Search for Supersymmetry in proton-proton collisions at $\sqrt{s}$ =13 TeV with jets, b-jets and missing transverse momentum

A thesis

submitted in partial fulfillment of the requirements

Of the degree of

Doctor of Philosophy

By

Aditee Rane
ID No.: 20143352

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH (IISER) PUNE

2019

Dedicated to my Mother...

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Aditee Rane)
ID No.: 20143352

Date:

CERTIFICATE

Certified that the work incorporated in the thesis entitled "Search for supersymmetry in multijet events with missing transverse momentum" submitted by Aditee Rane was carried out by the candidate, under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or institution.

(Dr. Seema Sharma)

Date:

Acknowledgements

I express my sincere gratitude to my supervisor, Dr. Seema Sharma for her valuable guidance and support over the five years of this thesis work. Apart from being an academic guide, I was also fortunate to have her as a supporting and motivating person in my life; her hardworking, sincerity towards work and self belief in taking up challenges will always be respected and remain inspiration for me. The timely completion of the thesis work would not have been possible without her foresight for prioritizing various tasks.

I am thankful to the CERN CMS collaboration for providing me an excellent opportunity to dwell into some of the fundamental questions about our universe. My thanks are due to the multijet+MHT aka RA2/b SUSY group and Fast Simulation group in the CMS collaboration. During the course of the SUSY analyses, it was a great pleasure and very learning experience interacting with Dr. Bill Gary, Dr. William Ford, Dr. Jeff Richman, Dr. Owen Long, Dr. Kenichi Hatakeyama, Dr. Keith Ulmer, Dr. Andrew Whitebeck, Dr. Rishi Patel, Dr. Kevin Pedro, Dr. Kevin Stenson, Dr. Sam Bein, Dr. Ahmad Borzou, Dr. Simon Kurz and Dr. Alexx Perloff. Their help, guidance, support and contribution at various stages is just unforgettable. The feedback and inputs from Dr. Lukas Vanelderen and Dr. Sezen Sekmen helped me to contribute meaningfully to the development of fast modeling of hadron showers in CMS calorimeters.

I would like to extend my gratitude to my Research Advisory committee members, Dr. Sourabh Dube and Dr. Arun Thalapillil to help me making this work successful ! Their independent assessment and review of my progress at regular intervals helped us to identify my weaknesses as well as helped me in keeping track of my timelines.

I owe a big thank you to the IISER-Pune IT team, who constantly put sincere efforts in providing and managing a very good internet facility at IISER, which made this work possible. I would also like to thank Fermilab LHC Physics Centre and Dr. Kenichi Hatakeyama (Baylor University) for providing an access to their computing facilities and full user support without which this project would not have been completed. The IndiaCMS collaboration has provided me an excellent platform to present my work regularly and interact with faculty as well as fellow students, whom I am thankful to enrich my research experience.

I appreciate and thank IISER Pune administrative, technical and nontechnical staff for supporting very good research facilities and culture for research work, help in addressing technical issues and well maintained hostel facility which pushed me to work really good.

I would like to acknowledge the IISER-Pune for financially sponsoring my research scholarship, and the Department of Science and Technology (DST), India for supporting various collaboration meetings, SERC school and DAE symposium, and travel to CERN. Together, these provided an excellent research environment on daily basis and gave me a wide exposure over all.

Last but never the least, I would like to thank my friends and my family for being an immense source of encouragement and happiness. I owe a lot to all my friends at IISER and outside, their interactions were refreshing while tired and supportive while little stressed which cherished my productivity across five years. It has been a wonderful experience growing, struggling and learning with Mr. Vinay Hegde, Mr. Kunal Kothekar, Mr. Shubhanshu Chauhan, Mr. Anshul Kapoor, Mr. Shubham Pandey, Ms. Angira Rastogi, Ms. Bhumika Kansal from EHEP-IISER group and sharing many happy, memorable moments in a friendship with Ms. Aakanksha

Kapoor, Ms. Chinmayee Mishra, Ms. Preeti Chauhan and Ms. Sucheta Mazumdar at IISER-Pune.

