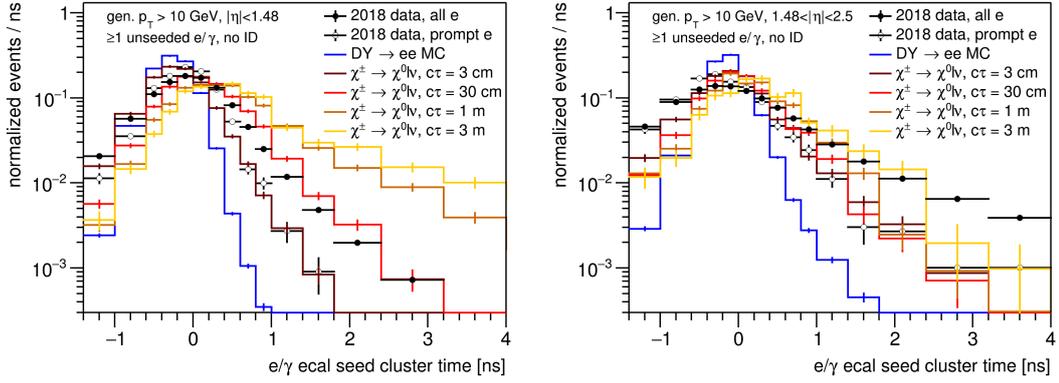


Figure 5.28: Time delay in ECAL signature in events with at least one  $e/\gamma$  reconstructed with the unseeded reconstruction angularly matched to a generator electron produced in the ECAL fiducial region. The coloured distributions are from the STHDM simulation, the blue distribution is from prompt simulation and the black and hollow dots are data. The black dots are for all the  $e/\gamma$  in data, while the hollow dots are for electrons from prompt signature in data, i.e. electrons from  $Z \rightarrow ee$ . All the plots are normalised to one to enable comparison of the relative distribution shapes. A mismatch is observed in the mean value and the standard deviation of the ECAL timing between the prompt simulation and prompt selection in data. The mismatch observed, however, is small compared to the delay observed in the ECAL signature for the STHDM model. The time delay in the ECAL signature for STHDM can be longer than 1 ns. Left - EB, Right - EE.

the larger path length, which can result in a larger standard deviation in the time measurement. The solid black dots show all the  $e/\gamma$ . The number of ECAL signatures falls exponentially with time delay. A relatively looser selection on the time delay will improve our signal selection while sharply reducing the number of events at the trigger level.

Figure 5.29 plots the time delay with similar features as Fig. 5.28. It has lower statistics than Fig. 5.28 but has a similar behaviour. The event frequency with unseeded  $e/\gamma$  signature with the maximum time delay in an event with at least one seeded, two unseeded  $e/\gamma$  with  $p_T > 10$  GeV that passes the ECAL quality selection for a signal electron,  $H/E < 0.2$  and  $\sigma_{i\eta i\eta} < 0.014(0.035)$  in EB(EE), has been integrated to obtain the Fig. 5.30 left. It shows the variation of rate and signal acceptance using the same mnemonic as similar plots before. The rate rises rapidly as the requirement for an event to contain at least one unseeded  $e/\gamma$  with an ECAL time delay is lowered from 2 ns to 1 ns. With a time delay selection of at least 1 ns, the  $\chi^\pm c\tau = 1$  m and 3 m have a significant number of events to have a higher signal acceptance while the benchmarks  $\chi^\pm c\tau = 3$  cm and 30 cm do not have enough events to benefit with respect to a  $p_T$  cut. This is concluded from the corresponding rate and signal acceptance plot with the sub-leading  $e/\gamma p_T$  in Fig. 5.30 right. A different approach is used for the benchmarks with lower  $\chi^\pm$  mean free path. It can also be observed from this distribution that the event rate increases by approximately 1.5 times for every 0.2 ns reduction in the threshold of ECAL time delay in the window from 2 ns to 1 ns. The ECAL time-delay-dependent HLT path is highly susceptible to ECAL calibrations, typically of the order of 0.2 ns. Hence, multiple HLT paths were created with a time delay threshold from 1 ns to 2 ns in steps of 0.2 ns as backup in case any path was pre-scaled to zero due to excessive event rate. The different paths with their rates, measured

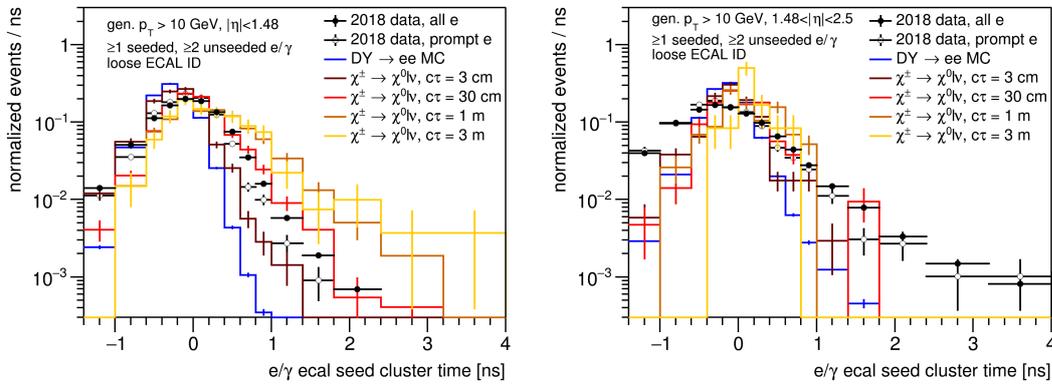


Figure 5.29: The distribution is similar in nature to Fig. 5.28 but is made after selections to be done before applying a threshold on time delay in the ECAL signature. The statistics in these distributions is low but the nature of the distributions shows similar behaviour as the previous figure. Left - EB, Right - EE.

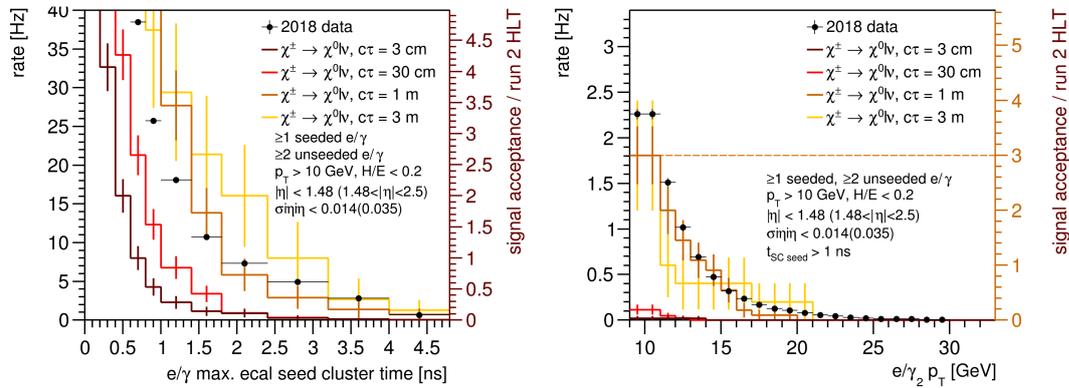


Figure 5.30: The data (black dot to be read with the left ordinate) shows the event rate and the simulation (coloured histogram to be read with the right ordinate) shows the signal acceptance compared to the Run 2 HLT. Left - The distribution shows the event rate and signal acceptance variation with the maximum ECAL time delay for an unseeded  $e/\gamma$ . The data rate rises rapidly as we lower the delay threshold from 2 ns to 1 ns and becomes too high below that. A good number of events with a significant time delay after the calorimeter identification and  $e/\gamma$  multiplicity selection is observed for  $\chi^\pm$   $c\tau = 1$  m and 3 m. Right - The rate and signal acceptance with a time delay cut of 1 ns plot with the sub-leading  $e/\gamma$   $p_T$ .

Table 5.5: Expected event rate of the new time delayed  $e/\gamma$   $e/\gamma$  HLT paths.

Name	Rate [Hz]
HLT_DiPhoton10Time0p8ns_v	13.5
HLT_DiPhoton10Time1ns_v	7.2
HLT_DiPhoton10Time1p2ns_v	4.7
HLT_DiPhoton10Time1p4ns_v	3.2
HLT_DiPhoton10Time1p6ns_v	2.1
HLT_DiPhoton10Time1p8ns_v	1.6
HLT_DiPhoton10Time2ns_v	1.1

using the recommended strategy by the CMS trigger group, is given in Table 5.5.

#### 5.4.2 Minor second moment of the ECAL shower shape

The secondary moments of the ECAL shower shape, discussed in Sec. 3.2.4, determine the circular symmetry of the ECAL energy deposit in the  $\eta \sim \phi$  plane. The ECAL crystals are oriented towards the proton collision vertex. As a result, a prompt  $e/\gamma$  signature will have a circularly symmetric energy deposit in the  $\eta \sim \phi$  angular direction. A displaced  $e/\gamma$  signature from the decay of a long-lived particle, e.g. the  $\chi^\pm$  on the other hand, will have an elliptical shape, as illustrated in Fig. 3.6. The secondary moments of the ECAL shower shape quantify the circular asymmetry. The smaller axis is measured by the minor second moment, referred to as  $s_{min}$ .

The  $s_{min}$  of the ECAL energy deposit is shown in Fig. 5.31 left(right) for EB(EE) in events with at least one reconstructed unseeded  $e/\gamma$  with  $p_T > 10$  GeV angularly matched to a generator electron with  $p_T > 10$  GeV using the conditions in Eq. 5.11. The prompt MC is shown in the blue histogram and all coloured histograms are the STHDM simulations for the different lifetimes of the  $\chi^\pm$ . The data is plotted with round dots. The hollow dots correspond to the prompt electron signature in data from the  $Z$ -boson while the solid dots correspond to all the unseeded  $e/\gamma$  with  $p_T > 10$  GeV representative of the event rate that can be obtained with a selection on the  $s_{min}$  variable. The prompt simulation and the prompt data peak together, signifying a faithful simulation of the  $s_{min}$  variable. There is a minor disagreement in its distribution shape between the prompt simulation and prompt selection in data, more significant in EE, which does not affect the final result. The signal benchmarks have a relatively similar distribution shape, making a selection on this variable inclusive of all benchmarks. The value of  $s_{min}$  for the signal model has a relatively larger number of events in the lower values compared to the prompt  $e/\gamma$  signifying an elliptical nature of the ECAL energy deposit, which was expected.

Figure 5.32 left(right) shows the  $s_{min}$  distribution in EB(EE) for all electrons after requiring at least one seeded and two unseeded  $e/\gamma$  with  $p_T > 10$  GeV,  $H/E < 0.2$  and  $\sigma_{i\eta i\eta} < 0.016(0.04)$  in EB(EE). The MC simulations for STHDM and DY are angularly matched to a generator electron from  $\chi^\pm$  and  $Z$ -boson, respectively. It has all the features similar to Fig. 5.31, with lower statistics. The above selections are representative of the selections to be applied before the  $s_{min}$  selection in sequential execution of the HLT path. The ECAL identification condition for  $\sigma_{i\eta i\eta}$  in an  $e/\gamma$  signature was loosened after observing higher signal electrons at low energy

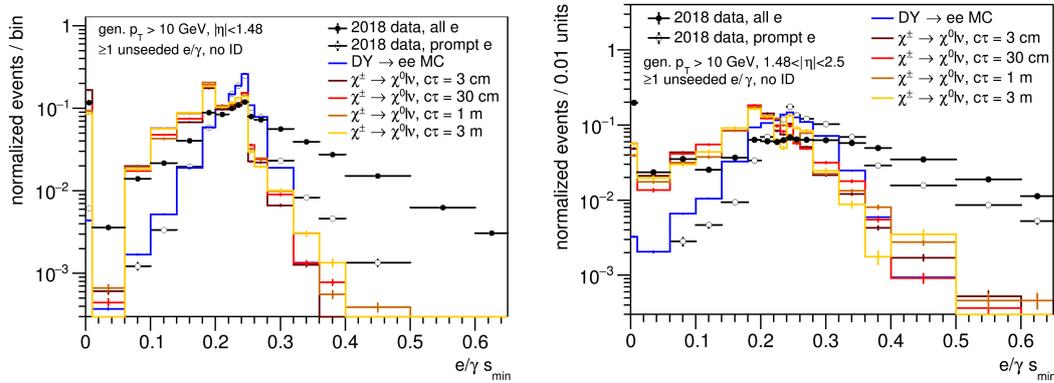


Figure 5.31:  $s_{min}$  of the ECAL signature in events with at least one  $e/\gamma$  reconstructed with the unseeded reconstruction angularly matched to a generator electron produced in the ECAL fiducial region. The coloured distributions are from the STHDM simulation, the blue distribution is from prompt simulation and the black and hollow dots are data. The black dots are for all the  $e/\gamma$  in data, while the hollow dots are for prompt electrons signature in data, i.e. electrons from  $Z \rightarrow ee$ . All the plots are normalised to one to enable comparison of the relative distribution shapes. A minor mismatch is observed for  $s_{min}$  between the prompt simulation and prompt selection in data. This mismatch is larger in the EE, however, does not affect the conclusion to a large extent. The signal simulation has larger number of events at lower values as expected compared to prompt data and simulation. Left - EB, Right - EE.

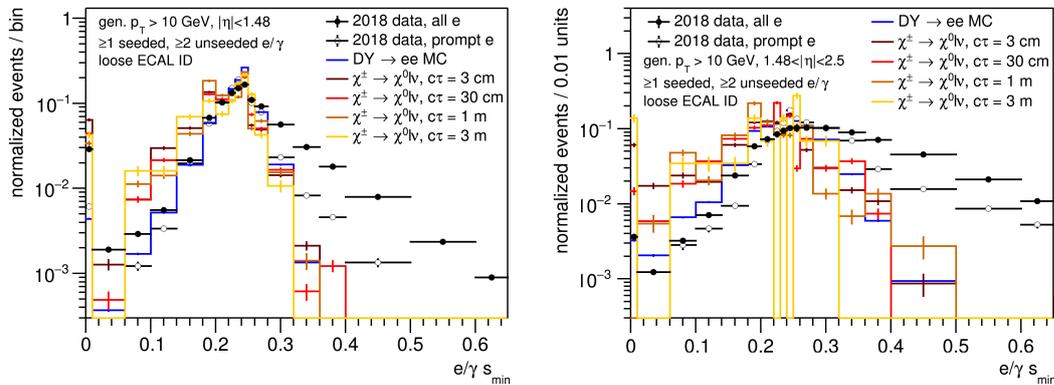


Figure 5.32: The distribution is similar in nature to Fig. 5.31 but is made after selections to be done before applying a threshold on the  $s_{min}$  of the ECAL signature. The statistics is these distributions is low but the distributions show similar behaviour as the previous figure. Left - EB, Right - EE.

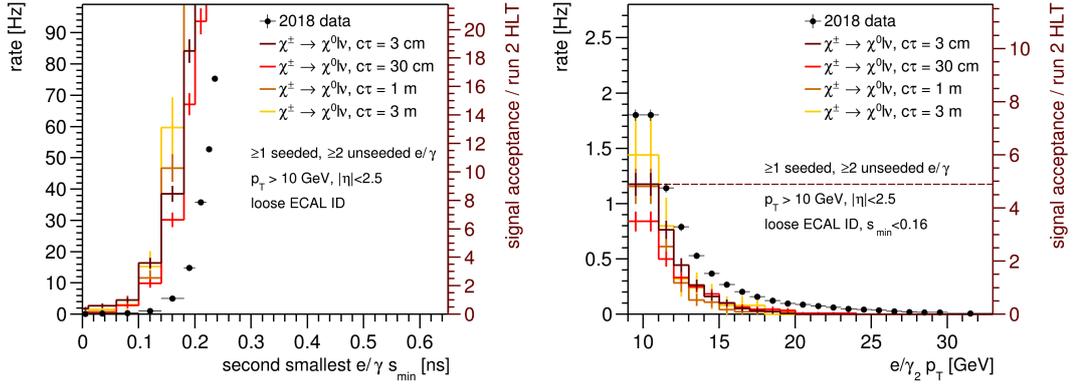


Figure 5.33: The data (black dot to be read with the left ordinate) shows the event rate and the simulation (coloured histogram to be read with the right ordinate) shows the signal acceptance compared to the Run 2 HLT. Left: The rate and signal acceptance change with increasing value of the second smallest  $s_{min}$  for events with at least 1 seeded and 2 unseeded  $e/\gamma$  with  $p_T > 10$  GeV. The rate definition is changed in this plot to integrate from lower to higher values. Right: The rate and signal acceptance with the selections in the left plot in addition to requiring  $s_{min} < 0.16$  shows fairly high signal acceptance gain with moderate trigger rate.

values.

The plot in Fig. 5.32 right can be integrated and scaled with the same strategy as followed before to obtain the rate and signal acceptance with selection on the  $s_{min}$  of the ECAL energy distribution. However, since the signal displaced electrons are expected to have lower values for  $s_{min}$  than the prompt electron, the integration is done from lower to higher values of the  $s_{min}$  value. Figure 5.33 left shows the rate and signal acceptance with the standard interpretation as discussed before as a function of the second lowest  $s_{min}$  of the ECAL energy deposit after requiring at least one seeded and two unseeded  $e/\gamma$  to have  $p_T > 10$  GeV,  $|\eta| < 2.5$ ,  $H/E < 0.2$  and  $\sigma_{i\eta i\eta} < 0.016(0.04)$  in EB(EE). The rate and signal acceptance rise rapidly with increasing values. The signal acceptance has a lower value at which this rise is observed compared to the rate. Event rate constraints limit the selection that can be made on this variable. Figure 5.33 right shows the rate and signal acceptance with the  $e/\gamma$  requiring a selection  $s_{min} < 0.16$ . A fairly large signal acceptance is expected to be between three to six times the standard run 2 triggers for this selection.

The ECAL calibrations and the PU conditions changed during the course of the study. The final trigger selections as shown in Table 5.6 were observed to be 10 Hz for  $s_{min} < 0.16$  and 3 Hz for  $s_{min} < 0.12$ . The main trigger of interest was the one with  $s_{min} < 0.16$  while the trigger

Table 5.6: Expected event rate of the new time delayed  $e/\gamma$   $e/\gamma$  HLT paths.

Name	Rate [Hz]
HLT_DiPhoton10sminlt0p12_v	3.2
HLT_DiPhoton10sminlt0p16_v	10.2

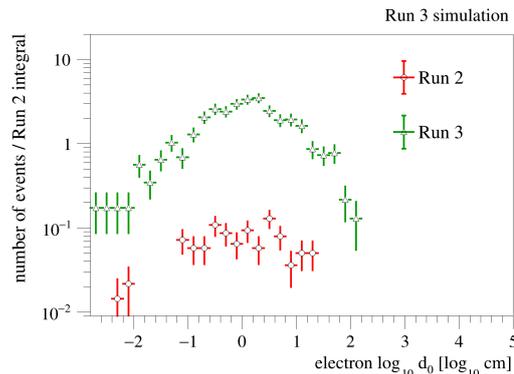


Figure 5.34: The gain in signal acceptance compared to the signal acceptance with the reference run 2 HLT is shown as a function of the base 10 logarithm of the generator electron  $d_0$ . The event account is scaled such that the signal acceptance in run 2, shown in red with circles, is normalised to one. The green histogram with triangular marks shows the corresponding signal acceptance with the new  $e/\gamma$   $e/\gamma$  HLT paths where the event count is scaled with the same factor as the events in the red histogram. An increase can be observed in the signal acceptance by over 30 times depending on the lifetime of  $\chi^\pm$ .

with  $s_{min} < 0.12$  was kept as a backup in case the trigger rate became too high and needed to be prescaled.

### 5.4.3 Gain in signal acceptance

The  $e/\gamma$   $e/\gamma$  triggers were created in the HLT framework. The STHDM simulation was re-emulated with the new triggers to measure the gain in signal acceptance compared to the run 2 triggers. Only di-electron final state events were considered while measuring the signal acceptance. However, no generator matching was required because offline reconstruction is expected to be more efficient than online trigger reconstruction if the events are accepted at the trigger level. The signal acceptance comparison was done with the generator electrons.

Figure 5.34 shows the gain in signal acceptance as a function of the generator electron  $d_0$ . The events are scaled by a factor such that the event acceptance with the run 2 reference is normalised to one. Such event scaling strategy enables interpretation of signal accepted by the new triggers in multiples of that in run 2. The events from all the simulation lifetimes of  $\chi^\pm$  are summed to obtain more significant statistics for all electron displacements. It is observed that at least one order magnitude more events are accepted by the low energy displaced di- $e/\gamma$  triggers compared to the run 2 reference for all displacements of the generator electron. It was also observed that most of the gain in signal acceptance comes from the lower transverse momentum requirement in the triggers. With these results, the triggers were integrated into the main CMS trigger menu for data collection at the beginning of LHC run 3.

There is a caveat to the signal acceptance numbers. The simulation relies on L1 conditions that are disabled in data. This results in a bias in all signal acceptance estimates. Moreover, the  $s_{min}$  cut is very aggressive. The loss of efficiency is recovered due to the low- $p_T$  threshold in the path. However such low- $p_T$  events would be lost by the L1 condition. A re-optimisation of this selection needs to be done taking into account the L1 conditions in future.

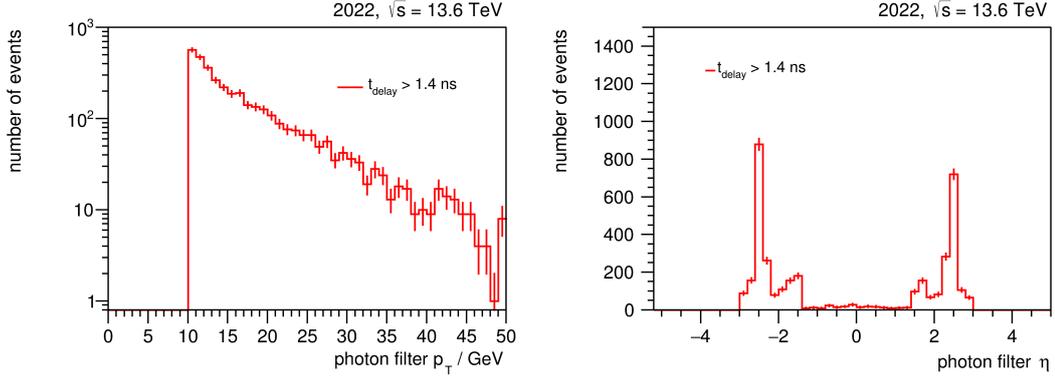


Figure 5.35: The  $p_T$  (left) and  $\eta$  (right) of online objects passing the last filter of HLT\_DiPhoton10sminltOp12\_v have been shown. The  $p_T$  accepts events with 10 GeV  $e/\gamma$  as expected. The  $\eta$  distribution has unexpected peaks in the EE at  $|\eta| > 2$ .

