UNIVERSITY OF CALIFORNIA SAN DIEGO

A search for new physics in events containing a Z boson, jets and missing transverse energy at a center of mass energy of 13 TeV with the CMS detector, and a new parallelizable algorithm for charged particle track reconstruction in the outer tracker of the CMS Detector in the HL-LHC

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

in

Physics

by

Balaji Venkat Sathia Narayanan

Committee in charge:

Professor Avraham Yagil, Chair
Professor Rommie E. Amaro
Professor Michael M. Fogler
Professor Frank K. Würthwein

2023

The Dissertation of Balaji Venkat Sathia Narayanan is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

எண்ணென்ப ஏனை எழுத்தென்ப இவ்விரண்டும்
கண்ணென்ப வாழும் உயிர்க்கு

திருக்குறள் 392

eṇṇeṉpa ēṉai eḻutteṉpa ivviraṇṭum
kaṇṇeṉpa vāḻum uyirkku

Thirukkural 392

The twain that lore of numbers and of letters give
Are eyes, the wise declare, to all on earth that live.

विद्या ददाति विनयं विनयाद् याति पात्रताम् ।
पात्रत्वात् धनम् आप्नोति धनाद् धर्मं ततः सुखम् ॥

हितोपदशः 6

vidyā dadāti vinayaṃ vinayād yāti pātratām |
pātratvāt dhanam āpnoti dhanād dharmaṃ tataḥ sukham ||

Hitopadeśa 6

Knowledge produces humility, humility produces worthiness. From worthiness riches are acquired, and from riches righteousness and hence happiness.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

At last, I've finally come around to writing the acknowledgments. One of the many ways I used to procrastinate in grad school (the one skill grad students learn without any help) is to read the acknowledgment, dedication and epigraph sections of the theses of previous grad students. As I write this at the culmination of my PhD years, I reflect at all those who have played a significant role in helping me cross the finish line. I can confidently say that in this journey no (hu)man is an island. It takes a "village" to raise a (an academic) child, and it is these human connections that matter the most in the PhD marathon.

Firstly, I am very grateful to my parents - Sathianarayanan S and Vidhya Sathianarayanan, for putting the efforts to educating me, ensuring that I get all possible opportunities, and always having my back. They have done the very best they can and ensured their son is at a better place than themselves.

Secondly, I am grateful to Prof. Avi Yagil, who took me as his graduate student in the Spring of 2018. Avi has been a wonderful advisor, with extreme amounts of patience and enthusiasm to train me into a well-rounded researcher. I am also very grateful to Prof. Frank Würthwein for his initial guidance in this journey. In addition, I am thankful to Prof. Rommie Amaro and Prof. Michael Fogler, who graciously accepted my request to serve in my dissertation committee.

Thirdly, I would like to thank our CERN postdocs Giovanni Zevi Della Porta (Gio) and Mario Masciovecchio with whom I worked closely for the supersymmetry analysis. I am very grateful to them hand-holding me through the ins and outs of doing a CMS analysis end to end, which in addition to scientific work, also involves interpersonal interactions with our collaborators and other members of the CMS community at various levels. They were always more than happy to help me mitigate any difficulties and firefight any situations that arose during this (almost) year and half long process from the initial discussion to the published paper. On this note, I am also grateful to our collaborators Marius Terörde, Sergio Sanchez Cruz and Pablo Martinez Ruiz del Arbol. The collaboration we had established turned out to be a fruitful one and we were able to quickly translate ideas to results and publish in a timely manner.

Fourthly, I would like to thank the Line Segment Tracking project group - Philip Chang, Yanxi Gu, Slava Krutelyov, Gavin Niendorf, Tres Reid, Matevž Tadel, Manos Vourliotis, Bei Wang and Prof. Peter Wittich. This project was started in 2019 and has come a long way from a blinking cursor in an editor window to initial results that are quite competitive with existing algorithms. As the years went on the collaboration grew from a small nucleus in UCSD to a multi-institutional group with its own funding, and our collective efforts have resulted in something greater than the sum of our individual contributions. Given that this project will continue till the commissioning of the HL-LHC, I hope to see much more improvements in the future and will be very thankful (and proud) if this becomes a part of the CMS software backend (CMSSW) in the HL-LHC.

Fifthly, I am very thankful to the tri-institutional "Surf-and-Turf" group of which Profs. Yagil and Würthwein (and hence us) are a part of - Dylan Gilbert and Samuel May (UCSD) who helped me in the initial stages of my journey, Bobak Hashemi (UCSD) who did the previous version of the analysis, and the GOAT Nick Amin (UCSB) who handled practically everything computing related for the group and was always helpful with every question - be it physics or computing, in addition to simply being an all round awesome person.

Sixthly, I am very grateful to the people who played important roles in my high school and undergraduate years. My initial inspiration to take up physics was motivated by my high school teacher Mr. Madhusudan who made the subject very interesting and enjoyable. This interest continued in the undergraduate program at the Indian Institute of Science (IISc) - Prof. Chandan Dasgupta, Prof. Rohini Godbole, Prof. Biman Nath, Prof. Diptiman Sen and Prof. V. Venkataraman (RIP), who were excellent instructors. My undergraduate summer in 2015 in the Giant Meterwave Radio Telescope with Prof. Yashwant Gupta was the first time I was exposed to research for an extended period of time. This was followed by another wonderful summer in Germany with Prof. Christoph Pfrommer at Heidelberg and the ensuing collaboration with Prof. Prateek Sharma back in IISc that resulted in my undergraduate thesis project in computational astrophysics. I am forever indebted to these people to prepare me for graduate school and a research journey.

There exists a life beyond work - coming to this realization is quite important in the PhD journey. I am very very thankful to the people who played an important role in the social non-academic part of these six years : Jyotsna Gidugu, Tapan Goel, Aravind Rao Karanam, Kaivalya Molugu, Nikhil Naik, Avaneesh Narla, Sindhana PS (classmates since 9th grade, through undergrad and finally here in UCSD), Nandagopal Ramachandran, Sabareesh Ramachandran, Varun Ramaprasad and Shubham Sinha.

Chapter 3 describes the Large Hadron Collider and the Compact Muon Solenoid Experiment. The figures used in this chapter are obtained from the following results : "Facts and figures about the LHC"
https://home.cern/resources/faqs/facts-and-figures-about-lhc, "CMS Luminosity–Public Results" *https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults*,
"Cutaway diagrams of CMS Detector" *https://cds.cern.ch/record/2665537*, "CMS DPG - Public Results" *https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK*,
"CMS Detector Slice" *http://cds.cern.ch/record/2120661*

Chapter 4 describes the basic principles involved with physics analyses with CMS. The figures used in this chapter are obtained from the following sources : "Event displays of top quark pair production in CMS collected in 2015" *https://cds.cern.ch/record/2712881* (2020)

Chapter 5 describes a search for supersymmetry in events containing a Z boson, jets and missing transverse energy in data collected during Run-II of the LHC. It is a partial reproduction of the paper "Search for supersymmetry in final states with two oppositely charged same-flavor leptons and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV", J. High Energ. Phys. 2021, 123 (2021).

Chapter 6 describes the Line Segment Tracking algorithm. It is a reproduction of the paper "Line Segment Tracking in the HL-LHC", arXiv:2207.08207

VITA

2013–2017 Bachelor of Science (Research), Indian Institute of Science, Bengaluru, India

2017–2018 Teaching Assistant, University of California San Diego

2018–2023 Graduate Research Assistant, University of California San Diego

2017–2023 Doctor of Philosophy in Physics (Expected), University of California San Diego

PUBLICATIONS

- The CMS collaboration., Sirunyan, A.M., Tumasyan, A. et al. Search for supersymmetry in final states with two oppositely charged same-flavor leptons and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV. J. High Energ. Phys. 2021, 123 (2021).
- Niendorf G, Reid T, Wittich P, Elmer P, Wang B, Chang P, Gu Y, Krutelyov V, Narayanan BV, Tadel M, Vourliotis E. Line Segment Tracking in the HL-LHC. arXiv preprint arXiv:2207.08207. 2022 Jul 17.

ABSTRACT OF THE DISSERTATION

A search for new physics in events containing a Z boson, jets and missing transverse energy at a center of mass energy of 13 TeV with the CMS detector, and a new parallelizable algorithm for charged particle track reconstruction in the outer tracker of the CMS Detector in the HL-LHC

by

Balaji Venkat Sathia Narayanan

Doctor of Philosophy in Physics

University of California San Diego, 2023

Professor Avraham Yagil, Chair

Two projects are presented in this thesis, the first pertaining to a physics analysis using the data from the CMS Experiment at the LHC, and the second pertaining to the development of a new parallelizable algorithm for reconstructing particle tracks in the

outer tracker of the CMS Detector at the upcoming High Luminosity LHC.

A search for supersymmetric phenomena beyond the standard model in a final state containing an on-shell Z boson, jets and missing transverse energy is performed using a data sample of proton-proton collisions at $\sqrt{s} = 13$ TeV corresponding to an integrated luminosity of 137 fb^{-1} collected by the CMS Experiment at the LHC between 2016 and 2018. The observed event yields are consistent with standard model predictions in the signal regions. These results are then interpreted to constrain the masses of supersymmetric particles in the context of the signal models. Gluino masses up to 1870 GeV, and chargino (neutralino) masses up to 750 (800) GeV are excluded at the 95% confidence level, which extends the reach over the previous results by a few hundred GeV.

The High Luminosity LHC (HL-LHC) will increase the instantaneous luminosity by a factor of five larger than the current levels achieved by the LHC. This high pile-up environment requires efficient and fast reconstruction of charged particles. A new algorithm called Line Segment Tracking takes a fundamentally different approach from existing iterative Kalman Filter based algorithms by doing a bottom-up reconstruction of tracks. Track stubs from adjoining detector regions are constructed, and stubs that are consistent with typical track trajectories are hierarchically linked to reconstruct complete tracks. Since the track stubs are produced locally and only require information from neighboring regions, they can be made in parallel. This motivates using architectures like GPUs to take advantage of the parallelism. The algorithm is currently implemented in the context of the CMS Phase-2 Outer Tracker in the HL-LHC, and targets NVIDIA Tesla V100 GPUs. Good physics and timing performance are obtained which paves way for further developments.

Chapter 1

Introduction

This thesis is a collection of two separate projects. The first project is a search for supersymmetry in a specific final state using data from the Compact Muon Solenoid (CMS) detector collected during the Run-II of the LHC. The second project details an ongoing development of a parallelizable particle tracking algorithm for the High Luminosity LHC, which will start collecting data in 2029.

Chapter 2 provides a brief physics overview about the standard model, its limitations, and supersymmetry, which is a beyond the standard model theory that attempts to mitigate some of these limitations.

Chapter 3 provides an overview of the Large Hadron Collider and the CMS Detector. It gets into brief detail about the various sub-detectors that make up the CMS Detector, and the trigger system. It also goes into detail about how various physics objects get reconstructed.

Chapter 4 provides a birds' eye view on how physics analyses are done using data produced by the CMS Detector. It describes the various tools, identification methods and unique quantities used in reconstructing particles like electrons, muons and jets. It also

describes the quantity called the missing transverse energy, its "measurement" and why this quantity is important for new physics searches. It then goes into detail about how a particle physics analysis is done, the various offline tools and techniques used for an analysis, like Monte-Carlo simulations, data driven background estimation techniques, and the statistical tools used to set exclusion limits.

Chapter 5 is supersymmetry search analysis in a final state containing the Z boson, jets and missing transverse energy. Like any other analysis paper/thesis chapter, it describes the signal processes, object selections, signal regions, background processes, background estimation techniques - both data driven and simulation based, overall validation of the background estimations, the results, and finally the interpretation of the results. This chapter closely follows the paper [15] and the related internal analysis note [16].

Chapter 6 details the Line Segment Tracking (LST) algorithm, which is the new parallelizable algorithm for the High Luminosity LHC. This chapter first describes the new HL-LHC detector with a pile-up of 200 and the new outer tracker modules and geometry in the CMS detector in the HL-LHC, in addition to getting in detail about the algorithm itself. It then describes the implementation of the algorithm on GPUs, after providing a brief overview of GPU computing and Cuda, and then delves into the various computational techniques involved in this implementation, followed by the timing performance and a reflection of what the future holds, and the various sub-projects in the horizon.

Chapter 2

The Standard Model and its extensions

2.1 Standard Model

The Standard Model of Particle Physics is currently the most exhaustive description of fundamental particles and their interactions. It provides descriptions for the interactions of three of the four fundamental forces - Electromagnetic, Strong and Weak, the fourth being Gravitational. In this framework (Figure 2.1), the elementary particles are fermions, with spin 1/2 while the force mediators are bosons with integer spin (1 or 0).

The fundamental spin 1/2 particles are either quarks or leptons (and their associated neutrinos). There are six types of quarks (and their anti-quarks) - up, down, charm, strange, top and bottom. The up, charm and top quarks have a charge of 2/3 while the down, strange and bottom quarks have a charge of -1/3, in a scale where the electron has a charge of -1. Since isolated quarks cannot exist, they are generally found in the form of mesons - which are made up of a quark and an anti-quark, and baryons - which are made up of three quarks [1]. A proton is a baryon made up of two up quarks and one down quark, resulting in a total charge of +1. A neutron is made up of one up quark and two down

[1]There have been recent developments with discoveries of pentaquarks which are outside the scope of this introduction

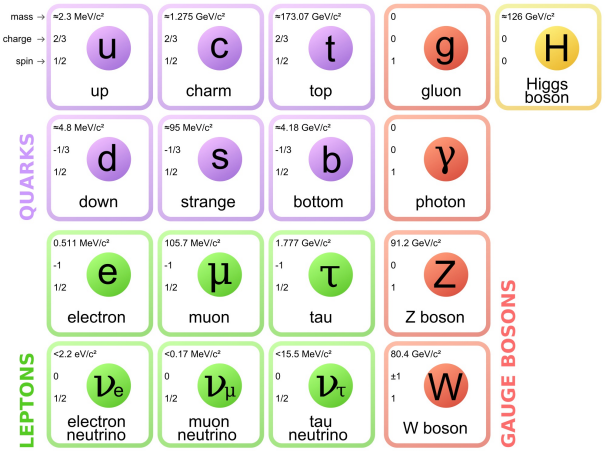


Figure 2.1. The Standard Model particles. Source : [1]

quarks, resulting in a total charge of zero. The leptons are of three flavors - electrons, muons and taus. Each of them have associated neutrinos - named electron, muon and tau neutrinos, which are predicted by the standard model to be massless.

The force mediator bosons are the gluon, the W boson, the Z boson, the photon, and the (most popular) Higgs Boson. The photon is the mediator for the electromagnetic interactions, the gluon for the strong interactions, and the W and Z bosons for the weak interactions. The Higgs Boson is the particle associated with the Higgs field, which mediates the spontaneous symmetry breaking required to provide W and Z bosons their mass.

2.2 Limitations of the Standard Model

The standard model has been shown to be very consistent with experimental results, and has been regarded as a landmark theory in particle physics. However, there are a few shortcomings of the standard model that render it incomplete.

Neutrino Masses

The standard model predicts that neutrinos are massless. However, we have observed neutrino oscillations [17], where neutrinos from one flavor can change themselves to become one of a different flavor. This happens because the flavor eigenstates of the neutrinos are actually linear combinations of the mass eigenstates, which evolve with their own propagation factors as the neutrinos travel through vacuum, and hence in the observation stage they take on a different avatar than the one at production. Since standard model neutrinos are massless, such an oscillation cannot exist. Hence, this discovery shows us that the neutrinos can have masses, with the latest upper bound being 0.8 eV [18].

Dark Matter Candidate

Observations of velocities of objects around galactic centers have shown a deviation from those predicted if the gravitational effects can be explained purely by the visible matter [19]. This "dark matter" only interacts gravitationally, and accounts for approximately 85% of the mass of the universe [20]. The standard model does not predict a particle with such properties, and (spoiler alert) searches for this particle at the particle level have not yielded any positive results yet.

Hierarchy Problem - Higgs Mass

A puzzling question after the Higgs Boson was discovered was why this new particle has a mass that is much much lighter (125 GeV) than the Planck mass (10^{19} GeV) - which is posited to be the scale at which quantum gravity effects become significant and the standard model would break, since one would expect that the quantum corrections to the Higgs boson mass will inevitably build up. However, if there has been some "fine-tuning" of the parameters of the quantum corrections, they can somehow perfectly cancel out and can result in a very low value of the Higgs mass, as was observed. The standard model does not tell us why such a fine-tuning should even happen on the first place.

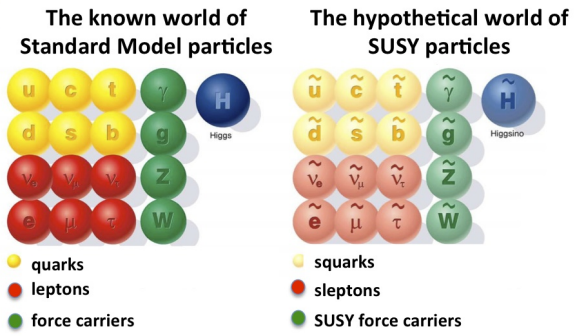


Figure 2.2. Standard model particles and their SUSY counterparts. Source : [2]

2.3 Supersymmetry - A Beyond the Standard Model Framework

Supersymmetry is a Beyond the Standard Model Framework that aims to address the limitations of the standard model. In the minimally supersymmetric standard model (MSSM) framework, every standard model particle is paired with a superpartner, as shown in Figure 2.2. The superpartners of the fermions have integer spins, while the superpartners of the bosons have half integer spins. The squarks (superpartner of the quarks) are called sup, sdown, scharm, sstrange, stop and sbottom, which have a spin of zero. The gauginos (superpartners of the gauge bosons) are called photino, gluino, zino and wino respecively. The Higgs boson also has a spin 1/2 partner called the higgsino. The generic terms "neutralino" and "chargino" are used to refer to neutral and charged boson superpartners respectively.

The hierarchy problem sees a natural correction in this framework when the additional contribution to the Higgs field from the stop squark (the superpartner of the top quark) mediated processes cancel out the top quark mediated quantum corrections (Figure 2.3), removing any appeal to fine-tuning of parameters in order to explain the Higgs mass. Supersymmetric fields can have a quantity called R-parity[21]. The standard model

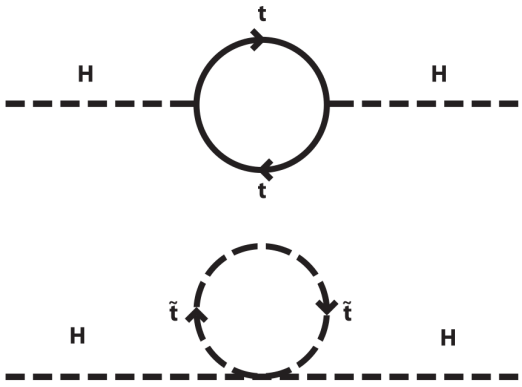


Figure 2.3. Stop squark mediated cancellations to the Higgs field quantum corrections. Source : [3]

particles will have $R = 1$ while the supersymmetric particles will have $R = -1$. If the fields are found to be R-parity conserving, then there will exist a lightest supersymmetric particle (LSP) that cannot decay to any other supersymmetric or standard model particle, and hence will be very stable. This particle is predicted to interact only through gravitational or weak interactions and will be a candidate for dark matter if electrically neutral.

While Supersymmetry sounds like the panacea for all the limitations that the standard model has, we have not seen any evidence of it yet. The physics analysis in Chapter 5 will be a search for specific supersymmetric models exhibiting a particular signature.

2.4 Parton Distribution Function

A proton is made up of three quarks in a sea of gluons. The quarks and gluons that form a part of the proton are called partons. The sea of gluons can spontaneously create and destroy quark-antiquark pairs, depending on the energy of the protons. When two protons collide, a host of internal processes occur which ultimately result in final

state products, produced as a result of inelastic collisions. Parton Distribution Functions (pdfs [2]) essentially provide probability distribution functions for the partons to have a fraction of the parent proton's momentum. For inelastic proton-proton scattering, the gluons will start to play a role (since they can create quark-antiquark pairs which can also interact). Hence, parton distribution functions get dependent on the magnitude of the momentum transfer (denoted by Q). These parton distribution functions are measured for a particular value of Q (dependent usually on the initial energy of the protons, and hence indirectly on the accelerator system) and then computed for other energy scales using the DGLAP evolution equations. Figure 2.4 shows the parton distribution functions at two different energy scales for the proton. We see that the heavier bottom quark starts to make an appearance as the momentum transfer value Q goes up to 100 GeV. This was why a Tevatron with a center-of-mass energy of 1 TeV was needed to discover the top quark. Earlier colliders could simply not have mustered up the energy to produce it. [3]

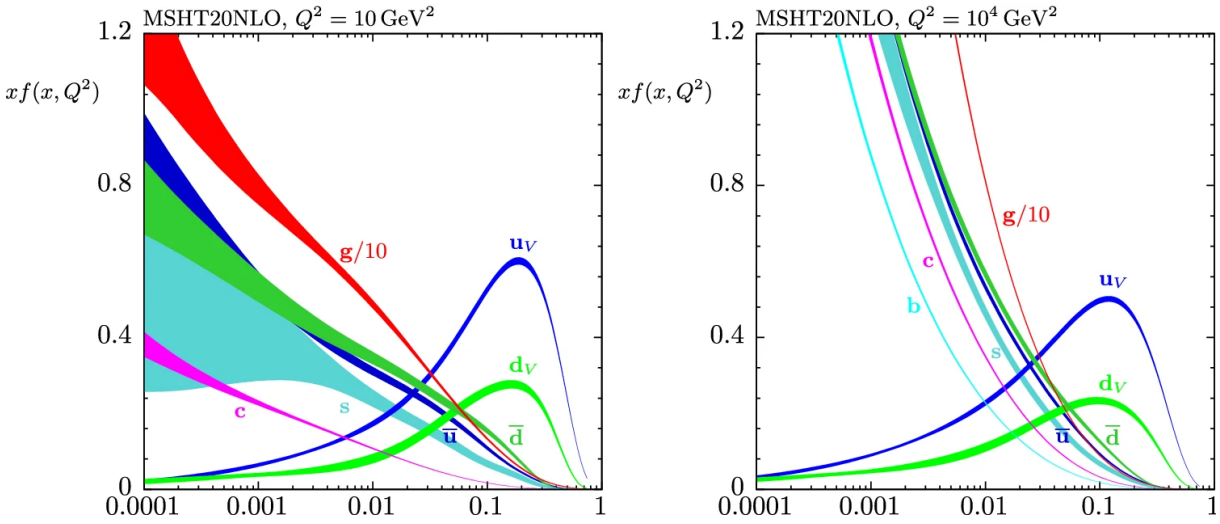


Figure 2.4. Parton distribution functions for the proton at two different Q values. Source : [4]

[2] Conveniently having the same acronym as probability density function, to confuse new grad students

[3] This statement is quite tangential because the Tevatron was a proton-antiproton collider, but the principle is that one needs to energize the protons/antiprotons to higher energies to have a nonzero probability of producing more massive quarks.

Chapter 3

The Compact Muon Solenoid Detector

3.1 The Large Hadron Collider

The Large Hadron Collider is a particle accelerator built by the European Organization for Nuclear Research (CERN [1]). It is the largest particle collider in the world. The accelerator is in the form of a 27km long ring. Four detectors are located along the ring - ATLAS (A Toroidal LHC Apparatus) and CMS (Compact Muon Solenoid) being the two general purpose detectors, ALICE (A Large Ion Collider Experiment) being a special purpose heavy ion collision based detector, and LHCb (LHC Beauty) being a special purpose detector for B-Physics.

A schematic of the LHC is shown in Figure 3.1. The first step consists of a linear particle accelerator Linac4 that generates negative hydrogen ions (H^-) having energy of 160 MeV. These are then fed through the Proton Synchrotron Booster, which strips the two electrons from the negative ions to create protons and accelerates then to 2 GeV. These protons are then fed into the Proton Synchroton (PS in the figure) which is a repurposed old particle accelerator from the 60s. This accelerates the protons to 26 GeV. Then these are fed into the Super Proton Synchroton (SPS in the figure), which is also a repurposed accelerator from 1976 which then accelerates the protons further to 450 GeV.

[1] French acronym for Conseil Européen pour la Recherche Nucléaire

These protons are then injected into the main LHC ring and accelerated to the target energy of 7 TeV (6.5 during Run II). These protons are then collided at the centers of the four detectors, creating a center-of-mass energy of 14 TeV (13 TeV in Run II). The protons are bunched together before injecting them into the main LHC ring. Each bunch contains 115 billion protons that are separated by 25 nanoseconds, which provides a collision rate of 40 MHz.

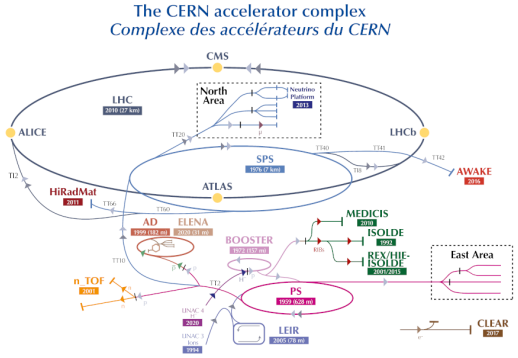


Figure 3.1. LHC Schematic. Source : [5]

In particle physics, the probability of a particular scattering or collision event is measured using a tool called the cross section (σ) and is measured in units of area. It is defined as the area of the target that the scattering particle sees (assuming a fixed target scattering). The unit **barn** is commonly used in this context, and corresponds to an area of 10^{-24} cm^2. Luminosity refers to the ratio of the number of events detected in a time interval to the cross section of the process that created these events, i.e.,

$$\mathcal{L} = \frac{1}{\sigma} \cdot \frac{dN}{dt} \tag{3.1}$$

The integrated luminosity $L = \int \mathcal{L} dt$ refers to the ratio of the total events detected in a time interval normalized by the cross section, and has units of inverse area. Across the years 2016-2018 when Run-II of the LHC happened, data corresponding to an integrated

luminosity of 137.2 fb^{-1} was collected. This means that a collision event with a cross section of $1/137 = 0.08$ fb (femtobarns) would have appeared once in these three years of data taking. Each proton bunch that collides in the center of the detectors every 25 ns corresponds to a luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$. Due to the bunched up nature of the collisions, we expect 30 inelastic collisions on average per 25 ns (also called a bunch crossing). These are referred to as pile-up. Figure 3.2 shows a distribution of pile-up recorded in collisions in 2018, which was the last year of Run-II of the LHC. We usually get a maximum of one "interesting " collision (i.e., one that is relevant to us physics wise) out of these 30. The interesting collision vertex is called the Primary Vertex, the others being called the pile-up vertices. Future developments involve increasing the luminosity such that the pile-up increases to 200. This will play an important role in the tracking algorithm described in chapter 6.

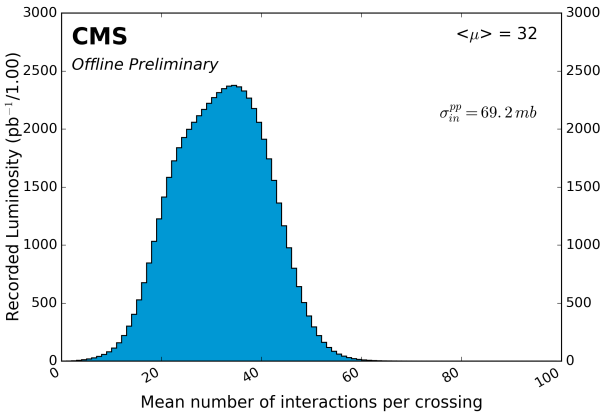


Figure 3.2. Distribution of pile-up in 2018. Source : [6]

3.2 The CMS Detector

The Compact Muon Solenoid detector [22, 23] is one of the two general purpose detectors in the Large Hadron Collider. it is situated in Cessy in France, diametrically opposite to the main CERN main campus in Meyrin, Geneva. Figure 3.3 shows a

photograph of the detector. It is 21 meters long and 15 meters tall [2]. The main feature of the detector is a 13 meter long, 5.9 meter inner diameter solenoidal superconducting magnet with a strength of 3.8 Teslas. Its function is to create a magnetic field along the proton beamline (inside the surface of this page in the Figure 3.3). Charged particles that get produced will then have helical trajectories under the influence of this magnetic field. The various detectors are stacked along the curved surface to detect these particles and reconstruct them.

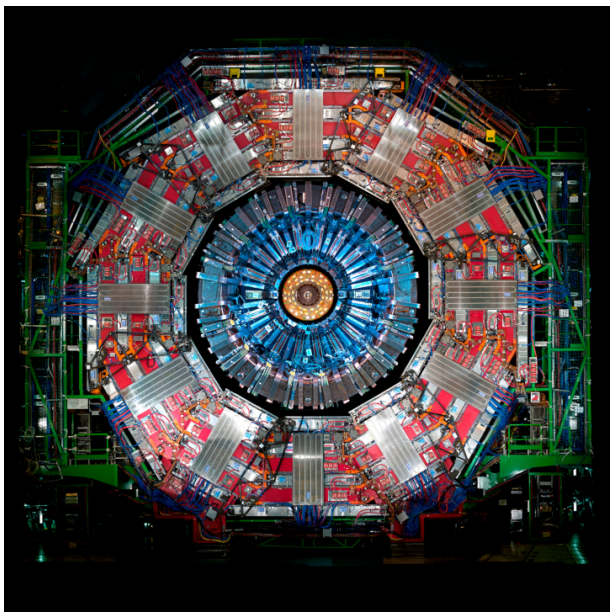


Figure 3.3. A photograph of the cross-section of the CMS Detector. Source : [7]

The coordinate system used by the CMS detector is centered at the nominal collision point inside the detector. The x-axis points radially inward towards the center of the LHC (to the right in Figure 3.3), the y axis points upwards, and the z axis points along the beam direction (inside the page). A spherical style coordinate system is used, with the azimuthal angle ϕ measured from the x axis in the x-y plane. The polar angle θ is measured with respect to the z-axis. However, the pseudorapidity $\eta = -\ln\tan\left(\frac{\theta}{2}\right)$ is the

[2]Despite being called "Compact". That is the scale in which these detectors operate

commonly used quantity in lieu of using θ directly. The transverse momentum (p_T) of the particles is measured along the x-y plane.

3.3 Components of the CMS Detector

Figure 3.4 shows a schematic cross-section of the CMS detector. The main components going radially outwards are the tracker, the calorimeters and the muon detectors, which will be covered in detail in the forthcoming sections. From this point of view, the detector is said to be split into a central barrel region, corresponding to the curved cylindrical surface, enclosed by two endcap regions on either end.

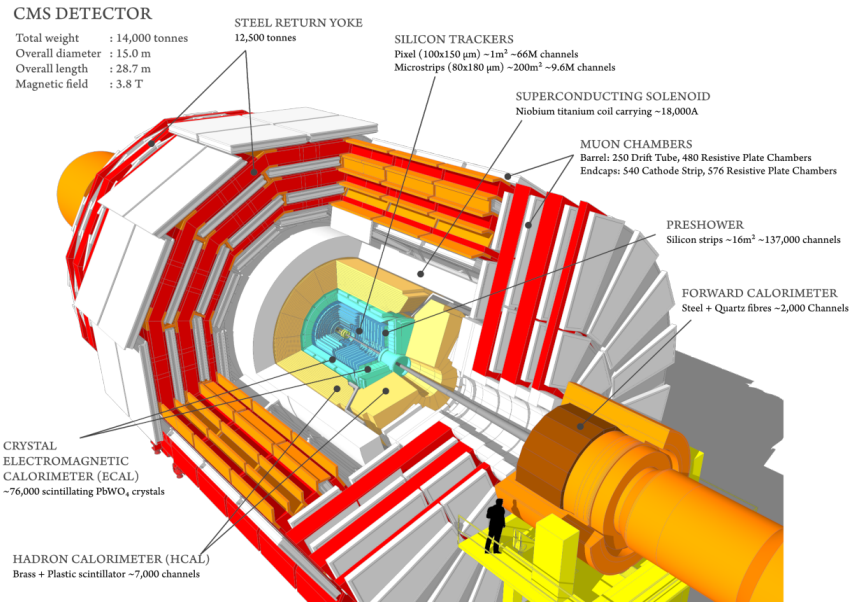


Figure 3.4. A cutaway view of the CMS detector, showing the individual detector components. Source : [8]

3.3.1 Tracker

The innermost sub-detector is the tracker [24]. A schematic in the r-z plane can be seen in Figure 3.5. It has an inner pixel detector made up of four layers ranging between 3cm and 16cm, and an outer strip detector consisting of 10 layers in the central barrel region till 1.1 m. The endcap regions have two disks for the inner pixel detector and 12

disks for the strip detector. When a charged particle passes through the tracker layers, it leaves small charge deposits called hits. Tracking algorithms in the software backend correlate hits consistent with the trajectories of particles in a solenoidal magnetic field and use the degree of bending to compute quantities like the momentum.

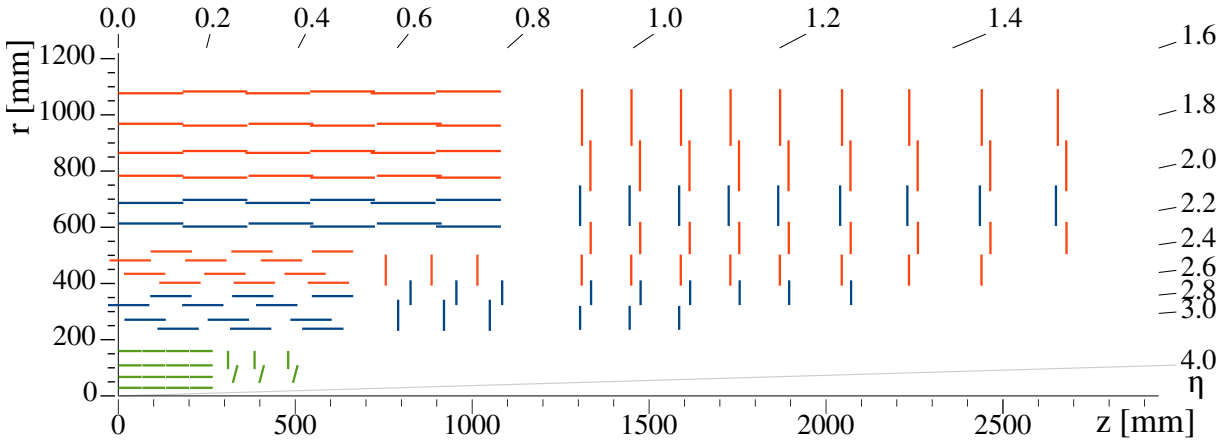


Figure 3.5. r-z schematic of the CMS Tracker. Source : [9]

Track reconstruction is a computationally intensive process and involves correlating hits that could have come from the same particle trajectory. The current tracking software backend uses a Kalman Filter based approach. Chapter 6 contains more details about current and future developments in track reconstruction.

3.3.2 Electromagnetic Calorimeter

Right outside the realm of the tracker is the Electromagnetic Calorimeter (ECAL). It is a hermetic, homogeneous calorimeter made up of nearly 75000 Lead Tungstate ($PBWO_4$) crystals. Lead Tungstate crystals have a relatively small radiation length [3] of 0.89cm and an equally short Moliere radius [4] of 2.2 cm, in addition to having a quick response (most of the energy is radiated within 25 ns). This means that a lot of these

[3]The average distance an electron travels before being stopped by the crystal
[4]The average radius of the scintillation produced by the interaction

crystals can be packed in close quarters resulting in good coverage and position resolution. The Tungstate crystals, being 22 cm in depth (25 radiation lengths) and one Moliere radius (2.2 cm) wide, are ideal for stopping electrons and photons completely, resulting in a perfect measurement of their energies.

To further distinguish between single high energy photons and showers of low energy photons, the ECAL has a preshower detector in front of it. These filter out the low energy photons and neutral pions and ensure only interesting high energy electrons and photons are measured by the calorimeter.

3.3.3 Hadronic Calorimeter

The Hadronic Calorimeter (HCAL) is present outside the ECAL whose function is to measure the energies of the hadrons. It has four components - barrel (HB), outer (HO), endcap (HE) and forward (HF). The HB and HO cover the barrel, with the HB within the HO. The HE covers teh endcap region, and the HF takes it even further in the endcaps, going all the way till $|\eta| = 5.0$. The calorimeters are made up of brass or steel absorbers that induce the showers, alternating with plastic scintillators connected to readouts. The barrel brass absorbers are wide enough to contain 6 interaction lengths (about 5cm). These relatively thick hadronic calorimeters ensure that almost all the hadrons are stopped and measured, leaving only the muons to traverse all the way outside the interior of the magnetic core. Measurement of hadrons is vital to the accurate measurement of missing transverse energy E_T^{miss}, which plays an important role in searches for Beyond the Standard Model (BSM) signatures.