And, finally I express heartfelt gratitude to my family without whom a PhD would have remained a dream. My brother, Mr. Shriram Rane, with all his unconditional love and support, has just been a call away whenever I needed ! My husband, Mr. Mayuresh Gadgil, suffered the most due to my deadlines and late working hours catching up with CERN timezones, but supported me tremendously in the most important last one-and-a-half years of this phase. My parents-in-law Mrs. Seema Gadgil and Mr. Suhas Gadgil helped me and supported me in every possible way to get this task done, and my cute sisters-in-law Mrs. Sneha Rane and Mrs. Manasi Shah, for being my very close friends.

I dedicate this thesis to my mother, Mrs. Supriya Rane, without whom my higher studies and a PhD would have remained a dream. Despite of all the blues and struggle in her own life, she made it possible that I have the freedom to choose what I want in my life, and gave me the strength to stand on my own !

Abstract

Standard Model of elementary particles (SM) explains various physical phenomena occurring in the world around us in the terms of participating fundamental particles and interactions among these particles. The particle family of the model is completed with the long sought Higgs boson of mass 125 GeV by the CMS and ATLAS experiments at the Large Hadron Collider in 2012. The SM, however, is largely accepted to be incomplete as it does not explain theoretical stability of the Higgs boson mass, or explain observed dark matter in the universe, or incorporate gravity to name a few. Many beyond SM theories are postulated to overcome limitations of the SM. One of such theories is Supersymmetry (SUSY). SUSY models predict a partner to every SM particle, which differs by spin half from the SM counterpart. These particles are expected to be more massive than their SM cousins.

This thesis presents a search for SUSY in a final state with multiple jets, b-jets and large missing transverse momentum. The search is performed using 137 fb^{-1}data collected by CMS experiment at LHC, over the years 2016, 2017 and 2018 using proton-proton collisions at centre of mass energy of 13 TeV. The SM events and different SUSY topologies which also give the same final state are considered. No signature for SUSY is found based on this analysis. The limits are put on different SUSY scenarios with squark and gluino pair production. Depending on model, gluinos with mass up to 2-2.3 TeV and squarks with mass 1.1-1.6 TeV are probed at 95% confidence level.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Scientific endeavours are mostly carried out to understand the phenomena observed in nature, and most of them as well influence, impact and benefit human kind in long term. The Large Hadron Collider (LHC) project at Geneva (Switzerland) is a collaboration of around 100 countries across the world, is one such unprecedented large scientific effort. It is established with an aim to answer some of the burning questions posed by particle physics and astrophysical observations.

Standard Model (SM) [1–3] is a theory of fundamental particles and their interactions which successfully explains a large amount of natural phenomena occurring in the universe around us. In the SM, the spin-half particles, leptons and quarks constitute the fundamental blocks of matter. Gauge bosons, W, Z, photon and gluons, carry one unit of spin, and are responsible for mediating various interactions among particles. The electromagnetic force experienced by particles carrying electric charge is mediated by photon. The weak force is mediated by W and Z bosons, in which W is responsible for phenomena like radioactive decays, while Z boson mediates neutral, weak interactions. The protons and neutrons within an atomic nucleus are bound together by strong force, overcoming repulsive electric forces among protons. The strong force is mediated by gluons. The masses of elementary particles are understood to originate from interactions with the Higgs boson field.