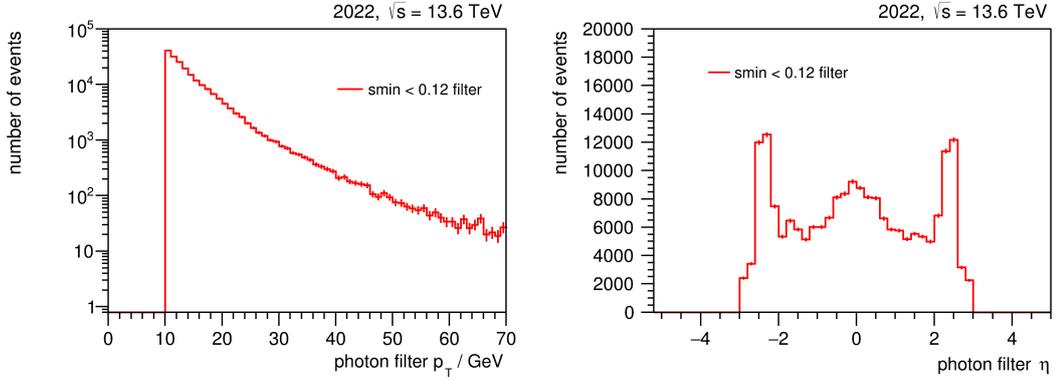


Figure 5.36: The  $p_T$  (left) and  $\eta$  (right) of online objects passing the last filter of HLT\_DiPhoton10Time1p4ns\_v have been shown. The  $p_T$  accepts events with 10 GeV  $e/\gamma$  as expected. The  $\eta$  distribution has unexpected peaks in the EE at  $|\eta| > 2$ .

#### 5.4.4 Observation in 2022 data and further changes

The data collected in early 2022 was monitored to assess the behaviour of the triggers. The HLT\_DiPhoton10sminlt\*\_v triggers showed unprecedented rates  $> 20$  Hz and were subsequently prescaled. The immense rate was a result of new ECAL calibrations that accepted a larger number of  $e/\gamma$  objects with low energy. The un-prescaled re-deployment of these triggers is pending due to insufficient time to study these triggers, as a fresh signal simulation sample was unavailable to measure the effect of further changes of selections on the triggers to constrain the rates.

The HLT\_DiPhoton10Time\*ns\_v group of triggers performed as expected and still collects data. These triggers also faced challenges from the ECAL calibrations. It was a reduced effect, though, and the additional rate was mitigated due to the increased opacity of the ECAL crystals, which reduced their timing and, subsequently, their rate.

Figure 5.35 and Fig. 5.36 plot the  $p_T$  and  $\eta$  of the events that pass the final filter of the

triggers `HLT_DiPhoton10sminlt0p12_v` and `HLT_DiPhoton10Time1p4ns_v` respectively. The  $p_T$  distribution is as expected. The  $\eta$  distribution shows unexpectedly large number of events in the forward region in EE at  $|\eta| > 2$ . These events were also verified to have a  $\phi$  dependence and are thought to be a result of some scattered radiation. Considering their azimuthal dependence, the events are ascribed to background contamination from an unknown source. The trigger purity can be increased, and the event rate can be lowered by removing these events. An additional selection  $\eta > 2.1$  was included in the following HLT menu that started collecting data in 2023.

## 5.5 Conclusion

Reducing the energy thresholds in the displaced lepton triggers was a challenging task. The additional rate obtained from reducing the trigger rate was countered by requiring selections specific to displaced and isolated electrons and muons in the triggers. The track information was not used for the electrons and formed a significant avenue for improvement in triggering displaced electrons. In addition to the rate, the timing of the menu with new triggers was also sought, and no significant deviation from the current timing of the menu was found. The signal acceptance gain with the new triggers was compared to a reference run 2 trigger, and in each scenario, a net gain in signal acceptance was observed. Multiple triggers were accepted to the CMS experiment trigger menu. The initial data collected with the new triggers was monitored.



# Chapter 6

## Data Scouting

*The effort to include the electrons and photons in the data scouting for low energy searches has been included in this chapter.*

In addition to the standard data taking, alternative data collection methods are employed at the HLT that employ strategies to go around the trigger constraints. The full detector information of a single event involves data collected from millions of readout channels. This results in an event size of about one MB per event. Writing events at the level-1 (L1) trigger rate of 100 kHz results in 100 GB of data per second. The high volume of data surpasses the bottleneck of readout rate to storage. Even if it were possible to output data at such a high rate, the CMS experiment would have to dedicate enormous financial, hardware and human resources to store and analyse this data. An innovative approach based on the prompt reconstruction of events reduces the event size to a  $\mathcal{O}(10)$  kB per event in data scouting which allows storage at an event rate of  $\mathcal{O}(30)$  kHz resulting in a data rate of  $\mathcal{O}(300)$  MB per second [152]. Scouting allows for a higher event rate and lower data rate than  $\mathcal{O}(1000)$  MB per second data collection rate of the standard HLT stream. Data scouting is implemented as an HLT path and runs in parallel with the other standard HLT paths. They share the same infrastructure.

The scouting mechanism exploits the high efficiency of object reconstruction at the HLT [229]. The HLT employs a computationally light version of the full CMSSW physics object reconstruction. It is highly efficient for the standard reconstruction of some objects. However, standard HLT paths reject events with these well-reconstructed objects using thresholds to fit the HLT event output rate constraints. For instance, the electron and photon reconstruction is  $> 80\%$  efficient for electrons and photons with  $p_T > 5$  GeV. However, the standard di-electron trigger requires  $p_T > 20$  GeV per electron due to the HLT rate constraint.

Only the physics objects reconstructed at the HLT and their corresponding identification variables are stored in the data scouting strategy while the rest of the detector data is discarded. Storing the detector physics objects reduces the data stored per event. The data output rate fits into the overall constraint of the HLT. This technique especially benefits physics studies involving low-mass resonances and low-energy particles. In the STHDM model, scouting should allow a study of low-energy and moderately displaced leptons.

The work in this chapter enables the integration of the electrons and photons to scouting in the LHC run 3. A publication is in preparation which summarises the physics possible with scouting and parking data. It contains a dedicated section on the new physics scope that the electron and photons in run 3 scouting bring.

Table 6.1: List of the initial L1 triggers that run the scouting HLT path grouped into categories based on the physics object that they target. A low-energy L1 trigger is accompanied by a high-energy counterpart as backup to the path in case too high rate is observed for the low-energy L1 trigger. The unprescaled paths accept all events that pass the trigger while the prescaled paths choose an event randomly out of a number of events equal to the prescale factor. An L1 trigger with prescale factor 0 implies it accepts no events.

<b>L1 trigger name</b>
UNPRESCALED (Prescale = 1)
L1_DoubleMu_15_7
L1_DoubleMu4p5er2p0_SQ_OS_Mass_Min7
L1_DoubleMu4_SQ_OS_dR_Max1p2
L1_HTT280er
L1_HTT320er
L1_HTT360er
L1_ETT2000
L1_HTT400er
L1_HTT450er
L1_SingleJet180
L1_SingleJet200
L1_DoubleJet30er2p5_Mass_Min300_dEta_Max1p5
L1_DoubleJet30er2p5_Mass_Min330_dEta_Max1p5
L1_DoubleJet30er2p5_Mass_Min360_dEta_Max1p5
L1_SingleEG36er2p5
L1_SingleLooseIsoEG28er2p1
PRESCALED (Prescale factor)
L1_DoubleMu_12_5 (0)
L1_HTT200er (1600)
L1_HTT255er (500)

## 6.1 Scouting strategy for electrons

The earlier scouting stream was implemented as a single HLT path called `DST_PFScoutingPixelTracking_v`. This path ran when any of the L1 trigger conditions shown in Table 6.1 was satisfied. The scouting HLT path contains a set of modules that run the HLT reconstruction for muons, electrons, photons, jets and missing energy in the detector. No selection was explicitly applied to the quality of the objects reconstructed and all the events that pass the L1 triggers were stored. The kinematics for the HLT objects were kept in the scouting data along with specific variables that have been typically used to separate the physics objects

of interest from the background. The detector data for the events in the scouting path have been discarded. The values of  $p_T$ ,  $\eta$ ,  $\phi$  of the ECAL supercluster (SC) have been stored for the electrons and photons. The variables  $\sigma_{in\eta}$ ,  $H/E$ ,  $I_H$ ,  $I_E$ ,  $r_9$ ,  $s_{min}$  and  $s_{maj}$  have been stored for both  $e$  and  $\gamma$ . The energy deposit in each crystal, the time of the energy deposit in the crystal, and the crystal identification number have also been stored for  $e$  and  $\gamma$ . For electrons, the track related variables  $1/E - 1/p$ ,  $I_T$ ,  $\Delta\eta_{seed}^{in}$ ,  $\Delta\phi_{in}$ , track charge, missing hits in the inner tracker,  $d_0$  and  $d_z$  were also stored. The best value of a track-related variable relating the SC to any of the tracks associated with it was stored for these variables. The variables have been described in detail in the Section 3.2.4.

The electron and photon reconstruction is identical to the online reconstruction used by the other HLT paths [187]. It uses the L1 trigger ECAL primitives to perform regional energy clustering in the ECAL [147]. A track reconstruction procedure based on the GSF algorithm in the tracker separates the electrons and photons. If a track is reconstructed, an electron is identified.

The above strategy for the scouting path has the advantage of having the maximum possible efficiency in the data stored for a physics case of interest. At the same time, there will be a large proportion of background which will be stored. A significant proportion of the background can be rejected with a low impact on the events of interest using quality selections on the reconstructed physics objects. Quality selections will also lower the rate. Hence larger input rate with L1 triggers will be possible, giving access to additional phase space and improving the physics program with scouting data. For instance, the kinematic thresholds on scouting electrons and photons can be lowered further by using the L1 triggers for double electrons. Furthermore, with additional available resources, the reconstruction for electrons and photons can be have a higher efficiency from unseeded reconstruction strategy.

Multiple changes have been integrated into the scouting strategy since its initial conception for LHC run 3. The changes include splitting the path based on the physics object targeted by the L1 triggers and applying quality selections to physics objects before storing the events. Almost all innovation of the scouting path for electrons and photons were done as part of this thesis and are detailed in this chapter.

Studies to commission the electron and photon collection in scouting need to show the effect of quality selections and new strategies on background rejection. Understanding the effect on isolated electrons and photons is also important. The effect of the object selection and change in strategy is shown with isolated electrons due to a shared interest in displaced electrons from the STHDM model. The effects should be analogous for photons due to the shared reconstruction method except for the presence of a track. These studies are performed with MC simulation samples. The effect of the changes on the rate of the scouting path and on the average time it takes to run is performed with dedicated data samples collected with unbiased triggers by the CMS collaboration. The accurate emulation of the CMS detector for the 2022 LHC run integrated changes from the developers who continuously tried to make the simulation as realistic to data-taking conditions as possible. The resulting changes happen very often and the HLT configuration needs to be re-run on the latest software version to take advantage of the improvements. Simulation samples with a low number of events are made available in the short term while the larger samples are kept until the simulation standards are more or less well known. Over time as the CMS collaboration starts collecting data and generating larger simulation samples with data-taking conditions, the smaller samples for commissioning the detector are not simulated anymore.

The MC simulation samples initially used to set up the electron collection in scouting are

Table 6.2: List of the MC samples simulated with the 2021 estimation of LHC run 3 running conditions.

Sl. No.	MC simulation sample
0	DYToLL_M-50_TuneCP5_14TeV-pythia8
1	DoubleElectron_Pt-1To300-gun
2	QCD_Pt-15to20_bcToE_TuneCP5_14TeV-pythia8
3	QCD_Pt-20to30_bcToE_TuneCP5_14TeV-pythia8
4	QCD_Pt-30to80_bcToE_TuneCP5_14TeV-pythia8
5	QCD_Pt-20To30_EMEnriched_TuneCP5_14TeV-pythia8
6	QCD_Pt-30To50_EMEnriched_TuneCP5_14TeV-pythia8
7	QCD_Pt-50To80_EMEnriched_TuneCP5_14TeV-pythia8

summarised in table 6.2. The  $\sqrt{s}$  of the  $p$ - $p$  collisions in this set of samples is 14 TeV. The `DYToLL_M-50` sample simulates the  $pp \rightarrow ll$  events where  $l$  is a lepton and the invariant mass of the lepton pair is  $> 50$  GeV. Most of the events in this sample have a  $Z$ -boson mediating the Drell-Yan process. Together with the `DoubleElectronGun` sample, the electrons from this sample are used to study isolated prompt electrons and will be referred to collectively as the signal electrons. The `DoubleElectronGun` sample uses a virtual gun in the GEANT4 simulator to fire two electrons with identical energy in angularly opposite directions. The energy of the electrons is a uniform distribution from 1 to 300 GeV. The electron-gun sample is especially useful to study the low-energy electrons since the SM signals/samples often have collinear electrons as decay products from a low-mass SM meson. Additionally, no simulation of a low-mass SM meson such as the  $J/\psi \rightarrow ee$  was available in 2021 and was done by me later. Prioritising the time constraint of the HLT menu deployment for CMS, I chose to go ahead with the `DoubleElectronGun` sample.

The QCD samples simulate the physics process  $pp \rightarrow XX$  where  $X$  can be a quark or a gluon resulting in a jet in the detector. The  $p_T$  limit of the parent particle that results in the jet is specified. Additionally, to obtain a larger number of electrons in the jets, the `EMEnriched` samples pre-select events with a potential large EM deposit while the `bcToE` samples require a parent  $b$ -quark or  $c$ -quark to decay to an electron. It was observed from phenomenology studies that the major background to the STHDM signal arises from leptons from jets from the  $b$ -quark and  $c$ -quark misidentified to be isolated. Optimising for the `bcToE` samples has the advantage of reducing the resulting background at the trigger selection stage. Collectively, electrons from these samples as well as electrons from pile-up (PU) vertices and multiple scattering are representative of the background electrons.

The data samples in table 6.3 were collected in 2018 to facilitate trigger studies with the new detector run conditions. As described in Section 5.1 before, the `ZeroBias` dataset is collected with a heavily prescaled trigger without any event selection at the HLT. This sample is useful in studies for new L1 triggers to estimate the rate of events. The `EphemeralHLTPHysics` is collected from events that pass any of all the L1 conditions without any HLT requirements. It is also collected by a prescaled trigger and is used to study the effect of HLT selections on the events.

In later studies, the  $J/\psi \rightarrow ee$  MC sample has been used with  $\sqrt{s} = 13.6$  TeV, which was

Table 6.3: List of the data samples simulated with the 2021 estimation of LHC run 3 running conditions.

Sl. No.	Data sample
0	ZeroBias
1	EphemeralHLTPhysics

the energy that the proton collisions at the LHC were able to reach in 2022. It is a SM signal process that requires detecting low-energy electrons properly. A minimum  $p_T$  selection on the  $J/\psi$  to be  $> 28$  GeV and on the daughter electrons to be  $> 15$  GeV is applied in the second sample in table 6.4. Applying the electron  $p_T$  selection in the second sample was necessary for more events to pass the trigger selection to study the electron identification variables.

## 6.2 Lower the scouting $p_T$ threshold with new L1 triggers

The last two rows in table 6.1 show the L1 triggers that initiate the electron reconstruction. The minimum threshold of energy required in the ECAL is 28 GeV. There exist L1 triggers with lower energy thresholds in the ECAL. The L1 trigger for two electrons, for example, requires only 12 GeV on the second-highest ECAL signature in the event. The L1 trigger for three electrons lowers the energy deposit requirement to 8 GeV. These triggers require a lower energy threshold in the ECAL and are sensitive to electrons from the decay of the  $\chi^\pm$ -particle in the STHDM model.

The double electron gun simulation was done with estimated running conditions of the LHC in 2021. The targeted  $\sqrt{s}$  for the new LHC run was 14 TeV. The PU vertices were not simulated for the double electron gun to keep the simulation simple for this sample. This also meant that it was not able to study certain identification variables, such as the isolation of the electrons.

The kinematic behaviour of the reconstructed electrons in the scouting collection from the double electron gun simulation is shown in Fig. 6.1. A large number of extra electrons are observed at low energy values. An inefficiency is observed in the pseudorapidity region between the ECAL barrel (EB) and end-cap (EE).

An angular association between the generator electron and the electrons in the scouting collection is necessary to estimate the reconstruction efficiency. Fig. 6.2 and Fig. 6.3 show the minimum angular difference between the generated electron and a scouting electron for the EB and EE, respectively. The difference between their pseudo-rapidity ( $\Delta\eta$ ) is symmetrical around zero. The difference between their azimuth ( $q\Delta\phi$ ) is multiplied by the charge of the generator

Table 6.4: List of the MC samples simulated with realistic LHC run 3 running conditions.

Sl. No.	MC simulation sample
1	JPsiToEE_Pt2To30_13p6TeV_TuneCP5_pythia8
2	JPsiToEE_JPsiPt28_EPt15To100_13p6TeV_TuneCP5_pythia8

electron, considering that the electron will bend in the CMS magnetic field. This distribution is ideally expected to be larger than zero. The angular matching condition is selected to be loose to have a maximal sample size to measure the reconstructed efficiency of electrons. The efficiency should hopefully not change depending on the selection unless the purity changes a lot.

$$\begin{aligned} |\Delta\eta| < 0.2, \quad -0.05 < q\Delta\phi < 0.3 \text{ in EB.} \\ |\Delta\eta| < 0.1, \quad -0.05 < q\Delta\phi < 0.2 \text{ in EE.} \end{aligned} \quad (6.1)$$

The angular condition is shown as the green distribution in the angular difference plots. A rough estimation of the remaining background can be obtained from the shoulder of the charge $\times\Delta\phi$  distribution in which the generator matching has been done (green) in Fig. 6.2 and Fig. 6.3 for EB and EE respectively. The background contamination after the generator matching is conservatively estimated to be less than 2% and 5% in EB and EE respectively.

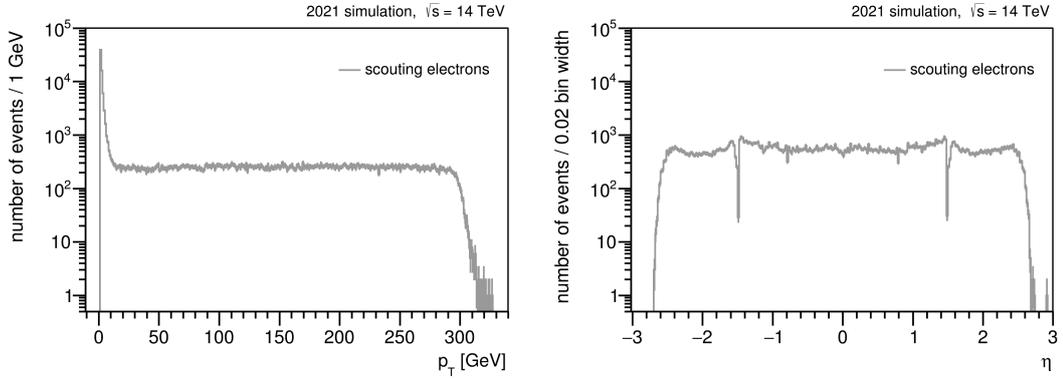


Figure 6.1: Kinematics of scouting electrons in the double electron gun samples show a large number of background electrons with low-energy and the  $\eta$  gap between the EB and EE.

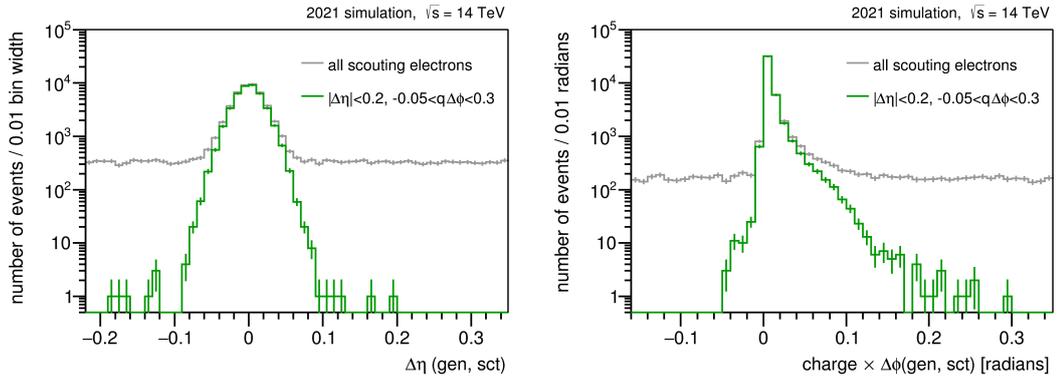


Figure 6.2: Distribution for scouting electrons in EB. Left: Minimum  $|\Delta\eta|$  between generator and scouting electrons with and without the angular selections applied to identify the signal electrons from background electrons. Right: The charge of the generator electron multiplied with  $\Delta\phi$  to take into account the charge dependent azimuthal bending due to the CMS magnet.

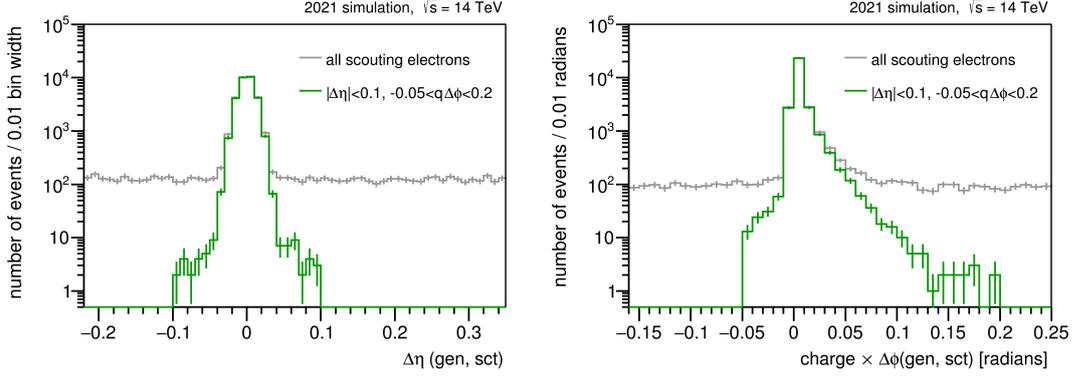


Figure 6.3: Distribution for scouting electrons in EE. Left: Minimum  $|\Delta\eta|$  between generator and scouting electrons with and without the angular selections applied to identify the signal electrons from background electrons. Right: The charge of the generator electron multiplied with  $\Delta\phi$  to take into account the charge dependent azimuthal bending due to the CMS magnet.

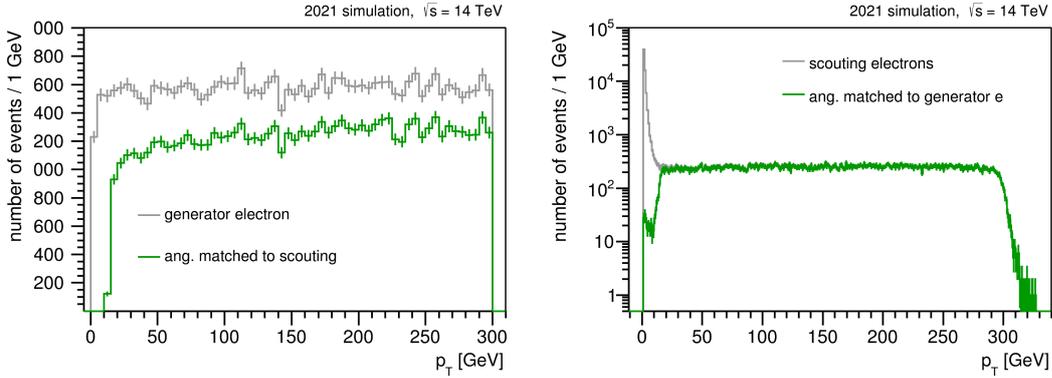


Figure 6.4: The result of angular matching between generator and scouting electrons shown in generator electrons (left) and scouting electrons (right).