A bit of history : The endcap calorimeters (HEs) were built by Russian groups in the collaboration, and the high quality brass used was procured by recycling World

War II era artillery shells manufactured by the erstwhile Soviet Navy. This recycling was also symbolic of erstwhile tools of destruction and war being repurposed for constructive scientific endeavors that benefit humankind as a whole.

3.3.4 Muon Detectors

The Muon detectors are the final set of detectors in the setup and play a vital role in the experiment [5]. They are interspersed between the magnetic return yoke, made up of steel. The strong magnetic field (3.8 T) completely saturates the steel yoke, producing a magnetic field in the opposite direction. Hence, the muons now start bending in the other direction in the detectors. There are four types of muon detectors used in the system, depending on the location. All of them work on muons knocking electrons off the atoms in an Ar/CO_2 gas mixture.

In the barrel, drift tube chambers are used. These drift tube chambers consist of 4 cm wide tubes which consist of stretched wire within a gas volume. When a muon travels through this gas filled region, it knocks electrons out of the gas atoms which are then detected by the positively charged wire. Cylindrical stacking of these tubes provide measurement of the radial distance from the interaction point.

In the two endcaps where the muon rate and the neutron induced background rate are high, cathode strip chambers are used. They consist of arrays of positively charged wires criss-crossed with arrays of negatively charged plates, immersed in a gas chamber. When a muon knocks an electron out, the electron drifts to the wires while the positively charged ion drifts to the plate. This gives the radial and the azimuthal coordinates [6].In the

[5]They are the reason the CMS is the Compact **Muon** Solenoid

[6]Note that the azimuthal coordinate in the barrel can be measured by the location of the drift tube since they are cylindrically placed. Hence we do not have the grid mechanism.

forward regions, Gas Electron Multipliers (GEMs) are now put to use (starting 2022). The GEMs also operate on a similar principle to the other detectors, but they are much longer than the others the longest one being more than 1m long, optimal for its positioning in the very forward regions. In addition, they can operate in a region with very low transverse magnetic field, as it is in the forward regions.

Resistive Plate Chambers are used in the barrel and the endcaps in addition to the drift tubes and the cathode strip chambers. While not being very precise on the position front, they provide excellent timing resolution. The quick and dirty measurement of the momentum of the muons play a role in the trigger system (Section 3.4)

3.4 Triggers

Proton bunches collide at 25 ns intervals, providing a frequency of 40 MHz. The sheer rate of collisions makes it impossible to reconstruct and store the avalanche of data produced by the detector for each event. To mitigate this, a two-tier trigger system makes quick on-the-fly decisions about whether an event is potentially "interesting" to store in memory.

The Level-1 (L1) trigger is the first level of the system. It is made up of FPGAs that make simple calculations of various particle momenta and energy based on the information from the detector components, and make a quick decision about preserving or discarding a collision event. FPGAs are quite useful for making very simple decisions very quickly. This level reduces the event rate by a factor of a million, and outputs events at the rate of 100 KHz, corresponding to an event every 10 microseconds.

The High Level Trigger (HLT) is the second level of the trigger system. It is purely a software based system that runs on a server farm. At this stage, event reconstruction happens with even higher precision. There are various HLT pathways that look for different signatures at the same time. For instance, a collision event that passed the L1 trigger might be checked for a myriad of signatures, like whether it has two electrons, or two muons, or an electron and a muon, and so on. If an event passes any of these various paths, it gets sorted and written to disk. At this stage, the total event rate (i.e., events that pass any one HLT path out of the myriad of paths) is less than 1 KHz.

3.5 Object Reconstruction

In a particle collision process, a myriad of exotic particles (Z boson, Higgs Boson etc) are produced, but almost all of them decay into secondary particles and so on, ultimately producing the most stable particles in the decay chains. The most common ones that live long enough to leave significant signatures in the detector components are electrons, muons, neutral hadrons, charged hadrons and photons. At the end of the day, the CMS detector only measures charge deposits from the pixel detectors and muon chambers, energy deposits from the calorimeters through the read-out electronics. It becomes very important to be able to tell which particle was produced from looking at the patterns of these deposits.

A rough guide of these signatures is shown in Figure 3.6. All charged particles leave a trail in the tracker. In fact, the fitted tracks from the tracker provide an estimate of the p_T for the charged particles. Electrons are roughly identified by the tracks, the associated ECAL deposit and nothing beyond that. Photons have ECAL deposits similar to the electrons but no associated tracks with them. Sometimes the track fitter might miss

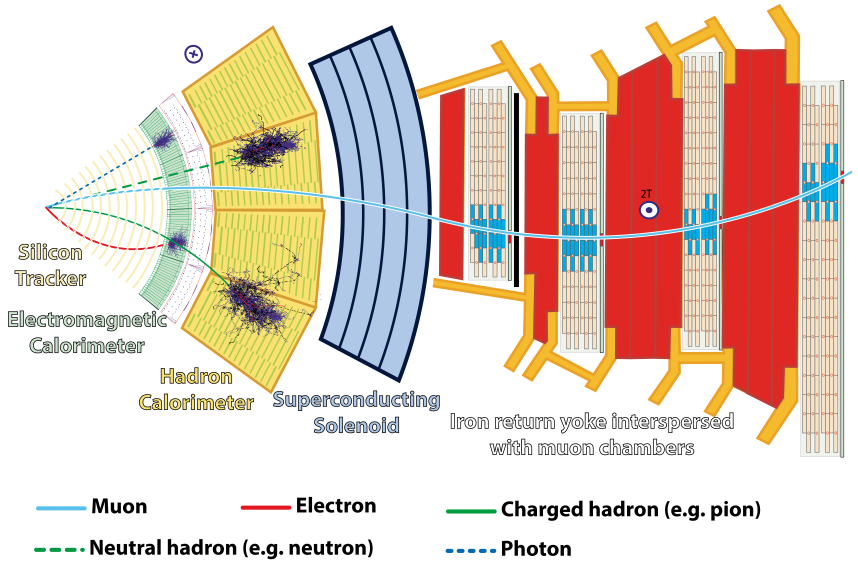


Figure 3.6. Object reconstruction mechanisms in the CMS Detector. Source : [10]

an electron track altogether, resulting in it getting mis-identified as a photon, and vice versa. Charged and neutral hadrons both have HCAL deposits, the difference between charged and neutral hadrons being that charged hadrons will have a small ECAL deposit and an associated track, while a neutral hadron will have neither. Muons only have an associated track and detections by the muon systems.

Particle Flow reconstruction algorithms [25] are used to use these basic correlated signatures from all the underlying sub-detectors to reconstruct objects. Accurate reconstruction, especially of hadrons is crucial for jet clustering and Missing Transverse Energy (E_T^{miss}) computation. Since E_T^{miss} (Section 4.3) is an important handle for new physics searches, it becomes imperative that reconstruction happens accurately. Machine Learning and tree based models have also started to play a vital role in object reconstruction.

3.6 Acknowledgments

This chapter described the Large Hadron Collider and the Compact Muon Solenoid Experiment. The figures used in this chapter are obtained from the following results : "Facts and figures about the LHC" *https://home.cern/resources/faqs/facts-and-figures-about-lhc*, "CMS Luminosity–Public Results" *https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults*, "Cutaway diagrams of CMS Detector" *https://cds.cern.ch/record/2665537*, "CMS DPG - Public Results" *https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK*, "CMS Detector Slice" *http://cds.cern.ch/record/2120661*

Chapter 4

Physics Using the CMS Detector

A typical particle physics analysis with data collected from the CMS Detector is usually geared towards verifying theories proposed by our theorist colleagues, measuring an experimental parameters like coupling constants or branching ratios, or finding new physics that will pique the interest of our theorist colleagues. As experimentalists, our goal is more aligned towards understanding how various physics objects are reconstructed (as seen in Section 3.5), being able to come up with efficient criteria that increases the sensitivity of our target "signal" process while reducing the influence of the "background" processes, and construct hypotheses in the context of the analysis which will then be verified with actual observed data.

4.1 Electrons and Muons

Precise measurement of the kinematic properties of leptons is crucial in the reconstruction of events. Unlike jets (Section 4.2) these are single particles which can be precisely measured in a straightforward manner, and when it comes to analyses where the presence of a Z boson is crucial, the di-leptonic channel is the preferred one, despite this process having a lower branching ratio.

4.1.1 Identification of Leptons - Lepton ID

Muons are reconstructed using the underlying information obtained from the trackers and the muon detectors which get incorporated in a threshold based identification mechanism called the "medium cut based muon ID" [26]. Since muons are fairly clean and straightforward to reconstruct, a simple cut based mechanism suffices for good identification. Some of the criteria used for this selection are

- A reconstructed track in the muon detectors
- A matching reconstructed track in the trackers (global muon criterion)
- χ^2 fit parameter for the global track

These criteria are designed to be efficient for prompt muons (muons from the interaction vertex) and from heavy quark (top/bottom) decays.

Electrons, on the other hand, are identified solely based on the tracker and ECAL deposits. In addition, they compete with photons which nearly have the same signature. Here a boosted decision tree (BDT) based approach [27] is used to produce confidence scores. Some of the variables used are

- Variables that deal with ECAL shower shape
  - $\sigma_{i\eta i\eta}$ - weighted width of ECAL shower along η
  - $\sigma_{i\phi i\phi}$ - weighted width of ECAL shower along ϕ
  - Unweighted widths of ECAL shower along η and ϕ
  - Cluster circularity
  - R_9 - the ratio of energy in a 3x3 set of ECAL towers to the ratio in a 5x5 set of ECAL towers

  - H/E - Ratio of ECAL energy deposit to the corresponding HCAL energy deposit

- Variables that deal with tracks and track-cluster matching

  - Track χ^2
  - $\Delta\eta_{in}$ - η difference between inner track and associated ECAL cluster
  - $\Delta\phi_{in}$ - ϕ difference between inner track and associated ECAL cluster
  - $\Delta\eta_{out}$ - η difference between outer tracker track and associated ECAL cluster
  - $\Delta\phi_{out}$ - ϕ difference between outer tracker track and associated ECAL cluster
  - E/p_{in} - Ratio between ECAL energy and inner tracker track momentum
  - E/p_{out} - Ratio between ECAL energy and outer tracker track momentum

Various working points- "loose", "medium" and "tight" are derived to maintain reasonable true and false positive rates. The loose working point is generally used for vetoing electrons while the tight working point is generally used to create a clean channel to target leptonically decaying processes. In addition to cutting on the BDT score, a "photon veto" is applied to prevent tagging $\gamma \to e^+e^-$ processes, which requires that the inner tracker not have any missing hits associated with an electron, since the conversion generally tends to occur after a photon has crossed some inner detector layers (where it would not register a hit).

To ensure that only those leptons from the primary vertices are used in analyses, cuts on the distances between the point of closest approach (PCA) of the lepton's track and the primary vertex (PV) are incorporated. These cuts are imposed on the x-y distance Δ_{xy}, the z-distance Δ_{z}, and the impact parameter significance SIP_{3D}, defined as the 3D displacement between the PCA and PV divided by its measurement uncertainty.

The overall lepton reconstruction efficiency is between 45% and 70% for electrons with $p_T > 25$ GeV and between 70% and 90% for muons with $p_T > 25$ GeV. In the lower p_T range of 15-25 GeV for electrons, the reconstruction efficiency is around 40%, and for muons in the p_T range of 10 to 25 GeV, the reconstruction efficiency is 55%.

4.1.2 Lepton Isolation

Lepton identification does not take into account whether the lepton is isolated or embedded inside jets or regions of hadronic activity. Analyses that depend on leptons require that the leptons not be produced from any sort of hadronic activity, but rather produced directly by the parent particles themselves, and hence only deal with those leptons that are well isolated.

Isolation is counter-intuitively measured as the sum total of the p_T of hadronic particles in a cone in $\eta - \phi$ space around the lepton. Hence, the lower the isolation value, the more isolated the lepton is

$$I_{\text{mini}} = \frac{\sum_R p_T(h^{\pm}) - \max(0, \sum_R p_T(h^0) + p_T(\gamma) - \rho\mathcal{A}(R/0.3)^2)}{p_T(\ell)} \tag{4.1}$$

where $\sum_R p_T(h^{\pm})$ is the sum of the p_T s of the charged hadrons, $\sum_R p_T(h^0)$ the sum of the p_T s of the neutral hadrons, $p_T(\gamma)$ the sum of the photons, and ρ is the event-specific average pile-up energy density per unit area. The effective area $\mathcal{A}$ is called the effective area. The last term is used to mitigate the effects of pile-up on the neutral hadron and photon energies. The isolation is computed in a cone of radius (R in the formula) that varies with the p_T as

$$R = \frac{10}{min[max(p_T, 50), 200]} \tag{4.2}$$

4.2 Jets

Jets are clusters of stable particles arising from the hadronization of a quark or a gluon from a collision process. Since free quarks cannot exist, they cluster together to form these jets. Since relativistic momentum and energy are always conserved, the total energy of the jet components is equal to the energy of the original quark that got produced in the process.

However, we do not get nicely clustered thin collimated beams of hadrons in nature. Instead, it is up to us to cluster particles using well-defined rules to create jets. The anti-k_T algorithm [28] is used in the CMS Experiment. This algorithm repeatedly clusters neighboring particles in pairs, within a cone of a particular ΔR threshold, and replaces them with a single "complex" particle. Once all the particles inside this cone are clustered (i.e., the nearest particle now is farther away than the cone size), this complex object is deemed to be a jet.

The standard jets that are used by analyses is the AK4 class of jets. They are clustered with the anti-k_T algorithm with a cone ΔR of 0.4. Since jets are treated physics wise as a single object, they are assigned kinematic values like p_T and mass, which will be used in selections. In addition, we can also get values like neutral energy fraction (fraction of total jet energy from neutral hadrons), charged energy fraction (fraction of jet energy from charged hadrons) etc, which will also be used as parameters for jet identification algorithms mainly to tag jets that could have been produced from certain processes like W Boson decays, or b-quark jets (to tag the $t\bar{t}$ process) and so on.

When the parent hadronizing particle (the particle that ends up producing quarks that end up as jets) has a large p_T, it is said to be "boosted". The quarks that gets produced in opposite directions in the rest frame of this process get collimated in the lab frame, resulting in a single "fat" jet instead of two distinct jets. In this scenario, we use the anti-k_T algorithm with a cone ΔR of 0.8 (roughly corresponding to 2 0.4 cones) to reconstruct these fat jets. When jets are clustered with this big of a cone, we can end up in scenarios where two closely spaced AK4 jets get clustered as a single AK8 jet. A special variable called n-subjettiness (τ_n) [29] is used to discriminate genuine fat jets from scenarios like these. τ_n roughly measures the closest distance between the constituent particles and a sub-jet, assuming the fat jet can be decomposed into n sub-jets. The variable $\tau_{nm} = \tau_n/\tau_m$ is generally used to estimate an apples-to-apples comparison between two scenarios where a jet can be decomposed to n sub-jets or m sub-jets. The smaller this distance, the higher the possibility that n sub-jets exist in lieu of m sub-jets. To tag instances where a single Z or W boson decays into a single fat jet in place of two jets, we generally require the $\tau_{21} = \tau_2/\tau_1$ variable to have a number greater than a threshold.

The masses of fat jets are measured using a technique called soft-drop [30] which removes spurious radiation from these jets and computes the mass of the remaining particles. This variable will be used to get an estimate of the mass of the parent particle that hadronized to form the fat jet.

When jets are reconstructed, care must be taken to associate jet constituents with the appropriate proton-proton collision, i.e, whether they come from the pile-up vertices or the primary vertex. After all, the primary vertex itself is identified to be that vertex whose associated objects have the largest value of $\sum p_T^2$. While particle flow algorithms

do their level best in associating hadrons to appropriate vertices, errors might still creep in. To correct for errors arising from these associations, jet energy corrections (JECs) parameterized by jet p_T and η are derived in data and simulation and used in physics analyses.

4.2.1 B-Tagging

Tagging Jets produced from a b-quark is important to tag $t\bar{t}$ processes, where the top quark decays to a bottom quark and a W boson. In addition, some hadrons containing b-quarks have long lifetimes that allow them to travel away from the collision vertices before decaying, which create tracks that point to a secondary vertex.

DeepCSV [31] is a machine learning based algorithm that takes track information and produces a score, which is a measure of confidence that a particular jet is a b-quark jet. From these, "tight", "medium" and "loose" working points are derived. The medium working point is derived such that the efficiency to correctly identify a b-jet is approximately 65% for a jet with a p_T of around 40 GeV, while having a mis-identification (false positive) rate of 1% for light flavor jets (u,d,s,g quark jets). The "tight" working point has a 0.1% mis-identification rate, and the "loose" working point has a 10% mis-identification rate. For analyses that want to specifically work with b-jets, the higher valued "tight" working point will be used, i.e., only those events that have jets that pass the tight working point will be considered. On the other hand, for analyses that want to reject the $t\bar{t}$ background, the "loose" working point will be used, i.e., any jet with even a semblance of a b-jet will be rejected.

4.3 Missing Transverse Energy : $\mathbf{E_T^{miss}}$

When two protons are collided head on (along the z direction), the initial total momentum in the x-y plane is zero. Hence we also expect the final x-y momentum of all particles produced in the event to be zero. However, due to measurement errors, or the production of particles like neutrinos or possibly new beyond the standard model particles which cannot be easily detected, a transverse total momentum imbalance may occur. This quantity is called missing transverse momentum ($\vec{p}_T^{\,miss}$), given by the following expression

$$\vec{p}_T^{\,miss} = -\sum_{all} \vec{p}_T \tag{4.3}$$

Its magnitude is called Missing Transverse Energy (E_T^{miss} or p_T^{miss}). Figure 4.1 shows a representative computation of this quantity in a $t\bar{t}$ production process. Since new physics particles like dark matter candidates are bound to escape undetected, E_T^{miss} provides a powerful handle for detecting them, and will be the "quantity of interest" in the Supersymmetry search outlined in Chapter 5.

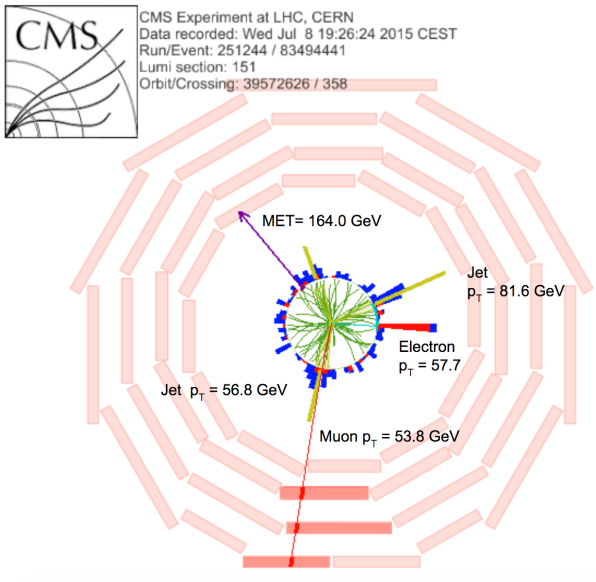


Figure 4.1. A representative portrayal of E_T^{miss} associated with a $t\bar{t}$ production. Source : [11]

E_T^{miss} is a double-edged sword. To be able to use E_T^{miss} effectively as an analysis variable, the measurement of momenta and energies of other particles need to precise, and the associated uncertainties and corrections properly quantified. While leptons are measured very accurately, the jets have some post-measurement corrections applied which are also propagated to the E_T^{miss} variable. The corrected E_T^{miss} is called the Type-1 E_T^{miss}, given by the expression

$$\vec{p}_T^{\,miss} = \vec{p}_T^{\,miss,\ raw} - \sum_{jets} \left(\vec{p}_{T,\ jet}^{\,corr} - \vec{p}_{T,\ jet}\right) \tag{4.4}$$

The reason we resort to using the missing transverse energy instead of the 3 vector missing energy (including the z plane) is that due to the fact that the actual quarks involved in a collision is not deterministic, we do not have a complete accurate picture of the happenings in the z-direction. All we have is a parton distribution function 2.4 that provides us with the probabilities of particular quarks being produced with a random fraction of the initial energy. In case of electron-electron collisions, like in the older LEP (Large Electron Positron Collider), the complete initial picture in all directions is known, since the electron is a point particle. In such scenarios the full missing energy is used, instead of the missing transverse energy

4.4 Isolated Tracks

Isolated Tracks (or Isotracks) are simply those tracks which are well isolated. They can be associated to other objects, like electrons, muons or charged hadrons. However this particular collection also has some isolated tracks that might be associated to these objects, but might fail some quality criteria in order to be included as any of the other objects. They are mainly used in the context of vetoes. For example, a requirement for a

lepton free final state will, in addition to requiring that there be zero leptons in the event, will also require that there be no isotracks that are matched to leptons in the event.

4.5 A Typical Particle Physics analysis

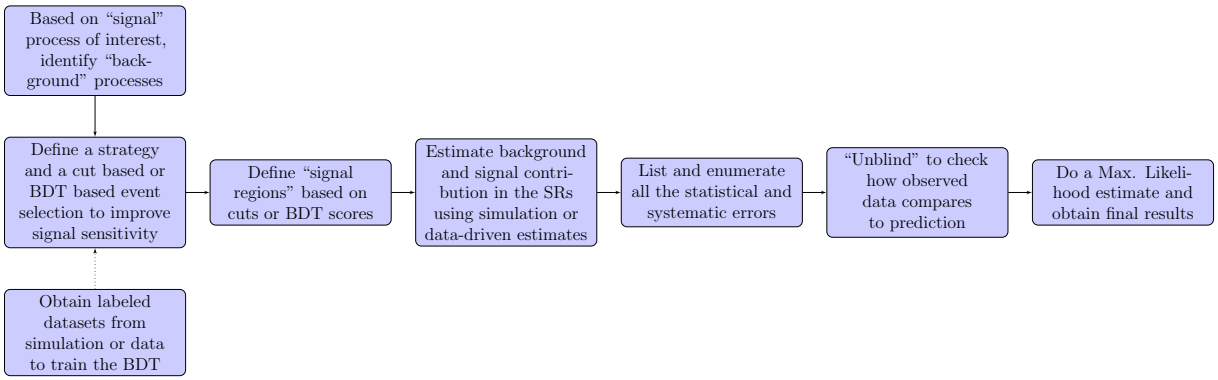


Figure 4.2. Steps involved in a particle physics analysis

Particle physics analyses have a goal that involves either trying to find evidence of a presently unobserved process and setting relevant bounds if these processes cannot be observed , or verifying or constraining parameters for processes that can already be observed. For instance, there are still analyses that continue to measure the mass of the Higgs Boson (discovered in 2012), or even the W boson [32] [1](discovered in 1983 by the UA1 and UA2 experiments at CERN) to reduce the associated uncertainties further. These target processes are called "signal processes", while processes that have final states similar to the signal processes are called "background processes".

Figure 4.2 shows the steps involved in a typical analysis. The signal and background processes are first identified. This is followed by a "signal strategy", which deals in basic physics object selections, and some baseline selections in parameter space (i.e., all possible

[1]The 2022 paper cited here used data from the CDF-II detector that operated at the Tevatron in Fermilab which was decommissioned in 2011, and has created ripples in the community because it measured a marked departure from the W boson mass predicted by the Standard Model and verified in earlier experiments

physics parameters) with the ultimate goal of improving the signal to background ratio. Once some baselines are established, signal regions are defined. The definitions for these regions can be based either on simple cuts and thresholds on parameters that are fundamental (like E_T^{miss}) or constructed (like the M_{T2} variable in Section 5.4.2), or can be based on the score produced by a complex classifier like a Boosted Decision Tree or a Neural Network. To engineer these thresholds, simulated datasets (Section 4.6) are often used. The entire analysis is done "blinded" i.e., real collision data is not used to define any strategies, so that predictions are not biased by actual physics observations.

Once these signal regions are finalized, the background and signal contributions in these regions are estimated. The most straightforward method involves simply using the simulated datasets. However, these are not very accurate due to limitations in simulations, especially of the detector components. However, they still serve a very purpose, since alternate methods do not exist for a large fraction of physics processes. The alternative to simulation is using a data-driven estimation. Data is obtained from a different region in parameter space and "transformed" into the background contribution to the signal regions. For instance, if $t\bar{t}$ is a major background in a same flavor opposite sign di-lepton final state (i.e., ee or $\mu\mu$), $t\bar{t}$ events from opposite flavor final states can be used with very little modification (since the probability of a $t\bar{t}$ process producing $e\mu$ and $ee + \mu\mu$ is the same). The advantage of using a data-driven method is that there are no limitations or approximations due to simulation, since actual data is being used. The caveat here is that such convenient methods are few and far between, and for some processes these regions might not even exist. The convention is to search if a data-driven method can be found, and if not, resort to using the simulated samples.

Once the predictions of total signal and total background in the signal regions are finalized, the analysis is "unblinded" and the predictions are checked with actual collision data in the signal region. After the unblinding, some important (and fame-worthy) checks are quickly done. For instance, to declare that there is evidence of the signal in the signal regions, the data is supposed to be 5σ away from the background, i.e., the likelihood that mere statistical fluctuations in data can explain the excess on top of the background contribution in the signal regions should be 2.866×10^{-7} or approx 1 in 3.48 million.

With the observation and the predictions in hand, a maximum likelihood analysis is done with the background only hypothesis and the signal + background hypothesis. Physics analyses tend to have different aims which can be targeted by the maximum likelihood analysis. For instance, a supersymmetry search analysis will aim to set bounds or improve existing bounds on the masses of the proposed SUSY particles, while a search for a very rare standard model process will want to set a limit on how much more data is required to see its evidence. On the other hand, analyses that measure some parameters would want to do a likelihood estimate using the signal + background hypothesis to check which value of the parameter ultimately maximizes the likelihood (or minimizes the negative log of the likelihood). Computing this likelihood involves taking into account all the uncertainties involved and choosing appropriate error models for them. In addition, a slew of other factors that can influence the analysis itself like binned vs unbinned, asymptotic estimation vs repeated Monte-Carlo simulation of toys, etc also influence the analysis

4.6 Monte-Carlo Simulation of datasets

Simulated datasets play a major role in background estimation and signal region optimization. These simulations are done from first principles and broadly involve three steps : Matrix element based hard scatter simulation, parton shower simulation, and detector response simulation.

The Monte-Carlo simulation method uses random number generators to sample from underlying matrix element squared distributions instead of performing computations to compute the cross sections and branching ratios. With reasonable computational power, a large corpus of simulated collisions can be generated relatively quickly, ultimately having distributions similar to the ones predicted by theory.

Hard scattering simulation is done through packages like `MADGRAPH` [33] or `POWHEG` [34] which calculate the interactions using perturbative expansions. These are usually done to next-to leading order (NLO), which while being reasonably good, does not capture the full physics of the underlying event. The input parton distribution functions of the protons are, however, from experimental results. Soft scattering interactions like parton shower generation from final state quarks, or initial and final state radiations, are not simulated from first principles in a perturbative expansion using event level generators. Software packages like `PYTHIA` [35] which simulate parton showers, hadronization, and initial and final state radiation, till the final particles, corresponding to the state just before detection. An up to date model of the CMS Detector is implemented using `GEANT4` [36] which primarily simulates the interaction of particles with detector material. Using this model, fake measurements are generated for the collision products which are then fed through the data pipelines to produce simulated data that mimics real collision data.

Since an arbitrarily large number of events for any process can be created, they need to be weighed appropriately while using them in analyses so that the final yields are consistent with the luminosity of the data which will be used. Hence, the simulated datasets carry a scale factor that provides the normalized yields to a luminosity of 1fb^{-1}, which then gets multiplied by the luminosity corresponding to the data taking period. In addition, various object level scale factors are also used to account for inaccuracies arising from modelling. Electrons and muons have associated scale factors that are derived for different p_T and η ranges, which are multiplied with the event weights to create an overall a-priori scale factor.

The above mentioned scale factors mainly tend to the inaccuracies arising from the detector simulation stage. Sometimes, additional post-facto scale factors might be required to account for generator level differences between data and simulation. These are derived from "control regions", which are regions in parameter/feature space similar to the signal regions, where the particular process for which the correction factor needs to be derived dominates. The standard method is to derive a simple ratio dependent scale factor by comparing data to simulation.

4.6.1 Fast simulation for multiple signal points

Signal models may sometimes have parameters that need to be determined experimentally. For example, in a supersymmetry signal model the mass of the supersymmetric particles is unknown. In scenarios like this, sufficient signal events for each possible parameter(s) need to be simulated in order to perform a maximum likelihood analysis (Section 4.7). A full simulation of the detector response to the final state particles might be computationally expensive in these cases. A simpler simulation framework called Fast Simulation (fastsim) [37] has been developed to address these scenarios. In a fastsim setup,

the detector geometry is much more simplified, with infinitely thin material layers. The emulation of the material is done by assigning thicknesses in radiation lengths instead of doing a full simulation of the particle interacting with the tracker material. Few more corners are cut with using empirical models for showering in the calorimeters. The entire object reconstruction pipeline is done using the results obtained from the simplified detector model, with some exceptions. These simplifications result in a factor of 20 reduction in time taken to simulate $t\bar{t}$ events (100 seconds with full simulation vs 5 seconds in fast simulation), with the fastsim results reproducing full simulation results to within 10%. These differences are accounted for in the analysis level by using fast sim to full sim scale factors prescribed by the collaboration.

4.7 Maximum Likelihood Analysis

The stated aims of an analysis, like measuring a parameter, or verifying theories, can be recast as a maximum likelihood problem. With parameter estimation, this is straightforward - estimate the requisite parameter that maximizes the likelihood of producing the observed data given an underlying model. When new physics (like supersymmetry) could not be found from a search style analysis, the underlying hypothesis testing problem - testing the background hypothesis vs the background + signal hypothesis against data to come up with bounds on the cross sections of the underlying theories, can also be recast as a maximum likelihood estimation, by estimating the "signal strength" i.e., the fraction of the signal that needs to be added to the background such that the likelihood of rejecting the signal + background hypothesis is greater than a threshold (say 95%). Much of this material has been sourced from Refs. [38] and [39].

4.7.1 Likelihood Function

We shall take a glimpse of the statistical procedures used by the CMS Collaboration in the context of constructing exclusion limits on new physics processes, in conjunction

with the physics analysis in Chapter 5. In such analyses, the unknown parameter of interest is the signal strength μ, accompanied by a set of nuisance parameters specified with the vector $\vec{\theta}$. The analysis is unblinded and the observed data D is seen in the channels, which by itself follows a Poisson distribution.

The Likelihood function is numerically equal to the probability that we have what we have, conditional on the value of the parameters. For a binned analysis like the one in Chapter 5 with one background (for simplicity), the likelihood can be written as

$$\mathcal{L}(\mu, \theta) = \mathbf{f}(\mathcal{D}|\mu, \vec{\theta}) = \prod_{i \in \text{bins}} \text{Pois}(n_i | b_i(\vec{\theta}) + \mu \cdot s_i(\vec{\theta})) \cdot p(\vec{\theta}) \tag{4.5}$$

where $\mathcal{L}(\mu, \theta)$ is the notation for the likelihood function, $\mathbf{f}$ is the probability that the dataset is observed given a strength μ and a nuisance parameter realization θ, and $p(\vec{\theta})$ the joint pdf of the nuisance parameters. The signal and background models $b_i(\vec{\theta}$ and $s_i(\vec{\theta}$ are taken to be deterministic functions of the nuisance parameters. We need to note that the likelihood function is not a probability, though it looks functionally similar to one. The task of maximum likelihood estimation is to find those values of μ and $\vec{\theta}$ that maximize this likelihood in some form. Since there are a lot of products involved which can blow up quickly, and a minimization problem is much straightforward to solve computationally due to extensive research having been done in the field of optimization, we work with the negative log likelihood

$$-\ln \mathcal{L}(\mu, \theta) = -\sum_{i \in \text{bins}} \Big[n_i \ln (b_i + \mu \cdot s_i) - \ln (n_i!) + \ln (p(\vec{\theta})) \Big] \tag{4.6}$$

4.7.2 Profile Likelihood and Hypothesis Testing

Hypothesis testing requires a test statistic based on the maximum likelihood parameters, and a criterion that would accept or reject a hypothesis based on a p-value derived from the test statistic. The profile likelihood ratio used by experiments at the LHC (developed for the Higgs discovery) is given by

$$\tilde{\lambda}(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\theta}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\theta})}, \mu \geq \hat{\mu} \tag{4.7}$$

where $\hat{\hat{\theta}}(\mu)$ is the conditional maximum likelihood estimate of $\vec{\theta}$ for a given value of μ, while $\hat{\mu}$ and $\hat{\theta}$ are the global maximum likelihood estimates for μ and $\vec{\theta}$ respectively. The test statistic q_μ is given by

$$q_\mu = q(\mu) = -2 \ln \tilde{\lambda}(\mu), \mu \geq \hat{\mu} \tag{4.8}$$

For discovery purposes, when we want to check the existence of a fluctuation in data that cannot be explained by background processes alone, we use the statistic

Every time the analysis is repeated (with new fresh data, all the way from scratch), a different value of q_μ will be obtained [2]. The distribution of this value is conditional on the value of the signal strength μ and the nuisance parameters θ, and is represented by $f(q_\mu|\mu, \theta)$. Using this distribution, we can compute a p-value p_μ to check how "far" the observed value (from this instance of the analysis) is, and take a decision. However, a modified frequentist approach is taken by the LHC experiments, with the construction of a quantity called the CL_s [40, 41], given by

$$CL_s = \frac{p_\mu}{1 - p_b} \tag{4.9}$$

[2] Frequentist stats 101

where p_μ and p_b are given by

$$p_\mu = \int_{q_{\mu,\text{obs}}}^{\infty} f(q_\mu|\mu, \hat{\hat{\theta}}(\mu, \text{obs}))dq_\mu \tag{4.10}$$

$$p_b = 1 - \int_{q_{\mu,\text{obs}}}^{\infty} f(q_\mu|0, \hat{\hat{\theta}}(0, \text{obs})dq_\mu \tag{4.11}$$

While setting exclusion boundaries, we solve a reverse problem wherein we find those value of μ such that $CL_s = 0.05$. Signal strengths greater than or equal to this value of μ will be rejected at the 95% level.

The p-value under the background-only hypothesis is computed

4.7.3 Estimating the distribution of the test statistic

The distribution of the test statistic $f(q_\mu|\mu, \theta)$ can be derived in a straightforward manner using Toy Monte-Carlo samples. In this method, the total distribution function **f**, which is functionally similar to the likelihood function, is used to generate fake data. Then the whole computation is done for each of these points to create distributions for a given value of μ. While this is the most accurate method, it is also computationally very expensive, and for scenarios like finding the appropriate value of μ given a p-value, this pretty much implies generates hundreds of distributions for different values of μ (to scan), each containing hundreds of points (to construct the distribution histograms).

When sufficient amount of data is used to construct the likelihood functions, certain asymptotic criteria apply, with which asymptotic functions can be created. In case of a single parameter of interest, like our case, it can be shown that [39]

$$q_\mu = \frac{(\mu - \hat{\mu})}{\sigma^2} + O(1/\sqrt{N}) \tag{4.12}$$

Here, the maximum likelihood estimate μ follows a Gaussian distribution centered at the "true mean" μ' and standard deviation σ, and N is the sample size. The standard deviation itself is obtained from the covariance matrix of parameters $\vec{nu} = (\mu, \vec{\theta})$ from the expression

$$\Sigma_{ij}^{-1} = -E\left[\frac{\partial^2 \mathcal{L}}{\partial \nu_i \partial \nu_j}\right] \tag{4.13}$$

We can then show that the distribution $f(q_\mu|\mu, \theta)$ is a non-central chi-square distribution for one degree of freedom. These asymptotic formulas are quite powerful for searches where a lot of candidate signal points need to be tested.

4.7.4 Testing the background-only hypothesis

Similar to the definition of q_μ in Equation 4.8, we define q_0 as

$$q_0 = -2 \ln \lambda(0) \tag{4.14}$$

since for the background-only hypothesis, the signal strength μ is zero[3]. The hypothesis test here will check for the p-value of a distribution $f(q_0|0, \hat{\hat{\theta}}(0, \text{obs}))$ for the observed value $q_{0,obs}$, producing a p-value condition

$$p_0 = \int_{q_{0,obs}}^{\infty} f(q_0|0, \hat{\hat{\theta}}(0, \text{obs}))dq_0 \tag{4.15}$$

From Equation 4.12, we can construct the asymptotic formula for q_0 to be

$$q_0 = \frac{\hat{\mu}^2}{\sigma^2} \tag{4.16}$$

Under the notion that the hypothesis is true, the special case of $E[\hat{\mu}] = 0$ will apply in which case one can see that $\hat{\mu}/\sigma$ is the expression for a standard normal variable, which

[3]The hypothesis $H_0 : \mu = 0$ creates the quantity q_0

implies that $\sqrt{q_0}$ follows a Gaussian. Hence the p-value can be computed to be

$$p_0 = 1 - \Phi(\sqrt{q_{0,obs}}) \tag{4.17}$$

where Φ is the Gaussian CDF. We can convert the p-value into a σ style number provided by the Gaussian based on $Z_0 = \Phi^{-1}(1 - \sqrt{q_{0,obs}})$. For the background-only hypothesis to be rejected, Z_0 should be at least 5, which will correspond to a p-value of 2.866×10^{-7}.