Despite of being a very successful theory supported by many collider and non-collider experimental observations, we know that the SM is still an incomplete theory. It does not have answers to many puzzles posed by nature like the composition of dark matter [4, 5], dominance of matter over anti-matter [6, 7], theoretical mass stability of recently discovered Higgs boson [8, 9], and incorporating fundamental gravitational force within it to name a few [10]. Believing the intuition based on development of the particle physics in past, it seems quite possible that the SM is only a partial manifestation of some more complete theory whose remaining constituents are still to be discovered. Many new theories are postulated to overcome shortcomings of the SM. Supersymmetry (SUSY) is one of such beyond SM (BSM) theory, which can possibly answer some of these unresolved questions. The SUSY postulates a counterpart for every SM field differing by a spin-half unit, and hence predicts many new particles. These particles, if exist in nature, are expected to be much heavier than their SM cousins as they have not yet already been observed. To produce such heavy particles in laboratory, we need very high energies, beyond the energy ranges exploited through accelerator experiments ever before.

At LHC, two proton beams are collided at centre of mass energy of 13 TeV, the highest collision energy ever achieved in a particle collider. When these relativistic protons collide, energy gets converted into mass, and heavy particles can be produced with mass of TeV range (if these exist in nature). The outcomes of collision are detected and recorded, at four different detectors installed on the LHC ring. The

Compact Muon Solenoid (CMS) is one of the two general purpose detectors designed to study outcome of proton proton collisions. A Toroidal LHC ApparatuS (ATLAS) is the another general purpose detector. This thesis describes a search for SUSY particles in multijets and missing transverse momentum final state using data equivalent of 137 fb^{-1} integrated luminosity, collected by CMS experiment over years 2016, 2017 and 2018 (called full Run 2 or legacy dataset).

The SUSY search strategies at both the ATLAS and CMS experiments can be broadly divided into categories like inclusive or generic searches based on final state signatures of jets, MET and zero or more leptons, and targeted searches leveraging on full mass reconstruction of SM particles (Z, W, H or top including the boosted object tagging). The inclusive searches mainly target processes with large production cross sections for gluinos and squarks [11–14]. The approach of not tuning to specific models for this search has benefit of simultaneous coverage to a large and diverse strongly produced SUSY phase space. While the zero or one lepton final states benefit from large branching fractions of SM particles (Ws or Zs) to quarks, the multilepton final states are relatively cleaner with lesser background contributions. Hence these searches performed in exclusive final states have different background composition, systematic uncertainties, and will provide a concrete cross-check in case of potential new discoveries. The sensitivity to stop pair production is enhanced by tagging the top-quarks in boosted regime, or constructing their masses using resolved decay products [15, 16]. Production cross section of both charged and neutral electroweakinos is very low and generally with small amount of hadronic activity. Such events are best triggered based on leptonic signatures which also makes the backgrounds smaller as compared to zero lepton searches. With well understood detectors and huge amount of data, the experiments are also extending searches to unusual topologies resulting from long lived particles, or soft leptons results from compressed particle spectra models [17, 18].

The SUSY search described in this thesis, is an extension of the analysis with 13 TeV collision data, first carried out with 2.3 fb^{-1} back in the year 2015 [13]. From 2015 to 2019, the analysis has gone through multiple evolutions to accommodate the increased amount of collected data and the changes in detector response over time. I directly contributed to the results published by this analysis with 35.9 fb^{-1} of CMS data in year 2017 [19] and 137 fb^{-1} of full Run 2 data in year 2019 [20]. These two milestone results included major evolutionary steps of going from 72 independent search regions to 174 independent search regions accommodating lowering jet multiplicity to explore compressed SUSY phase space. Such a detailed exploration of data in multiple dimensions of jets, b-jets, HT and MHT required revisiting the methods to estimate the SM backgrounds. I made a detailed study of the data driven method using tau-response templates hadronic tau background estimation in 2015 publication, and identified several lacunae related to kinematic distributions of muon and tau leptons which brought the estimated hadronic tau background yield closer to the expected, as summarized in chapter 6. In the full Run2 analysis based on 2016-2018 data, I worked on updating the two independent background strategies of event by event lost lepton estimation and tau template method for hadronic tau estimation by a new independent strategy using average transfer factors to obtain signal region yields starting from control regions in data. Hence, I integrated the estimation of these two major SM background components together (comprising around 50% of total background). I took a lead in understanding the challenges posed by data quality during legacy analysis, mainly devising the strategies to mitigate the effects of fake MHT due to successively degrading response of electromagnetic endcap detectors across the full dataset, and failure of a section of hadron calorimeter

endcap in 2018 data. I presented these studies in the CMS SUSY meetings as well as in CMS JetMET group, and these methods are used by various other analysis in the SUSY group. The current version of analysis has successfully extended gluino mass exclusion from 1600 GeV in year 2015 to 2310 GeV in 2019, with the full Run 2 data.