The result of generator level angular matching with scouting electrons with the conditions in Eq. 6.1 is shown in Fig. 6.4. The left distribution shows the  $p_T$  of all electrons at the generator level for the electron gun. The reconstruction efficiency distribution obtained by dividing the green distribution after the angular matching conditions in Eq. 6.1 with the grey distribution for all electrons at the generator level is shown in Fig. 6.5. The scouting path was efficient for  $e p_T > 30$  GeV due to the L1 trigger threshold. The right distribution in Fig. 6.4 shows the effect of angular matching selection on electrons in the scouting collection. The angular matching shows a similar behaviour as the generator electrons. The electrons are efficiently stored in the scouting collection for  $e p_T > 30$  GeV. Some additional electrons are found to be matched at very low energies. This could happen due to an improper angular matching with background electrons at low energies.

The reconstruction was efficient for electron  $p_T > 30$  GeV due to the L1 trigger condition being L1\_SingleIsoEG28er2p1 OR L1\_SingleEG36er2p5 for electrons. The L1\_SingleIsoEG28er2p1

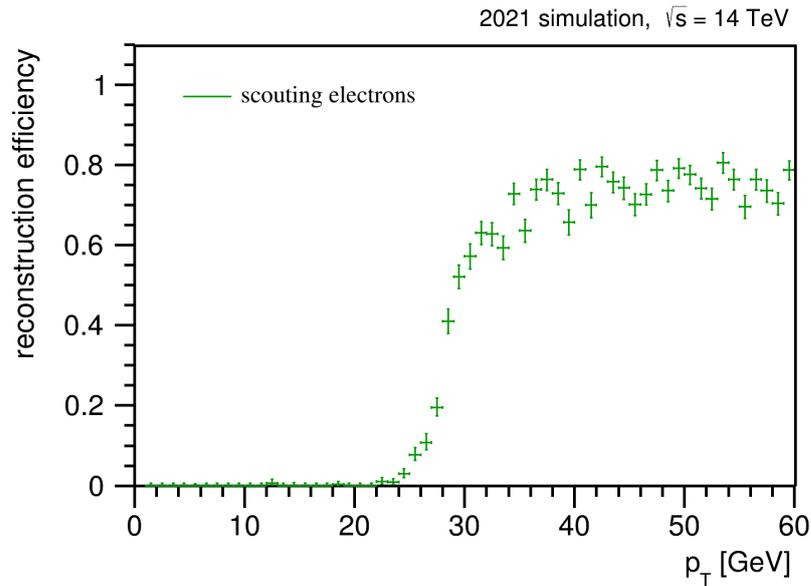


Figure 6.5: The reconstruction efficiency defined by dividing the good electrons by all generator electrons in Fig. 6.4 left distribution, shows scouting to be efficient for  $p_T > 30$  GeV. This effect is due to the L1 trigger.

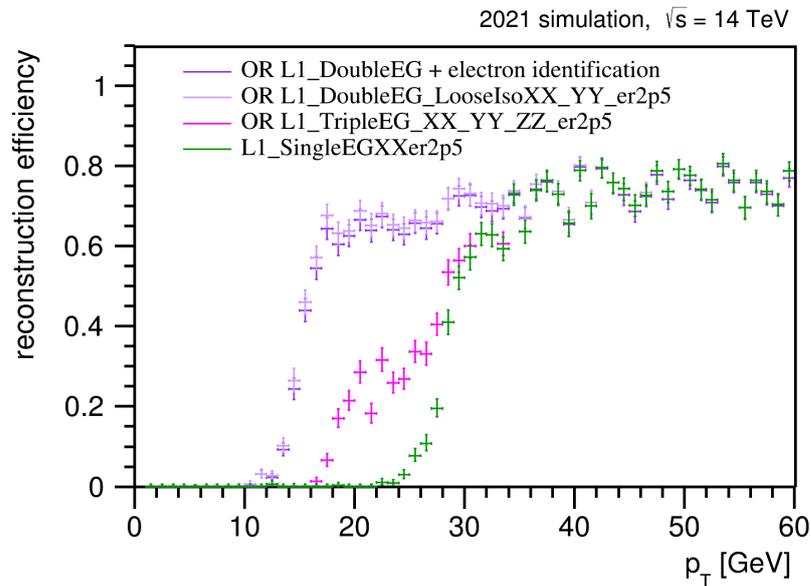


Figure 6.6: The inclusion of the L1\_DoubleEG triggers reduces the  $p_T$  threshold for the scouting to be efficient to  $< 20$  GeV.

requires a single ECAL tower with 30 GeV energy and  $|\eta| < 2.1$ . Similarly, the L1\_SingleEG36er2p5

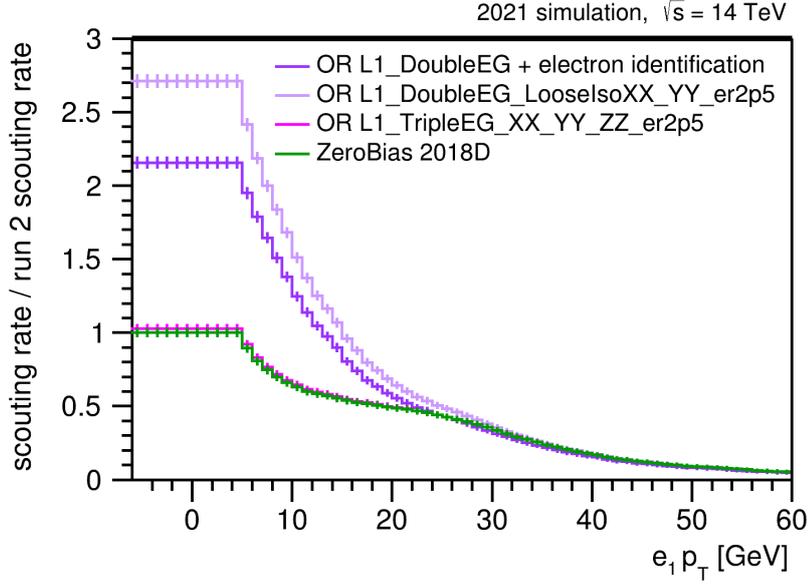


Figure 6.7: The inclusion of the L1\_DoubleEG triggers increased the rate by  $> 2.5$  times the L1 rate of the initial scouting menu. A basic shower shape based selection reduces the rate but the L1 input rate is still  $> 2$  times.

requires a single ECAL tower with 40 GeV energy and  $|\eta| < 2.5$ . The energy threshold is lower on di-electron/photon triggers. Lower energy thresholds can be obtained for electron and photon events in scouting using the L1 triggers with minimum thresholds that were un-prescaled. These were in the di-electron and tri-electron categories. In the di-electron category, the L1\_DoubleEG\_LooseIso22\_12\_er2p5 was tested and the L1\_TripleEG\_18\_17\_8\_er2p5, L1\_TripleEG\_18\_18\_12\_er2p5 and L1\_TripleEG16er2p5 were tested in the tri-electron category. The triple electron L1 rely on an extra electron from the surrounding activity such as from soft-scattering or PU and in principle should be avoided. It is worth studying to see the effect of the triple electron L1.

Figure 6.6 summarises the increase in efficiency with the lower L1 triggers. The di-electron L1 triggers provide better efficiency at low energy values while increasing the scouting throughput rate by a factor larger than 2.5, as shown in Fig. 6.7. The large rate strongly constrains its inclusion in the scouting menu. The rate study was performed using the zero bias dataset necessary for paths where the L1 conditions are re-emulated. Simultaneously, I also applied a very loose constraint on the shower shape variable of the electron and photon,  $\sigma_{i\eta i\eta} < 0.0126(0.0457)$  in EB(E), motivated by the run 2 identification criteria for electrons and demonstrated that the rate can be reduced further by rejecting background events while retaining a high efficiency for low-energy electron events. The rate with the identification selection was, however, still too high.

A parallel work based on my studies suggested using an L1 with the same energy thresholds but restricted to the ECAL barrel. Restriction to the ECAL barrel lowers the rate of the trigger considerably. For the L1\_DoubleEG\_LooseIso22\_12\_er2p5, the current L1 rate is 4.5 kHz compared to  $< 1$  kHz if it was restricted to EB. The L1\_DoubleEG\_LooseIso16\_12\_er1p5 has a rate of 2 kHz. Based on rate considerations, the L1 trigger list in Table 6.5 was added to

Table 6.5: List of the new L1 triggers added to scouting in 2022 for di-electron and di-photon events with reduced acceptance in  $\eta$  to control the rate of the scouting path.

New L1 triggers included in scouting
L1_DoubleEG_LooseIso16_LooseIso12_er1p5
L1_DoubleEG_LooseIso18_LooseIso12_er1p5
L1_DoubleEG_LooseIso20_LooseIso12_er1p5
L1_DoubleEG_LooseIso22_LooseIso12_er1p5

the scouting menu. The L1 menu with the new L1  $e/\gamma$  triggers was integrated into the HLT menu for scouting data collection for the first HLT menu implemented since the 2022 run of the LHC. Another important feature of this trigger is that it has no angular constraints on the two electrons, which provides selection in a phase space not accessible to the parking data collection method from the CMS detector.

### 6.3 Identification of electrons in the scouting collection

It was necessary to establish a concept of proof that effective electron discrimination could be done for isolated electrons from the background with the information stored in the scouting collection. The ability to identify good electrons that provide a detector signature of interest from the background physical processes, noise and fakes is called the identification of electrons. The  $e/\gamma$  L1 trigger conditions were dropped and all events were accepted from the MC simulation. The  $e/\gamma$  L1 triggers were removed to include scenarios where other L1 triggers triggered the event and required a well-reconstructed and isolated electron in the detector signature. This study was performed with the 2021 simulation of expected run conditions for the run 3 of the LHC.

The double electron gun and the  $Z \rightarrow ee$  sample were chosen as signal electrons while the QCD sample enriched with electromagnetic processes and the QCD sample with  $b$ -quark and  $c$ -quark jets decaying to electrons were used as background electrons. Additionally, the invariant mass of the electron pair in the  $Z \rightarrow ee$  sample was chosen to be in the  $Z$ -boson mass window. Choosing the di-electron invariant mass in the  $Z$ -boson mass range identifies a larger number of electrons from the decay of the  $Z$ -boson. Hence, it increases the proportion of isolated high-energy electrons in the sample and reduces the electrons coming from background sources. This method is also called increasing the purity of the sample. The electron identification used here was derived by the CMS collaboration for electron identification in run 2 [187]. Since the information stored in scouting is not identical to the offline reconstruction, an identification analogous to the run 2 electron identification has been used. Furthermore, it has been compared to another identification developed by me for low-energy electrons.

The identification efficiency is defined by dividing the distribution of electrons that pass a certain selection by the distribution of all the electrons in the sample. A good electron identification will reduce the number of electrons passing from the background sample while maintaining high efficiency for the signal electrons. The signal electron identification variables are studied based on if they are low-energy ( $< 20$  GeV) or high-energy ( $\geq 20$  GeV) to enable observing the behaviour differences between them. Fig. 6.8 shows the  $p_T$  distribution of all the categories of electrons considered in the identification study. No generator level matching has been performed

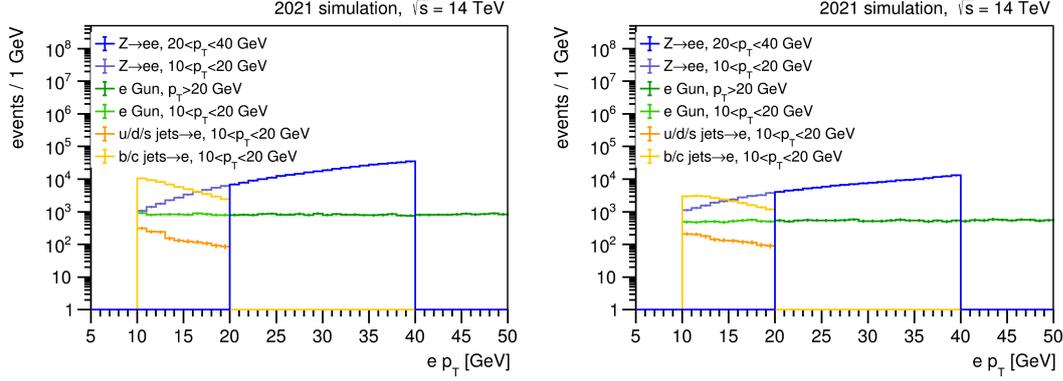


Figure 6.8: The  $p_T$  of scouting electrons with the selections for identification study. The signal electrons are from the  $Z \rightarrow ee$  (blue) and double electron gun (green) sample. They have been split into two regions based on the  $e p_T$ . Electron  $p_T < 20$  GeV (lighter shade) enabled a study of the low-energy electrons and comparison with high-energy electrons with  $e p_T > 20$  GeV (darker shade) in the signal MC samples. The background is from light jets (orange) and jets from heavy quarks (yellow). Only the low-energy electrons are considered for the background. The absolute number of events for each sample is shown in this plot. Left: EB, Right: EE.

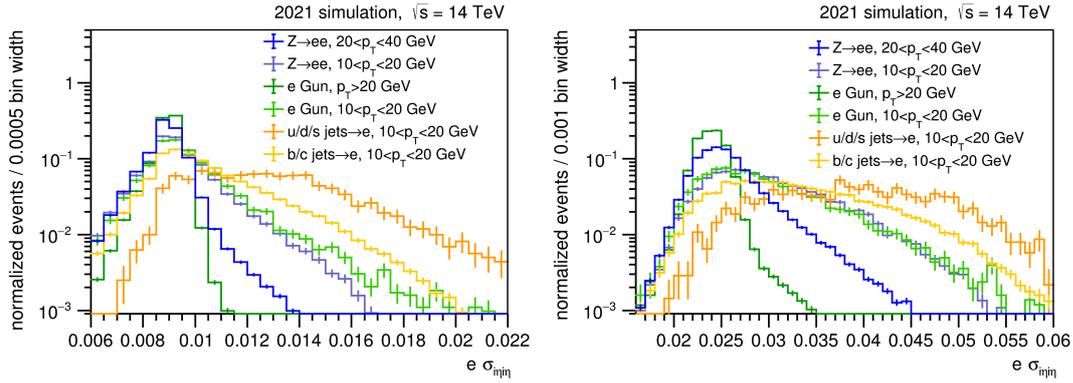


Figure 6.9: The  $\sigma_{i\eta i\eta}$  variable distribution is shown for the different background and signal samples as described in Fig. 6.8. The signal samples, shown in green and blue, are peaked while the background samples, shown in orange and yellow, are relatively spread. The low-energy signal electrons, shown in faded green and blue, are more spread compared to the high-energy signal electrons. A relatively loose selection on  $\sigma_{i\eta i\eta}$  will increase efficiency for low-energy electrons. Left - EB. Right - EE.

for the scouting electron variable distributions. This is primarily because the study was done very early to provide proof that the scouting collection's electrons can be well identified with the available set of variables. The variables at this stage were all derived from the ECAL SC.

Fig. 6.9 shows the distribution of the shower shape variable, the  $\sigma_{i\eta i\eta}$ , of the electron in the ECAL. It measures the spread in the shower shape of the energy deposition in the ECAL. There are two versions, one of which is cleaned for noise in the ECAL. This was found to work better for background discrimination by the electron and photon group of the CMS collaboration. For

high-energy electrons, the shower shape is centralised around one single ECAL crystal with a small amount of energy in the crystals surrounding the central crystal. The  $\sigma_{i\eta i\eta}$  has a sharp peak in its distribution corresponding to this behaviour for signal high-energy electrons. The background electrons tend to be non-isolated, with a fairly high amount of energy in the crystals surrounding the central crystal from other particles surrounding the electron. The distribution of  $\sigma_{i\eta i\eta}$  tends to be spread for background electrons. The energy deposit from a low-energy electron is fairly small in the central crystal. The energy deposit from PU interactions tends to reduce the centrality of the value of  $\sigma_{i\eta i\eta}$  for low-energy electrons. The PU simulation is absent for the double electron gun sample. As a result the energy deposit in the surrounding crystals of an  $e/\gamma$  is lower resulting in more peaked distribution for  $\sigma_{i\eta i\eta}$  in the double electron gun sample compared to  $Z \rightarrow ee$ . The enhanced noise related to the fast ageing of the crystals in the forward region is likely the main cause of this. A looser selection than the run 2 selection [187] increases the identification efficiency for low-energy electrons but with more background.

The behaviour of the low-energy electrons in the  $H/E$  distribution is shown in Fig. 6.10 left for EB and right for EE. Most of the signal electrons in the barrel have energy deposition in the hadronic calorimeter close to zero. The electrons in EE in the  $Z \rightarrow ee$  sample have a significant amount of energy deposit in the HCAL for the corresponding ECAL SC. This could be attributed to the high PU in the EE. The PU simulation is more faithfully done for the  $Z \rightarrow ee$  sample compared to the electrons in the electron gun sample. As the energy deposit in the ECAL reduces for low-energy electrons, the fraction  $H/E$  takes higher values due to the unchanged energy deposit from noise and PU in the HCAL. This results in the  $H/E$  distribution for low-energy electrons spreading to larger values than high-energy electrons. The  $H/E$  selection is kept loose. It is kept unchanged from a pre-selection value of  $H/E < 0.2$  during reconstruction. Since tracking is a computationally expensive algorithm, a pre-selection  $H/E < 0.2$  is required before running the track reconstruction. This is retained for electron identification.

The  $|\Delta\eta_{in}^{seed}|$  and the  $|\Delta\phi_{in}|$  distributions, shown in Fig. 6.11 and Fig. 6.12 respectively for EB - left and EE - right, are useful variables to separate the signal electrons from electrons in

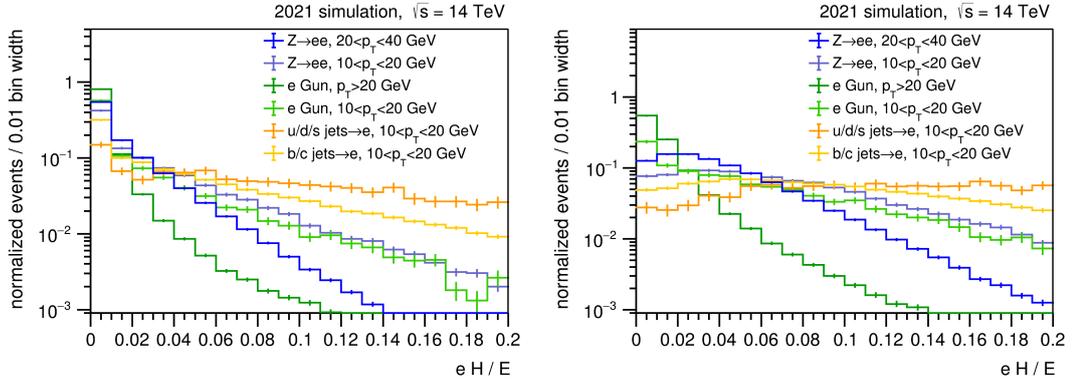


Figure 6.10: The  $H/E$  variable distribution is shown for the different background and signal samples as described in Fig. 6.8. The signal samples, shown in green and blue, have a larger number of events with lower HCAL energy deposit compared to the background samples, shown in orange and yellow. The low-energy signal electrons, shown in faded green and blue, have relatively higher values for  $H/E$  than high-energy electrons. A selection  $H/E < 0.2$  is retained for low-energy electron identification. Left - EB. Right - EE.

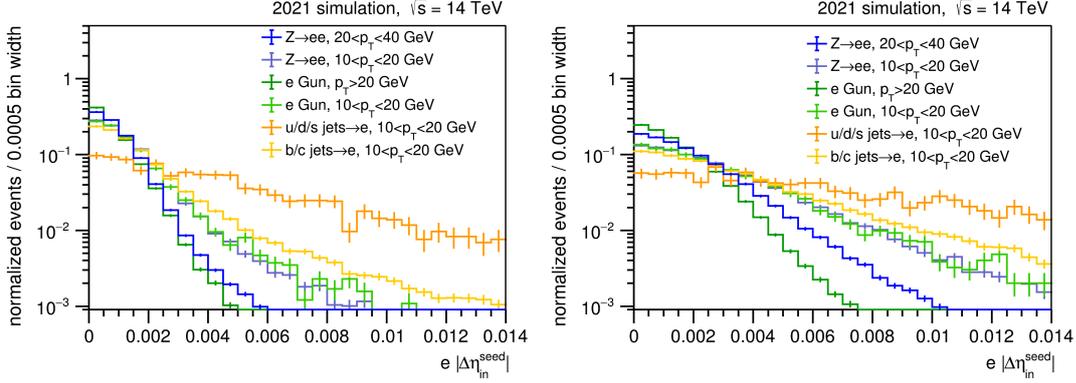


Figure 6.11: Comparison for the  $|\Delta\eta_{in}^{seed}|$  variable. The tracks with actual electrons in them such as those from signal and from  $b/c \rightarrow e$  jets have lower angular difference. This variable separates the signal electrons from the electrons in light jets. Left - EB. Right - EE.

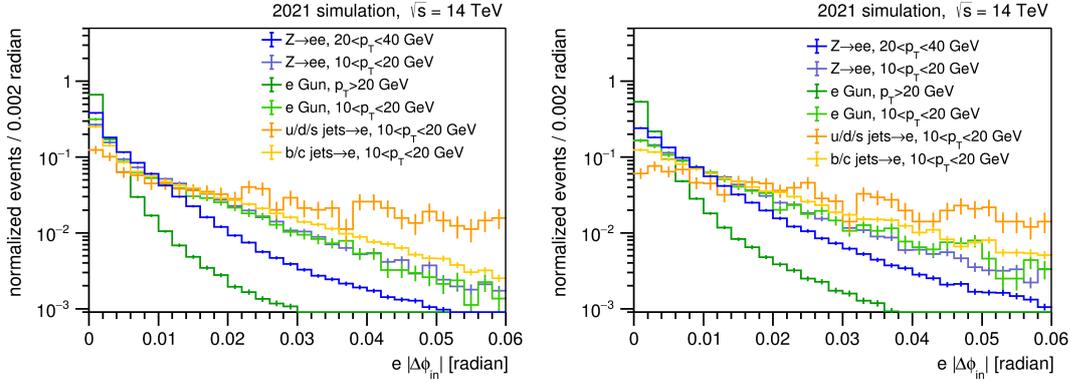


Figure 6.12: Comparison for the  $|\Delta\phi_{in}|$  variable. The tracks with actual electrons in them such as those from signal and from  $b/c \rightarrow e$  jets have lower angular difference. This variable separates the signal electrons from the electrons in light jets. Left - EB. Right - EE.

light jets. Considering the large background from light jets, a moderately tight selection is placed on these variables.