4.7.5 HiggsCombine

`HiggsCombine` [42] is a statistical analysis software package based on the packages `RooStats` and `RooFit` that was originally developed for the Higgs Discovery analyses. In its current avatar, it is a generic software package that can do anything maximum likelihood related. All it requires are data cards that outline the various processes involved, the predictions and observations in various histogram bins, and the associated uncertainties. Under the `Asymptotic` method, the standard output is the scale factor r that measures limits on the signal strength (same as μ above), using the CL_s criterion at the 5% level. It can run both asymptotic and toy based models for the maximum likelihood fits. This centralized tool makes interpretations simple and straightforward, reduces errors that arise from each analysis implementing their own maximum likelihood process, and hence provides for reproducibility.

4.8 Acknowledgments

This chapter described the basic principles involved with physics analyses with CMS. The figures used in this chapter are obtained from the following sources : "Event displays of top quark pair production in CMS collected in 2015" *https://cds.cern.ch/record/2712881* (2020)

Chapter 5

A search for supersymmetry in events containing a Z boson, jets and missing transverse energy

5.1 Introduction

As seen in Chapter 2, the standard model does not explain a host of phenomena, notably that of dark matter. While models like supersymmetry (SUSY) elegantly extend the standard model, the results of these extensions need to be experimentally verified. The material covered in this chapter is taken from the paper[15] to which the author contributed, and the related internal analysis note [16].

We present a search for the production of supersymmetric particles that follow a decay chain that involves the production of a Z boson, in addition to a supersymmetric particle that cannot decay further (called the Lightest Supersymmetric Particle). We use data collected over the entirety of Run 2 spanning over three years (2016, 2017, 2018), which amounts of 137 fb^{-1}. We interpret the results in the context of the simplified mass spectra (SMS) models, and in the case that the we do not see evidence of these processes, we set upper limits on the masses of the yet-to-be-observed particles.

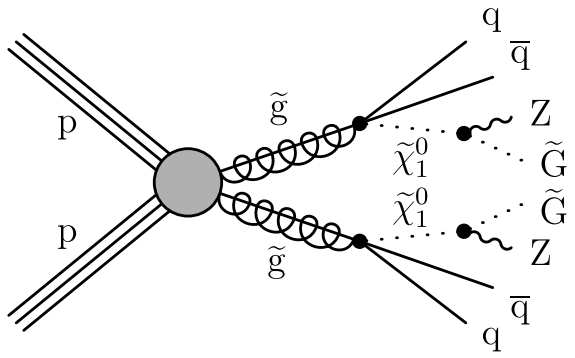
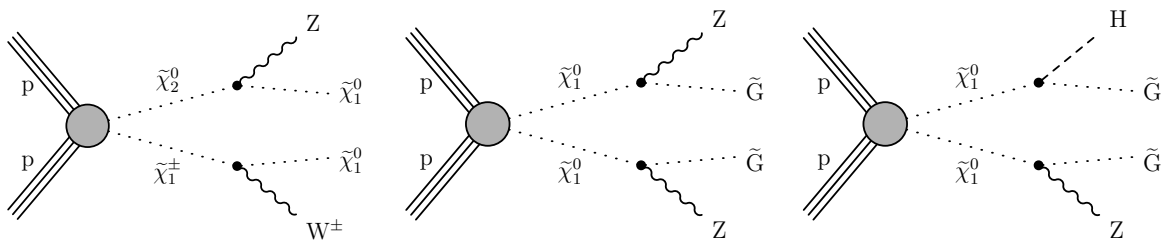


Figure 5.1. T5ZZ gluino production based SUSY model

We enumerate the target signal processes and confounding background processes, chalk out signal regions where the signal sensitivity (signal-background ratio) is maximized, enumerate the background yields and the expected signal yields in these signal regions, and then check the observed data against our predictions and form various conclusions based on the results.

5.2 Signal Processes

Figures 5.1 and 5.2 show the four target simplified SUSY models of this analysis. All of them are characterized by the presence of Z bosons and a pair of lightest supersymmetric particles (LSPs). Since LSPs being dark matter candidates cannot be detected via electromagnetic means, we expect that they will manifest in the form of missing transverse momentum. The T5ZZ model (Figure 5.1) is characterized by pair production of gluinos ($\tilde{\text{g}}$), which are the SUSY analogs of the gluons. Each gluino decays into a pair of quarks and a neutralino ($\tilde{\chi}_1^0$). We target the particualr final state where the neutralinos decay into Z bosons and gravitinos ($\tilde{\text{G}}$), which happens to be the lightest supersymmetric particle. To ensure a balance between abundance of events in the selection regions, while maintaining a good signal sensitivity, we require that one of the two Z bosons produced decays into



(a) TChiWZ model **(b)** TChiZZ model **(c)** TChiHZ model

Figure 5.2. Electroweakino production based SUSY models

two leptons, while the other Z boson decays hadronically. In addition, we expect that the gravitinos carry a significant amount of momentum, which will manifest in the form of a large quantity of missing transverse energy.

The three electroweakino production based models (Figure 5.2) are labelled TChiWZ, TChiZZ and TChiHZ. The TChiWZ model involves the production of a neutralino and chargino decaying into a Z boson and a W boson respectively, along with the LSP. We target the final state where the W boson decays into hadronic jets. In the TChiZZ and TChiHZ models, two neutralinos are produced, decaying to neutral bosons. In case of TChiZZ, We only target the final state where one Z decays into two leptons, with the other decaying hadronically. In the TChiHZ case, one of the neutralinos can decay into a Higgs Boson. One thing to note here is that while it is possible for the TChiZZ alone to exist in reality (without TChiHZ), the converse is not true, i.e., if the TChiHZ process exists in reality, the TChiZZ process should also exist as a neutralino that can decay into a Higgs Boson can also decay into a Z boson. This means that when we consider the TChiHZ process, we assume that the neutralino can decay into a Z or a Higgs with a branching fraction of 50%.

5.3 Object selections

Our analysis targets a di-lepton final state that reconstructs a leptonically Z boson. The fundamental requirement is that we need two well measured well isolated leptons and no additional leptons.

5.3.1 Electrons and Muons

Since we are dealing with a Z to dilepton decay, we want both the leptons to have the same flavor (di-electon or di-muon)[1]. In addition, we look for those leptons that pass the tight MVA ID cuts recommended by the CMS collaboration as we want them to be well measured. We also look for one lepton to have $p_T > 25$ GeV (designated the leading lepton), and the other lepton to have p_T between 20 GeV and the p_T of the leading lepton (designated the sub-leading lepton). In case of multiple candidates, the two highest p_T leptons are used. On the η front, we require the leptons to be well within the central regions ($|\eta| < 2.4$). We also skip the region ($1.4 < |\eta| < 1.6$) since this is the barrel to endcap transition region. The parent Z candidate is required to have a p_T of at least 55 GeV. To ensure the two leptons are well separated from each other in $\eta - \phi$ space, we require that the ΔR between these two be at least 0.1. For the isolation criterion, we use the mini relative isolation parameter.

Mini Relative Isolation (MiniRelIso)

We use the Mini Relative Isolation (Section 4.1.2) criterion to measure the isolation for the leptons. In this scenario, a lower isolation value means a better isolated lepton. We require this value to be less than 0.1 for the electrons and less than 0.2 for the muons.

We also require $|\Delta_{xy}| < 0.05$ cm, $|\Delta_z| < 0.1$ cm and $SIP_{3D} < 4$ to ensure the leptons are produced from the primary vertex.

[1]For the flavor symmetric background estimation (Section 5.5.1), we relax the same flavor requirement

5.3.2 Jets

Since we deal with both AK4 and AK8 jets, we have separate pre-selection criteria for them

AK4 Jets

We require jets to have $p_T > 35$ GeV, and $|\eta| < 2.4$. In addition, to prevent signal leptons clustered as jets from getting counted as jets, we only consider jets that are $\Delta R > 0.4$ away from the signal leptons. For b-tagging, we use the DeepCSV b-tag algorithm and use the working points provided by the collaboration. In some signal regions (Section 5.4.4 where a b-veto is needed, we lower the threshold to 25 GeV to be more conservative.

AK8 Jets

We require the AK8 fat jets to have $p_T > 200$ GeV, and $|\eta| < 2.4$. We also require them to be away from the signal leptons, this time with a $\Delta R > 0.8$ requirement. We also require the soft-drop mass to be between 65 and 105 GeV to ensure that the jet is produced from the decay of a W boson, in addition to requiring specific τ_{21} cuts

5.3.3 Photons

For the Drell-Yan background estimation (Section 5.5.2) we require single photon events (γ + Jets), for which we have specific selection criteria for the photons. Since the photon will mimic the Z boson candidate, we require that it has a p_T of at least 55 GeV, and $|\eta| < 2.4$. In fact, these requirements which were brought about by the triggers (Section 5.3.6) is what motivated us to have the di-lepton p_T and η requirement in the lepton pre-selections above (Section 5.3.1). Since we are more interested in the hadronic recoil, we are not very interested in the purity of the photon sample, and hence require very loose criteria, like not having a matching pixel track (pixel veto), and being $\Delta R > 0.2$

away from electrons that are at least 10 GeV in p_T. The latter cut is to reject conversions from electrons in W decays which are accompanied by real E_T^{miss}. We also require basic isolation cuts and require that the photon be away from the E_T^{miss} vector by 0.4 in ϕ, i.e., $|\Delta\phi_{\gamma,E_T^{miss}}| > 0.4$

5.3.4 Veto Leptons

As specified in this section, we require only two well measured leptons, and no additional leptons. We want this criteria to be as strong as possible, so we need to veto anything that remotely looks like an additional leptons. This and the next subsection (5.3.5) that talks about isotracks are objects meant for this very purpose.

For the veto leptons, we require them to have $p_T > 10$ GeV, $|\eta| < 2.4$, and pass only the loose MVA ID requirements provided by the collaboration in case of electrons, and medium ID requirements in case of muons. We also require the mini relative isolation (MiniRelIso) bto be less than 0.4

5.3.5 Isotracks

We require isotracks to have $p_T > 5$ GeV in case of leptons, and > 10 GeV in case of charged hadrons, and $|\eta| < 2.4$. Here we require the relative isolation to be $< min(0.2, 5/p_T)$

5.3.6 Triggers

As we saw in Section 3.4, triggers play a vital role in pre-selecting and filtering events passing some broad general criteria. We require events that pass the di-electon, di-muon and muon-electron High-Level Triggers (HLTs). In addition, we require events that pass the single photon triggers for the Drell-Yan background estimation (Section 5.5.2) and single muon triggers to check Monte-Carlo (MC) Modelling.

5.4 Signal Regions

5.4.1 Preselections

Since we target the Z boson decaying into two leptons, we require that the final state contains exactly two well measured same flavor leptons and well-isolated leptons, and no additional "veto leptons" or isotracks. In addition, we also require that the di-lepton mass be between 86 and 96 GeV (called the Z mass window). Since our analysis targets regions with relatively high $\mathrm{E_T^{miss}}$, we have a minimum $\mathrm{E_T^{miss}}$ cut of 50 GeV, mainly to remove low $\mathrm{E_T^{miss}}$ events that occur due to small mis-measurements. To ensure events containing $\mathrm{E_T^{miss}}$ due to jet mis-measurement do not enter the signal regions, we have the additional requirement that the $\mathrm{E_T^{miss}}$ direction be at least 0.4 radians in ϕ away from both the leading jets. We also veto any τ leptons with $\mathrm{p_T} > 20$ GeV.

5.4.2 The $\mathrm{M_{T2}}$ variable

Processes like the $\mathrm{t\bar{t}}$ production and decay have two possible sources of real $\mathrm{E_T^{miss}}$, each arising from the decay of a W boson (Figure 5.3a). In order to eliminate background contributions from $\mathrm{t\bar{t}}$, it becomes quite important to come up with a quantity that can help us cut out such W pair di-lepton decay events. The $\mathrm{M_{T2}}$ variable [43] is a powerful tool for this purpose, which is computed as

$$\mathrm{M_{T2}} = \min_{\vec{\mathrm{p}}_\mathrm{T}^{\mathrm{miss}(1)}+\vec{\mathrm{p}}_\mathrm{T}^{\mathrm{miss}(2)}=\vec{\mathrm{p}}_\mathrm{T}^{\mathrm{miss}}} \left[\max\left(m_\mathrm{T}^{(1)}, m_\mathrm{T}^{(2)}\right)\right], \tag{5.1}$$

where $\vec{\mathrm{p}}_\mathrm{T}^{\mathrm{miss}(i)}$ $(i = 1, 2)$ represent the two possible $\mathrm{E_T^{miss}}$ vectors arising from each of the W decays, and the $m_\mathrm{T}^{(i)}$ are the transverse masses that are obtained by pairing the i th lepton with $\vec{\mathrm{p}}_\mathrm{T}^{\mathrm{miss}(i)}$. This quantity exhibits an endpoint at the W boson mass) for the $\mathrm{t\bar{t}}$ events. Hence, a requirement that the $\mathrm{M_{T2}}$ be greater than the W boson mass is used in the signal regions.

5.4.3 Strong Signal Regions

Six orthogonal signal regions are formulated to target the T5ZZ process 5.1. They are split by the presence or absence of b-tagged jets, and in jet multiplicity, with SRA requiring only 2 or 3 jets, SRB requiring 4-5 jets, and SRC requiring 6 and higher. In addition, a high hadronic activity is required, as evidenced by the high H_T cuts. The M_{T2} cuts are used to limit contributions from the $t\bar{t}$ process (Section 5.4.2). The regions are split into several E_T^{miss} bins which are optimized to find a balance between enhancing the signal sensitivity (which warrant finer binning), and ensure each bin has enough statistics (which warrant coarser binning). These regions are listed in Table 5.1.

Table 5.1. Summary of strong signal region selections.

Region	N_{jets}	$N_{b\text{-}jet}$	H_T(GeV)	$M_{T2}(\ell\ell)$(GeV)	E_T^{miss} binning (GeV)
SRA b-veto	2–3	$= 0$	> 500	> 80	[100,150,230,300,∞]
SRB b-veto	4–5	$= 0$	> 500	> 80	[100,150,230,300,∞]
SRC b-veto	≥ 6	$= 0$	-	> 80	[100,150,250,∞]
SRA b-tag	2–3	≥ 1	> 200	> 100	[100,150,230,300,∞]
SRB b-tag	4–5	≥ 1	> 200	> 100	[100,150,230,300,∞]
SRC b-tag	≥ 6	≥ 1	-	> 100	[100,150,250,∞]

5.4.4 Electroweak Signal Regions

Three signal regions are formulated to target the three electroweakino mediated processes (Figure 5.2). The boosted and resolved VZ signal regions are meant to target the different kinematic regimes of the TChiWZ and TChiZZ processes. The SRVZResolved signal region is created for cases in which the hadronizing W or Z boson does not have a large boost, and hence the jets can be resolved. Hence we require the di-jet mass to be less than 110 GeV (targeting the W and Z decay processes), and jet multiplicity to be at least 2. In addition, we veto events with any b-tagged jets having $p_T > 25$ GeV. The SRVZBoosted region is meant to target that part those scenarios where the hadronizing

boson is boosted and hence decays into a collimated fat jet. Hence we require at least one fat jet with $p_T > 200$ GeV with the modified requirement that this jet be at least 0.8 radians in ϕ away from the fat jet. To prevent overlaps between these two regions, only events that fail the SRVZResolved requirement are checked if they pass the requirements for the boosted VZ signal region. The specially constructed HZ signal region targeting the TChiHZ process, where the Higgs decays into two b-jets. This signal region requires exactly two b-tagged jets, with the di-b-jet mass ¡ 150 GeV (to target the Higgs at 125 GeV). In addition, a special extension of the M_{T2} variable, where the transverse mass also includes the b tagged jet in addition to the lepton (called $M_{T2}(\ell b \ell b)$) is used, with the cut at 200 GeV, mainly to cut the top quark production process (ref Figure 5.3a), the top quark having a mass of 173 GeV. Table 5.2 lists these signal regions

Table 5.2. Summary of Electroweak signal region selections. Events in the SRVZBoosted region are required not to pass the requirements of the SRVZResolved region, in order to remove their overlap.

Region	N_{jets}	N_{fatjets}	$N_{\text{b-tags}}$	dijet mass (GeV)	M_{T2} (GeV)	E_T^{miss} binning (GeV)
SRVZBoosted	< 2	≥ 1	$= 0$	-	$--$	[100,200,300,400,500,∞]
SRVZResolved	≥ 2	–	$= 0$	$M_{jj} < 110$	$M_{T2}(\ell\ell) > 80$	[100,150,250,350,∞]
SRHZ	≥ 2	–	$= 2$	$M_{b\bar{b}} < 150$	$M_{T2}(\ell b \ell b) > 200$	[100,150,250,∞]

5.5 Background processes and estimation

The signal signature consists of two leptons that reconstruct a Z boson and a large missing transverse momentum. We broadly classify the background processes into three categories

1. Real E_T^{miss} + "fake" Z boson : The main contributors here are flavor symmetric processes like $t\bar{t}$ production (Figure 5.3a). The two leptons are produced from different vertices, but when we add up their relativistic momenta and compute the

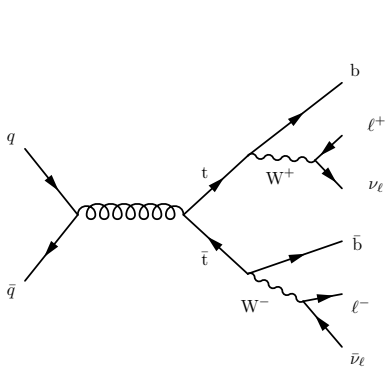
mass of the resulting fake parent, it falls in the Z mass window (86-96 GeV). The source of the E_T^{miss} here comes from the neutrinos that are produced from the decay of the W bosons. These backgrounds are called flavor Symmetric backgrounds because the leptons are produced from different vertices, and hence independently have an equal probability of being an electron or a muon.

2. Fake E_T^{miss} + real Z boson : The main contributor here comes from the Drell-Yan process (Figure 5.3b), where a Z boson is produced which decays into two leptons. The E_T^{miss} here is produced due to mis-measurement of the energies of the hadronic recoil quark jets.

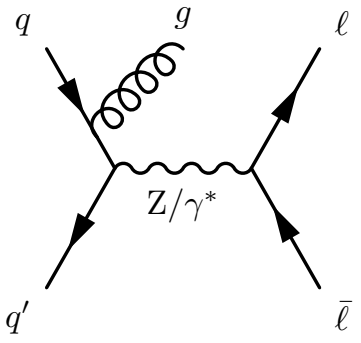
3. Real E_T^{miss} + Real Z boson : The main contributors here are real multi-lepton production processes. For example, in a WZ process (Figure 5.3c), the Z boson decays into two leptons, while the W boson decays into a lepton and a neutrino. If the lepton escapes detection, we can see a significant amount of E_T^{miss} produced in association with two leptons.

5.5.1 Flavor Symmetric Background Estimation

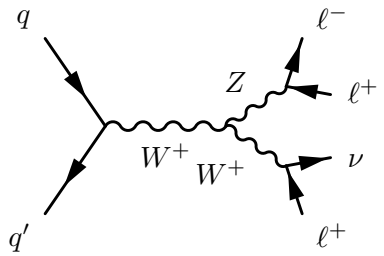
An important feature of the flavor symmetric backgrounds like the $t\bar{t}$ process (Figure 5.3a) is that the leptons come from different vertices, so the probability that a lepton can be an electron or a muon is independent of the other lepton. This means that on average, the number of same flavor (SF, $ee + \mu\mu$) background events from these processes in the signal regions should be the same as the number of different flavor (DF, $e\mu$) events. We take this feature to our advantage and devise a method to estimate the contribution using the data itself, instead of relying on imperfectly simulated Monte-Carlo (MC) samples.



(a) Fake Z + Real E_T^{miss}: $t\bar{t}$



(b) Real Z + Fake E_T^{miss}: Drell-Yan



(c) Real Z + Real E_T^{miss}: WZ

Figure 5.3. Background processes

However, the number of $e\mu$ events is not exactly equal to the number of $ee + \mu\mu$ events due to differences in reconstruction, identification and trigger efficiencies. . Hence, we define a transfer factor $R_{SF/DF}$ that accounts for these differences between the DF and SF events. We parametrize this factor into a product of two terms, one that relies only on the reconstruction efficiency differences, and the other that relies only on the trigger efficiencies. A detailed explanation of the derivation can be found in Appendix A[2].

For each signal region, we create a corresponding DF control region. Due to the paucity of entries within the di-lepton mass cut region (86 GeV $< M_{\ell\ell} <$ 96 GeV), we relax this requirement to $M_{\ell\ell} > 20$ GeV, use a transfer factor $\kappa = \frac{N(86<M_{\ell\ell}<96)}{N(M_{\ell\ell}>20)}$ to scale this estimate to the control regions. This factor κ is derived in different regions in parameter

[2]Thanks to our analysis collaborator Marius Terörde

space and used in the DF control regions. To summarize, the flavor symmetric contribution to the background can be written as

$$N_{SF} = N_{DF} \times \kappa \times R_{SF/DF} \tag{5.2}$$

κ estimation and validation

κ is derived in the following regions (without the E_T^{miss} binning)

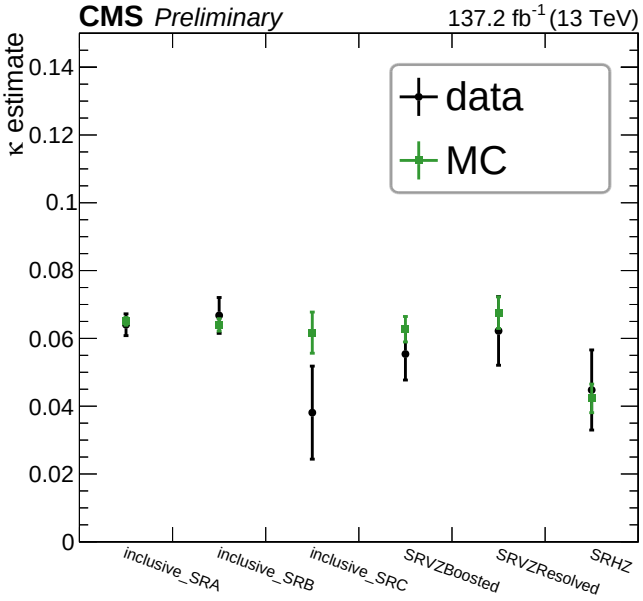
- Inclusive SRA - union of SRA b-veto and SRA b-tag
- Inclusive SRB - union of SRB b-veto and SRB b-tag
- Inclusive SRC - union of SRC b-veto and SRC b-tag
- SRVZResolved signal region
- SRVZBoosted signal region, with the AK4 jet veto relaxed
- SRHZ signal region

Since we have essentially ignored the b-tag vs b-veto, and the E_T^{miss} bins, we have additional regions where this effect is measured. We first define a "baseline" (at least 2 jets, $\Delta\phi(j_{12}, E_T^{miss})$ cuts, and $M_{T2} > 80$ GeV), and add additional cuts depending on the additional regions, which we call the "uncertainty estimation" regions.

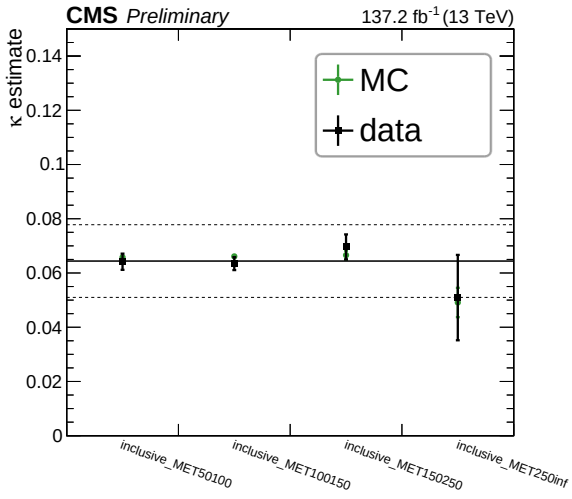
- Inclusive_MET50100 - baseline + 50 GeV$\leq E_T^{miss} \leq$ 100 GeV
- Inclusive_MET100150- baseline + 100 GeV$\leq E_T^{miss} \leq$ 150 GeV
- Inclusive_MET150250- baseline + 150 GeV$\leq E_T^{miss} \leq$ 250 GeV
- Inclusive_MET250Inf- baseline + $E_T^{miss} \geq$ 250 GeV
- Strong b-tag - baseline + at least one medium b-tagged jet

- Strong b-veto - baseline + medium b-jet veto

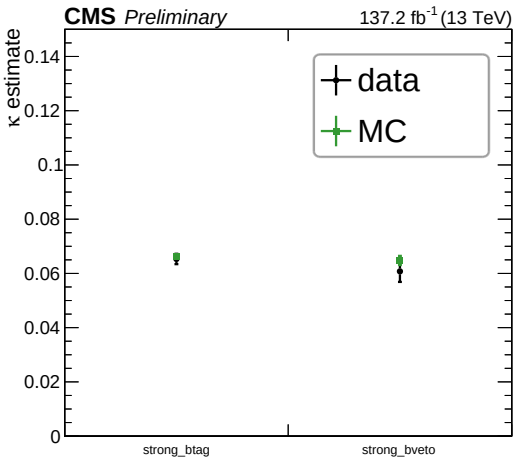
We derive a separate value of κ for each of the six regions mentioned above using Monte-Carlo samples corresponding to the flavor symmetric processes and in data. We then use the uncertainty estimation regions to derive systematic uncertainties to the κ values in addition to the statistical errors of the κ estimates.



(a) κ in six regions



(b) E_T^{miss} Uncertainty regions



(c) b-tag and b-veto regions

Figure 5.4. κ and associated uncertainty estimation

Figure 5.4 shows the κ estimates in the signal region, and derivation of the associated uncertainties in the E_T^{miss} and b-tag/b-veto uncertainty regions. There is good data-MC agreement in five of the six regions. For the inclusive SRC region, due to differences between data and Monte-Carlo estimates, we add the difference between the lower end of the Monte-Carlo estimate and the upper end of the data uncertainty bar to the statistical error. For the E_T^{miss} associated uncertainty, we fit the central value and derive the uncertainty to ensure the central values of all the E_T^{miss} bins are within this range. This uncertainty is our E_T^{miss} uncertainty. For the b-tag/b-veto region, the uncertainty is sub-dominant to the E_T^{miss} uncertainty, so we do not consider it.

Table 5.3 summarizes the results. The associated E_T^{miss} uncertainty obtained from Figure 5.4b is estimated to be 20.4%.

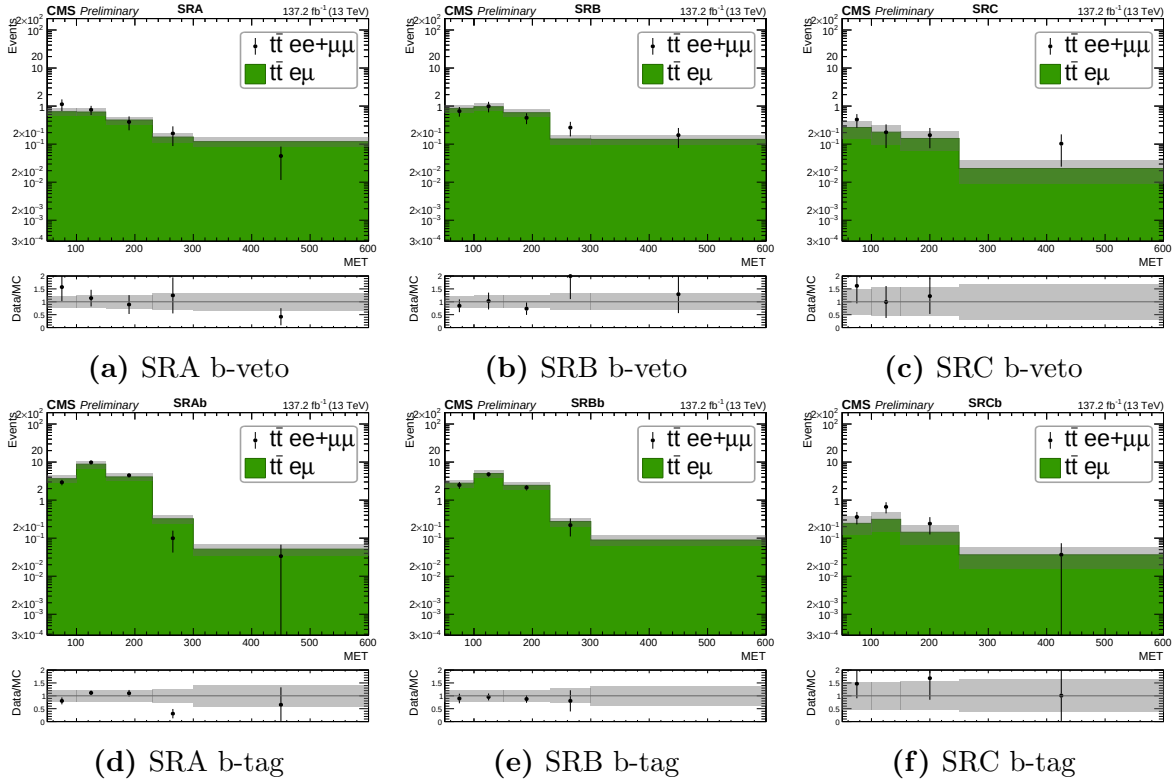
Table 5.3. The measured values of κ in the signal and the inclusive regions. The uncertainties here are the statistical uncertainties.

Region	κ value
Inclusive SRA	0.064 ± 0.003
Inclusive SRB	0.067 ± 0.005
Inclusive SRC	0.038 ± 0.018
SRVZBoosted	0.055 ± 0.007
SRVZResolved	0.062 ± 0.010
SRHZ	0.045 ± 0.012
Inclusive MET_50100	0.064 ± 0.003
Inclusive MET_100150	0.063 ± 0.002
Inclsuive MET_150250	0.070 ± 0.005
Inclusive MET_250Inf	0.051 ± 0.016
Strong b-tag	0.065 ± 0.002
Strong b-veto	0.061 ± 0.004

Closure test

We test our flavor symmetric estimation procedure using a closure test with flavor symmetric Monte-Carlo samples. The SF events ($ee + \mu\mu$) act as the "truth" and the

estimate obtained by the aforementioned method acts as the "prediction". The results of this test can be found in Figures 5.5 and 5.6, where we see good agreement between the $ee + \mu\mu$ distribution and the $e\mu$ based estimate.



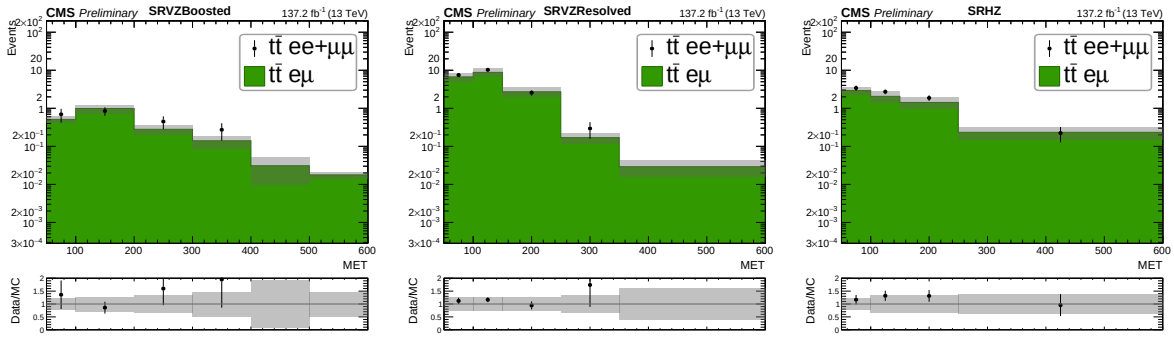
(a) SRA b-veto **(b)** SRB b-veto **(c)** SRC b-veto

(d) SRA b-tag **(e)** SRB b-tag **(f)** SRC b-tag

Figure 5.5. Flavor symmetric background estimation closure test in the strong signal regions

5.5.2 Drell-Yan Background Estimation

The common signature of the Drell-Yan contribution to the backgrounds is that the Z boson is real, while the associated $\mathrm{E_T^{miss}}$ is due to mis-measurement of the hadronic recoil (the gluon in Figure 5.3b). We use a data driven estimation of this background using γ+Jets data, since we are only interested in the mis-measurement of the jets in the recoil. A caveat with this method is that there is also a contribution from processes such as $W\gamma$ production, where the high $\mathrm{E_T^{miss}}$ contribution comes both from the neutrino and the lepton that was not measured. This is appropriately treated using Monte-Carlo



(a) SRVZBoosted **(b)** SRVZResolved **(c)** SRHZ

Figure 5.6. Flavor symmetric background estimation closure test in the electroweak signal regions

samples.

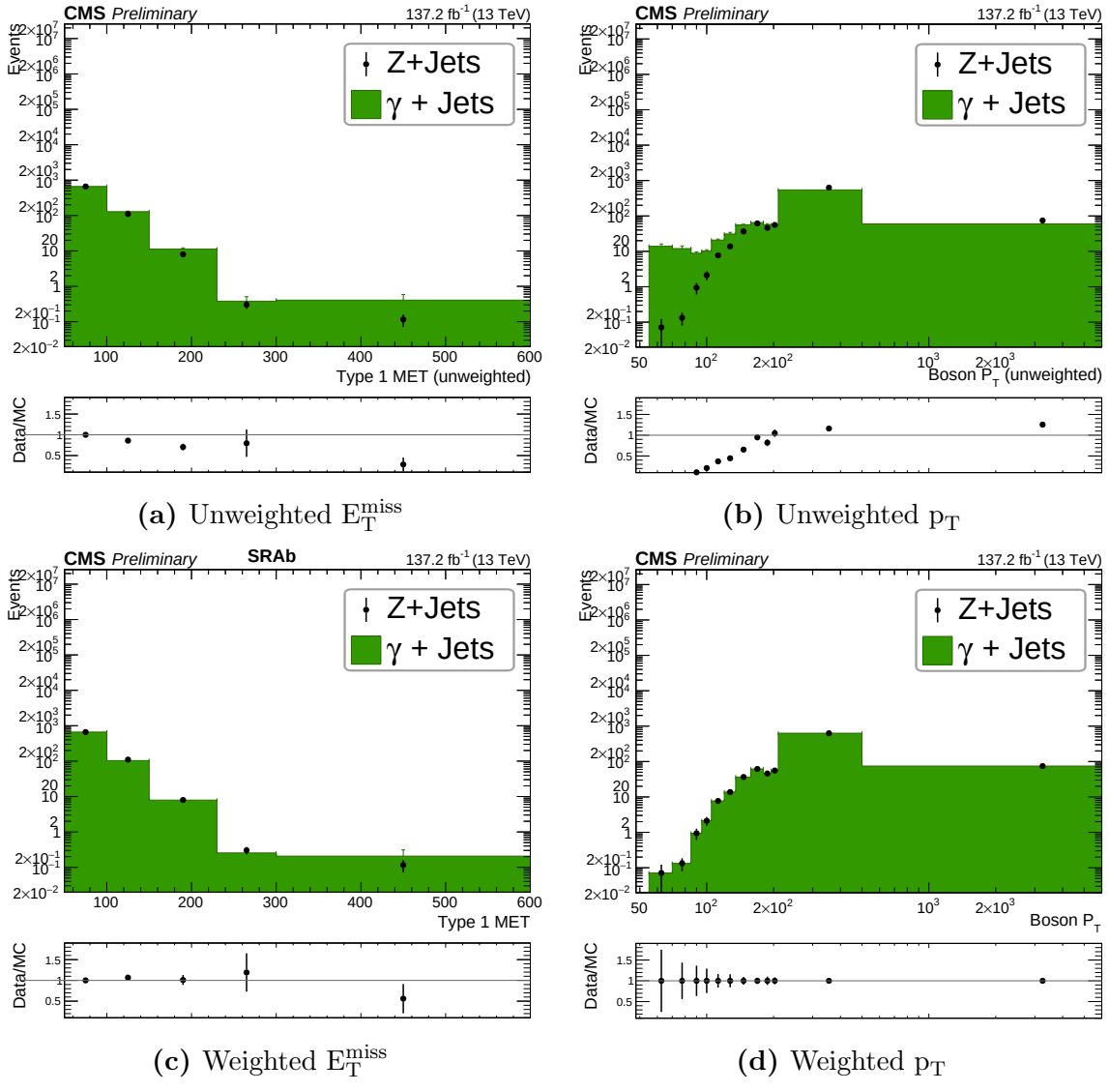
We pre-select events containing photon like objects summarized in Section 5.3.3. For each signal region, we define a control region where we will estimate the background from γ+Jets data. To estimate the M_{T2} for the control region cuts, we devise a simulation based method described below. Since the kinematics of a Z boson production is quite different from that of a photon production, we derive a set of weights to be applied to each event such that the Z boson p_T distribution shape is matched to the photon p_T distribution shape. These weights are derived in Monte-Carlo samples. To account for the difference in event yields due to different cross sections, we normalize the γ+Jets background such that this along with the other backgrounds matches the observed collision data in the signal regions in the 50-100 GeV E_T^{miss} bin. The background estimation in the signal region can be condensed as

$$N_{DY} = \mathcal{N} \cdot (N_{\gamma}^{'} - C \cdot N_{W\gamma}) \tag{5.3}$$

where $N_{\gamma}^{'}$ refers to the weighted γ+Jets data in the control region, $N_{W\gamma}$ the estimate of the Wγ tail and C its associated scale factor, and $\mathcal{N}$ the normalization factor derived from the 50-100 GeV E_T^{miss} bin.

Weights derivation

In the photon control regions (that correspond to the signal regions), we make distributions of the boson p_T using Drell-Yan and γ+Jets Monte-Carlo samples. For each p_T bin, we derive weights which are then applied to the γ+Jets data to derive the weighted estimate. An illustration of the steps is shown in Figure 5.7, where the effects on the E_T^{miss} distribution can be seen before and after reweighting for one signal region.



(a) Unweighted E_T^{miss} **(b)** Unweighted p_T

(c) Weighted E_T^{miss} **(d)** Weighted p_T

Figure 5.7. Distributions before and after reweighting

$M_{T2}(\ell\ell)$Simulation for the E_T^{miss} Templates

Since we require two leptons to compute the M_{T2}, we simulate the photon decaying to two visible leptons. We assume that the mother of the two simulated leptons has the mass of a Z boson (91 GeV) and the p_T of the photon. We obtain the angular distributions by sampling from a distribution uniformly distributed in $1 + cos^2(\theta)$, where θ is the angle in the frame where the mother particle is at rest. Once the two lepton directions are obtained this way, we compute the M_{T2} using these simulated leptons.

Templates closure

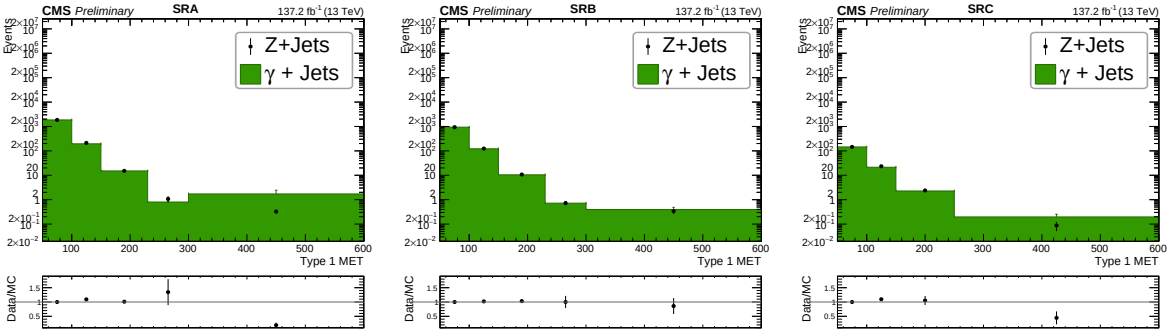
We perform a closure test in the control regions using Drell-Yan and γ+Jets Monte-Carlo samples to validate the weighting procedure and derive systematic uncertainties. These uncertaintes are computed bin-by-bin and is obtained as the maximum of the statistical uncertainty and the non-closure (i.e., the deviation of the ratio from 1). The results are shown in Figure 5.8 for the strong signal regions and Figure 5.9 for the electroweak signal regions, and tabled in Table 5.4.

Electroweak contamination in the E_T^{miss} template tails

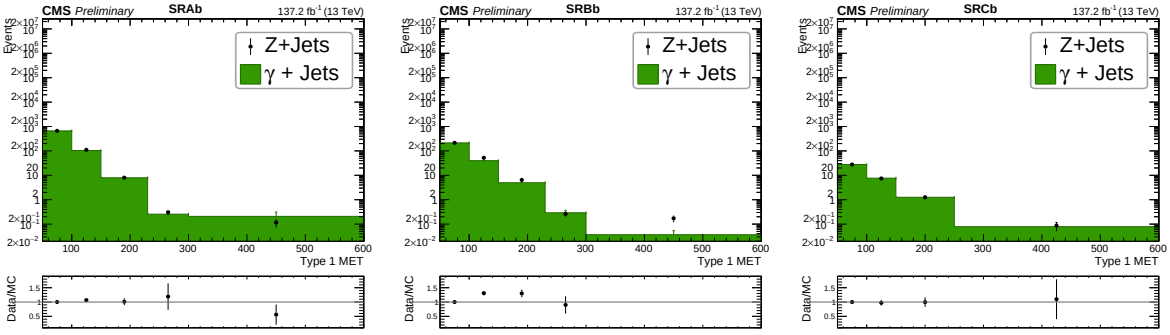
Processes like W+γ, W + Jets [3], etc where the W boson decays leptonically and the lepton is not measured, contribute to high values of E_T^{miss} which can manifest itself in the tails of the E_T^{miss} distribution in the photon control regions. These need to be accounted for, as seen in Equation 5.3. To check the Monte-Carlo modelling of these processes, and measure the scale factor C and systematic uncertainties, we choose a control region where these electroweak processes dominate. Such a control region has the following selections.

- Exactly one muon, $p_T > 25$ GeV
- Exactly one photon, $p_T > 55$ GeV

[3] the photon can be embedded in the jets, for instance

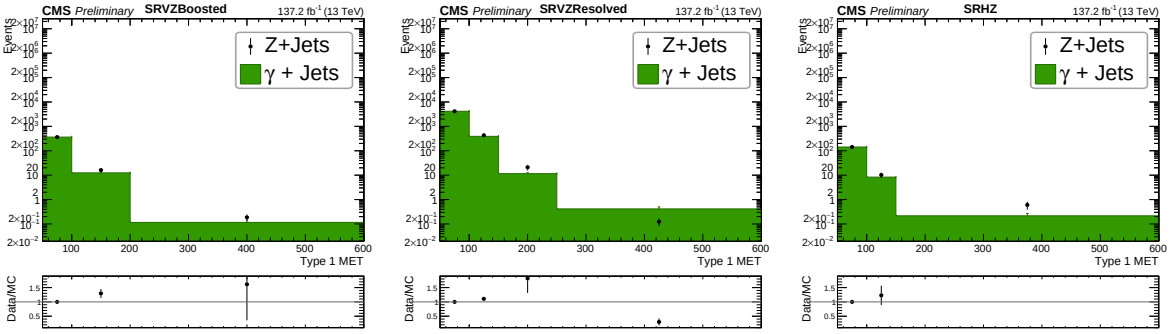


(a) SRA b-veto **(b)** SRB b-veto **(c)** SRC b-veto



(d) SRA b-tag **(e)** SRB b-tag **(f)** SRC b-tag

Figure 5.8. Templates closure plots in the control regions corresponding to the strong signal regions



(a) SRVZBoosted **(b)** SRVZResolved **(c)** SRHZ

Figure 5.9. Templates closure plots in the control regions corresponding to the electroweak signal regions

- $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}} > 50$ GeV
- $N_{\mathrm{jets}} \geq 2$
- $M_T(\mu, \mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}) > 30$ GeV

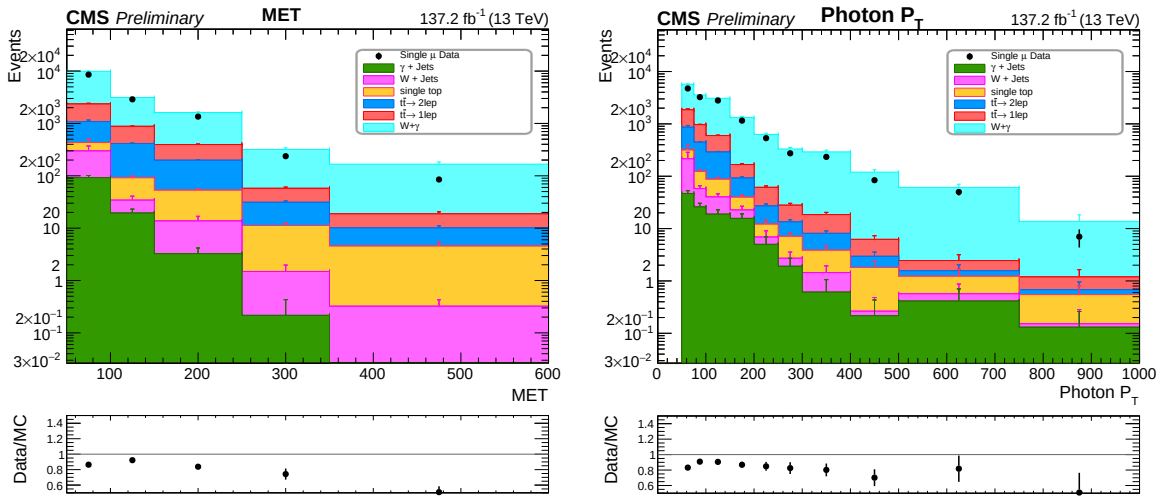
Table 5.4. Results of the MC closure test shown for all signal regions. All uncertainties in this table are statistical only.

SR	E_T^{miss} Bin (GeV)	Z Jets	γ + jets	Ratio
SRA b-veto	50.0-100.0	1854.33 ± 7.13	1854.33 ± 6.69	1.00 ± 0.01
	100.0-150.0	194.60 ± 2.15	212.39 ± 2.07	1.09 ± 0.02
	150.0-230.0	15.06 ± 0.59	15.21 ± 0.48	1.01 ± 0.05
	230.0-300.0	0.80 ± 0.14	1.08 ± 0.31	1.34 ± 0.45
	300.0+	1.72 ± 0.73	0.33 ± 0.07	0.19 ± 0.09
SRA b-tag	50.0-100.0	662.35 ± 8.55	662.35 ± 11.71	1.00 ± 0.02
	100.0-150.0	103.58 ± 3.07	110.53 ± 3.77	1.07 ± 0.05
	150.0-230.0	7.93 ± 0.65	8.01 ± 0.69	1.01 ± 0.12
	230.0-300.0	0.25 ± 0.08	0.30 ± 0.07	1.19 ± 0.46
	300.0+	0.21 ± 0.10	0.12 ± 0.04	0.56 ± 0.35
SRB b-veto	50.0-100.0	936.00 ± 4.60	936.00 ± 4.90	1.00 ± 0.01
	100.0-150.0	121.67 ± 1.53	124.26 ± 1.62	1.02 ± 0.02
	150.0-230.0	10.43 ± 0.43	10.76 ± 0.40	1.03 ± 0.06
	230.0+	1.11 ± 0.14	1.06 ± 0.13	0.95 ± 0.17
SRB b-tag	50.0-100.0	213.41 ± 3.10	213.41 ± 4.30	1.00 ± 0.02
	100.0-150.0	39.85 ± 1.18	52.16 ± 2.47	1.31 ± 0.07
	150.0-230.0	4.97 ± 0.34	6.47 ± 0.48	1.30 ± 0.13
	230.0+	0.33 ± 0.08	0.43 ± 0.08	1.32 ± 0.39
SRC b-veto	50.0-100.0	144.93 ± 1.95	144.93 ± 3.52	1.00 ± 0.03
	100.0-150.0	21.31 ± 0.65	23.37 ± 1.01	1.10 ± 0.06
	150.0-250.0	2.27 ± 0.21	2.39 ± 0.29	1.05 ± 0.16
	250.0+	0.20 ± 0.06	0.09 ± 0.04	0.45 ± 0.23
SRC b-tag	50.0-100.0	27.66 ± 0.88	27.66 ± 1.00	1.00 ± 0.05
	100.0-150.0	7.64 ± 0.45	7.42 ± 0.59	0.97 ± 0.10
	150.0-250.0	1.26 ± 0.15	1.25 ± 0.15	1.00 ± 0.17
	250.0+	0.08 ± 0.04	0.09 ± 0.04	1.10 ± 0.70
SRVZBoosted	50.0-100.0	363.55 ± 7.71	363.55 ± 11.65	1.00 ± 0.04
	100.0+	12.54 ± 1.06	16.29 ± 1.34	1.30 ± 0.15
SRVZResolved	50.0-100.0	4151.26 ± 39.42	4151.26 ± 46.75	1.00 ± 0.01
	100.0-150.0	389.81 ± 12.31	431.50 ± 15.13	1.11 ± 0.05
	150.0-250.0	11.54 ± 0.99	21.03 ± 5.56	1.82 ± 0.51
	250.0+	0.41 ± 0.09	0.12 ± 0.04	0.30 ± 0.12
SRHZ	50.0-100.0	142.41 ± 4.03	142.41 ± 8.71	1.00 ± 0.07
	100.0-150.0	8.32 ± 1.27	10.21 ± 2.33	1.23 ± 0.34
	150.0+	0.22 ± 0.05	0.60 ± 0.21	2.80 ± 1.20

- $\Delta\phi(j_{12}, E_T^{miss}) > 0.4$
- lepton and isolated track vetos for additional leptons and hadrons

The E_T^{miss} and Photon p_T distributions are shown in Figure 5.10. The $\gamma\mu$ data is off by around 30% on average from the Monte-Carlo predictions. This is due to imperfect Monte-Carlo modelling, and hence we assign the value of 0.7 to the scale factor C. The systematic uncertainty in this method is taken to be 30% of the total electroweak MC contribution in the control regions, i.e., $0.3 \times N_{W\gamma}$. There are some bins where this prediction can make the total Drell-Yan prediction negative, i.e., this can be higher than

the γ+Jets contribution, in which case those bins are set to zero.



(a) E_T^{miss} distribution

(b) Photon p_T distribution

Figure 5.10. Distributions of E_T^{miss} and Photon p_T in the electroweak process enriched single muon control region. The W+Jets sample only consists of those events in which the photon is a non-prompt one, while the W+γ consists only of prompt photon events

Systematic Uncertainties in the Templates prediction

The systematic uncertainties of the E_T^{miss} template prediction have the following sources

- The statistical uncertainty of the data γ + jets sample in each E_T^{miss} bin in the signal regions
- The systematic uncertainty arising from reweighting estimated in closure
- The normalization of the prediction in the 50-100 GeV E_T^{miss} bin
- The systematic uncertainty from the Electroweak subtraction in the E_T^{miss} tails

A complete summary of the Drell-Yan background estimate in each signal region along with the systematic uncertanties is summarized in Table 5.5.

Table 5.5. Summary of template predictions in all analysis bins together with the uncertainty from each source. The "ratio" number for each uncertainty shows that particular uncertainty divided by the total template uncertainty for a given bin.

SR	E_T^{miss} Bin	Prediction	Closure (ratio)	Normalization (ratio)	Statistical (ratio)	EWK Sub (ratio)
SRA b-veto	50-100	1284.02 ± 40.55	0.00 (0.00)	38.20 (0.94)	13.24 (0.33)	0.00 (0.00)
	50-100	1282.05 ± 40.92	0.00 (0.00)	38.17 (0.93)	13.22 (0.32)	0.00 (0.00)
	100-150	160.28 ± 17.02	14.59 (0.86)	4.77 (0.28)	6.30 (0.37)	1.26 (0.07)
	150-230	22.55 ± 4.83	1.15 (0.24)	0.67 (0.14)	3.56 (0.74)	1.21 (0.25)
	230-300	1.07 ± 0.88	0.47 (0.54)	0.03 (0.04)	0.18 (0.20)	0.51 (0.58)
	300+	3.04 ± 3.14	2.46 (0.78)	0.09 (0.03)	0.70 (0.22)	1.20 (0.38)
SRA b-tag	50-100	641.13 ± 28.47	0.00 (0.00)	26.78 (0.94)	8.73 (0.31)	0.00 (0.00)
	100-150	108.14 ± 9.91	7.25 (0.73)	4.52 (0.46)	4.39 (0.44)	0.53 (0.05)
	150-230	13.43 ± 2.82	1.61 (0.57)	0.56 (0.20)	1.61 (0.57)	0.84 (0.30)
	230-300	2.06 ± 1.46	0.95 (0.65)	0.09 (0.06)	0.65 (0.44)	0.35 (0.24)
	300-6001	1.24 ± 1.05	0.54 (0.52)	0.05 (0.05)	0.36 (0.34)	0.44 (0.42)
SRB b-veto	50-100	731.21 ± 31.45	0.00 (0.00)	29.10 (0.93)	10.76 (0.34)	0.00 (0.00)
	100-150	110.77 ± 7.47	2.33 (0.31)	4.41 (0.59)	4.88 (0.65)	0.33 (0.04)
	150-230	12.53 ± 2.20	0.71 (0.32)	0.50 (0.23)	1.58 (0.72)	0.54 (0.24)
	230-300	0.89 ± 0.62	0.15 (0.24)	0.04 (0.06)	0.19 (0.30)	0.36 (0.58)
	300+	1.03 ± 0.95	0.17 (0.18)	0.04 (0.04)	0.31 (0.33)	0.50 (0.53)
SRB b-tag	50-100	230.43 ± 16.60	0.00 (0.00)	15.84 (0.95)	4.53 (0.27)	0.00 (0.00)
	100-150	51.35 ± 16.61	15.82 (0.95)	3.53 (0.21)	3.16 (0.19)	0.52 (0.03)
	150-230	11.35 ± 3.99	3.43 (0.86)	0.78 (0.20)	1.46 (0.37)	0.44 (0.11)
	230-300	1.72 ± 1.09	0.67 (0.61)	0.12 (0.11)	0.55 (0.50)	0.24 (0.22)
	300+	0.48 ± 0.55	0.19 (0.34)	0.03 (0.06)	0.12 (0.21)	0.25 (0.46)
SRC b-veto	50-100	140.51 ± 14.80	0.00 (0.00)	13.27 (0.90)	5.96 (0.40)	0.00 (0.00)
	100-150	28.46 ± 5.33	2.73 (0.51)	2.69 (0.50)	3.31 (0.62)	0.03 (0.01)
	150-250	1.91 ± 0.64	0.30 (0.47)	0.18 (0.28)	0.31 (0.48)	0.37 (0.57)
	250+	0.54 ± 0.50	0.30 (0.60)	0.05 (0.10)	0.00 (0.00)	0.23 (0.46)
SRC b-tag	50-100	41.77 ± 7.25	0.00 (0.00)	6.83 (0.94)	2.22 (0.31)	0.00 (0.00)
	100-150	9.38 ± 2.07	0.89 (0.43)	1.53 (0.74)	0.92 (0.44)	0.12 (0.06)
	150-250	2.52 ± 0.91	0.42 (0.46)	0.41 (0.45)	0.50 (0.55)	0.24 (0.26)
	250+	0.00 ± 0.15	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
SRVZBoosted	50-100	329.22 ± 21.31	0.00 (0.00)	19.62 (0.92)	7.46 (0.35)	0.00 (0.00)
	100-200	15.46 ± 5.35	4.61 (0.86)	0.92 (0.17)	2.02 (0.38)	0.49 (0.09)
	200-300	0.00 ± 0.29	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
	300-400	0.00 ± 0.13	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.04 (0.28)
	400-500	0.02 ± 0.10	0.01 (0.07)	0.00 (0.01)	0.02 (0.24)	0.03 (0.27)
	500+	0.10 ± 0.13	0.03 (0.24)	0.01 (0.05)	0.07 (0.57)	0.02 (0.18)
SRVZResolved	50-100	4066.01 ± 83.44	0.00 (0.00)	73.21 (0.88)	35.96 (0.43)	0.00 (0.00)
	100-150	448.14 ± 51.84	47.95 (0.92)	8.07 (0.16)	14.97 (0.29)	4.96 (0.10)
	150-250	27.21 ± 23.33	22.36 (0.96)	0.49 (0.02)	2.74 (0.12)	4.87 (0.21)
	250-350	1.88 ± 2.48	1.31 (0.53)	0.03 (0.01)	0.42 (0.17)	1.53 (0.62)
	350+	1.76 ± 1.91	1.23 (0.64)	0.03 (0.02)	0.52 (0.27)	0.79 (0.41)
SRHZ	50-100	177.20 ± 15.59	0.00 (0.00)	14.36 (0.92)	5.39 (0.35)	0.00 (0.00)
	100-150	11.77 ± 4.39	3.98 (0.91)	0.95 (0.22)	1.34 (0.30)	0.20 (0.05)
	150-250	1.51 ± 2.78	2.71 (0.98)	0.12 (0.04)	0.44 (0.16)	0.12 (0.04)
	250-6001	0.22 ± 0.43	0.39 (0.90)	0.02 (0.04)	0.11 (0.25)	0.04 (0.10)

5.5.3 Real Z + E_T^{miss} Background Estimation

The main contributors to this background are di-lepton processes like WZ, ZZ, $t\bar{t}$ Z, and tri-lepton processes like ZZZ, WWZ etc. In the case of WZ (Figure 5.3c) for instance, the Z boson leptonically decays while the lepton from the W boson decay is not detected,

resulting in a huge E_T^{miss} (along with real E_T^{miss} from the neutrino). For the contributions from these processes, we use the simulated Monte-Carlo samples. To prevent overlap with the other data-driven methods, we only choose those events in the simulated samples with prompt electrons or muons consistent with the Z boson decay, and having the same Z parent, along with at least one prompt neutrino.

To check the accuracy of Monte-Carlo modelling and measure the data-MC scale factors, we validate our samples in control regions with three and four leptons, where these processes will contribute the most. From these regions we derive scale factors and uncertainties that arise from imperfect Monte-Carlo modelling of our samples.

Three and Four Lepton Control Regions[4]

We consider three and four lepton control regions that specifically target the ZZ, WZ and $t\bar{t}Z$ processes.We consider all possible opposite charge same flavor pairs made from leptons that pass the cuts. For the four lepton control sample, we require two such pairs, and the pairs are ordered by their proximity to the Z mass, with the one closer to the Z mass being the "primary Z candidate" and the other being the "secondary Z candidate". In addition, we require these regions to have at least two jets.

For the WZ three lepton control region, we have a b-jet veto with $p_T > 25$ GeV, $E_T^{miss} > 70$ GeV, and the transverse mass of the third lepton to be greater than 55 GeV. For the $t\bar{t}Z$ three lepton control region, we require at least two b-jets that pass the medium working points, $E_T^{miss} > 30$ GeV, and the separation in phi $\Delta\phi(j_{12}, E_T^{miss}) > 0.4$.

[4]Thanks to our analysis collaborator Dr. Sergio Sanchez Cruz

For the ZZ four lepton control region, we require four leptons, and no b-jets. We also require that the second di-lepton pair have a mass $M_{\ell\ell} > 40$ GeV to reduce contributions from low mass resonances and suppress combinatorial backgrounds from mismatched leptons.

The results of this exercise are shown in Figure 5.11. Table 5.6 shows the yields computed in the $86 < M_{\ell\ell} < 96$ GeV from which the scale factors are computed. These scale factors are derived individually for the three years (2016, 2017, 2018), since the Monte-Carlo production was done yearly, i.e., all the Monte-Carlo samples for a given year follow a given set of parameters, which are then tweaked for the production run for the subsequent years. The control regions do not isolate the diboson processes alone, though they dominate these regions. Hence to compute the scale factors, we need to subtract the yields from the other "background" processes from the data and divide this by the di-boson Monte-Carlo yields. After scaling these processes appropriately, we validate them in the same control regions by constructing distributions of other quantities of interest and checking the overall ratio. The results of this verification on the distribution of the $\mathrm{E_T^{miss}}$ is shown in Figure 5.12.

We take a 30% uncertainty on WZ and $\mathrm{t\bar{t}Z}$, a 50% uncertainty on ZZ to take into account the non-agreement of data with the predictions. We take a 50% uncertainty on all other rare samples used.

5.5.4 Validation Regions

As an overall check, and more specifically, a check on the performance of the data driven Drell-Yan background estimation process, we create a set of validation regions where

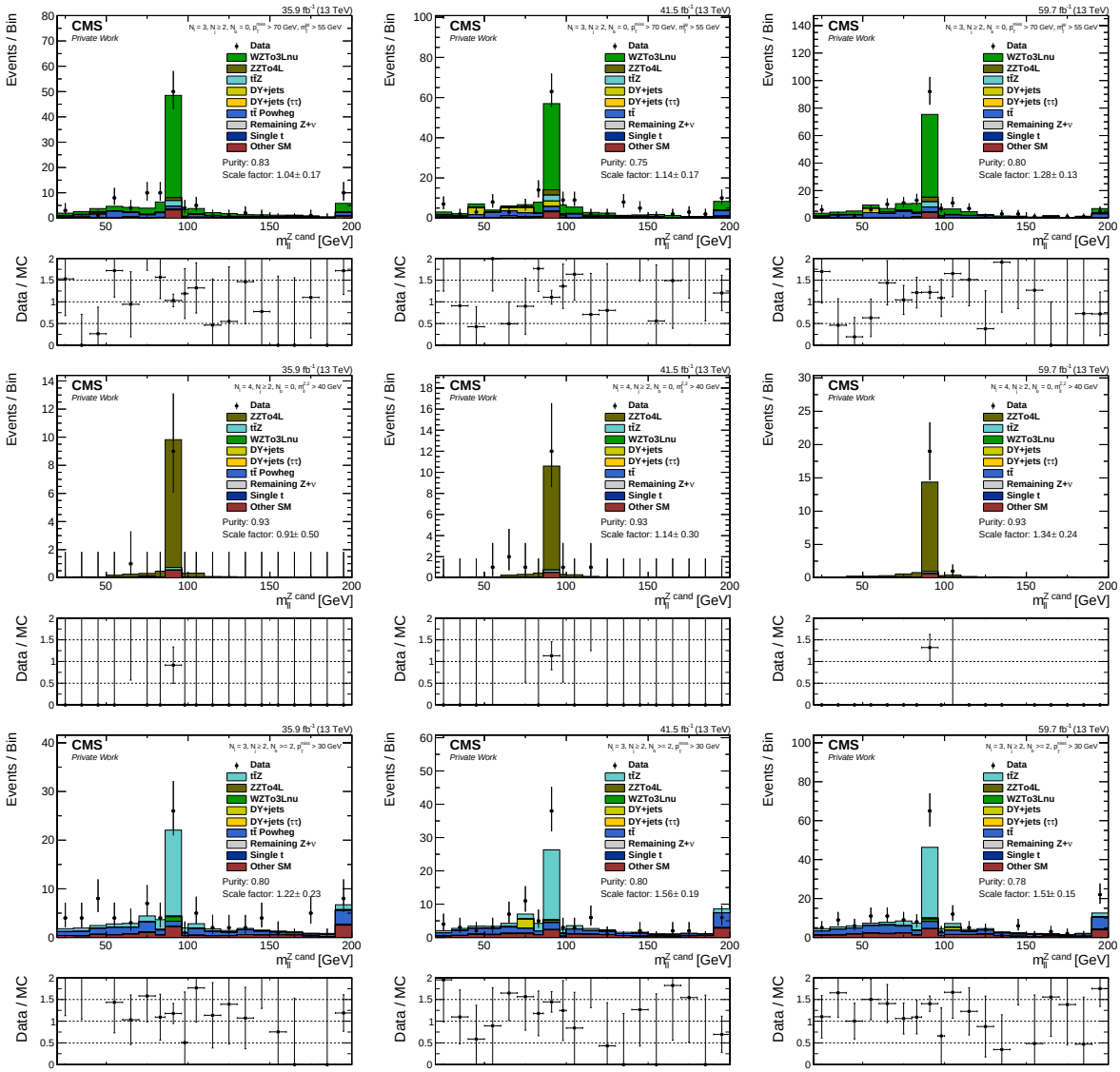


Figure 5.11. Distributions of $M_{\ell\ell}$ in the 3-lepton (WZ) control region (top), $M_{\ell\ell}$ in the 4-lepton (ZZ) control region (middle) and $M_{\ell\ell}$ in the $\mathrm{t\bar{t}Z}$ control region (bottom). From left to right: 2016, 2017, and 2018 data sets. Thanks to our analysis collaborator Dr. Sergio Sanchez Cruz

we are confident that there will be no signal and compare the overall data-background distribution here. These regions are obtained by inverting the $\Delta\phi(\mathrm{j}_{12}, \mathrm{E}_{\mathrm{T}}^{\mathrm{miss}})$ cuts. This ensures that these validation regions are completely orthogonal to the signal region, due to the fact that the $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ produced in this case will most likely be due to mis-measurement of jets. We repeat the methods used to estimate the backgrounds in the signal regions

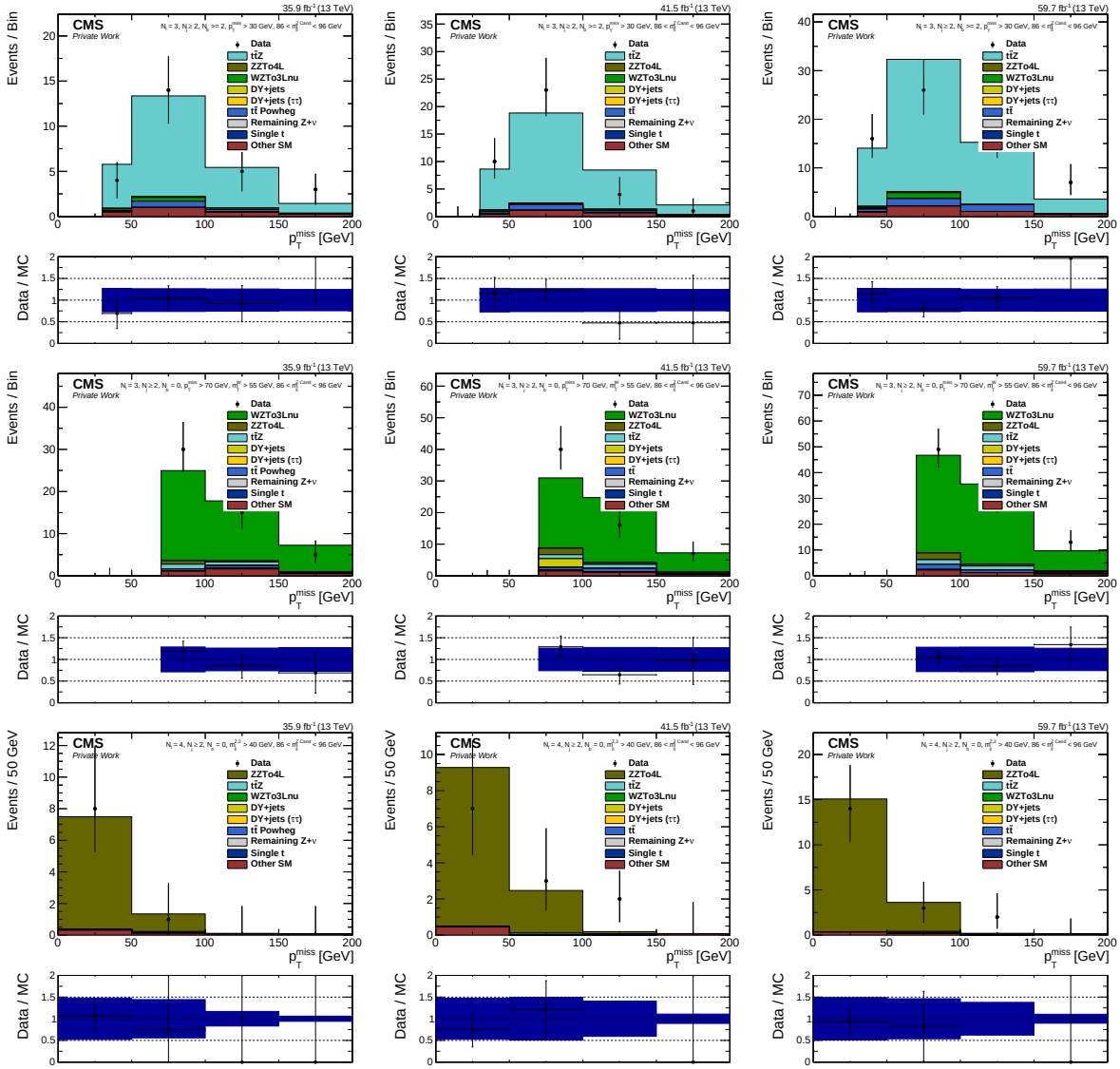


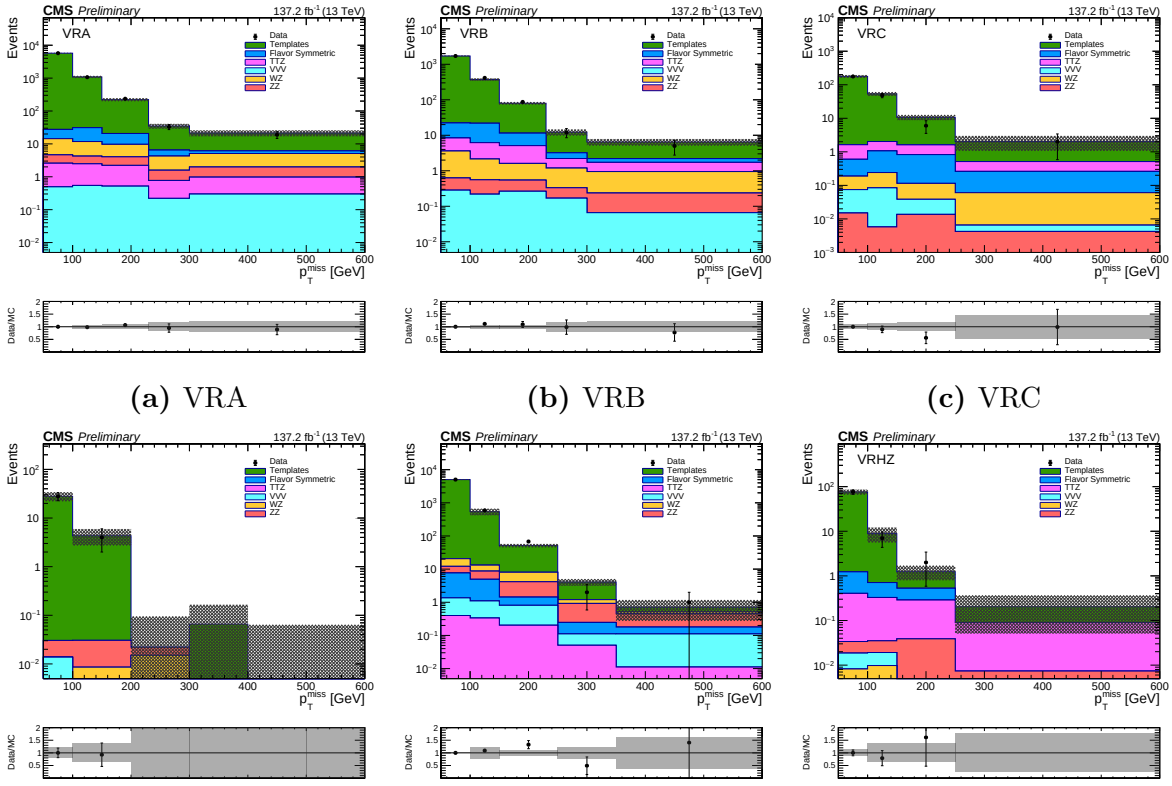
Figure 5.12. Distributions of E_T^{miss} in the WZ (top), ttZ (middle) and ZZ (bottom) control regions, split by year. From left to right : 2016, 2017, 2018. Thanks our analysis collaborator Dr. Sergio Sanchez Cruz

Table 5.6. 3-lepton, 4-lepton and $t\bar{t}Z$ control regions. Signal MC is WZ→3lν in the 3-lepton region, ZZ→4l in the 4-lepton region and $t\bar{t}Z$ →2l2ν in the $t\bar{t}Z$ region. All uncertainties are statistical uncertainties. Thanks to our analysis collaborator Dr.Sergio Sanchez Cruz

		WZ CR	ZZ CR	ttZ CR
2016	signal MC	40.44 ± 0.61	9.09 ± 0.30	17.67 ± 17.67
	background MC	8.05 ± 0.47	0.74 ± 0.13	4.39 ± 0.30
	data	**50**	**9**	**26**
	data-bkg.	41.95 ± 7.03	8.26 ± 4.11	21.61 ± 5.06
	(data-bkg.)/sig.	1.04 ± 0.17	0.91 ± 0.50	1.22 ± 0.23
2017	signal MC	42.92 ± 2.93	9.86 ± 0.09	20.92 ± 20.92
	background MC	14.09 ± 2.64	0.74 ± 0.15	5.38 ± 0.57
	data	**63**	**12**	**38**
	data-bkg.	48.91 ± 7.46	11.26 ± 3.41	32.62 ± 6.11
	(data-bkg.)/sig.	1.14 ± 0.17	1.14 ± 0.30	1.56 ± 0.19
2018	signal MC	60.23 ± 2.90	13.43 ± 0.12	36.30 ± 36.30
	background MC	15.15 ± 0.78	0.94 ± 0.15	10.01 ± 0.88
	data	**92**	**19**	**65**
	data-bkg.	76.85 ± 9.54	18.06 ± 4.36	54.99 ± 7.99
	(data-bkg.)/sig.	1.28 ± 0.13	1.34 ± 0.24	1.51 ± 0.15

in these validation regions. One caveat is that we do not split the regions by b-tag and b-veto. This means that we only have three strong validation regions called VRA (merger of SRA b-veto and SRA b-tag with inverted $\Delta\phi(\mathrm{j}_{12}, \mathrm{E}_{\mathrm{T}}^{\mathrm{miss}})$ cuts), VRB (merger of SRB b-veto and SRB b-tag with inverted $\Delta\phi(\mathrm{j}_{12}, \mathrm{E}_{\mathrm{T}}^{\mathrm{miss}})$ cuts), and VRC (merger of SRC b-veto and SRC b-tag with inverted $\Delta\phi(\mathrm{j}_{12}, \mathrm{E}_{\mathrm{T}}^{\mathrm{miss}})$ cuts), and three electroweak validation regions namely VRWZBoosted (SRVZBoosted with inverted $\Delta\phi(\mathrm{j}_{\mathrm{fat}}, \mathrm{E}_{\mathrm{T}}^{\mathrm{miss}})$), VRWZResolved (SRVZResolved with inverted $\Delta\phi(\mathrm{j}_{12}, \mathrm{E}_{\mathrm{T}}^{\mathrm{miss}})$) and VRHZ (SRHZ with inverted $\Delta\phi(\mathrm{j}_{12}, \mathrm{E}_{\mathrm{T}}^{\mathrm{miss}})$). The $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ distribution plots for the validation regions are shown in Figure 5.13.