The ECAL isolation for electrons, Fig. 6.13 left - EB and right - EE, is defined in a cone  $\Delta R = 0.3$  around the ECAL SC. In the signal events there is no substantial energy deposit around the electron so the isolation is close to zero. In background events, such as those coming from jets, there is a large background of other particles. The energy deposit in the background crystals for non-isolated low-energy electrons is high and thus the isolation is low. So all the isolation variables are good discriminators of isolated low-energy electrons, especially in the sample  $b/c \rightarrow e$ , where actual electrons are present and the track parameters do not discriminate as well. The isolation value for the electron gun cannot be trusted as the PU simulation has not been included.

Similarly, the isolation of the electron track, Fig. 6.14 left - EB and right - EE, in the tracker also serves as an important discriminator. The tracker isolation is especially effective in

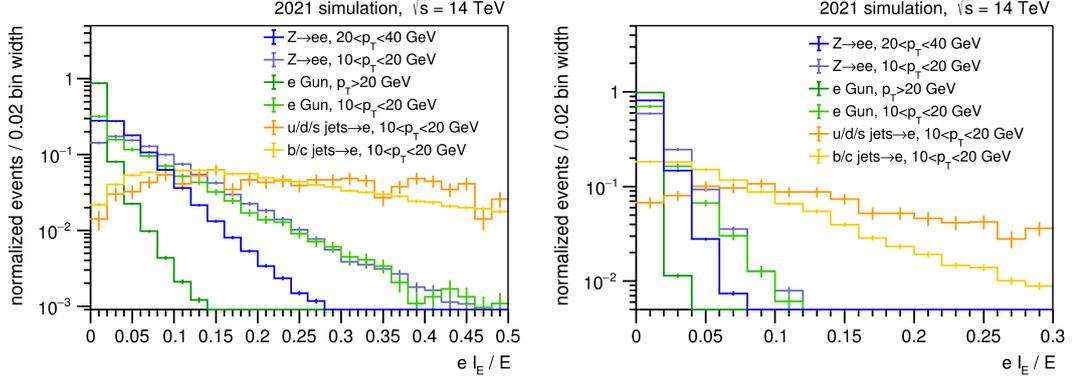


Figure 6.13: Comparison for the relative isolation in ECAL, the  $I_E/E$  variable. The signal electrons, shown in green and blue, are relatively well isolated compared to the background electrons. The  $I_E/E$  for low-energy electrons is relatively high than the high-energy electrons. Left - EB. Right - EE.

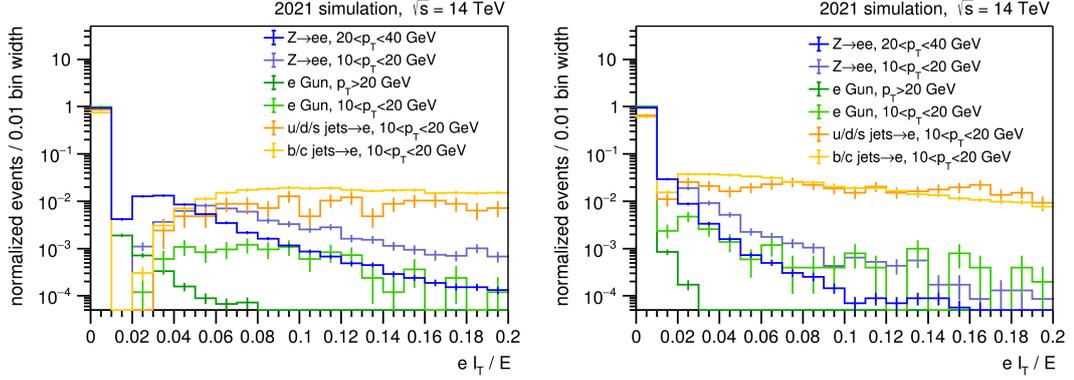


Figure 6.14: Comparison for the relative isolation in the tracker, the  $I_T/E$  variable. Left - EB. Right - EE.

separating the low-energy electrons from the electrons in jets from  $b$ -quarks and  $c$ -quarks.

Isolation in HCAL, shown in Fig. 6.15, is not used. The HCAL noise in an angular cone  $\Delta R = 0.3$  is expected to be a significant proportion of the ECAL energy deposit from the low-energy electron. A tight selection on this HCAL isolation might just be effective in removing events with high HCAL noise. The selection on  $H/E$  and selection on other isolation variables is estimated to achieve the required separation from the QCD and other background events.

Fig. 6.16 shows the identification efficiency comparing the standard identification selections to the selections for low-energy electrons for EB (left) and EE (right). Table 6.6 lists the selections on the individual variables chosen manually with a visual estimation of the best discrimination point between the signal and background electrons and with the standard identification selections as a reference. The identification efficiency is obtained by dividing the electrons that pass a certain identification with all the scouting electrons in the scouting collection. The darker distribution in all colour categories shows the low-energy identification selection while the lighter

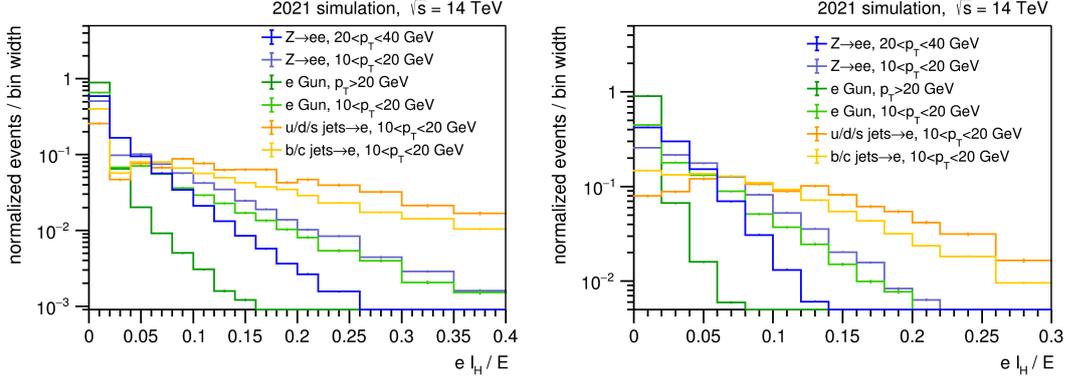


Figure 6.15: Comparison for the relative isolation of the electron in the HCAL, the  $I_H/E$  variable. The noise and energy deposit from PU is substantial compared to the low-energy of the electrons. The discriminating power of this variable decreases for electrons with lower energy. The isolation in HCAL is high for the electrons in the double electron gun sample due to absence of PU in the simulation. Left - EB. Right - EE.

Table 6.6: Identification selection on all variables demonstration for higher efficiency for low-energy electrons in the scouting collection.

variable	EB		EE	
	run 2 id.	low- $p_T$ scouting id.	run 2 veto id.	low- $p_T$ scouting id.
$\sigma_{i\eta i\eta} <$	0.013	0.015	0.035	0.045
$H/E <$	0.13	0.2	0.13	0.2
$ \Delta\eta_{in}^{seed}  <$	0.01	0.008	0.015	0.012
$ \Delta\phi_{in}  <$	0.07	0.06	0.1	0.06
$I_E/E <$	0.5	0.25	0.5	0.1
$I_T/E <$	0.2	0.001	0.2	0.001
$I_H/E <$	0.3	-	0.3	-

distribution shows the standard selection used for high-energy electrons. It is inferred that the low-energy identification retains most of the efficiency in signal electrons while reducing the electrons from the background QCD events by a significant factor in the region  $p_T > 10$  GeV. The identification also performs better in the ECAL barrel. This study serves as proof that a dedicated low-energy electron identification can be developed with the variables stored in the scouting collection.

## 6.4 From seeded to unseeded reconstruction

The standard  $e/\gamma$  reconstruction algorithm is initiated around an L1 ECAL trigger tower. An L1 tower sums the energy in the ECAL and defines a threshold on whether the energy deposit and distribution is significant. [147]. This methodology is referred to as the L1 seeded  $e/\gamma$

reconstruction.

A different reconstruction algorithm tries to make SCs in the entire ECAL without requiring seeding from an L1 trigger tower. This algorithm combines information from all the channels in the ECAL instead of a regional reconstruction and is thus time expensive. It significantly increases the efficiency of low-energy electron reconstruction at the HLT.

The JPsiToEEM2To50 MC sample with the 2023 realistic simulation conditions were used for this study. No L1 selection was applied to check and compare the  $e/\gamma$  reconstruction module's reconstruction ability in events not triggered by the L1  $e/\gamma$ . As shown in Fig. 6.17, unseeded reconstruction provides better efficiency. It does, however, result in increased time complexity of the scouting algorithm and rate of the scouting path beyond the limit provided at the HLT for the CMS experiment. A plan was developed in the scouting working group of the CMS collaboration and was put into use to take advantage of the unseeded reconstruction.

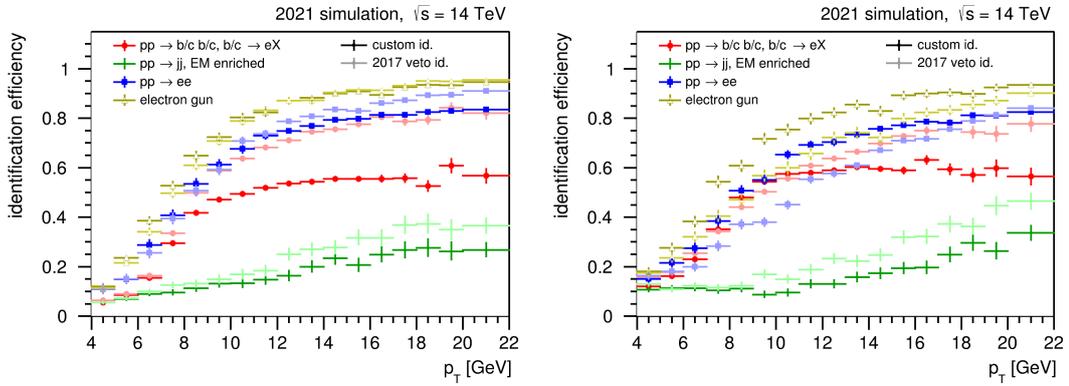


Figure 6.16: Identification efficiency for scouting electrons based on selections in table 6.6. The blue and yellow histograms represent signal samples while the red and green histograms represent the background samples. The darker histogram in each sample is the low- $p_T$  scouting id. while the lighter histogram is the run 2 electron id. The identification with the low- $p_T$  scouting id. rejects background better with a small reduction in the signal efficiency for low energy electrons compared to the run 2 identification. Left - EB. Right - EE.

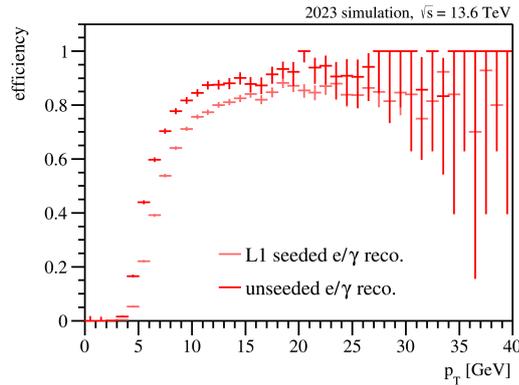


Figure 6.17: Increased efficiency with unseeded reconstruction.

The scouting path was split into three paths based on the L1 triggers. The single L1  $e/\gamma$  triggers were put into a path called `DST_Run3L1SingleEG30_PFScoutingPixelTracking_v`. The double L1  $e/\gamma$  triggers were put into a path called `DST_Run3L1DoubleEG16EG12_PFScoutingPixelTracking_v`. All other L1 triggers were kept in the original scouting path called `DST_Run3_PFScoutingPixelTracking_v`. The scouting reconstruction module was kept unchanged between the three paths. However, electron identification was imposed for the  $e/\gamma$  scouting paths before running the scouting reconstruction module. The L1 seeded  $e/\gamma$  reconstruction was run and at least one electron with  $p_T > 30$  GeV was required for `DST_Run3L1SingleEG30_PFScoutingPixelTracking_v` and at least two electrons with  $p_T > 12$  GeV were required for `DST_Run3L1DoubleEG16EG12_PFScoutingPixelTracking_v`. The electrons were additionally required to pass  $\sigma_{i\eta i\eta} < 0.014$  (0.035) in the EB (EE). The identification was based on the work in the previous section on electron identification and, in addition to reducing the rate and time of the scouting menu, increased the purity of the scouting sample. The electrons in the path `DST_Run3_PFScoutingPixelTracking_v` did not pass through any other selection. The rate reduction was achieved by rejecting background events with the identification in the  $e/\gamma$  triggered L1 paths.

Table 6.7 shows the rate reduction and time change with the scouting reconstruction split to three HLT paths based on the L1 triggers. The rate of the path is reduced and the time remains relatively unchanged for higher efficiency for the low-energy electrons. Later, to reduce the time of the HLT paths, a similar idea was used for the remaining group of L1 triggers. This work was done in collaboration with Bennett Paul Greenberg from Daniel R. Marlow's group at Princeton University. I created the HLT configuration for this while he was involved in testing the rate and time of the paths. The efficiency plot with the  $J/\psi$ -meson was produced while he did a simultaneous efficiency plot with the  $\eta$ -meson.

## 6.5 Identification of low-energy SM particles

The scouting  $e/\gamma$  paths should be able to detect the SM particles under the constraint of the L1 energy thresholds of the ECAL and the resolution for low-energy electrons. A search for the  $\eta$ -meson decaying to  $2\mu$  and  $2e$  is targeted by a collaborating group. It should be possible, in particular, to use the  $J/\psi$ -meson to understand low-energy electrons in scouting. Electrons from decay of the  $J/\psi$ -meson would allow for energy and angular resolution measurement of the scouting electrons. They would also enable a comparison between the performance of the electrons in data scouting compared to the standard electrons used by the CMS collaboration for physics analyses. A comparison between data and MC simulation is also important to characterise and standardise the scouting electrons.

The MC simulation sample for  $pp \rightarrow J/\psi \rightarrow ee$  with electron  $p_T$  threshold greater than 15 GeV as given in tab. 6.4 has been used. A maximum energy requirement of 100 GeV is also applied on the electrons to have higher event statistics in the energy range of interest for scouting

Table 6.7: Rate and time with scouting path split by the L1 triggers.

Path	Rate (Hz)	HLT menu time (ms)
Unsplit	30281	$476 \pm 5$
Split	22138	$471 \pm 4$

electrons. The maximum energy requirement has a minimal effect due to the exponentially decaying electron energy distribution.

The  $J/\psi$ -meson, as a result of the minimum  $p_T$  requirement, is boosted. The opening angle of the decay electrons is narrow. Fig. 6.18 shows the  $\Delta R$  between the daughter electrons of the  $J/\psi$ -meson. Most of the simulated events have  $\Delta R < 0.15$ . The events that pass the scouting paths are boosted even further due to a higher energy threshold than the simulated threshold which is why that choice for simulation threshold is okay. The  $\Delta R$  distribution for the events that pass the scouting path peaks at 0.1. Since the L1 ECAL trigger towers construct in the cone  $\Delta R < 0.3$ , most events pass the single  $e/\gamma$  L1 trigger. The events that pass the double  $e/\gamma$  L1 trigger have a wider opening angle, so the angular resolution at the L1 can reconstruct them as separate particles. Consequently, the events also have a lower boost compared to the single  $e/\gamma$  L1 trigger. Very few events pass the combination of non- $e/\gamma$  based L1 triggers.

The  $J/\psi$ -meson simulation contains background from PU and fakes from imperfect reconstruction algorithms. It is essential to filter the electrons from  $J/\psi$ -meson decay in the scouting collection to develop a proper identification for the electrons of interest. An angular match between the available generator-level truth information and the scouting collection is done following a similar strategy as in Section 6.2. At least one electron must be reconstructed in the scouting collection for the selection study. Fig. 6.19 and Fig. 6.20 show the  $\Delta\eta$  (left) and the charge of the generator electron ( $q$ ) multiplied by the  $\Delta\phi$  (right) for the EB and EE respectively. The  $\Delta\eta$  is almost symmetric about zero. A secondary peak is observed at about 0.1 from those events where only one electron was reconstructed and hence both the generator electron are filled using the single electron. The angular matching constraint allows a maximum of one scouting electron to be matched to the generator electron which is typically smaller than 0.1 and removes the potential duplication that could occur because of the proximity between the two electrons. The difference of 0.1 arises due to the angle between the generator electrons. The condition for generator-based angular selection for the electron SC is chosen to be Eq. 6.2.

$$\begin{aligned} |\Delta\eta| < 0.005 \text{ and } -0.06 < q \times \Delta\phi < 0 \text{ for EB.} \\ |\Delta\eta| < 0.03 \text{ and } -0.06 < q \times \Delta\phi < 0.01 \text{ for EE.} \end{aligned} \quad (6.2)$$

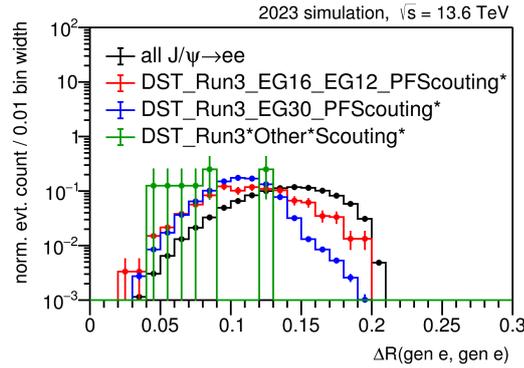


Figure 6.18: The  $\Delta R$  distribution between the generator electrons from the  $J/\psi$ -meson show that most events pass the single  $e/\gamma$  L1 trigger. The green distribution shows the events collected by scouting paths with L1 triggers excluding the single and the double  $e/\gamma$  trigger.

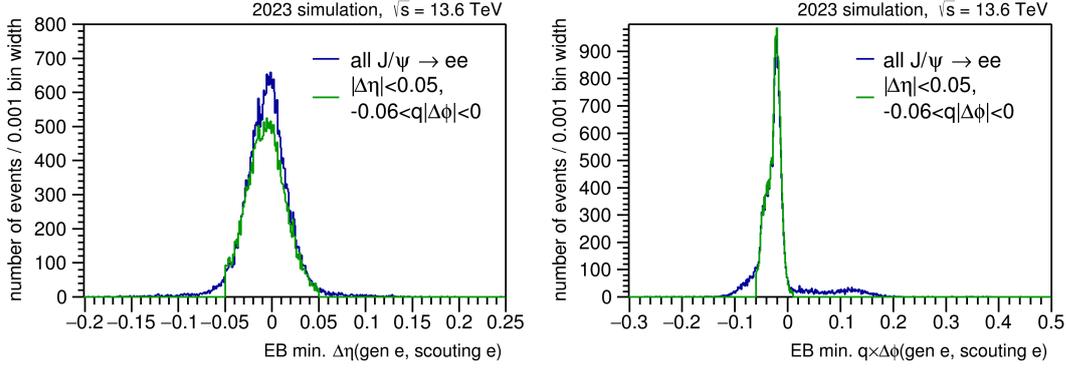


Figure 6.19: For EB: The difference between the generator  $\eta$  and scouting SC  $\eta$  (left), generator charge multiplied with difference between the generator  $\phi$  and the scouting  $\phi$  (right) before and after the chosen generator matching conditions.

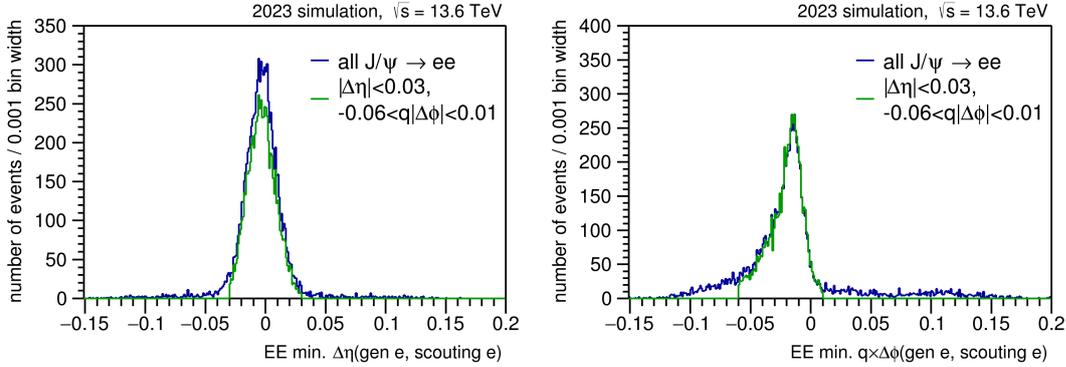


Figure 6.20: Same as Fig. 6.20 for EE.

To obtain an angular match with a generator electron, the electrons in both generator and scouting collection are sorted by highest  $p_T$  first. For every generator electron, a search is done for a scouting electron in the angular cone specified by Eq. 6.2. If a match is found, it is stored and the associated scouting electron is removed from further angular match with the next generator electron.

Fig. 6.21 to Fig. 6.27 show the distribution of variables for the ECAL SC in EB (left) and EE (right). At least one reconstructed electron is required in the scouting collection. A tighter selection will reduce the statistics in the plot; hence such a loose selection is made to maximise the statistics of the plot. The green histogram in each distribution is for scouting electrons for whom an angular match is found in the generator electrons and hence are signal enriched. The blue histogram is for all scouting electrons in the collection. No other event selection is applied. The distributions are normalised so that the comparison in the shape of the distributions can be made efficiently.

The identification selection is done visually based on an estimate of a value which provides the best separation between generator matched and all scouting electrons. The  $\sigma_{i\eta\eta}$  selection is

chosen when a sharp drop in signal electrons is observed. The  $H/E$  variable is useful to identify good SCs with low PU deposition around them. The selections on  $1/E - 1/p$ ,  $|\Delta\eta_{in}^{seed}|$  and  $|\Delta\phi|_{in}$  are used to identify SCs with good tracks in them. The ECAL and tracker isolation are dropped because the isolation is defined in a  $\Delta R < 0.4$  cone where the other electron from the  $J/\psi$ -meson decay is present. These electrons are of relatively low energy,  $p_T < 50$  GeV, and hence are unlikely to make it to the HCAL. The isolation in HCAL is used to separate events with high PU around the  $J/\psi$ -meson. It is also necessary in data to reduce the events with large hadronic activity such as jets.