We see a good agreement between the observed data and the predictions in the validation regions, while the VRWZBoosted region suffers from a lack of statistics.



(a) VRA **(b)** VRB **(c)** VRC

(d) VRWZBoosted **(e)** VRWZResolved **(f)** VRHZ

Figure 5.13. E_T^{miss} distributions in the validation regions

5.6 Results

The results of the analysis in the signal regions are shown in Figures 5.14 for the strong signal regions and 5.15 for the electroweak signal regions. Tables 5.7 and 5.8 tabulate the results in the E_T^{miss} bins in the strong and electroweak signal regions respectively. We see no significant deviation from the standard model.

In the strong signal regions, the 150-230 GeV E_T^{miss} bin in the SRA b-tag region has the largest disagreement where 31.4 ± 3.8 events are expected and 42 events are observed, corresponding to a local significance of 1.4 standard deviations. In the electroweak signal regions, the last ($E_T^{miss} > 350$ GeV) bin in the SRVZResolved region where 6.3 ± 2.2 events are expected and 2 events are observed, corresponding to a local significance of 1.2 standard deviations.

5.7 Interpretation

While we do not observe any significant deviation that could be indicative of the presence of processes beyond the standard model, we can interpret them in the context of the simplified mass spectrum models. Our objective is to now set some bounds on the masses of the potential supersymmetric particles in the context of the signal processes, assuming they get discovered at some point.

5.7.1 Signal Samples

Signal data samples for various mass points of the supersymmetric particles in the the four signal processes (T5ZZ, TChiWZ, TChiZZ, TChiHZ) are simulated using the Fastsim method (Section 4.6.1). The T5ZZ model has two free parameters - the mass of the gluino and the mass of the neutralino, with the mass of the gravitino taken to be 1 GeV (lightest supersymmetric particle). In the TChiWZ case, the two free parameters are the masses of the $\tilde{\chi}_2^0$ (or the $\tilde{\chi}_1^\pm$) and the $\tilde{\chi}_1^0$. In the TChiZZ and TChiHZ cases, the only free parameter is the mass of the $\tilde{\chi}_1^0$. In the T5ZZ case, the fast simulation samples contain events with the mass of the gluino simulated in 50 GeV intervals from 800 to 1800 GeV, and the mass fo the $\tilde{\chi}_1^0$ also in 50 GeV intervals from 0 to 1600 GeV. In the TChiWZ case, the mass of the $\tilde{\chi}_2^0/\tilde{\chi}_1^\pm$ particle has been simulated in 25 GeV intervals from 100 to 1000 GeV, while the mass of the $\tilde{\chi}_1^0$ is also simulated in 25 GeV intervals from 0 to 600

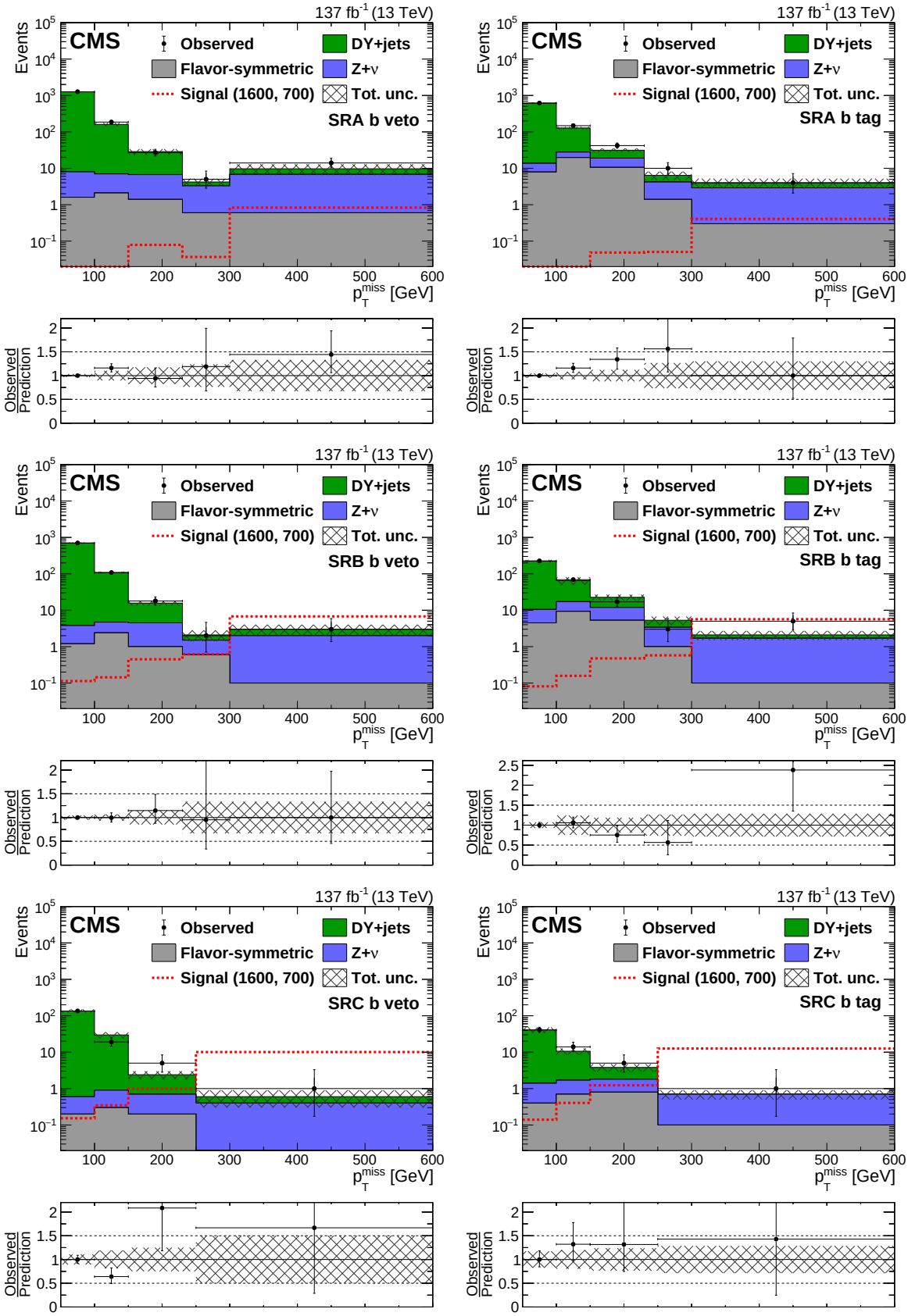


Figure 5.14. Results of the search in the strong signal regions

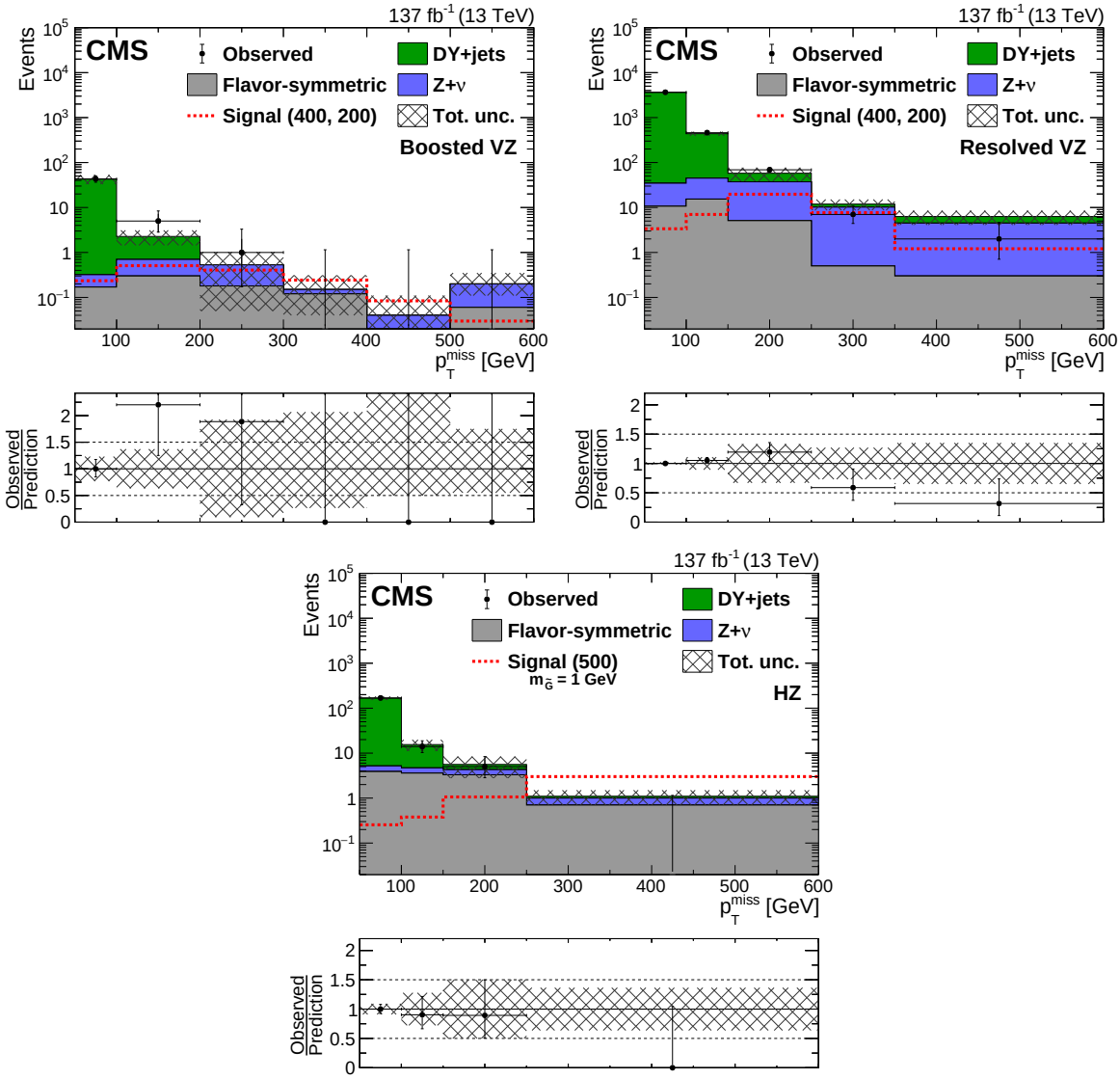


Figure 5.15. Results of the search in the electroweak signal regions

Table 5.7. Predicted and observed event yields in the strong signal regions for each E_T^{miss} bin as defined in Table 5.1. Uncertainties include both statistical and systematic sources. The E_T^{miss} template prediction in each SR is normalized to the 50-100 GeV bin in data

Signal Region	SM processes	E_T^{miss} Bins [GeV]				
SRA b-veto		50–100	100–150	150–230	230–300	>300
	Drell-Yan	1253 ± 41	153 ± 16	22.0 ± 4.9	0.9 ± 0.8	2.9 ± 3.0
	Flavor-symmetric	1.6 ± 0.5	2.1 ± 0.6	1.4 ± 0.5	0.6 ± 0.3	0.6 ± 0.2
	Z+ν	6.4 ± 1.2	4.9 ± 0.9	5.3 ± 1.0	2.7 ± 0.5	6.2 ± 1.2
	Total background	1261 ± 41	160 ± 16	28.8 ± 5.0	4.2 ± 1.0	9.6 ± 3.2
	Observed	1261	186	27	5	14
SRA b-tag		50–100	100–150	150–230	230–300	>300
	Drell-Yan	602 ± 28	99.9 ± 9.3	12.3 ± 2.6	2.2 ± 1.6	1.1 ± 1.0
	Flavor-symmetric	7.9 ± 1.8	19.7 ± 4.4	10.6 ± 2.4	1.4 ± 0.4	0.3 ± 0.2
	Z+ν	5.8 ± 0.9	8.1 ± 1.2	8.4 ± 1.2	2.8 ± 0.5	2.6 ± 0.6
	Total background	616 ± 28	128 ± 10	31.4 ± 3.8	6.3 ± 1.7	4.1 ± 1.2
	Observed	616	148	42	10	4
SRB b-veto		50–100	100–150	150–230	230–300	>300
	Drell-Yan	696 ± 31	103.6 ± 7.1	11.2 ± 2.1	0.6 ± 0.6	1.0 ± 0.9
	Flavor-symmetric	1.2 ± 0.4	2.4 ± 0.7	$1.0^{+0.3}_{-0.4}$	0.6 ± 0.3	$0.1^{+0.2}_{-0.1}$
	Z+ν	2.6 ± 0.5	2.3 ± 0.4	3.5 ± 0.6	0.9 ± 0.2	1.9 ± 0.4
	Total background	700 ± 31	108.2 ± 7.1	15.7 ± 2.3	2.2 ± 0.7	3.0 ± 1.0
	Observed	700	108	18	2	3
SRB b-tag		50–100	100–150	150–230	230–300	>300
	Drell-Yan	215 ± 16	48 ± 16	10.7 ± 3.8	1.9 ± 1.3	0.4 ± 0.5
	Flavor-symmetric	$4.5^{+1.1}_{-1.2}$	9.3 ± 2.2	5.3 ± 1.3	$1.0^{+0.3}_{-0.4}$	$0.1^{+0.2}_{-0.1}$
	Z+ν	6.0 ± 1.1	7.9 ± 1.4	6.6 ± 1.2	2.4 ± 0.4	1.6 ± 0.3
	Total background	225 ± 16	65 ± 16	22.7 ± 4.2	5.3 ± 1.4	2.1 ± 0.6
	Observed	225	69	17	3	5
SRC b-veto		50–100	100–150	150–250	>250	
	Drell-Yan	135 ± 14	28.8 ± 5.6	1.7 ± 0.5	0.2 ± 0.2	
	Flavor-symmetric	0.2 ± 0.1	0.3 ± 0.2	0.2 ± 0.1	$0.0^{+0.1}_{-0.0}$	
	Z+ν	0.4 ± 0.1	0.6 ± 0.2	0.5 ± 0.2	0.4 ± 0.1	
	Total background	135 ± 14	29.7 ± 5.6	2.4 ± 0.6	0.6 ± 0.3	
	Observed	135	19	5	1	
SRC b-tag		50–100	100–150	150–250	>250	
	Drell-Yan	39.6 ± 7.1	8.9 ± 2.0	2.0 ± 0.7	0.0 ± 0.2	
	Flavor-symmetric	0.4 ± 0.3	0.7 ± 0.4	0.8 ± 0.5	0.1 ± 0.1	
	Z+ν	1.0 ± 0.2	1.0 ± 0.2	1.0 ± 0.2	0.6 ± 0.2	
	Total background	41.0 ± 7.1	10.7 ± 2.1	3.8 ± 0.9	0.7 ± 0.2	
	Observed	41	14	5	1	

GeV. In the TChiZZ case, the mass of the $\tilde{\chi}_1^0$ has been simulated from 125 GeV to 1000 GeV and in the TChiHZ case, simulated from 125 GeV to 1300 GeV.

Table 5.8. Predicted and observed event yields in the Electroweak signal regions for each E_T^{miss} bin as defined in Table 5.2. Uncertainties include both statistical and systematic sources. The E_T^{miss} template prediction in each SR is normalized to the 50-100 GeV bin in data.

Signal Regions	SM processes	E_T^{miss} Bins [GeV]					
SRVZBoosted		50–100	100–200	200–300	300–400	400–500	>500
	Drell-Yan	42.7 ± 9.9	1.6 ± 0.8	0.0 ± 0.5	$0.0^{+0.1}_{-0.0}$	$0.0^{+0.1}_{-0.0}$	$0.0^{+0.1}_{-0.0}$
	Flavor-symmetric	$0.2^{+0.2}_{-0.1}$	0.3 ± 0.2	$0.2^{+0.2}_{-0.1}$	0.1 ± 0.1	$0.0^{+0.1}_{-0.0}$	0.1 ± 0.1
	Z+ν	0.2 ± 0.2	0.4 ± 0.2	0.3 ± 0.1	$0.0^{+0.1}_{-0.0}$	$0.0^{+0.1}_{-0.0}$	0.1 ± 0.1
	Total background	43.0 ± 9.9	2.3 ± 0.8	0.5 ± 0.5	$0.2^{+0.2}_{-0.1}$	$0.0^{+0.1}_{-0.0}$	0.2 ± 0.1
	Observed	43	5	1	0	0	0
SRVZResolved		50–100	100–150	150–250	250–350	>350	
	Drell-Yan	3613 ± 80	394 ± 46	21 ± 18	1.7 ± 2.4	1.8 ± 1.9	
	Flavor-symmetric	$10.7^{+3.0}_{-2.9}$	15.4 ± 4.2	5.1 ± 1.5	0.5 ± 0.2	0.3 ± 0.2	
	Z+ν	24.0 ± 4.1	29.5 ± 5.6	32.2 ± 6.5	9.7 ± 2.2	4.2 ± 1.1	
	Total background	3648 ± 80	439 ± 47	58 ± 19	11.9 ± 3.2	6.3 ± 2.2	
	Observed	3648	461	69	7	2	
SRHZ		50–100	100–150	150–250	>250		
	Drell-Yan	163 ± 15	10.8 ± 4.1	1.3 ± 2.5	0.1 ± 0.3		
	Flavor-symmetric	3.9 ± 1.4	3.6 ± 1.3	3.3 ± 1.2	0.7 ± 0.3		
	Z+ν	1.3 ± 0.3	1.1 ± 0.2	1.0 ± 0.2	0.3 ± 0.1		
	Total background	168 ± 15	15.6 ± 4.3	5.6 ± 2.8	1.2 ± 0.4		
	Observed	168	14	5	0		

Table 5.9 summarizes the systematic uncertainties in the signal processes. The integrated luminosity uncertainty rages between 2.6 and 2.7%. The uncertainty of the lepton reconstruction procedure between data and Monte-Carlo simulations is 6% for the muons and 3% for the electrons. A further systematic uncertainty of 4% is added to account for differences between fast simulation and full simulation procedures. In addition we take a 3% uncertainty for the trigger efficiency. ISR modelling and Pile-up carry uncertainties that range from 0-2.5% and 1-2% respectively, while the E_T^{miss} modelling in fast simulation has an uncertainty ranging from 0 to 4%. Finally, the statistical uncertainties are of the order of 3-15%. These in total amount to around 18% in uncertainties.

Table 5.9. Systematic Uncertainties for the fast-sim signal samples

Source of uncertainty	Uncertainty (%)
Luminosity	2.6
Lepton reconstruction and isolation	5
Fast simulation scale factors	4
Trigger modeling	3
Jet energy scale	1-5
ISR modeling	0-2.5
Pileup	1-2
b tag modeling	0-5
Fast simulation MET uncertainty	0-4
Q^2 scale	1-3
Statistical uncertainty	3-15
Total uncertainty	10-18

5.7.2 Maximum Likelihood Based Interpretation

A maximum Likelihood fit was done for the signal + background hypothesis, i.e., the signal model ($b + \mu \cdot s$) is fitted to the data ($\mathcal{D}$) under the likelihood function described in Section 4.7. Our goal is find the minimum value of μ for a given mass point such that the likelihood of the (s+b) hypothesis is less than 5%. If this strength μ is less than one, we can say with 95% confidence that given the amount of data we have, this particular mass point does not exist physically. All mass points where $\mu < 1$ can then be rejected at the 95% level.

We run the fit before unblinding with the background predictions alone, and then compare this to the actual fit with real observed data. An upward fluctuation in data compared to background predictions in the relevant signal regions means (Figures 5.14 and 5.15) that the observed contour will exclude less mass points than the predicted one, since there is now a slightly greater chance of SUSY being found because of the upward fluctuation. Similarly a downward fluctuation implies that the observed contour will exclude more mass points.

5.7.3 Interpretation of the T5ZZ model

Figure 5.16 shows the exclusion plot for the T5ZZ model. The rainbow background is the strength μ multiplied with the cross section. It esentially conveys the cross section that is excluded at the 95% level. The contours therefore map those mass points where the 95% excluded cross section is the actual cross section. All the six strong signal regions are included to make in the overall maximum likelihood fit. This ensures that regions like SRA b-tag/b-veto cover those models where the gluino and $\tilde{\chi}_1^0$ masses are close and hence less jet activity is expected, while regions like SRC b-tag/b-veto cover those models where the masses are far apart from each other, resulting in higher hadronic activity. The black contour line in the figure shows the limit region with the observed data. Our expected gluino mass limit limit is around 1.65 TeV when the $\tilde{\chi}_1^0$ neutralino mass is small, and around 1.8 TeV when the neutralino mass is large. The observed limit is slightly lower than the expected limit, which can be corroborated with the upward data fluctuations in the strong signal regions in the high E_T^{miss} bins.

5.7.4 Interpretation of the TChiWZ model

Figure 5.17 shows the exclusion plot for the TChiWZ model. The three regions (VZ Boosted, VZ Resolved and HZ) are used. We use the HZ region to improve our sensitivity to the part of phase space where the Z decays into two b-jets, since only this region allows b-jets. The VZ boosted region plays a role in the models where the $\tilde{\chi}_1^0$ is almost massless, since the W boson carries away all the energy and gets boosted. The $\tilde{\chi}_2^0/\tilde{\chi}_1^{\pm}$ is assumed to have a wino like cross section, and nearly degenerate in mass. The two VZ regions contribute almost all of the acceptance to this model. The expected upper limit on the $\tilde{\chi}_2^0$ mass is around 640 GeV when the $\tilde{\chi}_1^0$ LSP is almost massless. The observed limit closely mirrors the expected for low $\tilde{\chi}_2^0$ masses and then goes up as the $\tilde{\chi}_2^0$ mass increases, culiminating in a limit of around 750 Gev when the $\tilde{\chi}_1^0$ is massless. This upward trend

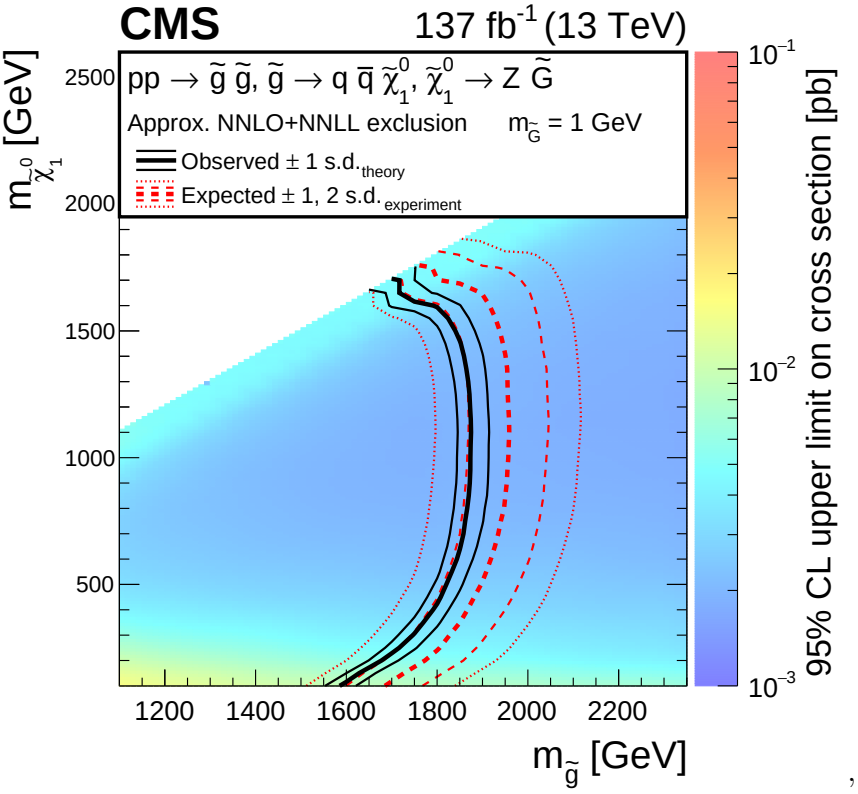


Figure 5.16. Exclusion plot for the T5ZZ model. Mass points to the left of the black (red) contour are observed (expected) to be excluded

in the observed limit compared to the expected limit can be attributed to the massive downward fluctuations in data in the high E_T^{miss} bins of the VZ Resolved region, and the empty high E_T^{miss} bins in the other VZ and HZ signal regions, as seen in Figure 5.15.

5.7.5 Interpretation of the TChiZZ model

Figure 5.18 shows the 1D exclusion plot fo the TChiZZ model. The x axis shows the mass of the $\tilde{\chi}_1^0$. The y axis conveys the same information as the rainbow background in Figures 5.16 and 5.17, i.e., the cross section eliminated at the 95% level. The magenta line shows the theoretical cross section. In the 1D space, the "contour" where the theoretical cross section meets the 95% exclusion cross section is a point, which is the intersection of the pink line with the dashed black line (expected) or solid blue line (observed). All the three electroweak signal regions (VZ Boosted, VZ Resolved and HZ) are included, with

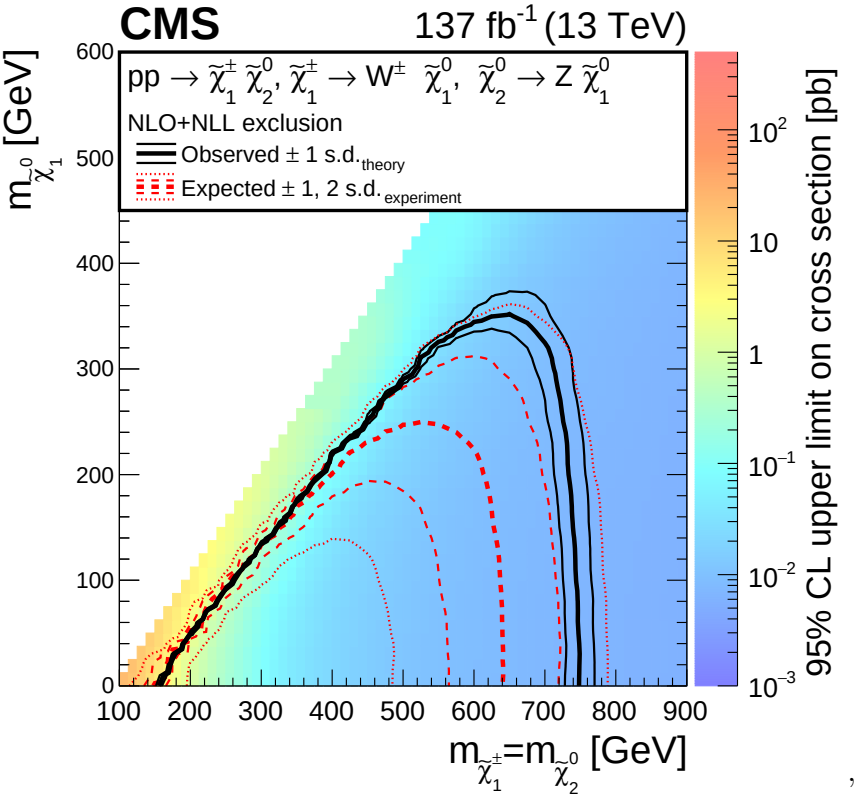


Figure 5.17. Exclusion plot for the TChiWZ model. Mass points within the black (red) contour are observed (expected) to be excluded

the HZ region for signal acceptance in the cases where the Z decays to two b-jets. The expected upper limit on the $\tilde{\chi}_1^0$ neutralino mass is at 700 GeV, while the observed limit is higher at 800 GeV. Similar to the TChiWZ case, the downward fluctuations in the data in the observed results (Figure 5.15) in the high E_T^{miss} bins cause the upward fluctuation in the limits.

5.7.6 Interpretation of the TChiHZ model

Figure 5.19 shows the 1D exclusion plot for the TChiHZ model. This process cannot be produced with a 100% branching ratio, because a neutralino that can decay to a Higgs boson can also decay to a Z boson. Hence we also need to consider signal contributions from the TChiZZ model. The maximal branching fraction to the final state is 50%, achieved when the $\tilde{\chi}_1^0$ decays with 50% probability to either Z or Higgs boson. In

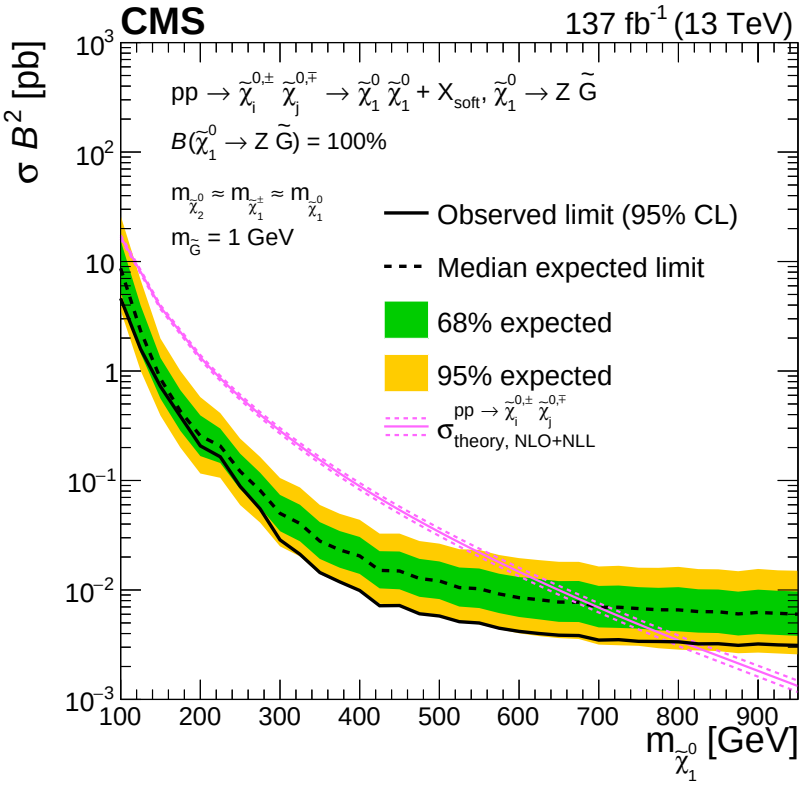


Figure 5.18. Exclusion plot for the TChiZZ model. Regions where the blue solid (black dashed) line reaches below the theoretical cross section (magenta band) in the exclusion plot are observed (expected) to be excluded.

this scenario, the final state has a 25% probability of occurrence. The 1D exclusion plot essentially produces limits for this scenario. All the three electroweak signal regions are used to compute the limits. The main contributor here is the HZ signal region, since we exclusively target the H $\rightarrow b\bar{b}$ process, the other signal regions covering the edge cases when the b-tag fails. The upper limit is expected to be around 520 GeV, and the observed is around 640 GeV.

5.8 Summary and Conclusions

A search for beyond-the-standard model processes was undertaken using the data collected by the CMS detector during the three years of Run-II (2016, 2017, and 2018).

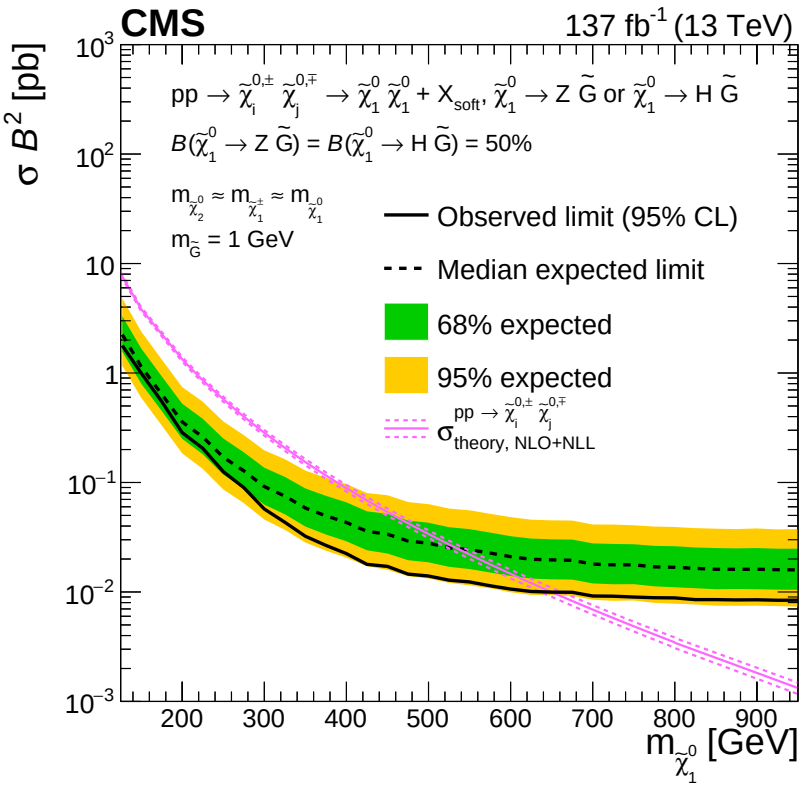


Figure 5.19. Exclusion plot for the TChiHZ model assuming a 50% branching ratio. Regions where the blue solid (black dashed) line reaches below the theoretical cross section, shown in magenta, in the exclusion plot are observed (expected) to be excluded.

Six strong signal regions were created to target the strong gluino mediated process T5ZZ, and three electroweak signal regions were created to target the electroweakino mediated processes TChiWZ, TChiZZ and TChiHZ. The main background contributors - Flavor Symmetric, Drell-Yan and Z+ν processes were identified. Data-driven background estimations for the flavor symmetric and Drell-Yan processes were designed, while for the Z+ν process, Monte-Carlo simulated samples were directly used, with scale factors derived in multi-lepton control regions. The observed event yields were found to be consistent with the standard model expectations. These results were then used to set upper limits on the production cross sections and masses of the supersymmetric processes. Gluino masses up to 1870 GeV, and chargino (neutralino) masses up to 750 (800) GeV were excluded, extending the reach over previous results by a few hundred GeV.

5.9 Acknowledgments

This chapter described a search for supersymmetry in events containing a Z boson, jets and missing transverse energy in data collected during Run-II of the LHC. It is a partial reproduction of the paper "Search for supersymmetry in final states with two oppositely charged same-flavor leptons and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV", J. High Energ. Phys. 2021, 123 (2021).

We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC and thank the technical and administrative staff at CERN and other CMS institutes for their contributions to the success of the CMS Collaboration. In addition, we acknowledge the computing centers and personnel of the Worldwide LHC Computing Grid for delivering the computing infrastructure essential to our analyses. Finally, we acknowledge the enduring support for the construction and operation of the LHC and the CMS detector.

On a smaller scale, the author thanks the opposite sign analysis sub-group - Marius Terörde, Dr Sergio Sanchez Cruz, Dr Pablo Martinez Ruiz del Arbol, Dr Giovanni Zevi Della Porta and Dr Mario Masciovecchio for the fruitful collaboration that made this analysis possible.

Chapter 6

Line Segment Tracking : A new parallelizable particle tracking algorithm for the HL-LHC

6.1 Tracking 101

The first step in the reconstruction of charged particles is reconstructing their tracks. The tracker provides accurate estimates of momentum and energy of the particles and plays an important role in the reconstruction of the parent objects. In addition, tracking plays an important role in vertex reconstruction algorithms.

As we saw in Section 3.3.1, the tracker consists of an inner pixel detector and an outer strip tracker. These detectors are made of layers of pixels or strips. When a charged particle passes through these layers, it leaves charge deposits (called hits). Track reconstruction is essentially a pattern matching problem that correlates the hits in various layers to reconstruct particle trajectories, under the hypothesis that a charged particle has a helix like trajectory in a solenoid magnetic field. Once the trajectories are reconstructed, the momenta of the particles can be calculated. Charged particle track reconstruction is also the most expensive and the most time consuming step in the object reconstruction pipeline.

The algorithm currently in use in the CMS Computing Backend is an iterative Kalman Filter based algorithm [44]. Kalman Filtering is an algorithm that uses sequential noisy measurements to update the parameters of a system. In the context of tracking, a Kalman Filter essentially works by considering the hits in the layers as observations, and the track parameters as the system parameters. Assuming a track till layer N-1 has been reconstructed with a set of parameters, a prediction is created for layer N. The track state is then updated using the hit position closes to the prediction. While this procedure provides excellent track reconstruction capabilities, it is at its heart an iterative process. The speed of the algorithm scales linearly with the possible number of tracks (ultimately the number of hits) in a collision event.

6.2 Marching ahead : Tracking Challenges in the The High Luminosity LHC

The High Luminosity LHC (HL-LHC) [45] will be the successor to the LHC and will start operations in 2029. In this accelerator, the instantaneous luminosity, i.e., the number of protons that collide with each other in a bunch crossing, will be increased by a factor of 5 with respect to current LHC operations, leading to pile-up values as high as 200.

Since track reconstruction is a combinatorial problem, the time taken to reconstruct all tracks in a collision event increases exponentially with the pile-up, as shown in Figure 6.1 ([12]). We see that we need approximately two orders of magnitude more time to be able to efficiently reconstruct tracks using the current setup in the HL-LHC era. Given that the current online track reconstruction time is of the order of a few milliseconds, we are already looking at a few seconds (1-2 seconds) range with the current system.

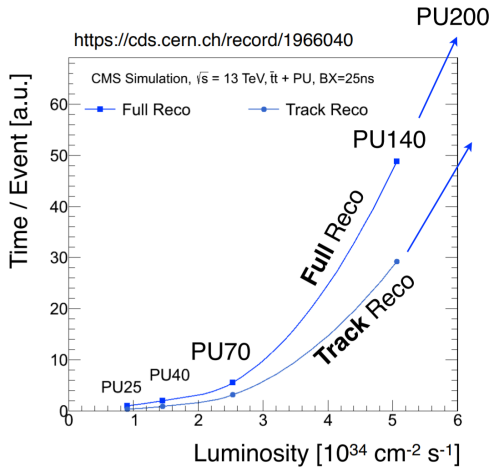


Figure 6.1. Track reconstruction time estimates with pile-up. The y-axis units are arbitrary. This plot should be used only for an apples-to-apples comparison of track reconstruction times at different pile-ups. Source : [12]

In addition, computational performance of single thread processors is plateauing while demands are increasing.

Not all hope is lost, because this is also the era of multi-core processors and heterogeneous co-processors like Graphical Processing Units (GPUs) and Field Programmable Gate Arrays (FPGAs). Smartly devised algorithms and code can be made to take advantage of these new paradigms and produce the speeds that computational loads of this era demands. The fastest supercomputer today [46], called the Frontier at Oak Ridge National Laboratories [47] has approximately 9400 compute nodes, each containing 4 GPUs and one AMD EPYC 64 core CPU. To take advantage of this new age in computing, there have been changes to detector design (Section 6.2.1) in addition to tracking algorithms and software.

6.2.1 The CMS Outer Tracker in the HL-LHC

Our tracking algorithm (Section 6.3) specifically targets the reconstruction of the tracks in the outer strip trackers. Figure 6.2 shows the r-z profile of the outer tracker,

which is composed of bi-layer p_T modules shown in Figure 6.3. A single p_T module has two silicon layers that are $\approx$ 2 to 4 mm apart. The advantage of using these closely spaced modules is that the two hits corresponding to the same trajectory can be correlated quickly and in parallel, resulting in small track stubs that provide more details than a single hit can provide. These track stubs along with their extra information can then serve as basic building blocks in track reconstruction. In addition, as the right cartoon in Figure 6.3 shows, hit correlation based on simple p_T lower bounds can be done fairly quickly, resulting in the reduction of combinatorics. Essentially, these very closely spaced modules act as basic filters that, in addition to reducing the combinatorics considerably, can provide more information than just two hits.

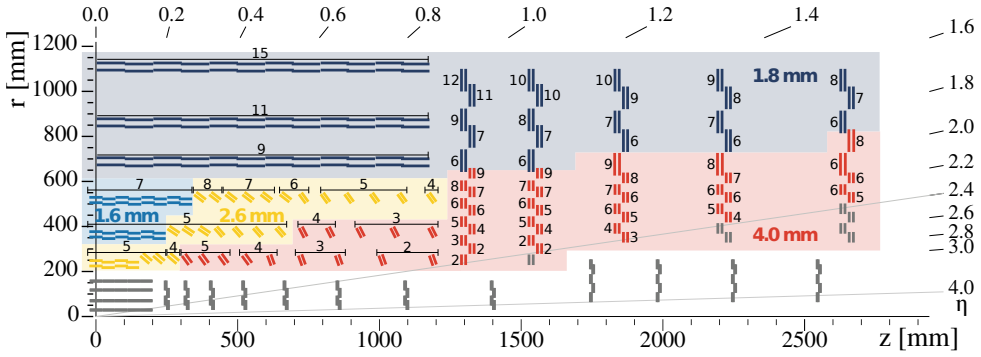


Figure 6.2. r-z profile of the outer tracker of the CMS Detector in the HL-LHC. Source : [13]

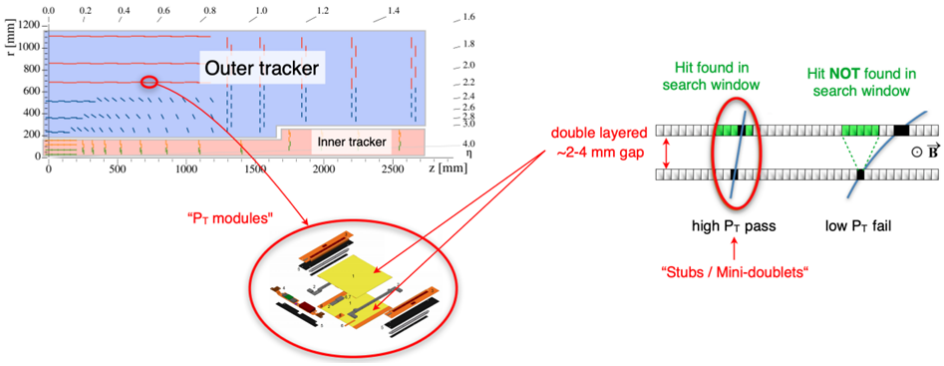
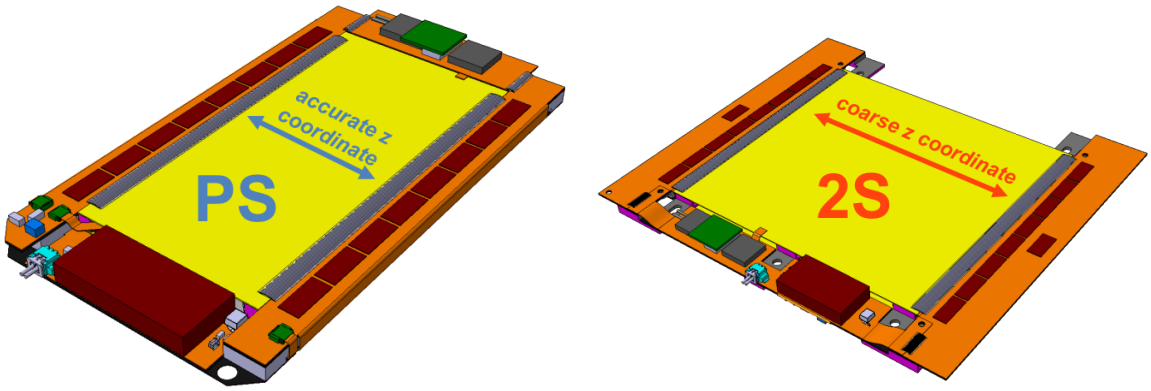


Figure 6.3. Outer tracker geometry and module structure in CMS at the HL-LHC. Source : [13]

The outer tracker is comprised of two types of modules, called PS (Pixel-Strip) and 2S (Strip-Strip) modules. PS modules are present in the inner regions of the tracker and are show in red, blue and yellow in Figure 6.2, while 2S modules are present in the outer regions of the tracker and are shown in bluish-gray in Figure 6.2. Figure 6.4a shows the cross section of a single PS module. The Pixel side is comprised of 32×960 macro pixels, each 1.5mm by 100μm in size which provide accurate measurements of the hit position. The strip side is comprised of 2×960 strips, each 2.4cm by 100μm in size. Figure 6.4b shows the cross section of the 2S module. Both of its sides are comprised of 2×1016 strips, each 5cm by 90μm in size.



(a) PS module **(b)** 2S module

Figure 6.4. The two module types that make up the detector. Left : PS, Right : 2S. Source : [13]

A consequence of the bi-layer p_T module based stub finding is that both the top and bottom surfaces of a sensor must be connected to the readout electronics. The price paid for this is that the module is segmented into two halves, and these two halves are read out independently on both ends. From a reconstruction perspective, this implies that tracks that cross the lower side on one half and the upper side on the other half are lost,

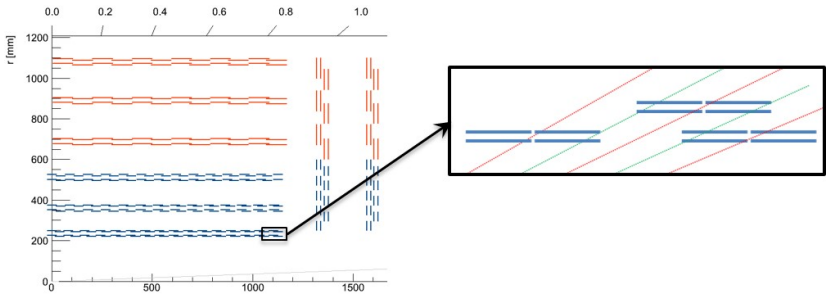
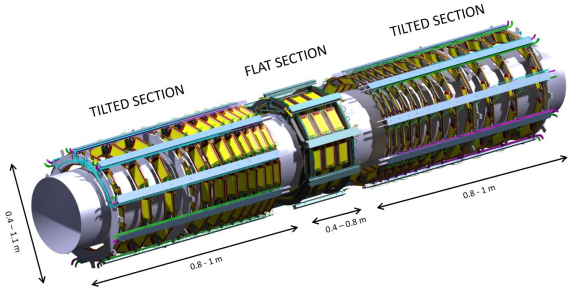
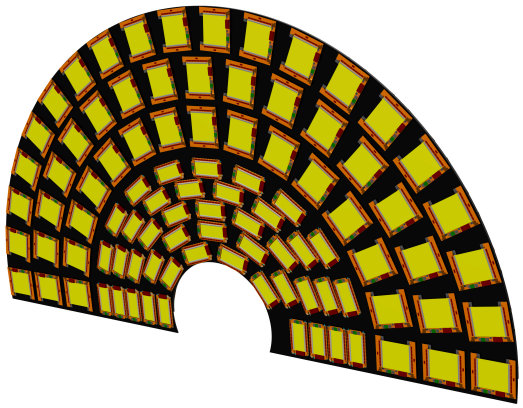


Figure 6.5. Tracks that don't hit the same half of the module in both layers are lost. This is why we have tilted modules

as shown in Figure 6.5. This tends to happen mainly in the edge of the barrel region, and contributes to an inefficiency of 30% in the first barrel layer. To mitigate this, we have a "transition region" where modules are progressively tilted, as can be seen in the first three barrel layers in Figure 6.2. A realistic picture of the tilted and barrel modules in a layer in 3D is shown in Figure 6.6a. In Figure 6.6a, we also see how the flat and tilted



(a) Flat and Tilted modules in one barrel layer.



(b) One half of an endcap ring

Figure 6.6. Left : 3D view of the modules in the barrel, showing the flat and tilted modules. Right : Module arrangement in one endcap ring. Source : [13]

modules are arranged in 3D space. The flat modules are arranged along the curved surface of the cylinder of the barrel. The tilted modules gradually transition from fully flat to fully vertical, which coincide with the endcap rings (Figure 6.6b).

6.3 Line Segment Tracking in the Outer Tracker

Line Segment Tracking (LST) is a parallelizable algorithm that reconstructs tracks in a bottom-up fashion in the outer tracker. We start with the correlated hit pair stubs which we call **Mini-doublets**, and link them together to create four hit objects called **Segments**, we then link these to create longer and longer objects till we reconstruct entire tracks that span the whole length of the outer tracker. Since the reconstruction of a higher order object only depends on the lower order objects in the immediate vicinity, the reconstruction is highly local. For instance, the correlation of two hits to create a Mini-doublet only requires hyper-local information about hits in adjacent modules, and so all such track pairs can be done in one step. Similarly linking them upwards into Segments and Triplets (6 hit object) only requires information in the immediate neighbourhood - a small region of the detector, and hence such linkages can also be done in parallel, and so on.

Mini-doublets, which are our fundamental building blocks, get linked to form segments. Two segments with a common mini-doublet form a triplet. Two triplets with a common mini-doublet form a quintuplet (10 hits). In addition, we also incorporate tracking seeds from the inner pixel tracker produced by other algorithms like Patatrack. These seeds can be linked with a Triplet or a Quintuplet to form a Pixel Triplet or a Pixel Quintuplet. These track seeds enable us to reduce the combinatorial fakes to an even greater extent, since we have good quality "pure" seeds from the pixels, in addition to providing a link between the inner pixel and the outer strip detectors.

Line Segment Tracking was originally inspired by the XFT algorithm used at the CDF Experiment at the Tevatron, Fermilab [48]. A similar algorithm was used in the

context of grouped layer tracker layout design [49] which showed promising results in reducing combinatorics.

6.3.1 Mini-Doublets : The fundamental building block

The circled track stub created in the right cartoon in Figure 6.3 is called a Mini-doublet (MD). In addition to having information about the two hits, this object also holds information about the slope (i.e., slope of the circle in 2D space) consistent with a particle being produced in the vicinity of the origin. This slope contains additional error terms that take care of multiple scattering, displaced tracks and uncertainties in the "luminous' i.e., the collision region.

The two hits of the mini-doublet are designated "anchor hit" and "outer hit". The anchor hit generally tends to be one that is well measured. In case of a mini-doublet in a PS module the pixel hit is the anchor hit, while in the 2S module, since both hits are equivalent, the hit on the lower side is taken to be the anchor hit.

6.3.2 Combinatorics and Physics considerations

Figure 6.7 shows the hit occupancy in the six barrel layers in a typical Pile-up 200 collision event. In these barrel modules, we have hits varying from around 36,000 in the innermost layers to 6000 in the outermost layer (Table 6.1). If hits are naïvely linked to form mini-doublets, we will be dealing with approximately a million of them in every event in the barrels alone. Basic physics and geometric facts, like both the hits need to come from the same half of the module, and the fact that the slope angle should be consistent with a track with $p_T > 0.8$ GeV reduces the mini-doublet multiplicities down to a manageable 22,000 mini-doublets in the barrel.

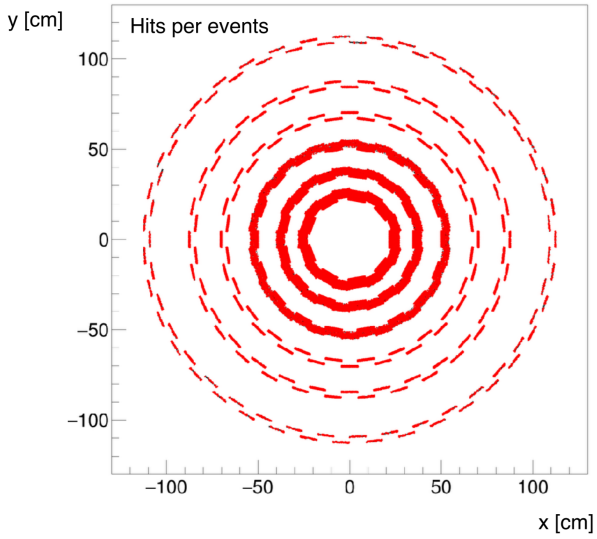


Figure 6.7. x-y map of the barrel of the Phase-2 outer tracker showing the hit occupancy in each layer

Table 6.1. Hit multiplicities and Mini-doublet multiplicities before and after selections in the barrel for a typical PU200 collision event

Layer	1	2	3	4	5	6
Hits	18K	23K	24K	27K	19K	11K
Mini-doublet no selection	24K	40K	50K	59K	33K	12K
Mini-doublets	7507	5106	4179	4927	4714	2814

6.3.3 Segments and Module Maps

Segments (also called Line Segments) are created by linking mini-doublets from modules across different layers. Figure 6.8 shows a representative diagram of two mini-doublets linking to form a segment. A crucial part of the segment creation process is that mini-doublets in a given module should be considered for linking only with those mini-doublets in the immediate neighborhood, so that unphysical connections are not considered. A neighborhood with too few connections means that not all tracks will be reconstructed, and a neighbourhood with too many connections means that a lot of time and computational resources will be spent considering a good number of unphysical connections, which can also result in fake reconstructed tracks.

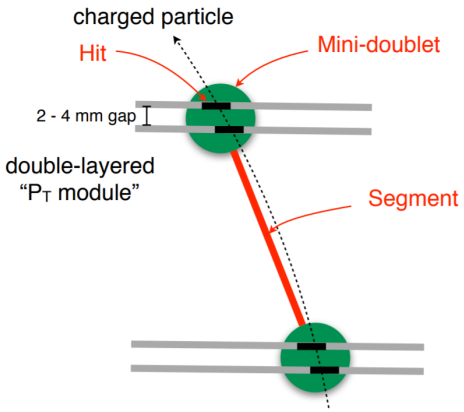


Figure 6.8. Two Mini-doublets (Green dots) linked using the module map to create a Line Segment (Red line)

We construct a "module map" which is a lookup table detailing the possible connections for each module. To find the optimal number of connections for a given module, we take its four corners and shoot helices with p_T of 0.8 GeV from these corners. We then list all the modules in the adjoining barrel and/or endcap layers these helices intersect. These will form the boundaries of the connected region in 3D space. All modules enclosed by this connected region form the list of connected modules for a given module. We have taken advantage of the fact that the lower the p_T, the higher the bending of the tracks, so two 0.8 GeV helices shot in the opposite directions create a region of modules, within which all helices of higher p_T will go through. This first principles approach ensures that we take all possible edge cases into account, which would have been quite difficult if we relied only on simulations to compute the map.

Additional selections for the segments involve geometrical and slope based ones similar to what we have for the mini-doublets. In addition to ensuring that the overall slope parameter needs to be consistent with a track with p_T ¿ 0.8 GeV, we also look for equivalency criteria, i.e., whether the parameters corresponding to the two contributing

mini-doublets are the same.

6.3.4 Higher order objects : Triplets and Quintuplets

Triplets are formed by combining two segments with a common mini-doublet, as shown in Figure 6.9. The requirement of a common mini-doublet greatly reduces the combinatorics and spurious connections. In addition, we require selections such as the third mini-doublet's anchor hit should be consistent with a straight line drawn using the two inner mini-doublets in the r-z plane [1], in addition to slope consistency selections similar to the ones we have seen in the mini-doublet and segment steps.

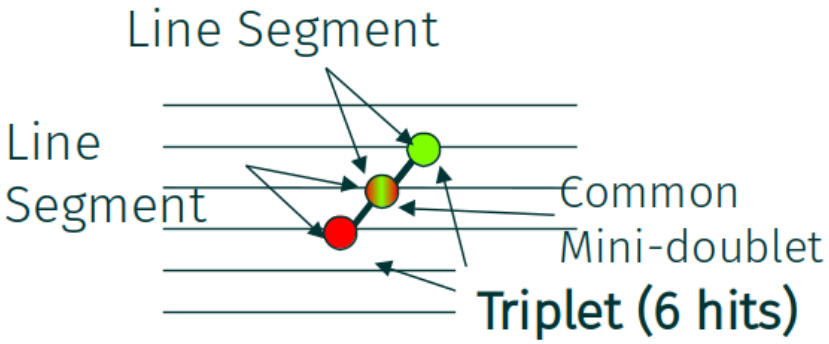


Figure 6.9. A Triplet is formed from two segments having a common mini-doublet

Two triplets with a common mini-doublet can be joined together to form a quintuplet, as shown in Figure 6.10. As seen in the triplet case, the mini-doublet requirement reduces combinatorics. These quintuplets are the longest objects exclusively constructed using the hits in the outer tracker. Just like in the triplet case, we apply a straight line consistency based selection in the r-z plane, using the first and third mini-doublet anchor hits (or the first two hits alone in some cases) and checking the consistency of the outer two anchor hits to this straight line. Since we have triplets as our two constituting objects here,

[1] A helix in 3D space can be projected into a straight line in r-z plane($r = \sqrt{x^2 + y^2}$)

we can also use the fact that the three anchor hits of the triplet can uniquely construct a circle in the x-y plane. Hence we have the additional selections of requiring that the circles formed from the anchor hits of the inner and outer triplets have radii close to each other, and checking the consistency of the two outer hits to the circle formed by the inner triplet. The consistency is checked by computing the χ^2 of the outer two hits and the thresholds are derived empirically for different categories of quintuplets and provide further reduction of the spurious linkages.

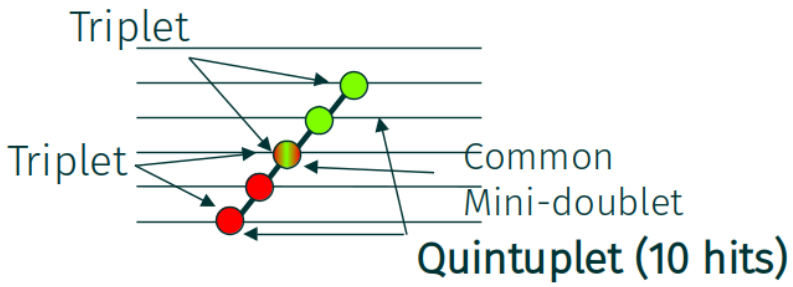


Figure 6.10. A Quintuplet is formed from two triplets having a common mini-doublet

6.3.5 Incorporating the Pixel Tracks : Pixel Triplets and Pixel Quintuplets

In addition to hit information from the outer tracker, we also have access to the track stubs reconstructed by the inner pixel tracker. We incorporate these as line segments, henceforth called Pixel Line Segments (pLS). The triplets and quintuples produced in the outer tracker can be linked to these pixel line segments to produce objects such as Pixel Triplets (pT3s) and Pixel Quintuplets (pT5s). The linking of the pixel track stubs with the outer objects provides a viable method to reduce the fake tracks even further, since we expect the pixel track stubs to be very clean.

Pixel Quintuplets (pT5s) are formed by linking Pixel Line Segments with Quintuplets, as shown in Figure 6.11. The selection parameters we use are similar to what we have incorporated in the Quintuplet selection process, i.e., χ^2 like parameters that check for consistency in the r-z and x-y planes. In the $r - z$ plane, we extrapolate the straight line that we get from the pixel line segment parameters, and check the consistency of the five outer anchor hits. Similarly in the x-y plane, we use the track parameters of the pixel line segments to get the circle center and radius (from the reconstructed p_T) of the circle and check the consistency of the quintuplet. With quintuplets being the longest and cleanest objects we have, and with the addition of the pixel line segments their purity increases even further, which makes the Pixel Quintuplets our primary candidate for our final track candidate collection (Section 6.3.7).

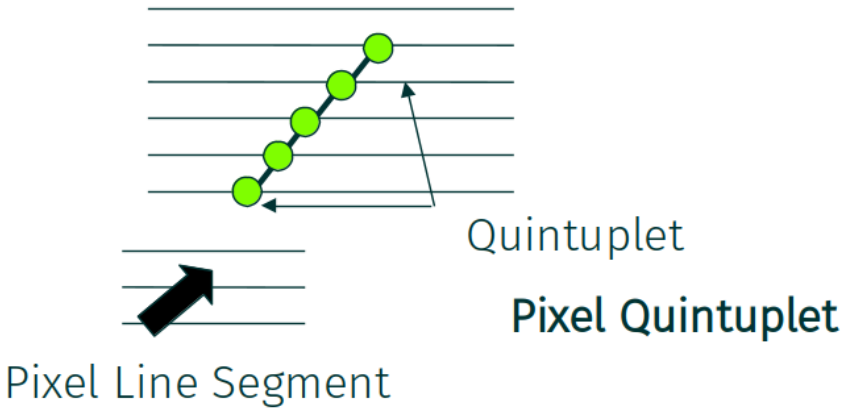


Figure 6.11. A Pixel Quintuplet is formed by linking pixel track stub with a triplet

Similar to the Pixel Quintuplets, we create Pixel Triplets (pT3s) by linking Pixel Line Segments with outer tracker triplets, as shown in Figure 6.12. In principle, we can create two pT3s for a given pT5 (since a pT5 has two pT3s). To prevent such overlaps, we only use those triplets that are not part of any quintuplet. Similar to the pT5 selections,

we select pT3s based on the the triplet's anchor hits are to the pixel line segment based on the χ^2 distributions in the $r - z$ and x-y plane.

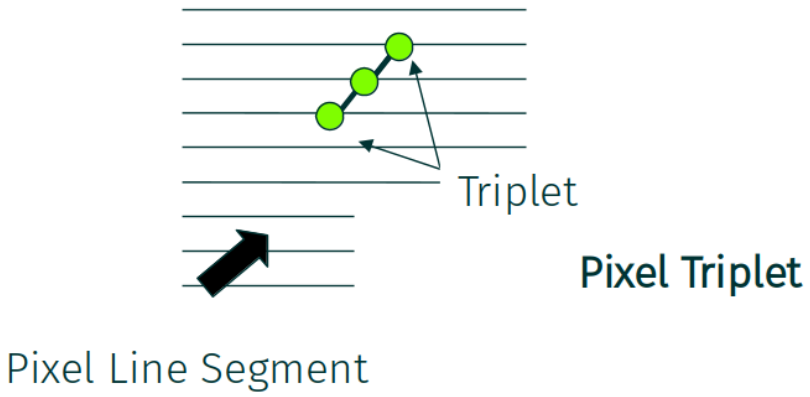


Figure 6.12. A Pixel Triplet is formed by linking a pixel track stub with a triplet

6.3.6 Duplicate cleaning

We employ an elaborate duplicate cleaning procedure that reduces the overall number of objects created. Right after all objects of a particular type (i.e., quintuplets or pixel triplets) are created, they undergo a first pass cleaning procedure. At this level, two objects that share more than 70% of hits are deemed to be duplicate, and the object with a lower χ^2 value is retained. A second pass is done before the track candidate addition process (Section 6.3.7) is done, in which ΔR values are computed between these objects and if two objects are within a threshold (usually 10^{-4}), the one with the lower χ^2 value is retained. These two steps ensure a very extensive cleaning of the individual object collections.

6.3.7 Track Candidates

Our final collection that will be the output of the algorithm is the Track Candidates. The final track candidate collection consists of Pixel Quintuplets, Pixel Triplets, Quintuplets and Pixel Line Segments, added in this sequence, which is essentially a decreasing order of

"purity" (i.e., fraction of fakes). The Pixel Quintuplets, being the longest and cleanest objects are added as is and will form the bulk of the track candidate collection. To ensure track duplicates between different objets are not added to the final collection, we "cross-clean" the objects. This implies that when we add Pixel Triplets to the track candidate collection, we ensure only those pixel triplets that are not near a Pixel Quintuplet in $\eta - \phi$ space are added (we have already cleaned the pixel triplets with each other in the duplicate cleaning steps). This ensures that we only add those Pixel Triplets added that exclusively correspond to those tracks which could not produce a Pixel Quintuplet possibly due to a missing mini-doublet and/or hit somewhere. Similarly, when we add Quintuplets, we only add those that are not near a Pixel Quintuplet or a Pixel Triplet. Quintuplets serve the dual purpose of providing coverage for those tracks that for some reason did not get reconstructed in the pixel tracker (hence no pixel line segment seeds), and more importantly, provide coverage for displaced tracks, which will be crucial for physics studies involving long lived objects. Finally, the Pixel Line Segments that follow a strict set of criteria (they need to not be a part of any other object, they should be appreciably far away in $\eta - \phi$ space with the other objects etc) reconstruct tracks in the extremely forward region ($|\eta| > 2$).

Cross-cleaning procedures involve methods similar to the ones used for duplicate cleaning. We look for proximity in the $\eta - \phi$ space by computing the ΔR between the track objects. For objects that involve pixels, we also look if the same set of pixel hits are used in the process.

6.4 Physics performance

We measure physics performance and fine-tune cuts using simulated datasets. We simulate a $t\bar{t}$ production process in an 200 pile-up environment, consistent with the

luminosities that will be observed in the HL-LHC. We then write out the hits from the simulated outer tracker modules, along with the "truth" information about the simulated tracks as reference. A track candidate is said to be matched to a true track if at least 75% of the associated hits match, i.e., a Quintuplet is said to be matched to a track if 8 or more hits match. This matching procedure is called MultiTrack Validation (MTV).

The efficiency is defined as the fraction of the total truth tracks that get matched to a track candidate, i.e.,

$$e = \frac{\text{Number of tracks that have at least one MTV matched track candidate}}{\text{Total Number of Tracks}} \tag{6.1}$$

Efficiency is measured in bins of p_T, η and ϕ, where the numerator and denominators in Equation 6.1 get adjusted accordingly, i.e., in case of a p_T distribution, the denominator only considers tracks with a true p_T value corresponding a particular p_T bin.

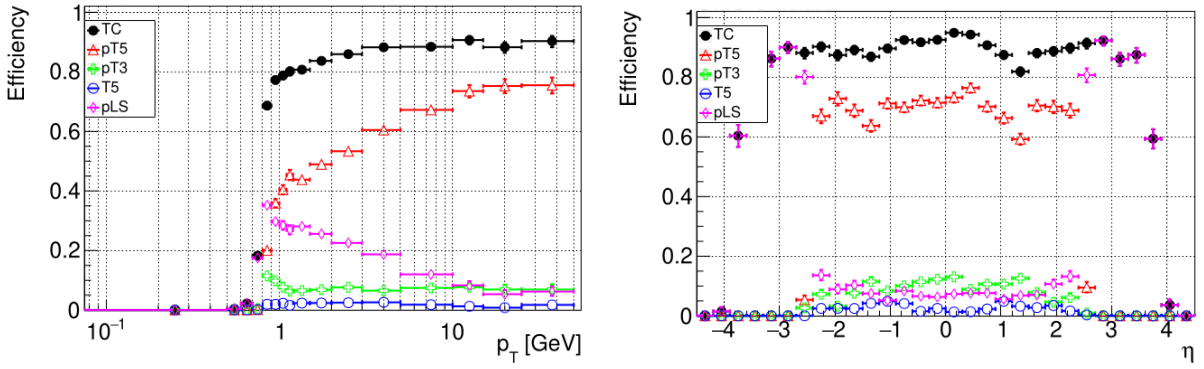
Fake rate is calculated as the fraction of track candidates that cannot be MTV matched to a true tracks, while duplicate rate is calculated as the fraction of track candidates that are duplicates. A track candidate is marked as a duplicate if it can be matched to a simulated track which has at least one other track candidate matched to it, i.e., if two track candidates are matched to a simulated track, both tracks are marked as duplicates. These quantities can also be measured in p_T, η and ϕ bins, just like the efficiency.

6.4.1 Efficiency performance of Line Segment Tracking

Figure 6.13 shows the track candidate efficiency (in black dots) as a function of track p_T and η. Since we only reconstruct tracks with $p_T > 0.8$ GeV, we see an efficiency

turn-on in its neighborhood. The η distribution has been constructed with a $p_T > 0.9$ GeV selection, to target the region after the turn-on. Hence it is relatively flat.

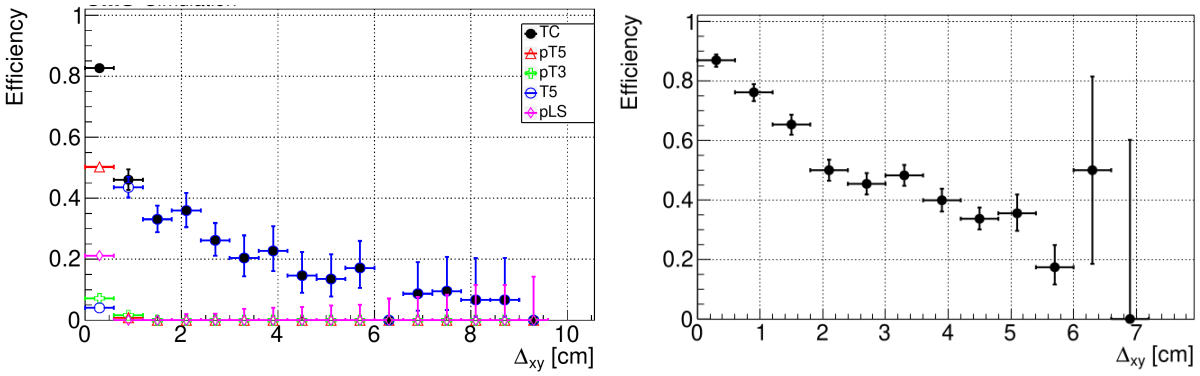
The different colors show the contribution to the total efficiency from the four components. The Pixel Quintuplets contribute the highest to the efficiency, while the unlinked Pixel Line Segments contribute in the forward ($|\eta| > 2$) regions.



(a) Efficiency vs p_T **(b)** Efficiency vs η

Figure 6.13. Efficiency on $t\bar{t}$ + PU200 sample

To measure the performance of displaced track reconstruction, we compute the efficiency as a function of the x-y distance from the interaction point (Δ_{xy}). Figure 6.14a shows the efficiency of the displaced tracks in the $t\bar{t}$ +PU200 sample, while Figure 6.14b shows the efficiency in a sample of displaced muon tracks originating from a point displaced from the origin in a 5cm cube (Figure 6.14b). We see that in the more realistic $t\bar{t}$ +PU200 case, the Quintuplets play a major role in reconstructing displaced tracks. This efficiency is given to us "for free" at this point, because we did not do anything extra to target displaced track reconstruction. These promising results show that this can be improved further.

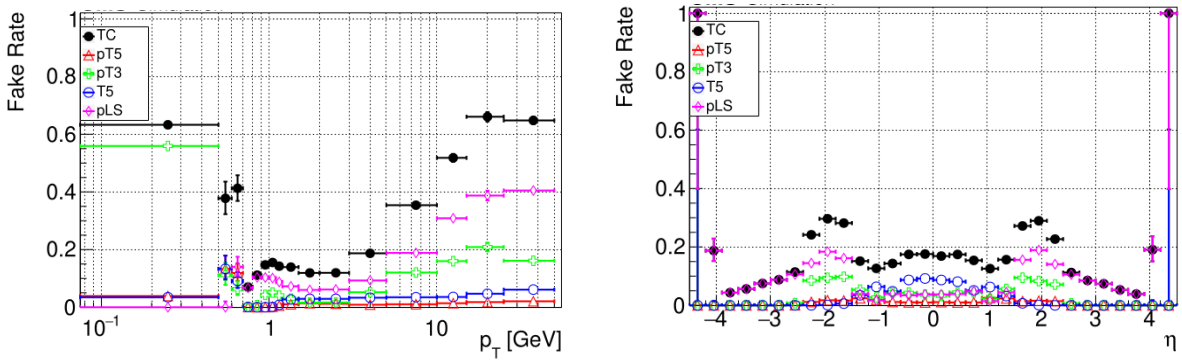


(a) Efficiency vs Δ_{xy} in $t\bar{t}$ + PU200 sample **(b)** Efficiency vs Δ_{xy} in 5cm muon cube sample

Figure 6.14. Efficiency with displaced tracks - PU200 and muon cube

6.4.2 Fake and Duplicate Rates performance

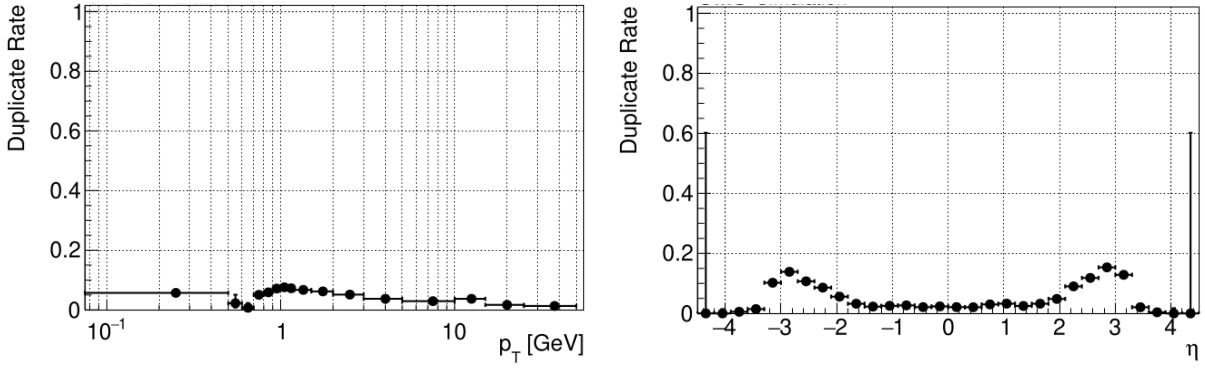
The fake rates as a function of p_T is shown in Figure 6.15a and as a function of η in Figure 6.15b, along with the breakdown into individual components. The Pixel Line Sgments contribute the highest to the fake rates in the endcap ($|\eta| > 1.6$) while in the forward region ($|\eta| > 2.5$) all the fakes are attributable to them. while in the neighborhood of $\eta \approx 0$, the quintuplets contribute the most. These fake rates can be reduced further with fine-tuning of selection parameters and a full fit of hit patterns.