Multiple tracks can be reconstructed for a single SC starting from the primary vertex, see Fig. 6.29. The tracks provide better angular resolution for low-energy electrons than the SC. This will be illustrated later in data. Track identification is performed to reduce the track multiplicity for every SC as well as increase the quality of the tracks considered in the event. A good track is identified by studying the quality of reconstruction of the track and match the energy and angular difference with respect to the SC. A similar angular match philosophy is used to study

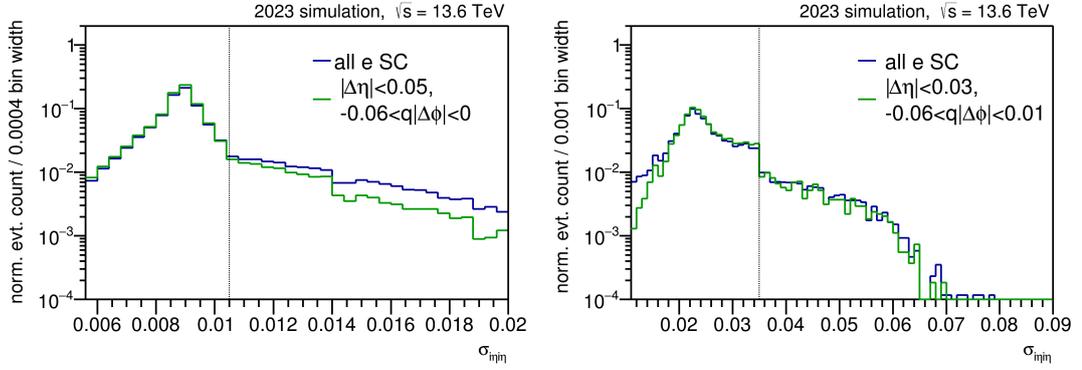


Figure 6.21: The ECAL SC shower shape  $\sigma_{inj\,inj}$  variable cleaned for noise in ECAL with selection for the  $J/\psi$ -meson. The vertical line shows the value where the selection cut is made. The variable needs to have a value less than this cut. Left - EB, Right - EE.

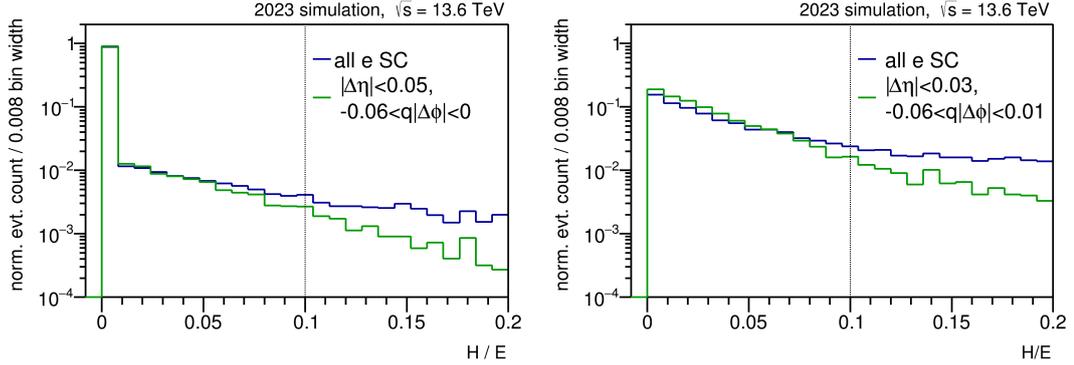


Figure 6.22: The ECAL SC  $H/E$  variable for proportion of energy deposit in the HCAL with selection for the  $J/\psi$ -meson. The vertical line shows the value where the selection cut is made. The variable needs to have a value less than this cut. Left - EB, Right - EE.

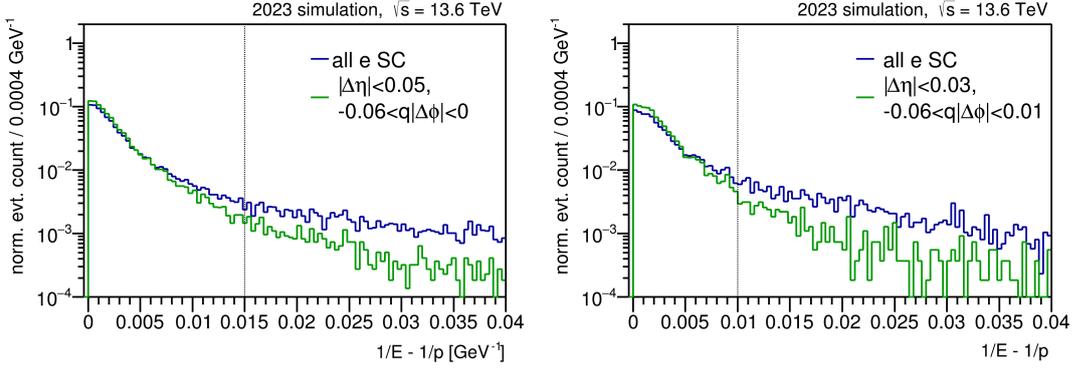


Figure 6.23: The ECAL SC  $1/E - 1/p$  variable with selection for the  $J/\psi$ -meson. The vertical line shows the value where the selection cut is made. The variable needs to have a value less than this cut. Left - EB, Right - EE.

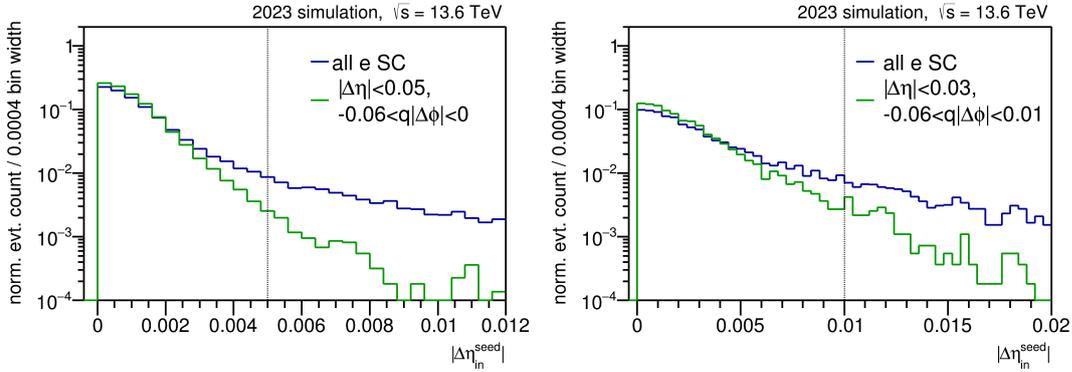


Figure 6.24: The ECAL SC  $\Delta\eta_{in}^{seed}$  angular difference with track variable with selection for the  $J/\psi$ -meson. The vertical line shows the value where the selection cut is made. The variable needs to have a value less than this cut. Left - EB, Right - EE.

these variables for the  $J/\psi$ -meson.

Fig. 6.30 and Fig. 6.31 show the angular differences  $\Delta\eta$  (left) and  $q \times \Delta\phi$  (right) for the EB and EE respectively. Each SC has a minimum of one track required as the baseline. The minimum  $|\Delta\eta|$  between a generator electron and a reconstructed scouting track for each generator electron is chosen for the distribution. Similarly, the minimum  $q \times \Delta\phi$  is chosen to fill the right distribution. An immediate comparison with the SC angular difference variables in Fig. 6.19 and Fig. 6.20 reveals that the tracks represent the angular parameters of the generator electron at least one order of magnitude better than the SC for low-energy electrons.

$$\begin{aligned} |\Delta\eta| < 0.0015 \text{ and } -0.004 < q \times \Delta\phi < 0.002 \text{ for EB.} \\ |\Delta\eta| < 0.003 \text{ and } -0.006 < q \times \Delta\phi < 0.003 \text{ for EE.} \end{aligned} \quad (6.3)$$

A similar angular matching philosophy is followed for the tracks as the SC. The generator electrons and the scouting SCs are sorted with the highest energy first. For each generator

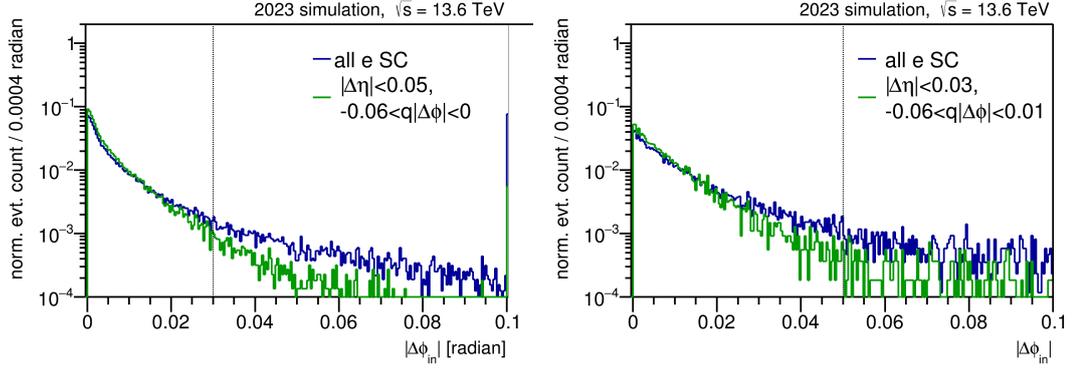


Figure 6.25: The ECAL SC  $\Delta\phi_{in}$  angular difference with track variable with selection for the  $J/\psi$ -meson. The vertical line shows the value where the selection cut is made. The variable needs to have a value less than this cut. Left - EB, Right - EE.

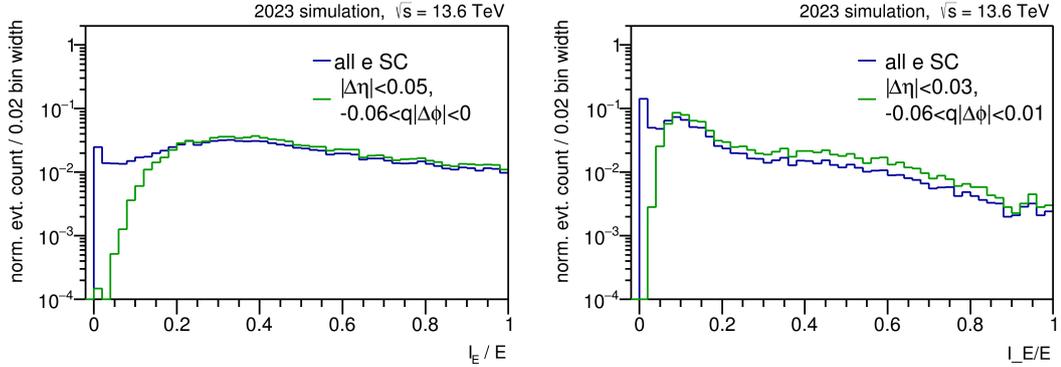


Figure 6.26: The ECAL SC relative isolation  $I_E/E$  variable. No selection is applied on this variable for the  $J/\psi$ -meson. Left - EB, Right - EE.

electron, a track is searched for in each SC using the conditions mentioned in Eq. 6.3. If a track is found to satisfy the angular conditions, it is stored and the corresponding scouting electron (i.e the SC) and all its tracks are removed from the list before generator matching is performed for the next electron.

The generator matched scouting electron tracks are plotted in the green histogram distributions while the blue distribution includes all the tracks for each SC. The distributions are normalised to enable comparison between them.

The reconstruction is efficient for track  $p_T > 12$  GeV, see Fig. 6.32. A large number of fakes are observed for  $p_T < 12$  GeV. A selection  $p_T > 12$  GeV is used to reduce the multiplicity of tracks for each SC.

The ECAL is a better estimator of energy than the track. The track often loses energy due to the missing bremsstrahlung photons. Especially with low-energy electrons, the tracks have a larger azimuthal bending due to the CMS magnetic field. They lose a larger amount of energy from bremsstrahlung due to the longer path in the tracker layers. Tracks for such electrons are

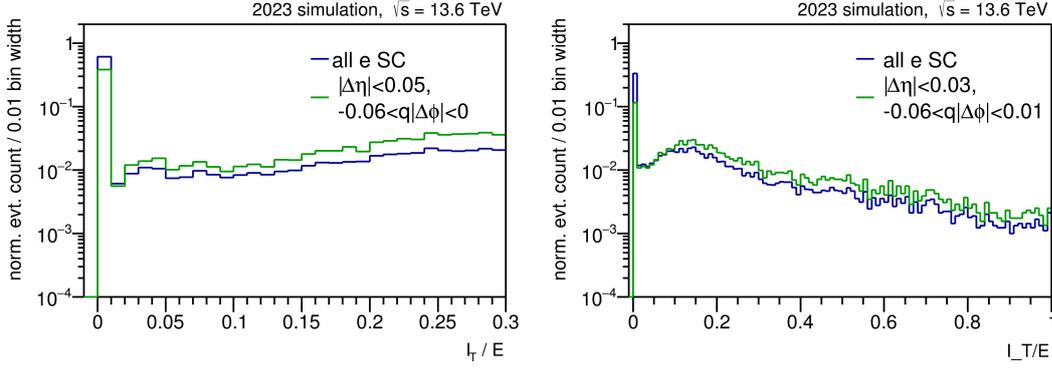


Figure 6.27: The ECAL SC relative isolation  $I_T/E$  variable. No selection is applied on this variable for the  $J/\psi$ -meson. Left - EB, Right - EE.

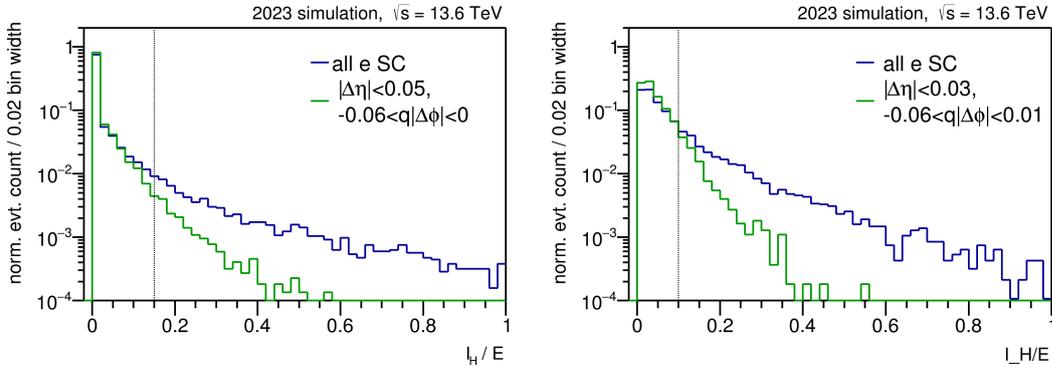


Figure 6.28: The ECAL SC relative isolation in HCAL  $I_H/E$  variable with selection for the  $J/\psi$ -meson. The vertical line shows the value where the selection cut is made. The variable needs to have a value less than this cut. Left - EB, Right - EE.

reconstructed with lower energy and larger mismatch with the parent particle than the SC. A comparison in the energy of the SC and the track is shown in Fig. 6.33 for EB (left) and EE (right). The distribution peaks at a value of about 5% less than one. The selection on this variable is currently kept loose at  $|\Delta E/E| < 1$  but there is scope for improvement with this variable. An asymmetric cut around zero should also be considered for this variable.

The track  $\chi^2/d.o.f.$  of the track quantifies the quality of track reconstruction. Ideally, it is required to be close to one. The distribution of the track  $\chi^2/d.o.f.$  for the electron track is shown in Fig. 6.34. As expected, the distribution peaks at one and includes most generator-matched tracks. At higher values, the number of tracks that are generator matched reduces exponentially by at least one order of magnitude. The track is required to pass a selection  $\chi^2/d.o.f. < 3(2)$  in EB(EE).

The azimuthal difference between the SC and the track, shown in Fig. 6.35 for EB (EE) in the left (right), is expected to be symmetric about zero with peaks slightly off-centre due to the higher angular resolution of the track compared to the angular resolution of the SC. The low-energy

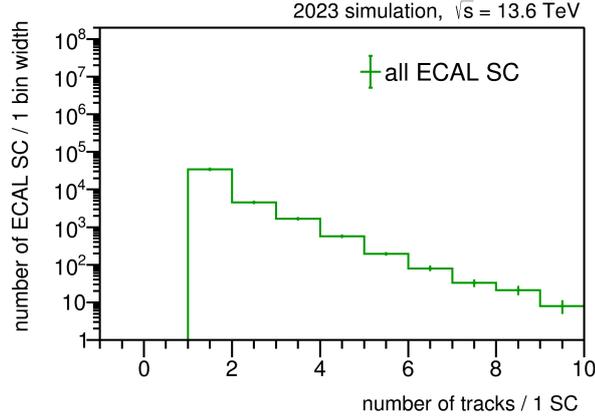


Figure 6.29: Every SC can have multiple tracks reconstructed in association to it. The number of tracks for every SC falls exponentially. However, some ECAL SC can have more than 10 tracks reconstructed in association to it. The correct track corresponding to the electron that resulted in the SC is essential.

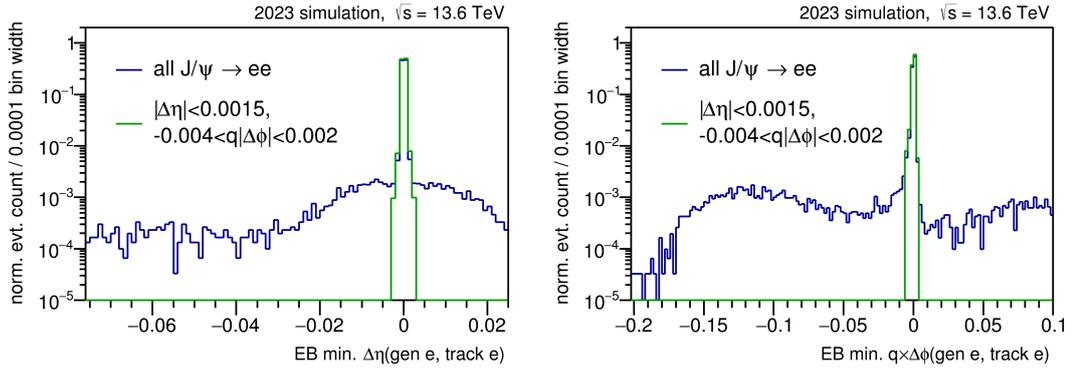


Figure 6.30: For EB: The minimum difference between the generator  $\eta$  and the track  $\eta$  (left) for the scouting electron, generator charge multiplied with difference between the generator  $\phi$  and the track  $\phi$  (right) before and after the chosen generator matching conditions.

electrons bend in the magnetic field by at least one crystal. The ECAL SC azimuth is calculated assuming the production of a high-energy electron from the proton collision vertex. Hence, the azimuthal angle detected by the ECAL SC for low-energy electrons is mis-reconstructed by a factor of one crystal. However, assuming that the  $J/\psi$  is produced almost at the primary vertex, the azimuth of the track is expected in the same direction as the SC. Many backgrounds and fakes are reconstructed with significantly larger azimuthal angular differences with the SC. A selection of  $|\Delta\phi(\text{SC}, \text{track})| < 0.06$  is used on this variable, which roughly corresponds to the azimuth of one ECAL SC.

No large difference is observed in the  $\eta$  difference between the SC and the track, see Fig. 6.36 left (right) for EB (EE). So currently, no selection is used on this variable but there is scope for

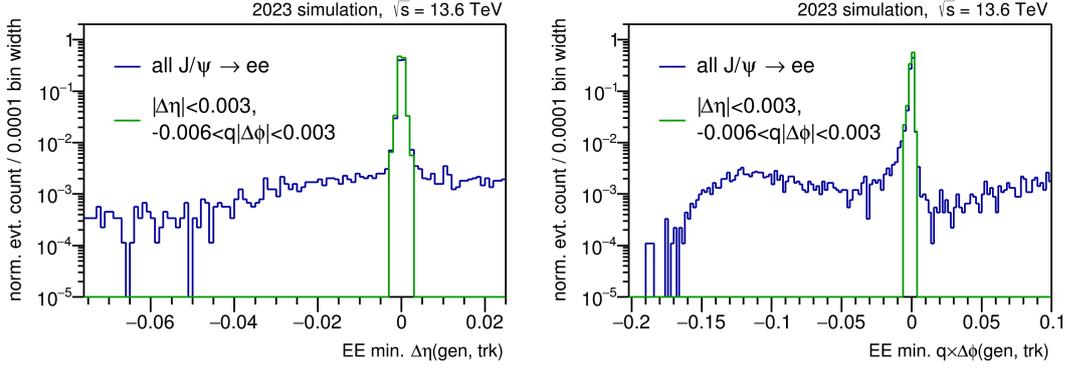
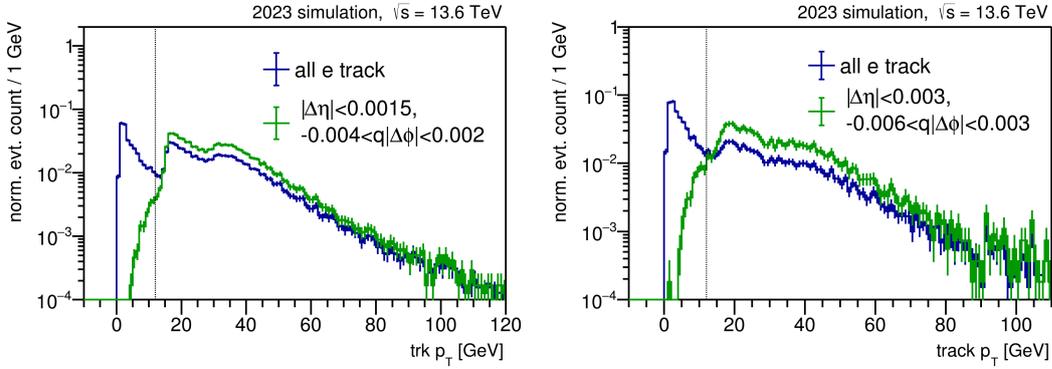


Figure 6.31: Same as Fig. 6.30 for EE.

Figure 6.32: The gen-matched tracks show that track  $p_T$  is efficient after  $p_T > 12$  GeV. The vertical line shows the value where the selection cut is made. The variable needs to have a value greater than this cut. Left - EB. Right - EE.

improvement.

Generally, the variables have only been selected with visual differences in the  $J/\psi \rightarrow ee$  MC. There is scope for further optimisation with background processes which will be done in the next iterations but lie beyond the scope of this thesis due to time constraints. There still exist multiple tracks for each SC after these selections. The track with the lowest  $|\Delta E(\text{SC}, \text{track})/E_{\text{SC}}|$  is associated with the SC while the others are dropped. This results in a unique track for each scouting ECAL SC.

Events with oppositely charged pair of electrons with the identification selections discussed in this section and passing either the single L1 or double L1  $e/\gamma$  scouting paths are considered for the di-electron invariant mass plot. Only the highest energy opposite charged electron pair is considered for the invariant mass plot. The invariant mass plot can be made using two strategies. Fig. 6.37 left shows the invariant mass obtained with the kinematic parameters of the SC. In the second strategy, the  $\eta$  and  $\phi$  are taken from the track while the energy is taken from the ECAL SC. The mass is set to 0.5 MeV. Fig. 6.37 right shows the invariant mass plot with the second strategy. A clear improvement in resolution is observed for the  $J/\psi$  mass peak. The resolution

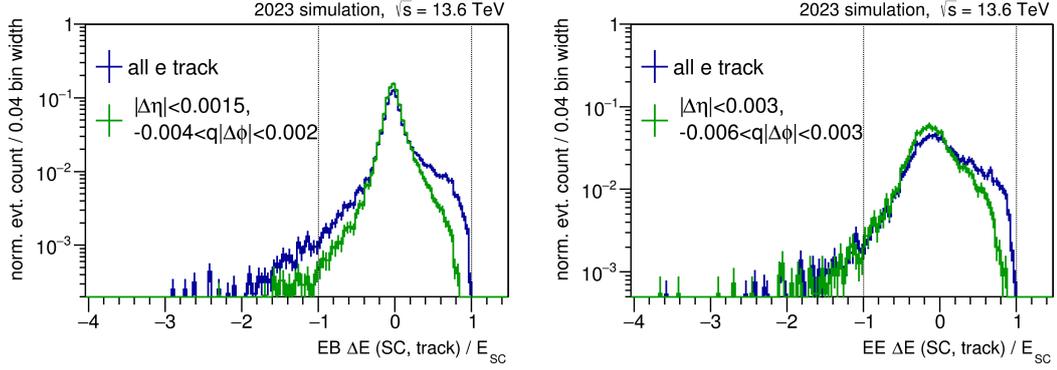


Figure 6.33: The minimum difference between the energy of the track and the SC for each SC is plotted divided by the energy of the SC. A loose selection is placed on this variable. The vertical line shows the value where the selection cut is made. The variable needs to have a value between the left and the right vertical lines. Left - EB. Right - EE.