(a) Fake rate vs p_T **(b)** Fake rate vs η

Figure 6.15. Fake rates on $t\bar{t}$ + PU200 sample

Figure 6.16a shows the duplicate rates as a function of p_T, and 6.16b shows the duplicate rate as a function of η. Since we do rigorous duplicate cleaning and cross cleaning in an efficient way, the duplicate rates are quite low. A low duplicate rate ensures that the final track candidate collection is compact and repeated reconstruction is reduced.

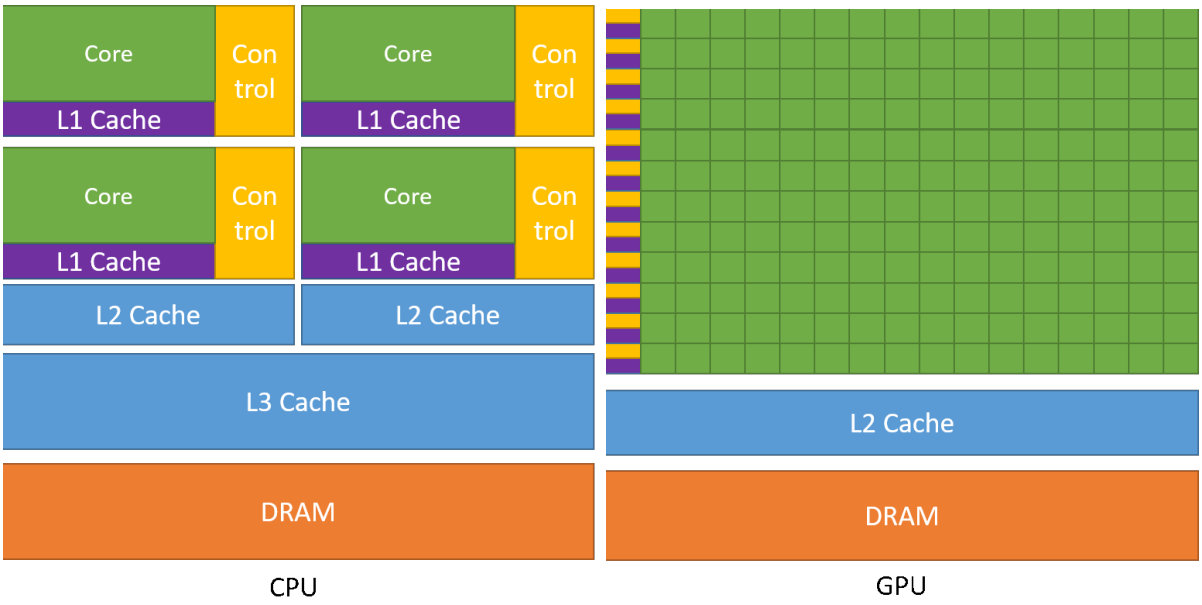


(a) Duplicate rate vs p_T **(b)** Duplicate rate vs η

Figure 6.16. Duplicate rates on $t\bar{t}$ + PU200 sample

6.5 Introduction to GPUs

Graphical Processing Unites (GPUs) are co-processors that are mainly used to render graphics and play a vital role in the video games industry. Graphics rendering involves mathematically intensive operations like ray tracing which involve humongous amounts of matrix operations. To specifically achieve these, GPUs were designed with a lot of compute cores but very little memory cores (Figure 6.17b). A path-breaking innovation in 2001 [50] showed that these co-processors can be used to do general matrix multiplication, which kick-started the phenomenon now called General Purpose Computing on Graphical Processing Units (GPGPUs) and launched a new realm of parallel computation. With the rising popularity of deep learning which relies on fast successive matrix multiplications, hardware manufacturers are keeping up with increased demands for speed and usability. Nvidia is currently the leader in GPGPUs with their computation specific line of cards



(a) CPU architecture **(b)** GPU architecture

Figure 6.17. CPU vs GPU Architecture. Source : [14]

(formerly called the Tesla line of cards), with AMD entering the race and hitting it big with the Frontier Supercomputer [47] which uses AMD Instinct GPUs.

GPUs have numerous compute cores compared to CPUs, as shown by the green grids in Figure 6.17. However, they compromise on caches and transport. In addition, GPUs are slower, with clock speeds in the neighborhood of 1 GHz, compared to CPUs having clock speeds of 2-5 GHz. If we program GPUs such that the compute cores can do work on existing data while the caches wait for new data, we can achieve tremendous speed-ups.

6.5.1 Cuda

Cuda, formerly an acronym for Compute Unified Device Architecture, is a proprietary programming framework developed by Nvidia that targets their line of GPUs. Cuda is at the heart of deep learning frameworks like PyTorch and Tensorflow. This framework

provides us with various high level C++ extensions that abstract away the (proprietary secretive) processes of memory creation, computation and transfers. At the time of writing, the latest Cuda version i1.8, which targets their line of GPUs all the way up to their latest Hopper architecture (named after Grace Hopper). Cuda is being made more accessible and easier to use with every successive version, starting with Unified Memory in Cuda 6 which eliminated the need for explicitly allocating memory in the GPU and manually copying data back and forth,or thread collectives in Cuda 9 that overcomes a previously known limitation wherein 32 threads would run together in lockstep (called a warp), and so all of them would need to wait even if one thread in this warp slows down. These new innovations bring GPU programming much closer to conventionally used CPU style programming.

The fundamental programming unit of a Cuda program is the **kernel**. A kernel specifies what every thread in a parallel computation operation does. Since GPUs work best when compute cores each get their own little task to do, and do not communicate with each other, a single universal prescription can be written. This is called the Single Instruction Multi Thread (SIMT) programming paradigm. The compute task is divided into threads, with each thread doing one task in parallel. The kernel essentially describes what a single thread should do. A block is a group of threads, and the entire group of blocks is called the grid. An entire block is executed by one streaming multiprocessor [2] (which can also execute more blocks in parallel). Blocks cannot be split between streaming multiprocessors. The blocks have a special property which is that the threads in the block can access a small amount of common memory called the shared memory. For instance, if the job requires that some memory be shared between some threads, in addition to the threads using their own memory (a hierarchical structure of sorts), they can be grouped

[2]A streaming multiprocessor is essentially a collection of compute cores with its own memory bus and common cache and shared memory

into the same block and take advantage of the shared memory. Each thread gets its own ID, which can then be used to identify which part of the work needs to be done by this thread.

6.6 Line Segment Tracking on the GPUs

Since Line Segment Tracking requires that objects be built hierarchically upwards, i.e., mini-doublets create segments, which then create triplets and so on, it can be seen as a a sequential list of parallel operations. The entire mini-doublet creation process can be parallelized, resulting in mini-doublets created across all modules of the detector. This will be followed by the segment creation process, and so on. Figure 6.18 shows a flowchart that details this process. Since each object creation process can be parallelized, they get their own kernel in addition to a kernel that deals with pre-processing the hits and loading them into GPU memory. The final collection of track candidates is then copied to the CPU. Efficient memory management (Section 6.6.1) and Multi-streaming (Section 6.6.2) play major roles in improving the computational performance.

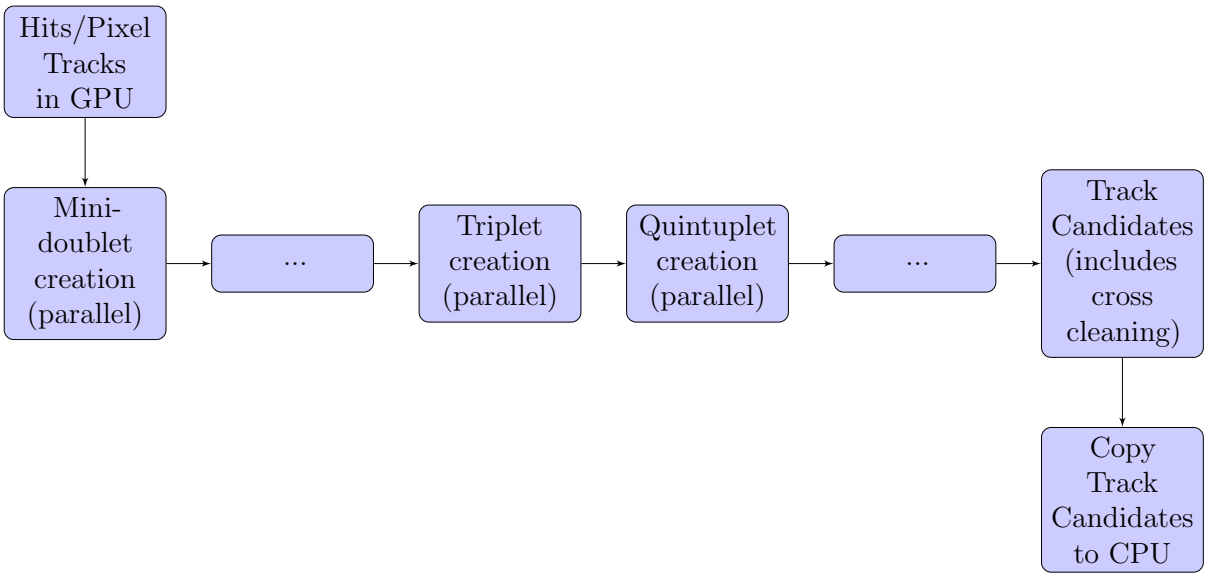


Figure 6.18. GPU implementation flowchart. Only the more important steps are explicitly shown, the others are represented by dots (. . .)

6.6.1 Computational and memory considerations

Memory Management - Structure of Arrays (SoA)

Since GPUs work on the Single Instruction Multi-thread (SIMT) framework where multiple cores perform the same computational task in parallel on different chunks of data, we can take full advantage of the architecture if we store the data be stored such that the compute cores get to use all the data that streams into the caches. We can maximize cache hit rates and memory transfers from the main GPU memory if adjacent compute cores (that work on adjacent units of memory) get to use their respective data without any delays. To this end, we store the data in a Structure of Arrays (SoA) data representation framework [51].

In the SoA framework, we store the data corresponding to a particular variable continuously for all objects in a single array in memory. For example, if we want to store the p_T information for mini-doublets, we store this for all mini-doublets continuously in an array (as opposed to creating a class object for each mini-doublet, which will have the p_T value as one of its members). This would mean that when the p_T values are required in a particular step in the computation process , all the active compute cores will be readily able to access this data with very high cache hit rates.

Reducing memory allocation overheads - Caching Allocator

Memory allocation in the GPU is a very time intensive task, and with the amount of memory allocations we have, it adds up in time. To reduce the huge timing overheads, we developed a caching allocator based on the CUB library developed by Nvidia [52], which exponentially allocates memory based on the current demand, i.e., if 5MB of memory is requested, the caching allocator makes 8MB available, which means that when we get to allocating memory that involves the remaining 3MB, we get it with zero overhead (when

they happen). Since our framework has memory allocation and deletion spread all over, as opposed to all of it happening in one place in the beginning of the code, the caching allocator provides around 32% improvement in compute time.

Reducing memory footprint - Pre-Calculation and Half Precision

Since we have a non-trivial number of intermediate objects in our journey from hits to track candidates (mini-doublets, segments etc), we require ample memory to store all the information pertaining to these objects as intermediates in construction. The total memory consumption adds quickly. In addition, we cannot rely on containers like STL Vectors to dynamically create and destroy memory, as they carry significant timing overheads. So, we pre-calculate the amount of memory that will be needed by the objects. For example, if we know beforehand that in 9.99% of the time, the total number of mini-doublets that can be created in a module is less than 100, we can allocate 100 memory locations for the mini-doublet quantities for each module. In fact, we can hyper-optimize the allocation by splitting the detector into various regions and computing the typical occupancies of objects in these regions. These small optimizations add up and reduce the memory footprint by factors of 2 or 3, bringing the total memory required down from 4-5 GB to 1-2GB per event.

We can also take advantage of Cuda's 16 bit half precision data types. To ensure efficient conversion of full precision 32 bit floats to half precision 16 bit floats and vice-versa, Cuda provides convenient functions `__float2half` and `__half2float`. Using half precision reduces the memory footprint for some objects by around 50%. Currently an event occupies approx 700MB in memory. With newer GPUs having Tensor cores that can efficiently work on 16 bit floats directly, a future plan of work will involve trying to take advantage of these Tensor cores.

6.6.2 Streams and Multi-streaming

Classic Multi-streaming implies event level parallelization, which provides for tracks in multiple collision events to get reconstructed in parallel but the sheer amount of compute cores required for object creation in our case limits this application. However, if we can use the times the GPU is not active while in the process of working on one event (i.e., GPU downtimes) to do part of the work from another event, we can constantly give work to the GPUs and can reduce the overall track reconstruction time averaged over many events. This relies on the caveat that the requisite input data are already present in the GPU, which is where the innovations used to reduce the memory footprint discussed in the previous section is of use. Figure 6.19 shows the timeline of GPU usage when collision events get processed sequentially. The situation we have is that all the compute cores in the GPU are used for a period of time, followed by an interval of little to no work, which is when preparations for the next cycle of high intensity workload happens. For example, after the GPU has constructed all the triplets, the compute cores have no work till the preparations for Quintuplet reconstruction are done. Our implementation of multi-streaming attaches events to streams and distributes the work such that during this downtime some other work gets done, for example during the downtime between Triplet and Quintuplet reconstruction, mini-doublets for the another event can get created. Figure 6.20 shows how tasks from different streams (each row corresponds to a stream here) get tessellated. However, here are significant overlap regions where the GPU needs to work on two kernels, which means that compute cores get split between these kernels which result in a reduction in performance (and increase in time) of individual kernels, but the overall timing gets greatly improved.

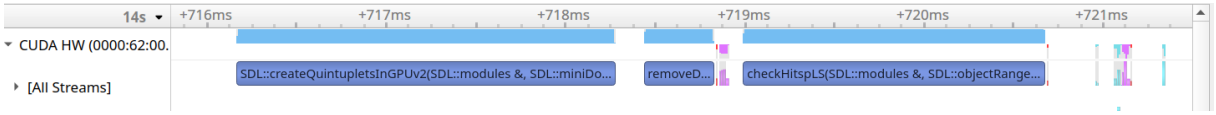


Figure 6.19. Single stream GPU workload timeline

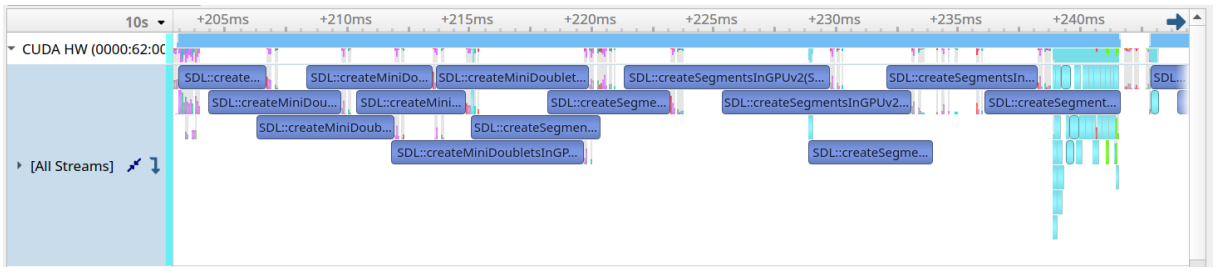


Figure 6.20. Eight streams GPU workload timeline

6.6.3 Timing Performance

Table 6.2 shows the complete timing performance with individual object breakdowns when the Line Segment Tracking GPU is used to reconstruct tracks above 0.8 GeV in a typical $t\bar{t}$ +PU200 event. Our best timing performance is 34 ms/event on a single stream and 26 ms/event on eight streams. We do not include the expensive initial copy of hits from host memory to GPU memory, nor do we include the final copying of track candidates back to the host memory. The timing performance is comparable to the latest results from the performance of the CMS Track Pattern recognition algorithms [53], which also reconstructs tracks with p_T above 0.8 GeV. Our Line Segment Tracking implementation is also comparable on the cost front with the CMS track pattern recognition algorithm scaled to 64 cores, since two 32 core Intel Skylake Gold Xeon Processors (commonly used in the multi-CPU efforts) have a similar price to a Tesla V100 GPU.

Table 6.2. Line Segment Tracking timing performance - $t\bar{t}$ + PU200, Tesla V100 GPU

Number of streams	Average time per event (ms)
1	33.7
2	27.3
4	26.2
6	26.3
8	25.7

6.7 What the future holds

While we have achieved good reconstruction efficiency, with comparable fake rates on $t\bar{t}$ +PU200 samples comparable to the reference CMS Track Pattern recognition algorithms, we still have a long way to go to be competitive. In addition, things are constantly in motion with new GPU architectures bringing new improvements and we need to keep up with the times, given that HL-LHC will not be online till 2029.

6.7.1 General Improvements

On the general physics improvement front, we will revisit some of the physics selections, looking at improving the various cuts to reduce the combinatorial background and improve the efficiency even further. On this front, our future work will also go towards improving the reconstruction efficiency of displaced tracks.

On the optimization and computational performance front, we have currently only scratched the surface. We hope to work towards efficient computation of mathematical parameters for physics selection, in addition to improving memory throughput by using half precision floats directly in the computation stage itself. As outlined in section 6.6.1, the newer GPUs have dedicated Tensor cores that can work with these half precision floats efficiently. Plans to improve memory coalescing and optimizing register usage will help in improving our timing performance further. Our final target is to deploy this in the software backend for real time and offline reconstruction in time for HL-LHC.

6.7.2 Alpaka : Multi architecture code

A serious limitation involved in writing GPU specific code is that porting it to other architectures essentially involves a complete refactoring of the code. Currently, the line of devices that Cuda targets is so narrow (only Nvidia line GPUs) that Cuda compatible

code does not run on any other architectures, including AMD GPUs. The cost of porting is paid in graduate student hours (not very expensive) and also carries with it a very high probability of making crucial errors (very expensive).

Alpaka [54, 55, 56] is a header only library that aims to provide parallelism by abstracting the nitty-gritty details away. To this end, Alpaka creates generic parallelization constructs like threads, blocks and grids for the computational units, and global, shared and register memory for the memory units. Software applications are written using these constructs instead of directly invoking vendor or architecture specific constructs. Alpaka backends then provide the necessary translation layers between the abstract modules and the architecture specific modules. Since these can be done at compile and run-time, a single executable can be used to run on multiple architectures. Since this model effectively separates the "algorithm" from the architecture specific "implementation", it is easy to extend the library to run on newer parallel programming paradigms that target newer hardware. For example, at the time of writing this thesis there is no equivalent Cuda like API for AMD GPUs. When such a thing is released, Alpaka just needs a new backend, and a Line Segment Tracking Alpaka implementation can run without a hitch.

6.7.3 Graph Neural Networks

The Line Segment Tracking algorithm is essentially a rules based approach to linking smaller track stubs to create larger ones. It is up to us to ensure that all exceptions are taken care of. The base framework can, however, be used for a machine learning based approach to track reconstruction using Graph Neural Networks.

Graph Neural Network approaches to track reconstruction is an active area of research[57, 58]. The advantage with approaching this from the context of Line Segment

Tracking is that a lot of the building blocks are already present conceptually, like constructing Mini-doublets first, then to Line Segments, followed by Triplets, etc. Technical developments like module maps have also been conceptualized and done in this context. In this possible approach using Graph Neural Networks, we can start from the mini-doublets rather than the hits. By changing the starting points, we have already incorporated a bit of physics knowledge which we do not know if the neural network will learn if we start from the hits. Similarly, the module map that we use has been derived from first principles without the assistance of simulations or the neural network itself, which by itself will be a crucial improvement over relying on incorrect simulations or hoping that the magic of nodes and edges and gradient descent [3] will learn. We think that with these starting points and additional technical details, the training can be sped up and the quality of results can be improved.

6.8 Acknowledgments

This chapter described the Line Segment Tracking algorithm. It is a reproduction of the paper : "Line Segment Tracking in the HL-LHC", arXiv:2207.08207.

This work is supported by the U.S. National Science Foundation under Cooperative Agreements OAC-1836650 and PHY-2121686 and grant NSF-PHY-1912813.

[3] and hyperparameter tuning using the time-tested and cost-efficient method of Graduate Student Descent

Appendix A

$R_{SF/DF}$ Determination

The $R_{SF/DF}$ factor is derived to account for the differences in reconstruction and trigger efficiencies arising between the different flavor (eμ) and same flavor (ee + $\mu\mu$) events, so that the different flavor yields can be used to estimate the flavor symmetric backgrounds. This factor is roughly parametrized into two - one depending only on the selection and reconstruction differences, and the other depending only on the trigger efficiencies

A.1 Parametrization

The following convention is used to label variables. A quantity without an upper index denotes all efficiencies (trigger, selection and reconstruction), a "*" denotes only reconstruction and selection efficiencies are applied, "hard" implies the quantity at the particle level (before selection and reconstruction), and "T" implies trigger efficiencies. In addition, we assume that the two leptons are independently reconstructed and hence the total efficiency of a di-lepton process is just the product of the two efficiencies.

The ratio of the total efficiencies between the muons and electrons $r_{\mu/e}$ is given as

$$r_{\mu/e} = \frac{\epsilon_\mu}{\epsilon_e} \approx \sqrt{\frac{\epsilon^*_{\mu\mu}\epsilon^T_{\mu\mu}}{\epsilon^*_{ee}\epsilon^T_{ee}}} \tag{A.1}$$

$$r_{\mu/e} \approx \frac{\epsilon^*_\mu}{\epsilon^*_e}\sqrt{\frac{\epsilon^T_{\mu\mu}}{\epsilon^T_{ee}}} \approx r^*_{\mu/e}\sqrt{\frac{\epsilon^T_{\mu\mu}}{\epsilon^T_{ee}}} \tag{A.2}$$

The total number of ee and $\mu\mu$ events can be estimated as

$$\begin{aligned} N_{ee} &= \epsilon^T_{ee} N^*_{ee} && \text{(A.3)}\\ &= \epsilon^T_{ee}(\epsilon^*_e)^2 N^{hard}_{ee} && \text{(A.4)}\\ &= \frac{1}{2}\epsilon^T_{ee}(\epsilon^*_e)^2 N^{hard}_{OF} && \text{(A.5)}\\ &= \frac{1}{2}\epsilon^T_{ee}\frac{\epsilon^*_e}{\epsilon^*_\mu} N^*_{OF} && \text{(A.6)}\\ &= \frac{1}{2}\frac{1}{r^*_{\mu/e}}\frac{\epsilon^T_{ee}}{\epsilon^T_{e\mu}} N_{OF} = \frac{1}{2}\frac{1}{r_{\mu/e}}\frac{\sqrt{\epsilon^T_{ee}\epsilon^T_{\mu\mu}}}{\epsilon^T_{e\mu}} N_{OF} && \text{(A.7)} \end{aligned}$$

$$\begin{aligned} N_{\mu\mu} &= \epsilon^T_{\mu\mu} N^*_{\mu\mu} && \text{(A.8)}\\ &= \epsilon^T_{\mu\mu}\epsilon^{*2}_{\ \mu} N^{hard}_{\mu\mu} && \text{(A.9)}\\ &= \frac{1}{2}\epsilon^T_{\mu\mu}\epsilon^{*2}_{\ \mu} N^{hard}_{OF} && \text{(A.10)}\\ &= \frac{1}{2}\epsilon^T_{\mu\mu}\frac{\epsilon^*_\mu}{\epsilon^*_e} N^*_{OF} && \text{(A.11)}\\ &= \frac{1}{2} r^*_{\mu/e}\frac{\epsilon^T_{\mu\mu}}{\epsilon^T_{e\mu}} N_{OF} = \frac{1}{2} r_{\mu/e}\frac{\sqrt{\epsilon^T_{ee}\epsilon^T_{\mu\mu}}}{\epsilon^T_{e\mu}} N_{OF} && \text{(A.12)} \end{aligned}$$

The same flavor complete yield can be computed as

$$\begin{aligned} N_{SF} &= \frac{1}{2}\left(r_{\mu/e} + \frac{1}{r_{\mu/e}}\right)\frac{\sqrt{\epsilon^T_{ee}\epsilon^T_{\mu\mu}}}{\epsilon^T_{e\mu}} N_{OF} && \text{(A.13)}\\ &= \frac{1}{2}\left(r_{\mu/e} + \frac{1}{r_{\mu/e}}\right)\mathrm{R_T} N_{OF} = \mathrm{R_{SF/DF}} N_{OF} && \text{(A.14)} \end{aligned}$$

The factors $r_{\mu/e}$ and $\mathrm{R_T}$ will be measured in data.

A.2 $r_{\mu/e}$ Derivation

The measurement is performed in a Drell-Yan control region with $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}} > 50$ GeV and at least two jets, in addition to a di-lepton mass cut 60 GeV ¡ $M_{\ell\ell} < 120$ GeV. This region is enriched in ee and $\mu\mu$ pairs from boson. This factor $r_{\mu/e}$ is known to depend on lepton $\mathrm{p_T}$ and η, and is derived separately for 2016, 2017 and 2018. The dependency function is as follows

$$
\begin{aligned}
r_{\mu/e}(\ell) &= r^0_{\mu/e} \cdot f(\mathrm{p_T}) \cdot g(\eta) \\
f(\mathrm{p_T}) &= (a_1 + b_1/\mathrm{p_T}) \\
g(\eta) &= a_2 + \begin{cases} 0 & |\eta| < 1.6 \\ c_1 \cdot (\eta - 1.6)^2 & \eta > 1.6 \\ c_2 \cdot (\eta + 1.6)^2 & \eta < -1.6 \end{cases}
\end{aligned}
$$

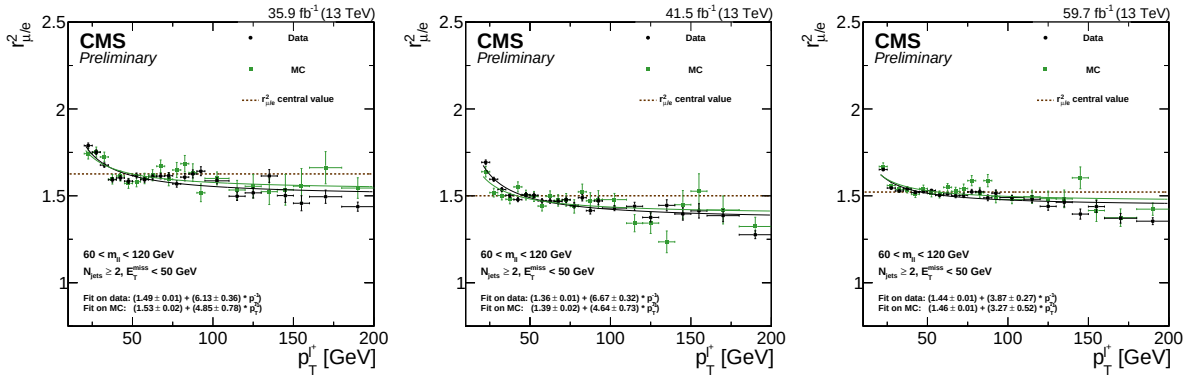
The variables $r^0_{\mu/e}$, a_1, a_2, b_1, c_1 and c_2 will be derived in data. Separate fits are done for $f(\mathrm{p_T})$ and $g(\eta)$ respectively. Fits are performed on $r^2_{\mu/e}$ distributions

$$
r^2_{\mu/e}(\ell^+, \ell^-) = r_{\mu/e}(\ell^+) \cdot r_{\mu/e}(\ell^-) \tag{A.15}
$$

By marginalizing over the negative lepton, the factor $r_{\mu/e}(\ell^+)$ is given by

$$
r^2_{\mu/e}(\ell^+) = r_{\mu/e}(\ell^+) \cdot \int r_{\mu/e}(\ell^-) d\ell^- = r_{\mu/e}(\ell^+) \cdot \overline{r_{\mu/e}}
$$

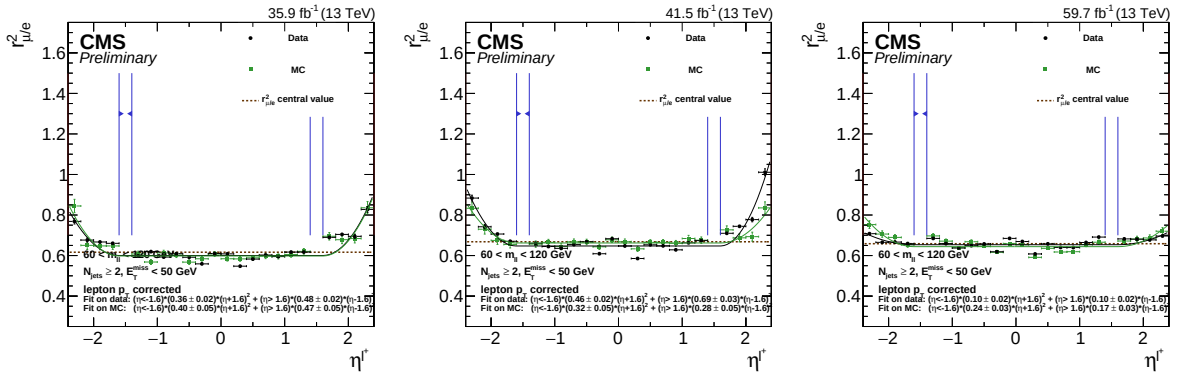
The fits are done to $r^2_{\mu/e}(\ell^+)$ against the $\mathrm{p_T}$ of the positive lepton, split by three years (2016, 2017, 2018). These fits are shown in Figure A.1. Consistent results can be seen by performing the fits separately with data and Monte-Carlo samples Figure A.2 shows the fit against η split by year to derive the parameters for the quantity $g(\eta)$. To remove the $\mathrm{p_T}$



(a) 2016 **(b)** 2017 **(c)** 2018

Figure A.1. Fit of $r^2_{\mu/e}(\ell^+)$ vs $\mathrm{p_T}$ of positively charged lepton for the three Run-II years

dependencies, each di-electron event is weighed by the momentum factors $f(p_{T,1}) \cdot f(p_{T,2})$. This affects the normalization of $g(\eta)$ by a factor of $\overline{r_{\mu/e'}}^{-4}$, with an additional factor of $\overline{r_{\mu/e''}}^{2}$ from fitting $r^2_{\mu/e}$ to η. These two factors by themselves do not completely account for



(a) 2016 **(b)** 2017 **(c)** 2018

Figure A.2. Fit of $r^2_{\mu/e}(\ell^+)$ vs η of positively charged lepton for the three Run-II years

$r_{\mu/e}$, since they were fitted essentially on $r^2_{\mu/e}$. A pre-factor, $r^0_{\mu/e}$ is derived as the central value of $r_{\mu/e}$. This is derived by computing the ratio of the square root of ee and $\mu\mu$ events in the Drell-Yan control region after weighting them by $f(p_{T,1}) \cdot f(p_{T,2}) \cdot g(\eta_1) \cdot g(\eta_2)$. Table A.1 lists the results of this computation and the complete set of parameters split by years

Table A.1. Results of the fits of $r_{\mu/e}$ as a function of p_T and η of the positive lepton in the Drell–Yan control region. The same quantities derived from simulation are shown for comparison. Only statistical uncertainties are given.

Year		$r^0_{\mu/e}$	a_1	b_1	a_2	c_1	c_2
2016	Data	1.277±0.001	1.493±0.008	6.135±0.364	0.600±0.001	0.356±0.022	0.476±0.024
	Simulation	1.279±0.002	1.528±0.018	4.854±0.783	0.598±0.003	0.403±0.052	0.467±0.050
2017	Data	1.226±0.001	1.356±0.008	6.665±0.325	0.647±0.002	0.462±0.024	0.690±0.027
	Simulation	1.220±0.001	1.389±0.017	4.637±0.728	0.662±0.004	0.316±0.052	0.284±0.050
2018	Data	1.234±0.001	1.437±0.006	3.870±0.266	0.653±0.001	0.097±0.015	0.099±0.015
	Simulation	1.240±0.001	1.464±0.012	3.274±0.517	0.644±0.002	0.242±0.033	0.172±0.031

A.2.1 $r_{\mu/e}$ Closure Test

As a closure test, each ee event in the Drell-Yan enriched region is weighed by $r^2_{\mu/e}$ and a new ratio $r^{\text{corr}}_{\mu/e}$ is computed by dividing the number of $\mu\mu$ events with the weighted ee event yield. This ratio should be very close to 1 and independent of the p_T and η of the leptons. Figure A.3 shows the dependence of $r_{\mu/e}$ without any parametrization, while Figure A.4 shows the same with the parametrization applied. A small correlation in the p_T distributions is seen in the low p_T region, since the p_T of the "leading" and "sub-leading" leptons are highly correlated (by definition). Since only a small fraction of events has this correlation, this is ignored and considered in the systematic uncertainties. Three different systematic uncertainties are taken : A flat 5% uncertainty, an uncertainty that depends on η, given by $(5\% \cdot \frac{|\eta|-1.2}{1.2})$ and a p_T dependent uncertainty, given by $(5\% \cdot \frac{110-p_T}{90})$, which account for the residual dependencies observed in the Figures A.4- A.6. The plots contain a single uncertainty band that covers all these uncertainties together. In the plots with uncorrected $r_{\mu/e}$, a flat 10% uncertainty is taken for illustration purposes alone.

Since the measurement of $r_{\mu/e}$ is performed in a control region different from the signal regions, the extrapolation into the signal regions is validated in Figures A.5 and A.6. This method is also validated by studying the dependency of $0.5(r^{corr.}_{\mu/e} + 1/r^{corr.}_{\mu/e})$ on some features of the events using the parametrization. Figure A.7 shows the results of these

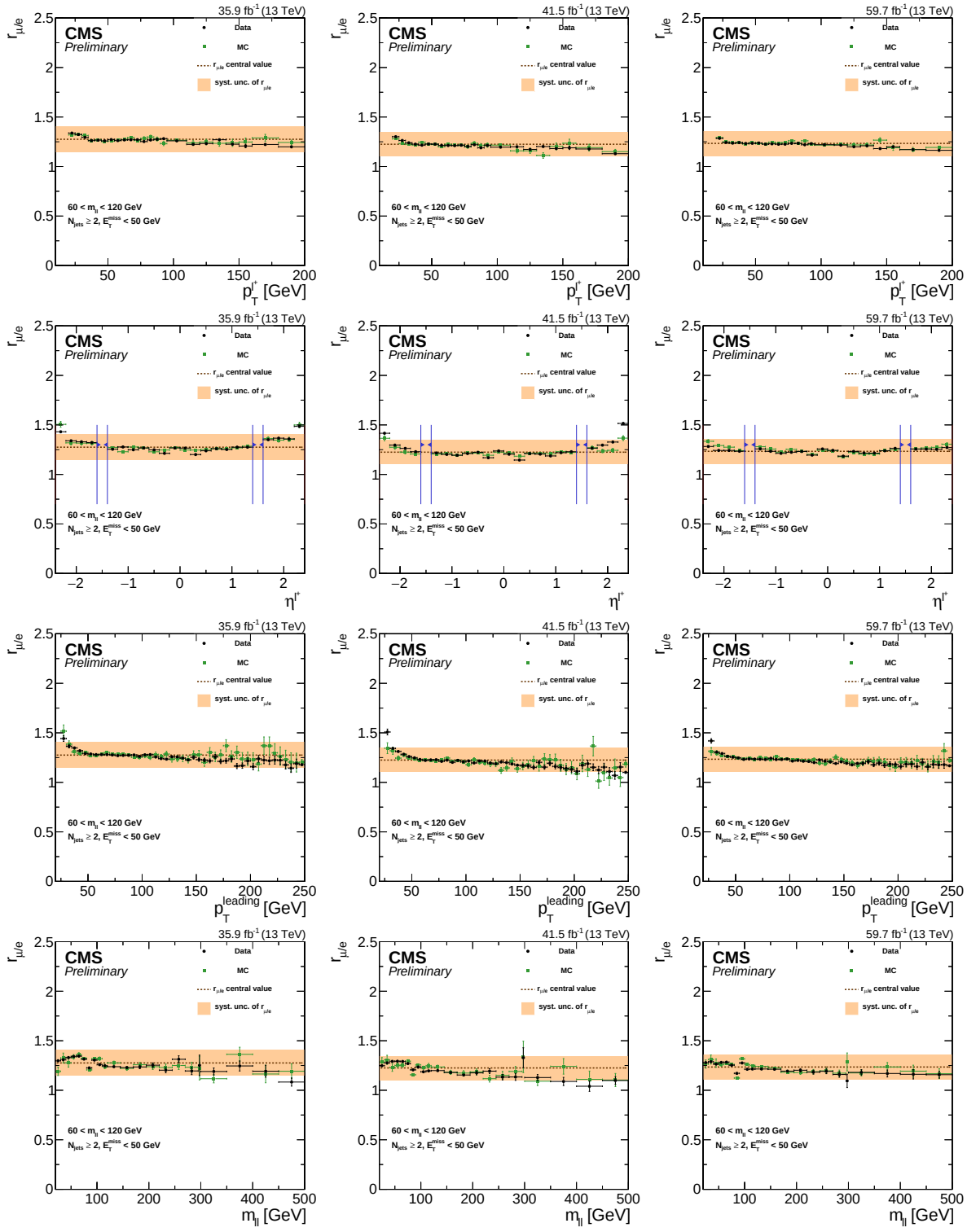


Figure A.3. From top to bottom: $r_{\mu/e}$ vs $\mathrm{p_T}$ and η of the positive lepton, leading lepton $\mathrm{p_T}$ and $M_{\ell\ell}$, without parametrization. From left to right: 2016, 2017 and 2018 data sets.

studies on some features. This quantity is very stable, showing no trends with respect to N_{jets}, $\text{E}_{\text{T}}^{\text{miss}}$, and M_{T2}, validating the extrapolation from the Drell–Yan control region into the signal regions.

A.3 R_T Derivation

The variable R_T depends on the trigger efficiencies, which are measured using a control sample of dilepton events collected with Particle Flow $\text{E}_{\text{T}}^{\text{miss}}$ triggers with a 120 GeV threshold. The efficiency is calculated as the fraction of events in this sample that also pass the corresponding di-lepton triggers (ee, $\mu\mu$, $e\mu$).

$$\epsilon_{trigger} = \frac{\text{Lepton pair } \cap PFMET \text{ trigger } \cap \text{ Dilepton trigger}}{\text{Lepton pair } \cap PFMET \text{ trigger}} \tag{A.16}$$

No minimum jet multiplicity is required, but all events with $N_{\text{jets}} > 2$, $\text{E}_{\text{T}}^{\text{miss}} > 150$ GeV and $\text{M}_{\text{T2}} > 80$ GeV are vetoed to exclude the signal regions. In addition a minimum $\text{E}_{\text{T}}^{\text{miss}}$ of 130 GeV is required to keep the PFMET triggers efficient and be orthogonal to the Drell-Yan control region. Separate efficiency factors are derived for each year to account for any trigger differences from one year to another.

Table A.2 shows the resulting trigger efficiencies in data and Monte-Carlo simulations. The observed efficiencies vary from 86% in 2017 $e\mu$ data to 98% in 2016 $\mu\mu$ Monte-Carlo samples. A systematic uncertainty of 3% for each trigger efficiency is assigned, which covers most of the disagreements between data and Monte-Carlo simulations. This results in an uncertainty of about 4% on R_T which is enough to account for disagreements between data and simulation and between the three years. Figure A.8 shows the dependence of R_T on different observed features. NO significant dependence of R_T on any event feature is observed and the 4% uncertainty is seen to be sufficient to cover fluctuations in

data and remaining trends in simulation. .

Table A.2. Trigger efficiency values for data and simulation with opposite sign, $p_T >$ 25(20) GeV and $\mathrm{E_T^{miss}} > 130$ GeV.

	2016 data set					
	Data			Simulation		
	numerator	denominator	$\epsilon_{trigger} \pm \sigma_{stat}$	numerator	denominator	$\epsilon_{trigger} \pm \sigma_{stat}$
ee	6890	7521	0.916±0.003	7549	8087	0.933±0.001
$\mu\mu$	8022	8455	0.949±0.003	9431	9624	0.980±0.001
eμ	13663	15357	0.890±0.003	15273	16654	0.917±0.001
R_T	1.048±0.043			1.043±0.041		
	2017 data set					
	Data			Simulation		
	numerator	denominator	$\epsilon_{trigger} \pm \sigma_{stat}$	numerator	denominator	$\epsilon_{trigger} \pm \sigma_{stat}$
ee	7342	8101	0.906±0.003	9337	9775	0.955±0.001
$\mu\mu$	7754	8879	0.873±0.004	9735	10285	0.947±0.001
eμ	13735	16009	0.858±0.003	17710	19075	0.928±0.001
R_T	1.037±0.044			1.024±0.040		
	2018 data set					
	Data			Simulation		
	numerator	denominator	$\epsilon_{trigger} \pm \sigma_{stat}$	numerator	denominator	$\epsilon_{trigger} \pm \sigma_{stat}$
ee	10231	11082	0.923±0.003	12300	12960	0.949±0.001
$\mu\mu$	10631	11495	0.925±0.003	13279	13927	0.953±0.001
eμ	19087	21251	0.898±0.002	23848	25740	0.926±0.001
R_T	1.029±0.042			1.027±0.040		

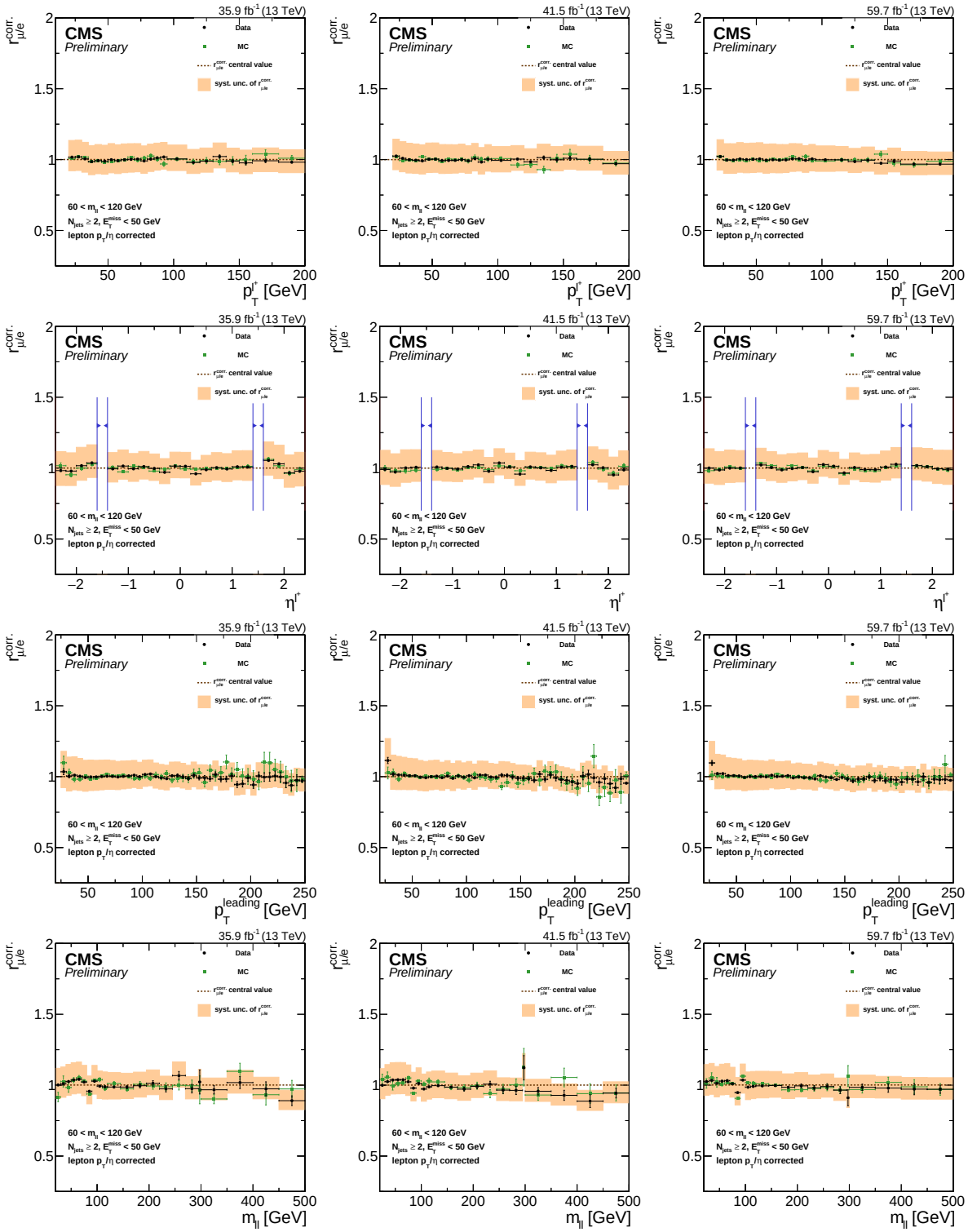


Figure A.4. From top to bottom: $r^{\text{corr}}_{\mu/e}$ dependency with parametrization applied on the p_T and η of the positive lepton, leading lepton p_T and $M_{\ell\ell}$. From left to right: 2016, 2017 and 2018 data sets.

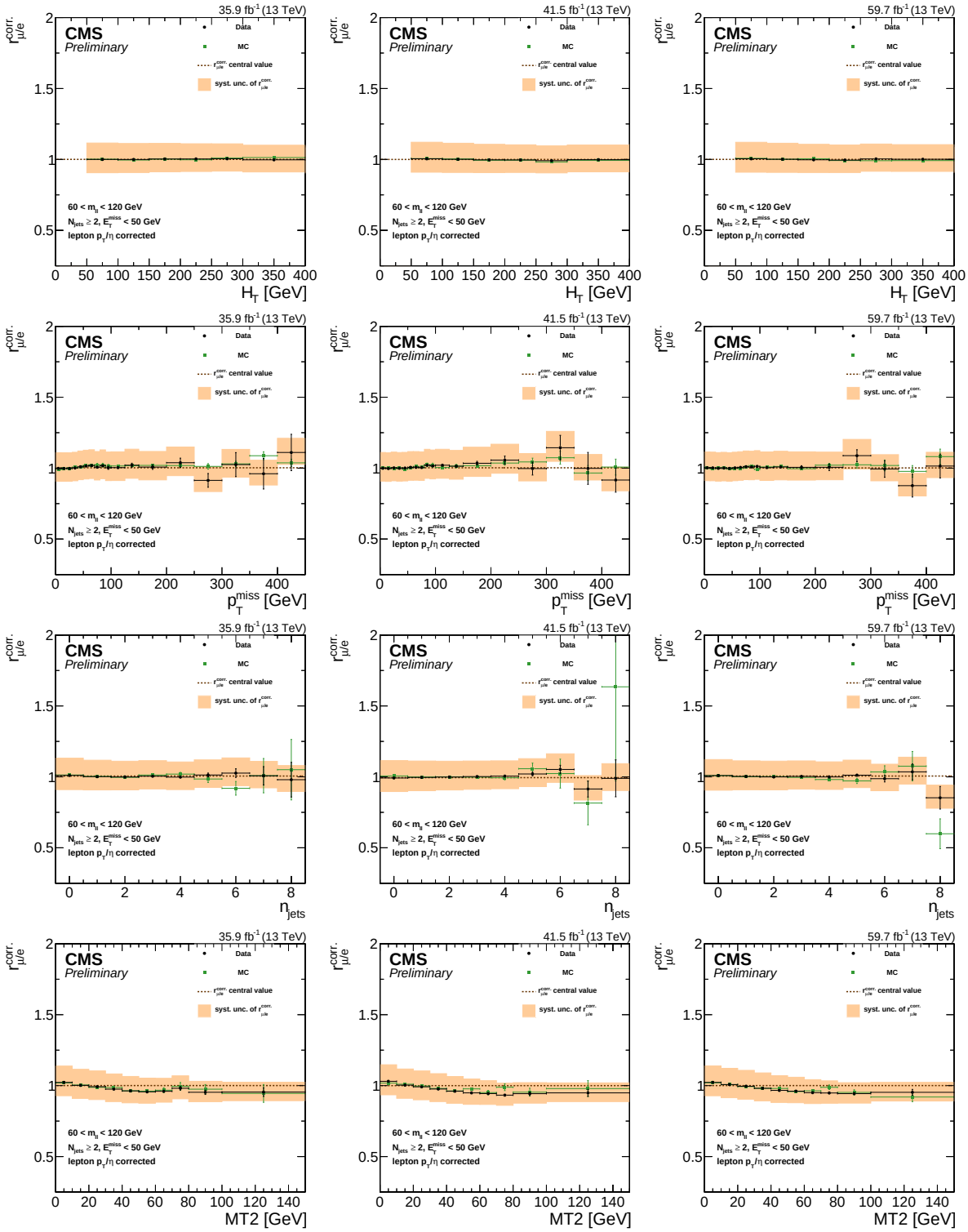


Figure A.5. From top to bottom: $r^{corr.}_{\mu/e}$ vs $\mathrm{H_T}$, $\mathrm{E_T^{miss}}$, N_{jets} and $\mathrm{M_{T2}}$ for data and MC with parameterization applied to the ee events. From left to right: 2016, 2017 and 2018 data sets. The systematic uncertainty is indicated by the orange band.

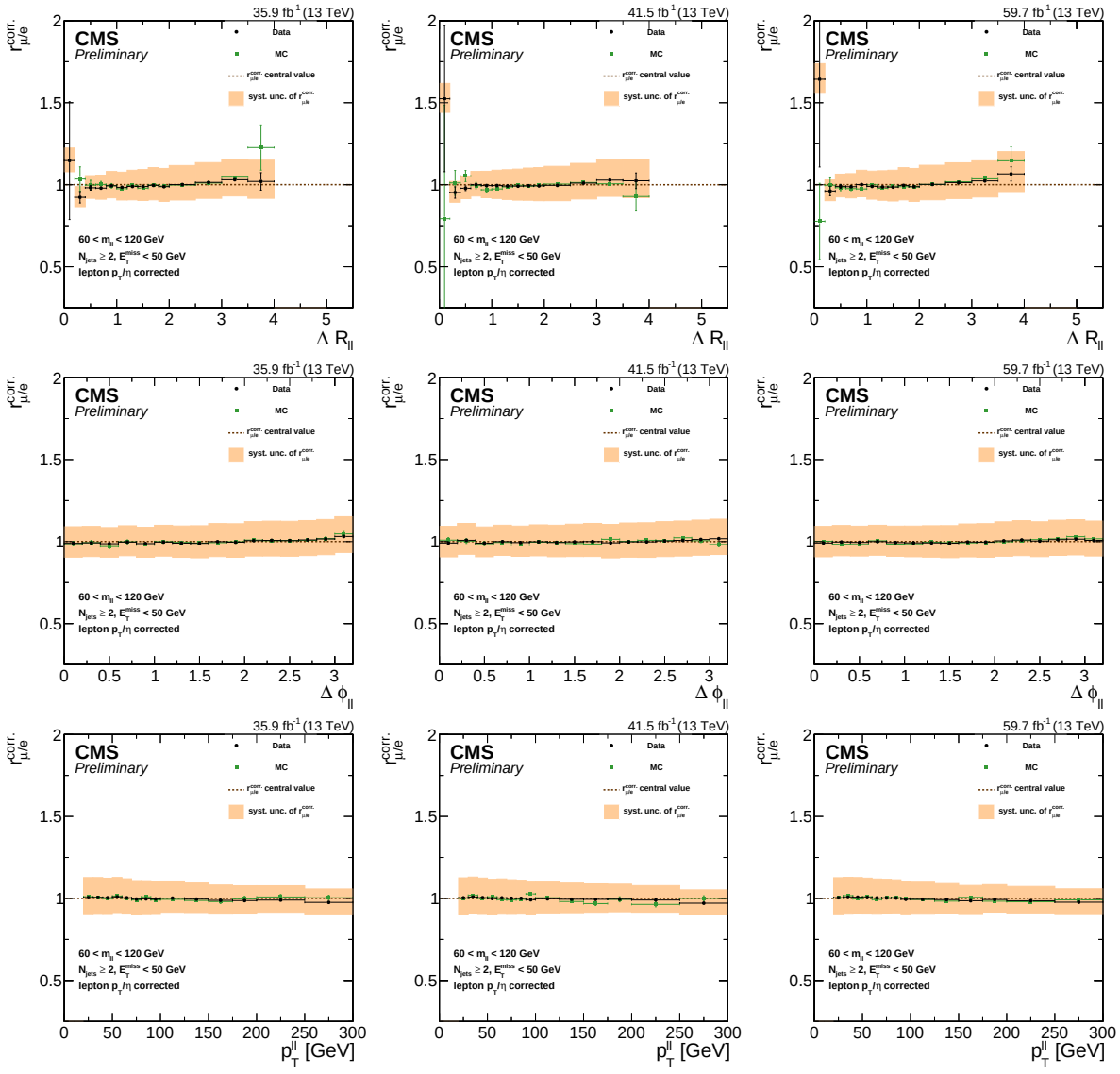


Figure A.6. From top to bottom: $r_{\mu/e}$ vs ΔR and $\Delta\phi$ between the leptons, and di-lepton p_T in data and MC with parametrization. From left to right: 2016, 2017 and 2018 data sets. The systematic uncertainty is indicated by the orange band.

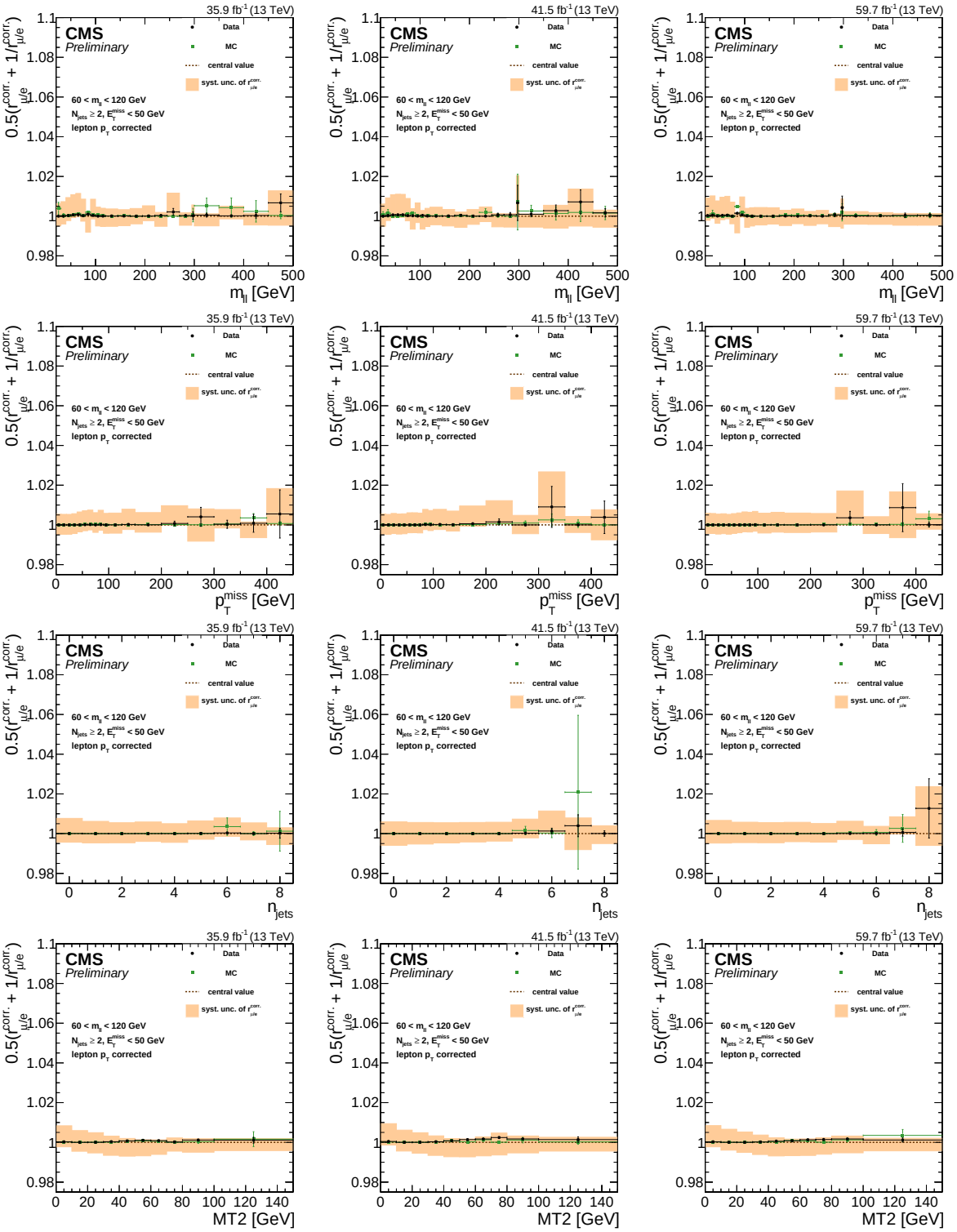


Figure A.7. From top to bottom: $0.5(r^{corr.}_{\mu/e} + 1/r^{corr.}_{\mu/e})$ dependence on $M_{\ell\ell}$, $\mathrm{E_T^{miss}}$, N_{jets} and $\mathrm{M_{T2}}$ for data and MC with parameterization. From left to right: 2016, 2017 and 2018 data sets. The systematic uncertainty is indicated by the orange band.

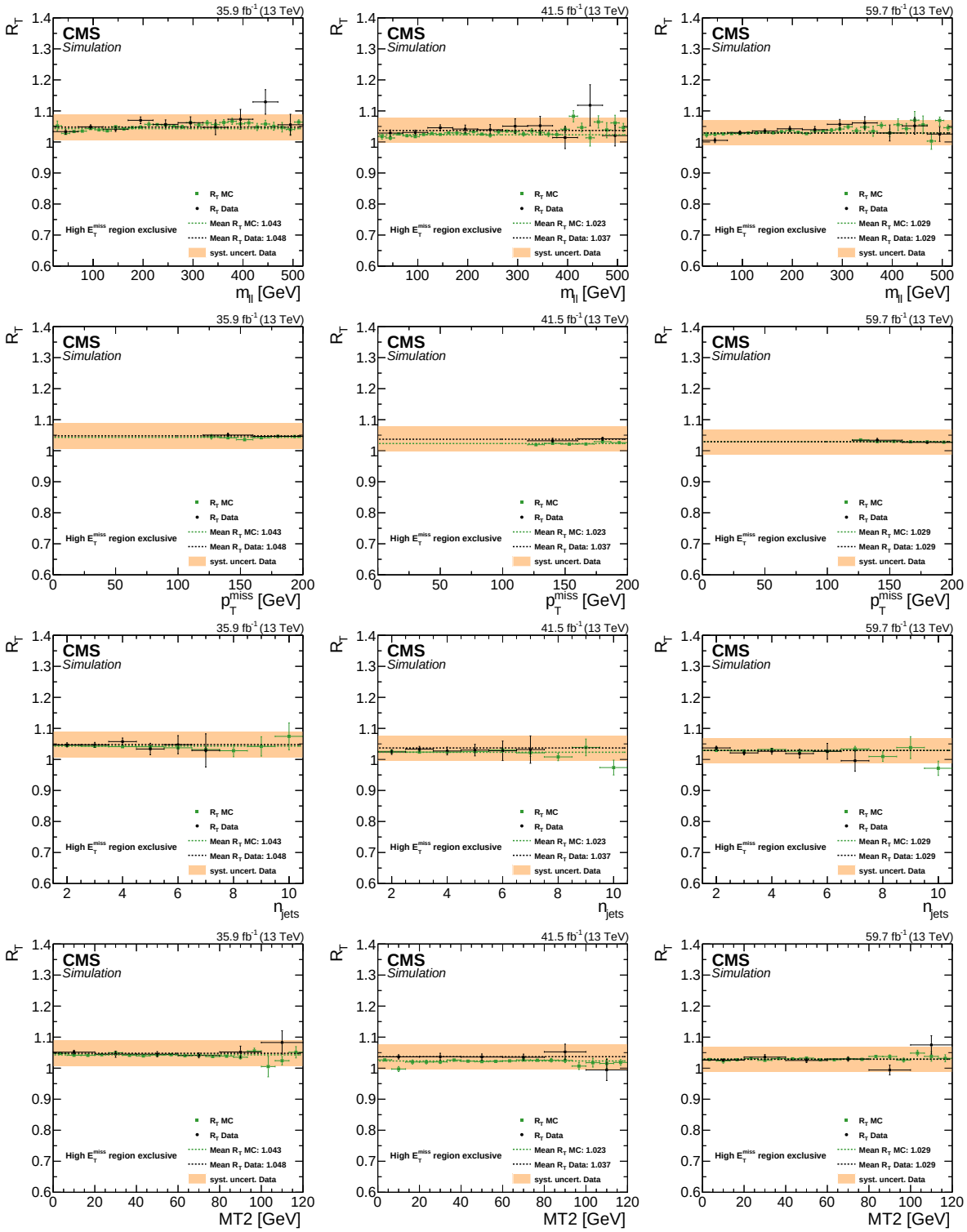


Figure A.8. From top to bottom: Dependency of trigger efficiency ratio on the $M_{\ell\ell}$, $\mathrm{E_T^{miss}}$, N_{jets} and $\mathrm{M_{T2}}$ for data and simulation. From left to right: 2016, 2017 and 2018 data sets. The 4% systematic uncertainty is indicated by the orange band.

Bibliography

[1] "The standard model: a beautiful but flawed theory - quantum diaries." https://www.quantumdiaries.org/2014/03/14/the-standard-model-a-beautiful-but-flawed-theory/.

[2] "Supersymmetric glue : The search for gluinos - cern." https://cms.cern/news/supersymmetric-glue-search-gluinos.

[3] "Unsolved problems in physics: The hierarchy problem." https://platosrealm.blog/2019/02/22/unsolved-problems-in-physics-the-hierarchy-problem/.

[4] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, "Parton distributions from LHC, HERA, tevatron and fixed target data: MSHT20 PDFs," *The European Physical Journal C*, vol. 81, apr 2021.

[5] "Facts and figures about the LHC—CERN." https://home.cern/resources/faqs/facts-and-figures-about-lhc.

[6] "CMS Luminosity - Public Results." https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[7] "Cms experiment." https://cds.cern.ch/record/1474902.

[8] "Cutaway diagrams of CMS detector." https://cds.cern.ch/record/2665537.

[9] "CMS Public Detector Performance Group Results." https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK.

[10] "CMS Detector Slice." http://cds.cern.ch/record/2120661.

[11] "Event displays of top quark pair production in CMS collected in 2015." https://cds.cern.ch/record/2712881.

[12] G. B. Cerati, "Tracking and vertexing algorithms at high pileup," tech. rep., CERN, Geneva, Oct 2014.

[13] "The Phase-2 Upgrade of the CMS Tracker," tech. rep., CERN, Geneva, Jun 2017.

[14] "Cuda C++ Programming Guide." https://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html.

[15] CMS Collaboration, "Search for supersymmetry in final states with two oppositely charged same-flavor leptons and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV," *Journal of high energy physics : JHEP.*, vol. 2021, no. 4, 2021-04.

[16] C. F. Madrazo, P. M. R. del Arbol, M. Masciovecchio, S. S. Cruz, C. R. Alvarez, B. S. Narayanan, M. Teroerde, , and G. Z. D. Porta, "Search for new phenomena in events with an opposite-sign same-flavor pair with the run 2 dataset," CMS Note 2018/277, 2020.

[17] Super-Kamiokande Collaboration, "Evidence for oscillation of atmospheric neutrinos," *Physical Review Letters*, vol. 81, pp. 1562–1567, aug 1998.

[18] KATRIN Experiment, "First direct neutrino-mass measurement with sub-ev sensitivity," 2021.

[19] F. Zwicky, "Die Rotverschiebung von extragalaktischen Nebeln," *Helvetica Physica Acta*, vol. 6, pp. 110–127, Jan. 1933.

[20] V. Trimble, "Existence and nature of dark matter in the universe.," *Annual Review of Astronomy and Astrophysics*, vol. 25, pp. 425–472, Jan. 1987.

[21] R. N. Mohapatra, "Supersymmetry and R-parity: an Overview," 2015.

[22] CMS Collaboration, "The CMS experiment at the CERN LHC," *JINST*, vol. 3, p. S08004, 2008.

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

[24] 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, Geneva: CERN, 1997.

[25] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector," *Journal of Instrumentation*, vol. 12, pp. P10003–P10003, oct 2017.

[26] CMS Collaboration, "Performance of the reconstruction and identification of high-momentum muons in proton-proton collisions at $\sqrt{s} = 13$ TeV," *Journal of Instrumentation*, vol. 15, pp. P02027–P02027, feb 2020.

[27] CMS Collaboration, "Electron and Photon performance using data collected by CMS at s = $\sqrt{s} = 13$ TeV and 25ns," *CMS-DP-2015-067*, Dec 2015.

[28] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *JHEP*, vol. 04, p. 063, 2008.

[29] J. Thaler and K. V. Tilburg, "Identifying boosted objects with n-subjettiness," *Journal of High Energy Physics*, vol. 2011, mar 2011.

[30] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, "Soft drop," *Journal of High Energy Physics*, vol. 2014, may 2014.

[31] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV," *Journal of Instrumentation*, vol. 13, pp. P05011–P05011, may 2018.

[32] CDF Collaboration, "High-precision measurement of the ¡i¿w¡/i¿ boson mass with the cdf ii detector," *Science*, vol. 376, no. 6589, pp. 170–176, 2022.

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

[34] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms," *JHEP*, vol. 11, p. 040, 2004.

[35] C. Bierlich, S. Chakraborty, N. Desai, L. Gellersen, I. Helenius, P. Ilten, L. Lönnblad, S. Mrenna, S. Prestel, C. T. Preuss, T. Sjöstrand, P. Skands, M. Utheim, and R. Verheyen, "A comprehensive guide to the physics and usage of pythia 8.3," 2022.

[36] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, F. Behner, L. Bellagamba, J. Boudreau, L. Broglia, A. Brunengo, H. Burkhardt, S. Chauvie, J. Chuma, R. Chytracek, G. Cooperman, G. Cosmo, P. Degtyarenko, A. Dell'Acqua, G. Depaola, D. Dietrich, R. Enami, A. Feliciello, C. Ferguson, H. Fesefeldt, G. Folger, F. Foppiano, A. Forti, S. Garelli, S. Giani, R. Giannitrapani, D. Gibin, J. Gómez Cadenas, I. González, G. Gracia Abril, G. Greeniaus, W. Greiner, V. Grichine, A. Grossheim, S. Guatelli, P. Gumplinger, R. Hamatsu, K. Hashimoto, H. Hasui, A. Heikkinen, A. Howard, V. Ivanchenko, A. Johnson, F. Jones, J. Kallenbach, N. Kanaya, M. Kawabata, Y. Kawabata, M. Kawaguti, S. Kelner, P. Kent, A. Kimura, T. Kodama, R. Kokoulin, M. Kossov, H. Kurashige, E. Lamanna, T. Lampén, V. Lara, V. Lefebure, F. Lei, M. Liendl, W. Lockman, F. Longo, S. Magni, M. Maire, E. Medernach, K. Minamimoto, P. Mora de Freitas, Y. Morita, K. Murakami, M. Nagamatu, R. Nartallo, P. Nieminen, T. Nishimura, K. Ohtsubo, M. Okamura, S. O'Neale, Y. Oohata, K. Paech, J. Perl, A. Pfeiffer, M. Pia, F. Ranjard, A. Rybin, S. Sadilov, E. Di Salvo, G. Santin, T. Sasaki, N. Savvas, Y. Sawada, S. Scherer, S. Sei, V. Sirotenko, D. Smith, N. Starkov, H. Stoecker, J. Sulkimo, M. Takahata, S. Tanaka, E. Tcherniaev, E. Safai Tehrani, M. Tropeano, P. Truscott, H. Uno, L. Urban, P. Urban, M. Verderi, A. Walkden, W. Wander,

H. Weber, J. Wellisch, T. Wenaus, D. Williams, D. Wright, T. Yamada, H. Yoshida, and D. Zschiesche, "Geant4—a simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250–303, 2003.

[37] S Abdullin and P Azzi and F Beaudette and P Janot and A Perrotta (on behalf ofthe CMS Collaboration), "The fast simulation of the cms detector at lhc," *Journal of Physics: Conference Series*, vol. 331, p. 032049, dec 2011.

[38] K. Cranmer, "Practical statistics for the lhc," 2015.

[39] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *The European Physical Journal C*, vol. 71, feb 2011.

[40] T. Junk, "Confidence level computation for combining searches with small statistics," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 434, pp. 435–443, sep 1999.

[41] A. L. Read, "Presentation of search results: The CL(s) technique," *J. Phys. G*, vol. 28, pp. 2693–2704, 2002.

[42] "CMS Higgs Combination toolkit." https://github.com/cms-analysis/HiggsAnalysis-CombinedLimit.

[43] C. G. Lester and D. J. Summers, "Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders," *Phys. Lett. B*, vol. 463, p. 99, 1999.

[44] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the cms tracker," *Journal of Instrumentation*, vol. 9, no. 10, 2014.

[45] I. B. Alonso, O. Brüning, P. Fessia, M. Lamont, L. Rossi, L. Tavian, and M. Zerlauth, *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*. CERN Yellow Reports: Monographs, Geneva: CERN, 2020.

[46] "November 2022 top500." https://www.top500.org/lists/top500/2022/11/.

[47] "spec sheet - oak ridge leadership computing facility." https://www.olcf.ornl.gov/wp-content/uploads/2019/05/frontier_specsheet.pdf.

[48] C. Ciobanu, J. Gerstenslager, J. Hoftiezer, R. Hughes, M. Johnson, P. Koehn, C. Neu, C. Sanchez, B. Winer, J. Freeman, S. Holm, J. Lewis, T. Shaw, T. Wesson, K. Bloom, D. Gerdes, J. Dawson, and W. Haberichter, "Online track processor for the cdf upgrade," *IEEE Transactions on Nuclear Science*, vol. 46, no. 4, pp. 933–939, 1999.

[49] V. Krutelyov, G. Cerati, M. Tadel, F. Wuerthwein, and A. Yagil, "Impact of tracker layout and algorithmic choices on cost of computing at high pileup," *PoS ICHEP2016, 194.*, 2016.

[50] E. S. Larsen and D. McAllister, "Fast matrix multiplies using graphics hardware," in *Proceedings of the 2001 ACM/IEEE Conference on Supercomputing*, SC '01, (New York, NY, USA), p. 55, Association for Computing Machinery, 2001.

[51] "How to manipulate data structure to optimize memory use on 32-bit intel architecture." https://www.intel.com/content/www/us/en/developer/articles/technical/how-to-manipulate-data-structure-to-optimize-memory-use-on-32-bit-intel-architecture.html.

[52] D. Merrill, "Cub v1. 5.3: Cuda unbound, a library of warp-wide, blockwide, and device-wide gpu parallel primitives," *NVIDIA Research*, 2015.

[53] "Performance of Phase-2 HLT Reconstruction and GPU offloading benchmarks," May 2021.

[54] A. Matthes, R. Widera, E. Zenker, B. Worpitz, A. Huebl, and M. Bussmann, "Tuning and optimization for a variety of many-core architectures without changing a single line of implementation code using the alpaka library," Jun 2017.

[55] E. Zenker, B. Worpitz, R. Widera, A. Huebl, G. Juckeland, A. Knüpfer, W. E. Nagel, and M. Bussmann, "Alpaka - an abstraction library for parallel kernel acceleration," IEEE Computer Society, May 2016.

[56] B. Worpitz, "Investigating performance portability of a highly scalable particle-in-cell simulation code on various multi-core architectures," Sep 2015.

[57] J. Duarte and J.-R. Vlimant, "Graph neural networks for particle tracking and reconstruction," in *Artificial Intelligence for High Energy Physics*, pp. 387–436, WORLD SCIENTIFIC, Feb 2022.

[58] A. Elabd, V. Razavimaleki, S.-Y. Huang, J. Duarte, M. Atkinson, G. DeZoort, P. Elmer, S. Hauck, J.-X. Hu, S.-C. Hsu, B.-C. Lai, M. Neubauer, I. Ojalvo, S. Thais, and M. Trahms, "Graph neural networks for charged particle tracking on FPGAs," *Frontiers in Big Data*, vol. 5, Mar 2022.