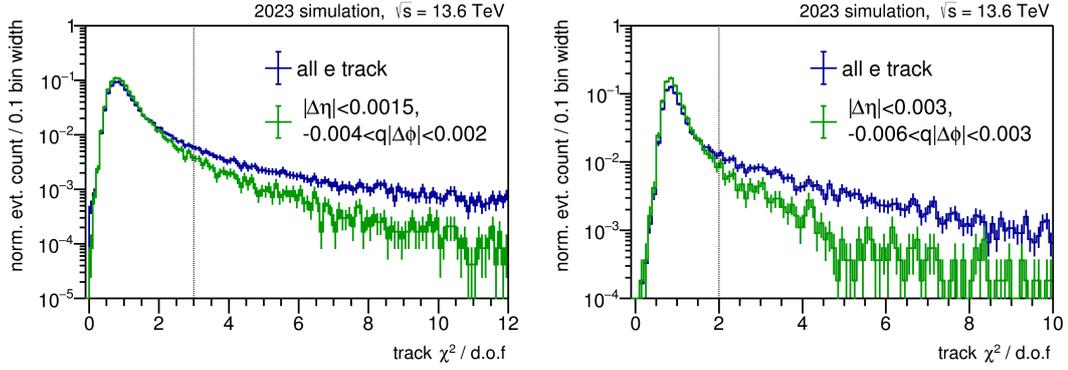


Figure 6.34: The track  $\chi^2/d.o.f$  quantifies the quality of track reconstruction. The vertical line shows the value where the selection cut is made. The variable needs to have a value less than this cut. Left - EB. Right - EE.

improvement came from using the angular parameters of the track and was originally used by me to motivate the CMS collaboration to add the electron track parameters to the scouting electron collection.

The suggestion to include the track parameters was accepted by the CMS collaboration and was included in the menu since beginning of run 3. It was critical improvement needed to obtain the low mass resonances with scouting electrons.

## 6.6 Observing the invariant mass peaks in data

The previous sections' results were all included for data collection in the CMS experiment by the beginning of LHC 2023 runs. Some of the lessons, such as the ones to include the track variables, were learnt during the commissioning runs of 2022. I made the observation of the  $J/\psi \rightarrow ee$  and

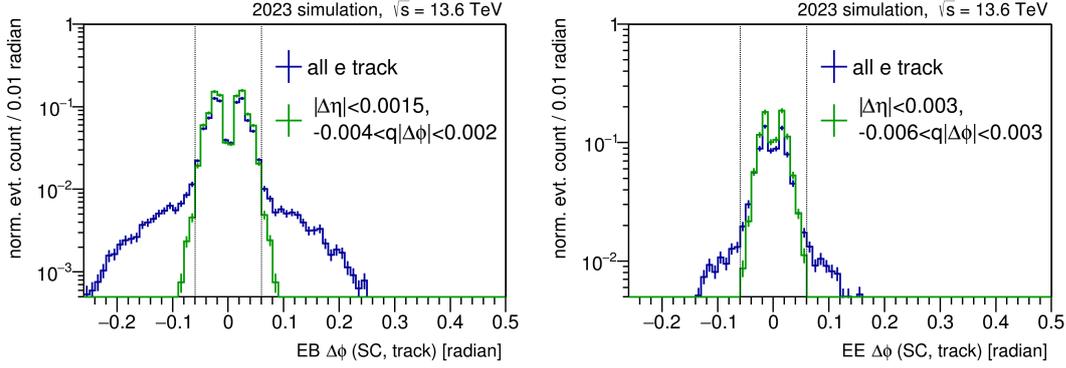


Figure 6.35: The minimum of the absolute azimuthal angular difference for a SC between the SC and an associated track. The vertical line shows the value where the selection cut is made. The variable needs to have a value between the left and the right vertical line. Left - EB. Right - EE.

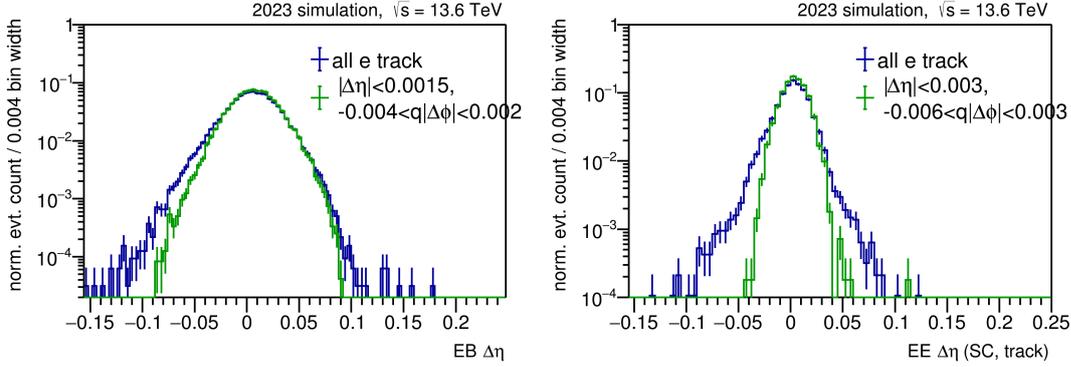


Figure 6.36: The minimum of the absolute pseudorapidity angular difference for a SC between the SC and an associated track. Left - EB. Right - EE.

the  $\Upsilon \rightarrow ee$  with the scouting data collected in 2023.

Table 6.8 and tab. 6.9 list the selections on the electron SCs and tracks in the scouting collection respectively. These were developed with the  $J/\psi$  MC. The track with the lowest  $|\Delta\phi(\text{SC}, \text{track})|$  after identification for tracks was chosen to be associated with the scouting electron. The di-electron invariant mass plot was made with events with at least one oppositely charged electron pair in them. Furthermore, the invariant mass of the highest energy oppositely charged electron pair was used to obtain the di-electron invariant mass plot shown in Fig. 6.38.

The four-momentum of the electrons was chosen using the second strategy using the track angular parameters discussed in the previous section to obtain the Fig. 6.38. The mass peaks for the  $J/\psi$ -meson,  $\Upsilon$ -meson and the  $Z$ -boson are clearly visible. The corresponding invariant mass with the four-momentum of the SC is shown in Fig. 6.39. Two peaks are visible for upsilon peak suggesting a better resolution with the SC parameters compared to the track. This is counter intuitive as normally the track would be expected to have better resolution. However,

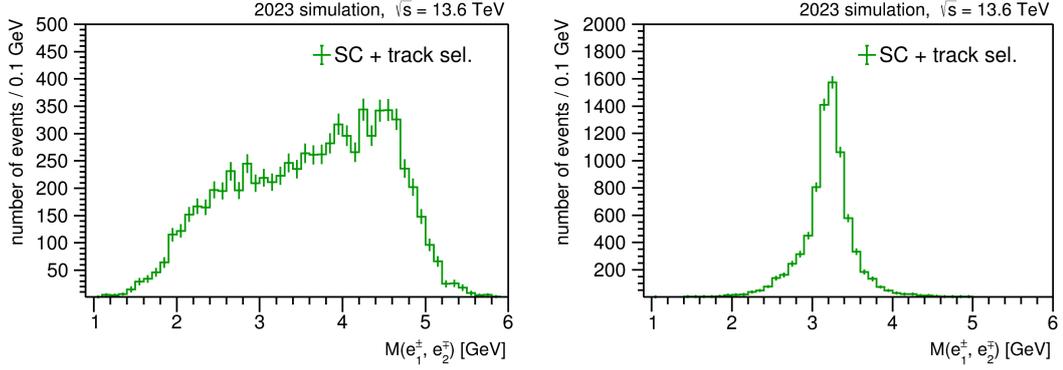


Figure 6.37: The leading and sub-leading oppositely charged di-electron invariant mass plot with, Left - Angular parameters of the ECAL SC, Right - Angular parameters of the selected track and energy from the ECAL SC.

Table 6.8: Selection on ECAL SC for the low-mass SM particles.

Variable	EB	EE
$\sigma_{in}i\eta <$	0.0105	0.035
$H/E <$	0.1	0.1
$I_H/E <$	0.15	0.1
$1/E - 1/p \text{ (GeV}^{-1}\text{)} <$	0.015	0.01
$ \Delta\eta_{in}^{seed}  <$	0.005	0.01
$ \Delta\phi_{in}  <$	0.03	0.05

Table 6.9: Selection on the tracks for the low-mass SM particles.

Variable	EB	EE
track $p_T \text{ (GeV)} >$	12	12
$ \Delta E(\text{SC, track})/E_{\text{SC}}  <$	1	1
$ \Delta\phi(\text{SC, track})  <$	0.06	0.06
$\chi^2/\text{d.o.f} <$	3	2

for high-energy electrons, significant losses from Bremsstrahlung result in a poor track angular resolution. The track reconstruction at the HLT is not able to recover all Bremsstrahlung losses as the SC. This is also evident with the  $Z$ -boson peak. However, the  $J/\psi$ -meson peak is not visible with only SC parameters.

The mass peaks in the di-electron invariant mass plot of the 2023 scouting data were individually fit to analytical functions. The background was estimated with a polynomial function. Each mass peak was analytically assumed to be a convolution of the Breit-Wigner distribution

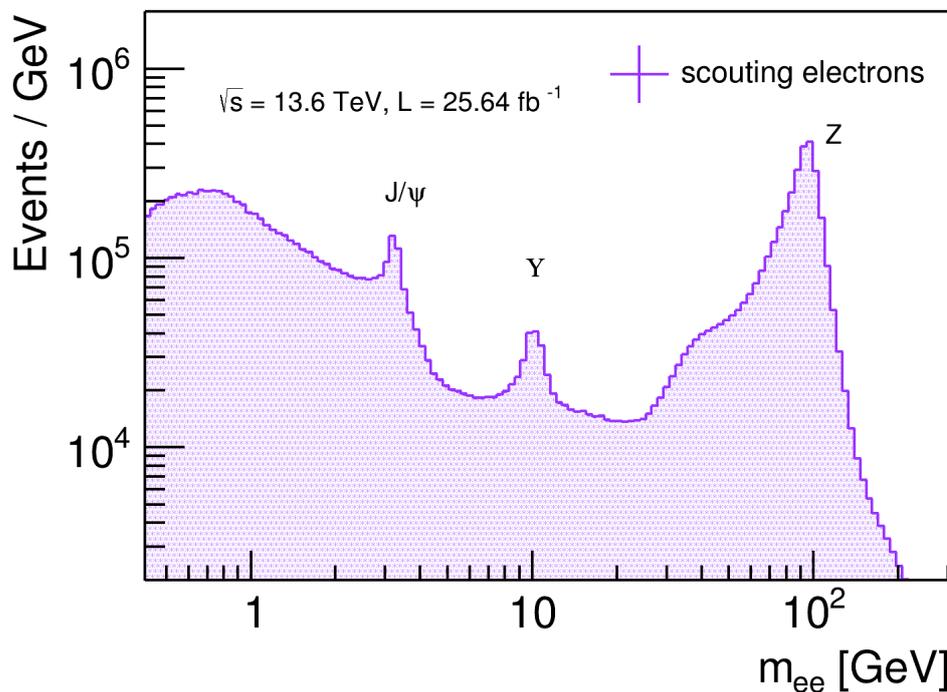


Figure 6.38: Di-electron invariant mass plot from the 2023 scouting data showing the mass peaks for  $J/\psi$ -meson,  $\Upsilon$ -meson and the  $Z$ -boson with  $\eta$  and  $\phi$  from the track and energy from the SC.

with Gaussian resolution effects. The mass peak for the  $\Upsilon$ -meson is asymmetric as it is a superposition of two of the excited states of the meson. Fig. 6.40 shows the fit to the meson mass peaks observed in data.

The mass of the  $J/\psi$  meson as obtained from the fit is  $3.3 \pm 0.2$  GeV. This is in the range of the PDG mass of the  $J/\psi$ -meson [8]. A total of  $44685 \pm 211$   $J/\psi$ -mesons are observed.

The mass of the  $\Upsilon$  excited states 1S and 2S are obtained at  $9.8 \pm 0.4$  GeV and  $10.6 \pm 1.0$  GeV respectively. Although, these values are in the range of the PDG mass of the  $\Upsilon$ -meson [8], an observation can be drawn that there is systematic calibration effect causing a shift towards higher values for the meson mass. This effect remains to be understood. A rudimentary calibration has been applied on the electrons and photons during the online reconstruction and it is likely required to derive a set of calibration for the scouting electrons and photons after the data collection. A total of  $54320 \pm 233$   $\Upsilon$ -mesons are observed.

This was the first-ever observation of low-mass invariant mass peaks with electrons in the scouting data and heralded the possibility of a trigger unbiased search for relatively low-energy physics with CMS electrons and photons. Other paths for low-energy electrons and photons use selections on the kinematic phase-space of the event to reduce the HLT rate for the specific physics they target. Triggers for low mass mesons usually require the  $\Delta R$  to be lower than a certain maximum value.

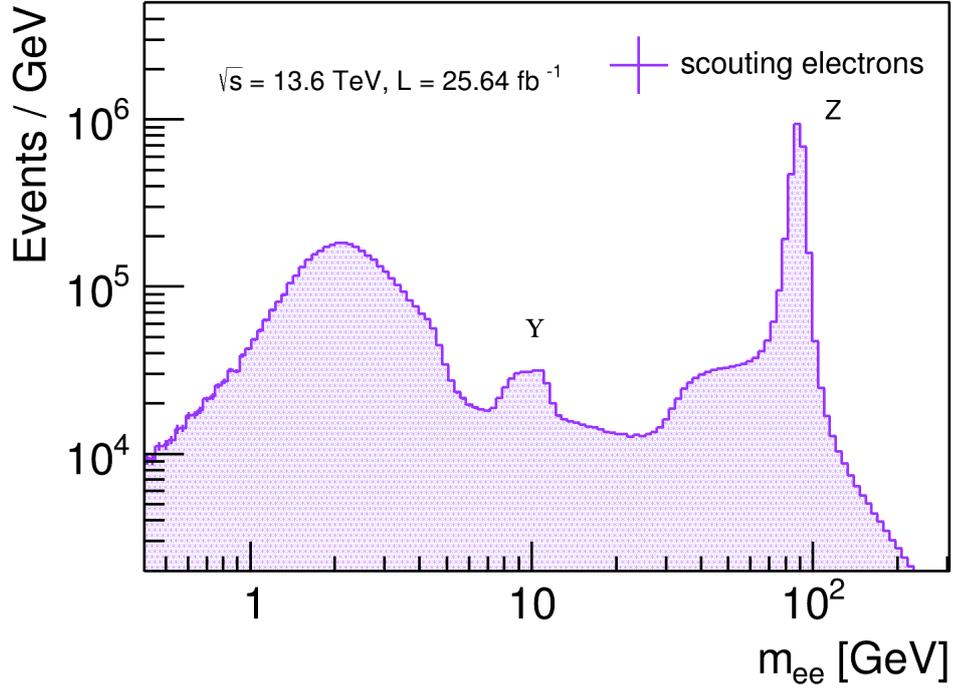


Figure 6.39: Di-electron invariant mass plot from the 2023 scouting data showing the mass peaks for  $J/\psi$ -meson,  $\Upsilon$ -meson and the  $Z$ -boson with the four momentum of the SC.

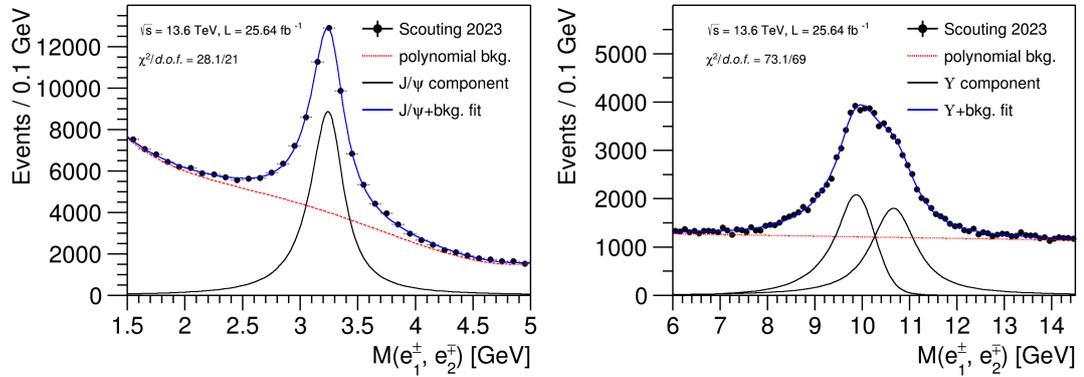


Figure 6.40: Polynomial background with Breit-Wigner convoluted with Gaussian distribution for mass peak fit to the di-electron invariant mass peaks for the  $J/\psi$ -meson (left) and  $\Upsilon$ -meson (right).

## 6.7 Trigger efficiency of scouting $e/\gamma$ paths.

The trigger efficiency was measured with a monitoring dataset. The monitoring dataset was collected with pre-scaled scouting and standard triggers. Unlike the scouting dataset, the mon-

itoring dataset collected the full event data allowing offline reconstruction. This enabled comparing the scouting electrons and photons with their offline analogue. Using the offline electrons and photons is advantageous because the background rejection with the offline algorithm is well established compared to the online reconstruction. The trigger efficiency is not dependent on the chosen reconstruction algorithm so the monitoring data allows us to study the trigger efficiency with offline objects.

To measure the trigger efficiency of scouting electron and photon triggers, an unbiased selection of events was obtained by choosing events that pass the trigger `DST_Run3_JetHT_PFScoutingPixelTracking_v*`. The results were compared to another unbiased selection of events collected with the trigger `DST_HLTmuon_Run3_PFScoutingPixelTracking_v*`. The results were established to be identical with a larger statistics for the events collected by `DST_Run3_JetHT_PFScoutingPixelTracking_v*`. The trigger efficiencies were obtained for barrel with end cap results to be added in future iterations of the study. The trigger used to make the event collection was called the reference trigger.

Offline electrons were selected with the veto and tight identification selection criteria analogous to the identification variable based selection discussed in the Run 2 electron and photon reconstruction in the CMS detector. [187]. Results were found identical within statistical constraints and the results with veto identification were made public owing to the limited statistics with the tight identification. Similarly the loose and tight identification were considered for the photons. To ensure that the  $e/\gamma$  was constructed at the trigger level and resulted in the event selection, a trigger filter object from the last filter in the trigger was required in an angular cone  $\Delta R < 0.06$  with respect to the offline electron. Only electrons with an angular match with the final trigger filter were selected in the event. The offline electrons were required to pass  $|\eta| < 1.44$  and  $p_T > 5$  GeV for barrel constraints and minimal energy requirements.

Events with at least one such electron was required for the target trigger `DST_Run3_EG30_PFScoutingPixelTracking_v*` and at least two such electron were required for the target trigger `DST_Run3_EG16_EG12_PFScoutingPixelTracking_v*`. The trigger efficiency is defined as,

$$\text{trigger efficiency} = \frac{\text{event passing target and reference trigger}}{\text{event passing reference trigger}}. \quad (6.4)$$

The trigger efficiency is computed with the highest energy electron or photon in the event for `DST_Run3_EG30_PFScoutingPixelTracking_v*`. Figure 6.41 shows the trigger efficiency for the single scouting  $e/\gamma$  trigger as a function of the  $e/\gamma$  path. There is a sharp rise in efficiency at 30 GeV which is the L1 threshold for the triggers. The efficiency gradually rises in efficiency until 40 GeV and plateaus after that.

## 6.8 Conclusion

The CMS data scouting strategy targets low-energy detector signatures. For the first time, it contains electrons and photons that can be used for analysis. For this to be possible, a series of changes was implemented. A new L1 triggers were added to the scouting path and scouting reconstruction for electrons and photons was changed to unseeded reconstruction. New strategy was used where the scouting path was split depending on the L1 trigger initiating the scouting reconstruction and identification selections were used in each path to fit into the HLT rate

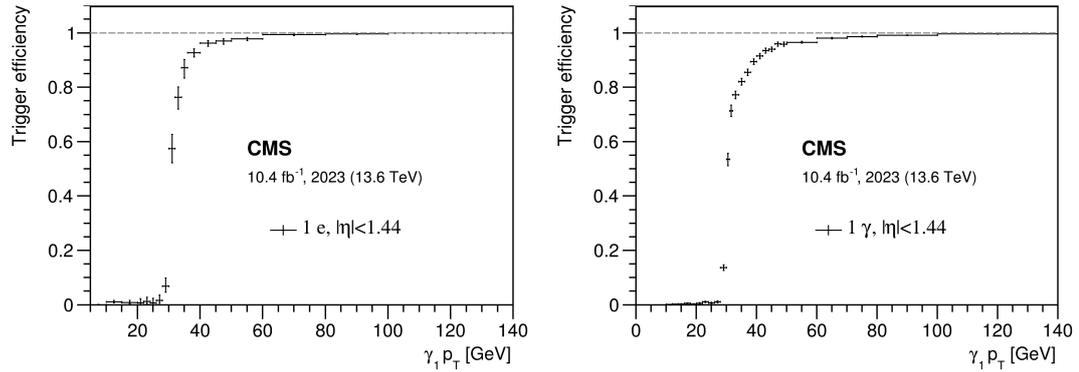


Figure 6.41: Trigger efficiency as a function of the object  $p_T$  in the ECAL barrel for the scouting path with - Left: single electrons, and Right: single photons.

and time constraints. Successfully obtaining the peaks for  $J/\psi$ -meson and  $\Upsilon$ -meson provided a concept of proof that the low-energy objects were of good enough quality to do physics with.

Further development of the electrons can be done by implementing a dedicated reconstruction algorithm for displaced electrons. The current reconstruction requires at least a hit in the pixel detector for an electron. This limits the  $d_0$  reach of the electron. The STHDM model can be used to develop and study the displaced electrons.

# Chapter 7

## Outlook

The new soft displaced lepton triggers enable an inclusive di-lepton search for compressed mass spectrum DM models. Building the triggers and including low-energy electrons and photons in data scouting was the first step in achieving this goal. There is subsequently a significant amount of work to be done to analyse and improve the data quality collected by the new soft displaced lepton triggers and data scouting.

The new triggers require monitoring channels throughout the data collection. The data-taking conditions and the detector behaviour fluctuate over the years the LHC operates. Some aspects can also fluctuate during a single year of data taking. The LHC machine includes a few technical stops during a single run for maintenance and can include minor technical upgrades. The factors in a single run that influence the trigger behaviour, e.g. the pile-up and bunch spacing, can change, which, if improperly monitored, can lead to problems in data collection, delayed trigger decisions and unexpectedly large data output rate.

The CMS experiment uses these technical stops to perform any necessary maintenance and upgrade the detector software to benefit from the lessons learnt during the data taking. The trigger menu is also upgraded to the most recent version to maximise the physics interests of analysts or bring changes to existing triggers for improved performance. During the runs, the detector undergoes operational wear and tear and radiation damage, which could reduce the physics performance of triggers. The changes to the detector and other effects cause changes to the physics efficiency at the trigger in the CMS experiment, necessitating regular monitoring of all the triggers. The monitoring component will run on a small portion of the data made accessible with the slightest possible delay. It will include a comparison with the previous trigger performance. Any significant changes will be observed and evaluated to understand the performance of the soft displaced lepton triggers. The change to trigger performance is often a significant component of the changes to the overall physics scenario targeted by that trigger.

The trigger efficiency measures the trigger performance. The first step in understanding the data collected by the soft displaced leptons would be to measure their trigger efficiency. It is defined as the ratio of events that pass a specific trigger to the number of events that are expected to pass it. The trigger efficiency defines the probability of accepting an event, including both the level-1 and high-level trigger stages. An orthogonal dataset will be used to measure the efficiency of the soft displaced lepton triggers. An orthogonal dataset is collected with an unbiased group of triggers relative to the trigger whose efficiency is being measured. The trigger efficiency will

enable us to understand if the triggers built efficiently detect their respective physics scenario.

The trigger efficiency calculation will also provide a first look at the data collected with the new triggers. An initial comparison with background and signal simulation can provide us with further insights to optimise the triggers for the STHDM model. The ultimate goal of collecting the data with new triggers is to set up a physics analysis of the data to search for STHDM.

While the soft displaced triggers have been built with the STHDM in mind, the selections in the triggers are generic on the physics objects and not on the model-dependent correlations between the physics objects. There would be other physics models that are likely to benefit from the inclusion of these triggers in the CMS physics program. I plan to examine other physics analysis scenarios that benefit from lower energy and displaced electrons, muons and photons.

The inclusion of low-energy electrons and photons in data scouting also provides scope for further low-energy explorations in the CMS experiment. The photon threshold is efficient for physics with single photon  $p_T > 30$  GeV, much lower than the standard trigger single photon threshold at  $p_T > 110$  GeV. The threshold is even lower for events with two photons in the CMS detector with data scouting. A comprehensive study of other physics searches that can benefit from low-energy photons in data scouting would improve the CMS physics program immensely.

The STHDM will further benefit from the inclusion of displaced electrons in data scouting. Displaced electrons in data scouting require dedicated tracking for displaced electrons at the trigger level. It is a largely unexplored aspect of the CMS experiment. A study to include displaced electron tracking would involve measuring physics efficiency improvement for displaced electrons. The STHDM model provides an ideal test candidate for this. The new algorithm will increase the time to process a single event at the trigger level. The increased time will need to be reduced either with new and innovative techniques or by increasing the processing power of the CMS trigger computer farm. It is possible that requiring displaced tracking for electrons will eventually reduce the event rate from the soft displaced leptons. The displaced electron tracking at the high-level trigger can potentially improve the physics reach of the CMS collaboration.

The trigger efficiency for the data scouting with single and double electrons or photons needs to be computed to understand the physics effectiveness of the events selected with data scouting. Since the electrons and photons have been included for the first time in the scouting program of the CMS collaboration, they will need to be studied carefully, e.g. have the energy resolution measured, before they are ready for use in physics analysis. The low-mass mesons detected in the scouting data provide a path to calibrate the low-energy electrons and photons.

A long-term goal is to improve the level-1 trigger to be sensitive to physics with displaced electrons and photons for the upcoming phase-2 upgrade of the CMS experiment.

# Chapter 8

## Conclusion

This thesis explored the potential of the CMS experiment and prepared data collection mechanisms to search for the singlet-triplet higgs portal dark matter (STHDM). Apart from the massive presence in cosmological structures, dark matter (DM) features are largely unknown. Understanding the behaviour of DM paves the way for our understanding of the primordial universe. Model building for DM phenomenology forms a central component of fundamental physics research. Some of the most popular DM models predict a new sector of quantum fields beyond the SM that could be responsible for the DM phenomena. Such fields could be produced at high-energy hadron colliders like the LHC machine.

The singlet-triplet higgs portal model for DM consists of a dark sector of three particles in the mass eigenstate. The charged dark partner,  $\chi^\pm$ , can be pair produced from a neutral electroweak current in the proton collisions. It subsequently decays via the charged electroweak current to give a displaced and oppositely charged di-lepton signature in the CMS detector with a small missing transverse momentum component. The leptons can be of different flavours and are displaced from the proton collision vertex owing to the lifetime of the  $\chi^\pm$  before it decays.

A phenomenological study of the signature of the STHDM in the CMS experiment was done. The kinematics of the STHDM model was compared to the most common background from heavy-flavour jets. A neural network based classification study concluded that lowering the transverse momentum selection on the leptons to 20 GeV will allow us to search for the STHDM over four magnitudes of the mean free path of the  $\chi^\pm$ . The mean free path of the  $\chi^\pm$  is directly correlated to the coupling in the dark sector. Effectively, such a search with the CMS experiment would enable us to probe four orders of magnitude of the coupling constant in the dark sector. The effort in this thesis took advantage of the upcoming run 3 LHC program to improve the acceptance of the data collected by the CMS experiment to be more sensitive to the STHDM model.

New triggers targeting low energy and displaced lepton signature were built for the lepton pairs  $\mu\mu$ ,  $e\mu$  and  $ee$ . The displaced electron signatures relied on the ECAL energy deposit owing to the lack of dedicated tracking for displaced electrons at the trigger level. Hence, the electron final state triggers were also called  $e/\gamma$  since an ECAL-only signature could be identified as a photon. Lowering the  $p_T$  threshold of existing displaced di-lepton triggers resulted in a massive rate increase. This rate was mitigated by additional constraints to the electrons and muon signatures in the trigger. The constraints used the displacement and the isolation of the lepton signature in the detector. Two of these triggers and a control trigger have been included in the

CMS experiment and have been collecting data since the beginning of data taking in 2022.

Furthermore, electrons and photons were added to the scouting data collection method of the CMS experiment. The data scouting uses the high efficiency of the event reconstruction at the trigger level to reduce the data size and store events with lower energy thresholds than the conventional triggers. It was observed that the electrons stored in this path had reduced efficiency to STHDM due to the missing displaced electron reconstruction at the trigger level. However, the prompt electrons enabled the observation of low-mass mesons in the scouting dataset. The observation of low-mass meson in the scouting dataset is the first result with the electrons in the CMS scouting dataset. It enables calibration of the low-energy electrons and photons in it. While the scouting dataset was limited in its usage for the STHDM model, it stores low-energy photons, which could enable interesting physics cases that are otherwise unavailable with the conventional data collection methods.

Lastly, the thesis concludes with all the work that needs to be done to take advantage of the efforts in this thesis. This work paves the way for exciting physics scenarios and DM searches in novel phase space for the CMS experiment.

# Appendix $\mathcal{A}$

## Logarithmic distribution of $d_0$

The  $d_0$  distribution of particles is directly correlated with the lifetime of the parent from whose decay they originate. It is an exponentially decaying distribution where the mean is directly proportional to the mean free path of the parent particle. The  $d_0$  distribution of particles in a detector is a convolution of the exponentially decaying distribution with a Gaussian smearing, which depends on the detector resolution for the particle track. In the lower length limit where the particle mean free path is shorter than the detector resolution, the Gaussian smearing dominates and in the higher length limit where the particle's  $d_0$  is large, the Gaussian smearing can be taken to negligible, the exponential decay behaviour becomes dominant.

The base 10 logarithm ( $\log_{10}$ ) of these extremities do not have a trivial distribution shape. However, it can be solved analytically. The lower and the higher length limit of the  $d_0$  distribution, i.e. the Gaussian distribution with mean zero and finite standard deviation and the exponential distribution, respectively, have been individually considered in the scenarios below.

A random variable  $X$  that follows an exponential distribution with mean  $\lambda$  is given by the

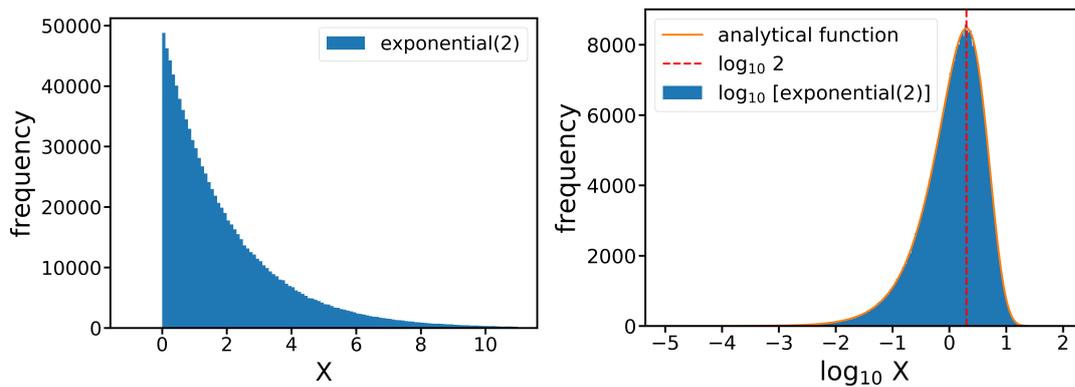


Figure A.1: Left: Randomly generated exponentially decaying distribution with mean 2, shown as the blue histogram. Right: The analytical form (orange line) matches with the randomly generated data (blue histogram) for the base 10 logarithm of the values of the exponential distribution. The peak in the distribution corresponds to the base 10 logarithm of the mean of the exponential distribution.

probability distribution function  $X \sim \lambda e^{-\lambda x}$ . The base 10 logarithm of the random variable  $X$  is given by,

$$\begin{aligned} Y = f(X) &= \log_{10} X \\ &\sim \lambda \ln 10 \, 10^{y-2\log_{10}\lambda} e^{-\lambda 10^{y-2\log_{10}\lambda}}. \end{aligned} \quad (\text{A.1})$$

Figure A.1 left shows one million randomly generated data points from an exponential distribution with a mean of two. Figure A.1 right shows the base 10 logarithm of the generated values in the left figure compared with the shape of the analytical function from Eq. A.1. The two distributions match within statistical fluctuations, verifying the analytical expression. It is important to note that the peak in the logarithm distribution corresponds to the base 10 logarithm of the mean of the exponential distribution,  $\log_{10} 2 \approx 0.301$ , in this case.

A random variable  $X$  that follows Gaussian distribution with mean zero and standard deviation  $\sigma$  is given by the probability distribution function  $X \sim e^{-x^2/2\sigma^2}/\sigma\sqrt{2\pi}$ . The base 10 logarithm of the random variable  $X$  is given by,

$$\begin{aligned} Y = f(X) &= \log_{10} X \\ &\sim 10^y \ln 100 e^{-10^{2y}/2\sigma^2}/(\sigma\sqrt{2\pi}). \end{aligned} \quad (\text{A.2})$$

Figure A.2 left shows one million randomly generated data points from a Gaussian distribution with mean zero and standard deviation  $\sigma$ . Figure A.2 right shows the base 10 logarithm of the generated values in the left figure compared with the shape of the analytical function from Eq. A.2. The two distributions match within statistical fluctuations, verifying the analytical expression. It is important to note that the peak in the logarithm of the Gaussian distribution corresponds to the base 10 logarithm of the standard deviation of the Gaussian distribution,  $\log_{10} 3 \approx 0.477$ , in this case.

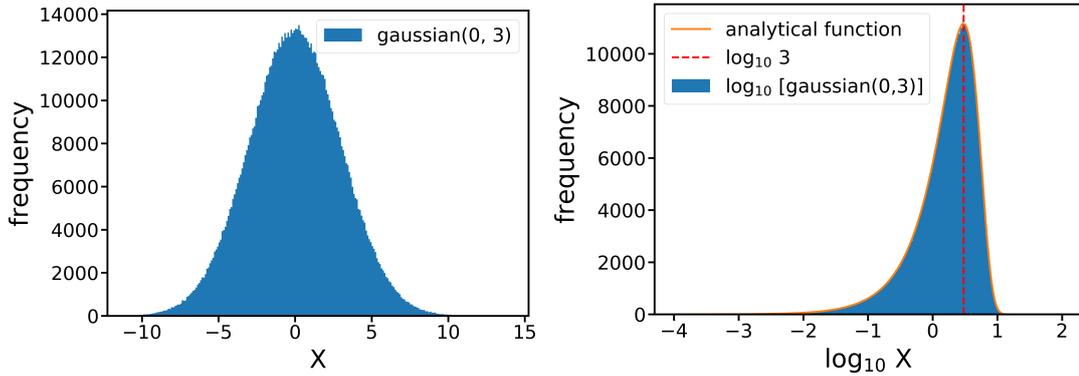


Figure A.2: Left: Randomly generated Gaussian decaying distribution with mean zero and standard deviation 3, shown as the blue histogram. Right: The analytical form (orange line) matches with the randomly generated data (blue histogram) for the base 10 logarithm of the values of the Gaussian distribution. The peak in the distribution corresponds to the base 10 logarithm of the mean of the Gaussian distribution,  $\log_{10} 3 \approx 0.48$ .

# Bibliography

- [1] M. Thomson. *Modern particle physics*. Cambridge University Press, New York, 2013.
- [2] E. Lyndon and B. Philip. LHC machine. *Journal of Instrumentation*, 3(08):S08001, Aug 2008.
- [3] S. Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.
- [4] S. Willenbrock. Symmetries of the Standard Model. In *Theoretical Advanced Study Institute in Elementary Particle Physics: Physics in  $D \geq 4$* , pages 3–38, Oct 2004.
- [5] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak interactions with lepton-hadron symmetry. *Phys. Rev. D*, 2:1285–1292, Oct 1970.
- [6] P. W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys. Lett.*, 12:132–133, 1964.
- [7] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, 1964.
- [8] R. L. Workman and Others. Review of particle physics. *PTEP*, 2022:083C01, 2022.
- [9] S. Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012.
- [10] G. Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1–29, 2012.
- [11] Y. Nagashima. *Elementary particle physics, volume 1: Quantum field theory and particles*. Wiley-VCH, Weinheim, 2010.
- [12] Y. Nagashima. *Elementary particle physics: Foundations of the Standard Model, volume 2*. Wiley-VCH, Weinheim, 2010.
- [13] D. J. Griffiths. *Introduction to electrodynamics*. Cambridge University Press, 4th edition, 2017.
- [14] M. Eibson. The historical development of Quantum Theory volume 1 parts 1 and 2: The Quantum Theory of Planck, Einstein, Bohr and Sommerfeld – Its foundation and the rise of its difficulties 1900–1925. *Physics Bulletin*, 36(6):260, Jun 1985.

- [15] K. Gottfried. P. A. M. Dirac and the discovery of quantum mechanics. *American Journal of Physics*, 79(3):261–266, Mar 2011.
- [16] L. D. Landau. Diamagnetism of metals. *Z. Phys.*, 64, 1930.
- [17] M. Gell-Mann. Symmetries of baryons and mesons. *Phys. Rev.*, 125:1067–1084, Feb 1962.
- [18] M. Gell-Mann. The Eightfold Way: A Theory of strong interaction symmetry. *OSTI Technical Report*, Mar 1961.
- [19] 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, Feb 1957.
- [20] A. Filimonova and S. Westhoff. Long live the Higgs portal! *JHEP*, 02:140, 2019.
- [21] M. H. Poincaré. Sur la dynamique de l'électron. *Rendiconti del Circolo Matematico di Palermo*, 21(1):129–175, Dec 1906.
- [22] E. Noether. Invariante variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1918:235–257, 1918.
- [23] S. K. Lamoreaux. Demonstration of the Casimir force in the 0.6 to 6  $\mu\text{m}$  range. *Phys. Rev. Lett.*, 78:5–8, Jan 1997.
- [24] P. A. M. Dirac. The quantum theory of the electron. *Proc. R. Soc. Lond. A*, 117:610–624, Feb 1928.
- [25] C. N. Yang and R. L. Mills. Conservation of isotopic spin and isotopic gauge invariance. *Phys. Rev.*, 96:191–195, Oct 1954.
- [26] R.P. Feynman and A. Zee. *QED: The Strange Theory of Light and Matter*. Princeton Science Library. Princeton University Press, 2014.
- [27] S. Weinberg. The making of the Standard Model. *Eur. Phys. J. C*, 34:5–13, 2004.
- [28] P. A. M. Dirac. The quantum theory of the emission and absorption of radiation. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 114(767):243–265, 1927.
- [29] R. P. Feynman. Space-time approach to quantum electrodynamics. *Phys. Rev.*, 76:769–789, Sep 1949.
- [30] B. R. Webber. Fragmentation and hadronization. *Int. J. Mod. Phys. A*, 15S1:577–606, 2000.
- [31] G. Arnison et. al. Experimental observation of isolated large transverse energy electrons with associated missing energy at  $\sqrt{s} = 540$  GeV. *Physics Letters B*, 122(1):103–116, 1983.
- [32] P. Bagnaia et. al. Evidence for  $z^0 \rightarrow e^+e^-$  at the CERN pp collider. *Physics Letters B*, 129(1):130–140, 1983.
- [33] The ATLAS, CDF, CMS, and D0 collaborations. First combination of Tevatron and LHC measurements of the top-quark mass. *arXiv*, Mar 2014.
- [34] S. Abachi et al. Observation of the top quark. *Phys. Rev. Lett.*, 74:2632–2637, 1995.

- [35] F. Abe et al. Observation of top quark production in  $\bar{p}p$  collisions. *Phys. Rev. Lett.*, 74:2626–2631, 1995.
- [36] M.C. Gonzalez-Garcia and M. Maltoni. Phenomenology with massive neutrinos. *Physics Reports*, 460(1-3):1–129, Apr 2008.
- [37] P. Fisher, B. Kayser, and K. S. McFarland. Neutrino mass and oscillation. *Annual Review of Nuclear and Particle Science*, 49(1):481–527, 1999.
- [38] P. F. de Salas, D. V. Forero, S. Gariazzo, P. Martínez-Miravé, O. Mena, C. A. Ternes, M. Tórtola, and J. W. F. Valle. 2020 global reassessment of the neutrino oscillation picture. *Journal of High Energy Physics*, 2021(2), Feb 2021.
- [39] K. G. Wilson. Confinement of quarks. *Phys. Rev. D*, 10:2445–2459, Oct 1974.
- [40] B. Bajc, J. Hisano, T. Kuwahara, and Y. Omura. Threshold corrections to dimension-six proton decay operators in non-minimal SUSY SU(5) GUTs. *Nucl. Phys. B*, 910:1–22, 2016.
- [41] N. Cabibbo. Unitary symmetry and leptonic decays. *Phys. Rev. Lett.*, 10:531–533, Jun 1963.
- [42] M. Kobayashi and T. Maskawa. CP-Violation in the Renormalizable Theory of Weak Interaction. *Progress of Theoretical Physics*, 49(2):652–657, Feb 1973.
- [43] H. Georgi and S. L. Glashow. Unity of all elementary-particle forces. *Phys. Rev. Lett.*, 32:438–441, Feb 1974.
- [44] J. C. Pati and A. Salam. Lepton number as the fourth ‘color’. *Phys. Rev. D*, 10:275–289, Jul 1974.
- [45] J. Ellis. Physics gets physical. *Nature*, 415, Feb 2002.
- [46] N. Craig and S. Koren. IR dynamics from UV divergences: UV/IR mixing, NCFT, and the hierarchy problem. *JHEP*, 03:037, 2020.
- [47] Z. Xing. Quark mass hierarchy and flavor mixing puzzles. *Int. J. Mod. Phys. A*, 29:1430067, 2014.
- [48] L. E. Ibanez. The second string (phenomenology) revolution. *Class. Quant. Grav.*, 17:1117–1128, 2000.
- [49] C. Rovelli. *Quantum gravity*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2004.
- [50] R. P. Feynman, F. B. Morinigo, W. G. Wagner, and B. Hatfield. *Feynman lectures on gravitation*. CRC Press Taylor and Francis group, 1996.
- [51] K. Akiyama et al. First M87 Event Horizon Telescope results. I. The shadow of the supermassive black hole. *Astrophys. J. Lett.*, 875:L1, 2019.
- [52] D. Finkelstein. Past-future asymmetry of the gravitational field of a point particle. *Phys. Rev.*, 110:965–967, May 1958.
- [53] A. Albrecht et al. Report of the dark energy task force. *arXiv*, Sep 2006.

- [54] E. Abdalla et al. Cosmology intertwined: A review of the particle physics, astrophysics, and cosmology associated with the cosmological tensions and anomalies. *JHEAp*, 34:49–211, 2022.
- [55] E. Corbelli and P. Salucci. The extended rotation curve and the dark matter halo of M33. *Monthly Notices of the Royal Astronomical Society*, 311(2):441–447, Jan 2000.
- [56] V. C. Rubin, W. K. Ford, Jr., and N. Thonnard. Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa through Sc. *Astrophys. J. Lett.*, 225:L107–L111, 1978.
- [57] K. G. Begeman, A. H. Broeils, and R. H. Sanders. Extended rotation curves of spiral galaxies: dark haloes and modified dynamics. *Monthly Notices of the Royal Astronomical Society*, 249(3):523–537, Apr 1991.
- [58] F. Zwicky. Die rotverschiebung von extragalaktischen nebeln. *Helv. Phys. Acta*, 6:110–127, 1933.
- [59] H. J. Rood, T. L. Page, E. C. Kintner, and E. R. King. The structure of the Coma cluster of galaxies. *Astronomical Journal*, 175:627, Aug 1972.
- [60] M. Viso. *Spitzer Space Telescope*, pages 2325–2329. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015.
- [61] J. P. Gardner et al. The James Webb space telescope. *Space Sci. Rev.*, 123:485, 2006.
- [62] M. C. Weisskopf, H. D. Tananbaum, L. P. van Speybroeck, and S. L. O’Dell. Chandra x-ray observatory (cxo):overview. *Proc. SPIE Int. Soc. Opt. Eng.*, 4012:2, 2000.
- [63] M. Grossmann A. Einstein. Entwurf einer verallgemeinerten relativitätstheorie und einer theorie der gravitation. *Monatshefte für Mathematik und Physik*, 26(A38), 1915.
- [64] A. Einstein. Lens-like action of a star by the deviation of light in the gravitational field. *Science*, 84:506–507, 1936.
- [65] R. Massey et al. Dark matter maps reveal cosmic scaffolding. *Nature*, 445:286, 2007.
- [66] 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.
- [67] A. A. Penzias and R. W. Wilson. A measurement of excess antenna temperature at 4080 Mc/s. *Astrophysical Journal*, 142:419–421, Jul 1965.
- [68] Planck Collaboration. Planck 2018 results. *Astronomy & Astrophysics*, 641:A6, Sep 2020.
- [69] T. M. C. Abbott et al. Dark energy survey year 3 results: Cosmological constraints from galaxy clustering and weak lensing. *Phys. Rev. D*, 105(2):023520, 2022.
- [70] M. Milgrom. A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis. *Astronomical Journal*, 270:365–370, Jul 1983.
- [71] J. D. Bekenstein. Relativistic gravitation theory for the MOND paradigm. *Phys. Rev. D*, 70:083509, 2004. [Erratum: Phys.Rev.D 71, 069901 (2005)].

- [72] M. Lisanti, M. Moschella, N. J. Outmezguine, and O. Slone. Testing dark matter and modifications to gravity using local Milky Way observables. *Phys. Rev. D*, 100(8):083009, 2019.
- [73] J. Zuntz, T. G Zlosnik, F. Bourliot, P. G. Ferreira, and G. D. Starkman. Vector field models of modified gravity and the dark sector. *Phys. Rev. D*, 81:104015, 2010.
- [74] C. Skordis, D. F. Mota, P. G. Ferreira, and C. Boehm. Large scale structure in Bekenstein’s theory of relativistic Modified Newtonian Dynamics. *Phys. Rev. Lett.*, 96:011301, 2006.
- [75] S. Tremaine and J. E. Gunn. Dynamical role of light neutral leptons in cosmology. *Phys. Rev. Lett.*, 42:407–410, 1979.
- [76] A. Boyarsky, O. Ruchayskiy, and D. Iakubovskiy. A lower bound on the mass of dark matter particles. *JCAP*, 03:005, 2009.
- [77] S. Profumo, L. Giani, and O. F. Piattella. An introduction to particle dark matter. *Universe*, 5(10):213, 2019.
- [78] P. Fayet. The Supersymmetric Standard Model. *Adv. Ser. Direct. High Energy Phys.*, 26:397–454, 2016.
- [79] A. A. Petrov and A. E. Blechman. *Effective Field Theories*. WSP, 2016.
- [80] E. Fermi. Versuch einer Theorie der  $\beta$ -Strahlen. I. *Zeitschrift fur Physik*, 88(3-4):161–177, Mar 1934.
- [81] B. Batell, M. Pospelov, and A. Ritz. Exploring portals to a hidden sector through fixed targets. *Phys. Rev. D*, 80:095024, 2009.
- [82] V. A. Mitsou. Overview of searches for dark matter at the LHC. *Journal of Physics: Conference Series*, 651(1):012023, Oct 2015.
- [83] F. Blekman, N. Desai, A. Filimonova, A. R. Sahasransu, and S. Westhoff. Soft displaced leptons at the LHC. *JHEP*, 11:112, 2020.
- [84] N. Arkani-Hamed, A. Delgado, and G. F. Giudice. The well-tempered neutralino. *Nucl. Phys. B*, 741:108–130, 2006.
- [85] S. Choubey, S. Khan, M. Mitra, and S. Mondal. Singlet-triplet fermionic dark matter and LHC phenomenology. *Eur. Phys. J. C*, 78(4):302, 2018.
- [86] A. Bharucha, F. Brümmer, and N. Desai. Next-to-minimal dark matter at the LHC. *JHEP*, 11:195, 2018.
- [87] M. Ibe, S. Matsumoto, and R. Sato. Mass splitting between charged and neutral Winos at two-loop level. *Phys. Lett. B*, 721:252–260, 2013.
- [88] A. Filimonova. *Reviving the thermal dark matter paradigm with long-lived particles*. PhD thesis, U. Heidelberg (main), U. Heidelberg (main), 2020.
- [89] A. Djouadi, Y. Mambrini, and M. Muhlleitner. Chargino and neutralino decays revisited. *Eur. Phys. J. C*, 20:563–584, 2001.
- [90] N. Arkani-Hamed, A. Delgado, and G.F. Giudice. The well-tempered neutralino. *Nucl. Phys. B*, 741:108–130, 2006.

- [91] J. Bramante, N. Desai, P. Fox, A. Martin, B. Ostdiek, and T. Plehn. Towards the final word on neutralino dark matter. *Phys. Rev. D*, 93(6):063525, 2016.
- [92] A. Canepa, T. Han, and X. Wang. The search for electroweakinos. *Annual Review of Nuclear and Particle Science*, Mar 2020.
- [93] V. Khachatryan et al. Search for long-lived particles that decay into final states containing two electrons or two muons in proton-proton collisions at  $\sqrt{s} = 8$  tev. *Phys. Rev. D*, 91(5):052012, 2015.
- [94] The CMS collaboration. Search for displaced leptons in the e-mu channel. *CMS Physics Analysis Summary CMS-PAS-EXO-16-022*, 2016.
- [95] The CMS collaboration. Performance of the CMS level-1 trigger in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *JINST*, 15(10):P10017, 2020.
- [96] G. Aad et al. Performance of the ATLAS muon triggers in Run 2. *JINST*, 15(09):P09015, 2020.
- [97] G. Aad et al. Performance of electron and photon triggers in ATLAS during LHC Run 2. *Eur. Phys. J. C*, 80(1):47, 2020.
- [98] I. Falconer. Corpuscles, electrons and cathode rays: J.J. Thomson and the ‘discovery of the electron’. *The British Journal for the History of Science*, 20(3):241–276, 1987.
- [99] J. Chadwick. Possible existence of a neutron. *Nature*, 129:312, 1932.
- [100] E. Rutherford. Collision of  $\alpha$  particles with light atoms. IV. an anomalous effect in nitrogen. *Phil. Mag. Ser. 6*, 37:581–587, 1919.
- [101] V. I. Veksler. A new method of acceleration of relativistic particles. *J. Phys.*, 9:153–158, 1945.
- [102] N. N. D. Gupta and S. K. Ghosh. A report on the Wilson cloud chamber and its applications in physics. *Rev. Mod. Phys.*, 18:225–290, Apr 1946.
- [103] S. Abachi et al. The D0 detector. *Nucl. Instrum. Meth. A*, 338:185–253, 1994.
- [104] P. A. Aarnio et al. The DELPHI detector at LEP. *Nucl. Instrum. Meth. A*, 303:233–276, 1991.
- [105] V. Khachatryan et al. Charged Particle Multiplicities in  $pp$  Interactions at  $\sqrt{s} = 0.9, 2.36,$  and 7 TeV. *JHEP*, 01:079, 2011.
- [106] Chatrchyan et. al. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008. Also published by CERN Geneva in 2010.
- [107] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock. *LHC design report: Volume 1*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004.
- [108] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock. *LHC design report: Volume 2*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004.
- [109] M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl. *LHC design report: Volume 3*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004.

- [110] E. Lopienska. The CERN accelerator complex, layout in 2022, 2022. General Photo.
- [111] Aad et. al. The ATLAS experiment at the CERN large hadron collider. *JINST*, 3:S08003, 2008. Also published by CERN Geneva in 2010.
- [112] K. Aamodt et al. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.
- [113] Alves et. al. The LHCb detector at the LHC. *JINST*, 3:S08005, 2008. Also published by CERN Geneva in 2010.
- [114] O. Adriani et al. The LHCf detector at the CERN large hadron collider. *JINST*, 3:S08006, 2008.
- [115] S. Alpigiani et al. Recent progress and next steps for the MATHUSLA LLP detector. In *2022 Snowmass Summer Study*, Mar 2022.
- [116] J. H. Yoo. The milliQan experiment: Search for milli-charged particles at the LHC. *PoS, ICHEP 2018:520*, 2019.
- [117] J. L. Pinfold. The MoEDAL experiment at the LHC - A new light on the high energy frontier. *PoS, EPS-HEP2017:327*, 2017.
- [118] G. Anelli et al. The TOTEM experiment at the CERN large hadron collider. *JINST*, 3:S08007, 2008.
- [119] H. Abreu et al. The FASER detector. *arXiv*, Jul 2022.
- [120] E. Engin. *Determination of the Parton Distribution Functions and Strong Coupling by Using Measurement of Jet Cross Sections at  $\sqrt{s} = 13$  TeV with the CMS detector with a combined QCD Analysis*. PhD thesis, Hamburg U., Hamburg, 2018.
- [121] J. M. Campbell, J. W. Huston, and W. J. Stirling. Hard interactions of quarks and gluons: A primer for LHC physics. *Rept. Prog. Phys.*, 70:89, 2007.
- [122] The CMS collaboration. 2023 proton-proton collisions at 13.6 TeV, 2023. Oct 2023.
- [123] D. Barney and S. Cittolin. CMS detector drawings, 2000. Jan 2000.
- [124] V. Karimäki, M. Mannelli, P. Siegrist, H. Breuker, A. Caner, R. Castaldi, K. Freudenreich, G. Hall, R. Horisberger, M. Huhtinen, and A. Cattai. *The CMS tracker system project: Technical design report*. Technical design report. CMS. CERN, Geneva, 1997.
- [125] The CMS collaboration. *The CMS tracker: addendum to the technical design report*. Technical design report. CMS. CERN, Geneva, 2000.
- [126] W. Adam et al. The CMS phase-1 pixel detector upgrade. *JINST*, 16(02):P02027, 2021.
- [127] G. F. Knoll. *Radiation detection and measurement*. John Wiley and Sons, New York, 3rd edition, 2000.
- [128] V. Benedikt. The CMS inner tracker: Transition from LHC Run I to Run II and first experience of Run II. In *European Physical Society Conference on High Energy Physics 2015*, 2016.
- [129] CMS collaboration. Tracker material budget plots, 2020. Material Budget for Tracker Phase 1 on 21.07.2020.

- [130] The CMS collaboration. *The CMS electromagnetic calorimeter project: Technical design report*. Technical design report. CMS. CERN, Geneva, 1997.
- [131] P. Bloch, R. Brown, P. Lecoq, and H. Rykaczewski. *Changes to CMS ECAL electronics: addendum to the technical design report*. Technical design report. CMS. CERN, Geneva, 2002.
- [132] R. Wigmans. *Calorimetry: Energy measurement in particle physics*, volume 107. Oxford University Press, 2000.
- [133] B. Marzocchi. Simulation of the CMS electromagnetic calorimeter response at the energy and intensity frontier. *Journal of Physics: Conference Series*, 1162:012007, Jan 2019.
- [134] P. Adzic et al. Energy resolution of the barrel of the CMS electromagnetic calorimeter. *JINST*, 2:P04004, 2007.
- [135] The CMS collaboration. *The CMS hadron calorimeter project: Technical design report*. Technical design report. CMS. CERN, Geneva, 1997.
- [136] The CMS collaboration. Calibration of the CMS hadron calorimeters using proton-proton collision data at  $\sqrt{s} = 13$  TeV. *JINST*, 15(05):P05002, 2020.
- [137] A. M. Sirunyan et al. Performance of missing transverse momentum reconstruction in proton-proton collisions at  $\sqrt{s} = 13$  TeV using the CMS detector. *JINST*, 14(07):P07004, 2019.
- [138] G. Acquistapace et al. CMS, the magnet project: Technical design report. *CERN Document Server*, May 1997.
- [139] The CMS collaboration. Muon reconstruction and identification performance with Run-2 data. *CERN Document Server*, 2020.
- [140] M. A. Akl et al. CMS Technical Design Report for the Muon Endcap GEM Upgrade. *CERN Document Server*, Jun 2015.
- [141] J. G. Layter. *The CMS muon project: Technical design report*. Technical design report. CMS. CERN, Geneva, 1997.
- [142] O. Prokofev et al. Aging tests of full scale CMS muon cathode strip chambers. *Nucl. Instrum. Meth. A*, 515:226–233, 2003.
- [143] F. Loddo et al. An RPC-based technical trigger for the CMS experiment. In *12th Workshop on Electronics for LHC and Future Experiments (LECC 2006)*, pages 284–288, Sep 2006.
- [144] F. Sauli. The gas electron multiplier (GEM): Operating principles and applications. *Nucl. Instrum. Meth. A*, 805:2–24, 2016.
- [145] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth. A*, 389:81–86, 1997.
- [146] A. Afaq et al. The CMS high level trigger system. *Nuclear Science, IEEE Transactions on*, 55:172 – 175, Mar 2008.
- [147] S. Dasu et al. The TriDAS project. Technical design report, vol. 1: The trigger systems. *CERN Document Server*, Dec 2000.

- [148] P. Sphicas et al. The TriDAS project. Technical design report, vol. 2: Data acquisition and high-level trigger. *CERN Document Server*, Dec 2002.
- [149] A. Tapper, D. Acosta, et al. CMS technical design report for the level-1 trigger upgrade. *CERN Document Server*, Jun 2013.
- [150] W. Adam et al. The CMS high level trigger. *Eur. Phys. J. C*, 46:605–667, 2006.
- [151] I. Z. Fernandez et al. High-Luminosity Large Hadron Collider (HL-LHC): Technical design report. *CERN Document Server*, 10/2020, Dec 2020.
- [152] D. Anderson. Data scouting in CMS. *PoS, ICHEP 2016:190*, 2016.
- [153] The CMS collaboration. Data Parking and Data Scouting at the CMS Experiment. *CERN Document Server*, Sep 2012.
- [154] R. P. Walker. Synchrotron radiation. *CERN Document Server*, 1994.
- [155] F. James. Monte Carlo theory and practice. *Rept. Prog. Phys.*, 43:1145, 1980.
- [156] B. R. Webber. Monte Carlo simulation of hard hadronic processes. *Ann. Rev. Nucl. Part. Sci.*, 36:253–286, 1986.
- [157] A. Buckley et al. General-purpose event generators for LHC physics. *Phys. Rept.*, 504:145–233, 2011.
- [158] M. A. Dobbs et al. Les Houches guidebook to Monte Carlo generators for hadron collider physics. In *3rd Les Houches Workshop on Physics at TeV Colliders*, pages 411–459, Mar 2004.
- [159] V. N. Gribov and L. N. Lipatov. Deep inelastic e p scattering in perturbation theory. *Sov. J. Nucl. Phys.*, 15:438–450, 1972.
- [160] G. Altarelli and G. Parisi. Asymptotic freedom in parton language. *Nucl. Phys. B*, 126:298–318, 1977.
- [161] R. D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015.
- [162] R. D. Ball et al. Parton distributions from high-precision collider data. *The European Physical Journal C*, 77(10), Oct 2017.
- [163] S. Dulat, T. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan. New parton distribution functions from a global analysis of quantum chromodynamics. *Phys. Rev. D*, 93(3):033006, 2016.
- [164] J. Alwall. 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.
- [165] R. Frederix, S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro. The automation of next-to-leading order electroweak calculations. *JHEP*, 07:185, 2018. [Erratum: *JHEP* 11, 085 (2021)].
- [166] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr, and B. Fuks. FeynRules 2.0 - A complete toolbox for tree-level phenomenology. *Comput. Phys. Commun.*, 185:2250–2300, 2014.

- [167] C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer, and T. Reiter. UFO - The Universal FeynRules Output. *Comput. Phys. Commun.*, 183:1201–1214, 2012.
- [168] T. Sjostrand, S. Mrenna, and P. Z. Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006.
- [169] J. L. Kneur and A. Neveu.  $\Lambda_{\overline{\text{MS}}}^{\text{QCD}}$  from renormalization group optimized perturbation. *Phys. Rev. D*, 85:014005, 2012.
- [170] I. Georghescu. Simulating parton showers. *Nature Rev. Phys.*, 3(2):73, 2021.
- [171] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton fragmentation and string dynamics. *Phys. Rept.*, 97:31–145, 1983.
- [172] B. Andersson. *The Lund model*, volume 7. Cambridge University Press, Jul 2005.
- [173] R. D. Field and S. Wolfram. A QCD model for e+ e- annihilation. *Nucl. Phys. B*, 213:65–84, 1983.
- [174] V. Khachatryan et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J. C*, 76(3):155, 2016.
- [175] J. Allison et al. GEANT4 developments and applications. *IEEE Trans. Nucl. Sci.*, 53:270, 2006.
- [176] J. Allison et al. Recent developments in GEANT4. *Nucl. Instrum. Meth. A*, 835:186–225, 2016.
- [177] S. Agostinelli et al. GEANT4—a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.
- [178] S. Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014.
- [179] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, Mar 1960.
- [180] The CMS collaboration. Commissioning CMS online reconstruction with GPUs. *CERN Document Server*, 2022.
- [181] T. Speer, K. Prokofiev, R. Frühwirth, W. Waltenberger, and P. Vanlaer. Vertex fitting in the CMS tracker. Technical report, CERN, Geneva, 2006.
- [182] K. Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of the IEEE*, 86(11):2210–2239, 1998.
- [183] R. Frühwirth, W. Waltenberger, and P. Vanlaer. Adaptive Vertex Fitting. Technical report, CERN, Geneva, 2007.
- [184] The CMS Collaboration. A candidate event featuring a Higgs decaying into two  $\tau$  leptons. CMS Collection, 2022.
- [185] A. M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017.

- [186] S. Chatrchyan et al. Time reconstruction and performance of the CMS electromagnetic calorimeter. *JINST*, 5:T03011, 2010.
- [187] Sirunyan et. al. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *Journal of Instrumentation*, 16(05):P05014, May 2021.
- [188] A. M. Sirunyan et al. Reconstruction of signal amplitudes in the CMS electromagnetic calorimeter in the presence of overlapping proton-proton interactions. *JINST*, 15(10):P10002, 2020.
- [189] The CMS collaboration. ECAL superclustering with Machine Learning. *CERN Document Server*, 2021.
- [190] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov. Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC. *Journal of Physics G: Nuclear and Particle Physics*, 31(9):N9–N20, Jul 2005.
- [191] A. M. Sirunyan et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at  $\sqrt{s} = 13$  TeV. *JINST*, 13(06):P06015, 2018.
- [192] A. M. Sirunyan et al. Performance of the CMS muon trigger system in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *JINST*, 16:P07001, 2021.
- [193] M. Cacciari, G. P. Salam, and G. Soyez. The anti- $k_t$  jet clustering algorithm. *JHEP*, 04:063, 2008.
- [194] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet user manual. *Eur. Phys. J. C*, 72:1896, 2012.
- [195] A. M. Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in  $pp$  collisions at 13 TeV. *JINST*, 13(05):P05011, 2018.
- [196] A. M. Sirunyan et al. Performance of missing transverse momentum reconstruction in proton-proton collisions at  $\sqrt{s} = 13$  TeV using the CMS detector. *JINST*, 14(07):P07004, 2019.
- [197] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. DELPHES 3, A modular framework for fast simulation of a generic collider experiment. *JHEP*, 02:057, 2014.
- [198] E. Alpaydin. *Introduction to Machine Learning, fourth edition*. Adaptive Computation and Machine Learning series. MIT Press, 2020.
- [199] E. Bols, J. Kieseler, M. Verzetti, M. Stoye, and A. Stakia. Jet flavour classification using DeepJet. *JINST*, 15(12):P12012, 2020.
- [200] A. Tumasyan et al. Identification of hadronic tau lepton decays using a deep neural network. *JINST*, 17:P07023, 2022.
- [201] A. Tumasyan et al. A new calibration method for charm jet identification validated with proton-proton collision events at  $\sqrt{s} = 13$  TeV. *JINST*, 17(03):P03014, 2022.
- [202] R. D. Sipio, M. F. Giannelli, S. K. Haghighat, and S. Palazzo. DijetGAN: A Generative-Adversarial Network approach for the simulation of QCD dijet events at the LHC. *JHEP*, 08:110, 2019.

- [203] J. A. Martínez, Q. Thong Nguyen, M. Pierini, M. Spiropulu, and J. R. Vlimant. Particle Generative Adversarial Networks for full-event simulation at the LHC and their application to pileup description. *J. Phys. Conf. Ser.*, 1525(1):012081, 2020.
- [204] The CMS collaboration. A deep neural network-based tagger to search for new long-lived particle states decaying to jets. *CERN Document Server*, 2019.
- [205] J. Duarte et al. Fast inference of deep neural networks in FPGAs for particle physics. *JINST*, 13(07):P07027, 2018.
- [206] W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5:115–133, 1943.
- [207] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.
- [208] V. Nair and G. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of ICML*, volume 27, pages 807–814, Jun 2010.
- [209] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors, 2012.
- [210] M. Aaboud et al. Search for electroweak production of supersymmetric states in scenarios with compressed mass spectra at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *Phys. Rev. D*, 97(5):052010, 2018.
- [211] A. M. Sirunyan et al. Search for new physics in events with two soft oppositely charged leptons and missing transverse momentum in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *Phys. Lett. B*, 782:440–467, 2018.
- [212] M. Aaboud et al. Search for long-lived charginos based on a disappearing-track signature in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *JHEP*, 06:022, 2018.
- [213] A. M. Sirunyan et al. Search for disappearing tracks as a signature of new long-lived particles in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *JHEP*, 08:016, 2018.
- [214] A. M. Sirunyan et al. Search for disappearing tracks in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *Phys. Lett. B*, 806:135502, 2020.
- [215] G. Aad et al. Search for heavy neutral leptons in decays of  $W$  bosons produced in 13 TeV  $pp$  collisions using prompt and displaced signatures with the ATLAS detector. *JHEP*, 10:265, 2019.
- [216] G. Aad et al. Search for displaced vertices of oppositely charged leptons from decays of long-lived particles in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *Phys. Lett. B*, 801:135114, 2020.
- [217] V. Khachatryan et al. Search for displaced supersymmetry in events with an electron and a muon with large impact parameters. *Phys. Rev. Lett.*, 114(6):061801, 2015.
- [218] A. Tumasyan et al. Search for long-lived particles decaying to leptons with large impact parameter in proton–proton collisions at  $\sqrt{s} = 13$  TeV. *Eur. Phys. J. C*, 82(2):153, 2022.
- [219] E. Ma and D. Suematsu. Fermion triplet dark matter and radiative neutrino mass. *Mod. Phys. Lett. A*, 24:583–589, 2009.

- [220] B. Fuks, M. Klasen, D. R. Lamprea, and M. Rothering. Gaugino production in proton-proton collisions at a center-of-mass energy of 8 TeV. *JHEP*, 10:081, 2012.
- [221] B. Fuks, M. Klasen, D. R. Lamprea, and M. Rothering. Precision predictions for electroweak superpartner production at hadron colliders with RESUMMINO. *Eur. Phys. J. C*, 73:2480, 2013.
- [222] L. Randall and D. Tucker-Smith. Dijet searches for supersymmetry at the large hadron collider. *Phys. Rev. Lett.*, 101:221803, Nov 2008.
- [223] E. Farhi. A QCD test for jets. *Phys. Rev. Lett.*, 39:1587–1588, 1977.
- [224] A. Banfi, G. P. Salam, and G. Zanderighi. Phenomenology of event shapes at hadron colliders. *JHEP*, 06:038, 2010.
- [225] F. Pedregosa et al. Scikit-learn: Machine learning in Python. *J. Machine Learning Res.*, 12:2825–2830, 2011.
- [226] M. Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [227] A. F. Agarap. Deep learning using rectified linear units (ReLU). *CoRR*, page NA, 2018.
- [228] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014.
- [229] The CMS collaboration. Data parking and data scouting at the CMS experiment. Sep 2012.