In addition to the SUSY search analysis, I invested significant time on FastSim calorimetry project. The project was aimed at improving the performance of Fast Simulation (FastSim), a widely used simulation technique in SUSY sample generation (50 K to few million events for a single point on SUSY mass scan). I started with basic kinematic comparisons of FastSim versus FullSim, where FullSim uses Geant4 based detail detector simulation and is well tested using test beam data. These comparisons lead to the understanding of the origin of observed discrepancies in FastSim when compared to FullSim. As a step towards improving the performance of Fast simulation, I developed a new parameterized form to model hadron shower starting position and its transverse shower development in FastSim. The performance of this new parameterized model is also validated against Geant4 and found to be better by orders of magnitude at many places as compared to existing FastSim models in use.

The thesis has elaborative discussion on the strategy used in the analysis work and also summarizes to the reader the SUSY phase space exclusion achieved based on this analysis. It also talks about the methodical investigation done to understand limitations of existing FastSim model and the progress made towards its improvement. The thesis is structured as follows. Chapter 2 gives a brief introduction to the SM and motivates why one should search for SUSY. Chapter 3 explains the basic experimental set up used at CMS detector to collect data. As detector conditions are evolving continuously due to exposure to radiation, it becomes important to understand data very well to not get mislead by any anomalous signal in detector. Later part of chapter 3 discusses how various issues with legacy data quality were handled during this SUSY search. Chapter 4 discusses role of Monte Carlo simulations in such SUSY search, and in particular various studies carried out to improve performance of Fast Simulation which has key role in SUSY sample generation. Chapter 5 starts with basic details of SUSY analysis discussed in this thesis and introduction to background estimation. Chapter 6 is devoted to details of two different background estimation strategies, tau template and average transfer factor, implemented as major parts of this thesis work. At the end, chapter 7 summarizes results and conclusions of this analysis.

Chapter 2

Standard Model and Supersymmetry

2.1 Standard Model of elementary particles

Standard Model (SM) has been an immensely successful theory in explaining the fundamental constituents of matter and interactions among them. The particles in this theory can be classified into three categories. Fermions form a group of elementary particles which are fundamental constituents of known matter. Gauge bosons is a set of particles which act as mediators for weak, electromagnetic and strong forces. The Higgs boson, the recently discovered scalar boson, interaction with which gives rise to nonzero mass to fermions and weak interaction gauge bosons. The theory is tested to a very high precision by a series of accelerator experiments carried at LEP, Tevatron and LHC [21–24]. The success of this theory is two fold. It explained and supported most of the experimental observations. Secondly, many of its predictions based on theoretical perspective later got confirmed through experimental observations. As an example, precision measurements at LEP [22] tested electroweak theory to per mill precision level.

2.1.1 Fermions

Fermions have characteristic half-integral spin as the name suggests. Leptons and quarks fall in this category, with three generations for each based on mass heirarchy [25]. Table 2.1 shows particle content with their fundamental properties for fermion class. There is an associated antiparticle to every particle, such that all quantum numbers for antiparticle are opposite to that of particle. Three lepton generations have individual associated quantum numbers, L_{e}, L_{μ} and L_{τ}. L_{e} =1 for first generation, L_{μ} =1 for second generation and L_{τ} =1 for third generation. In quark sector, quantum number called charm (C=1), strangeness (S=1), bottomness (b=1) and topness (t=1) are associated to quarks c, s, b and t respectively. For u and d quarks, relevant quantum numbers associated are the third component of isospin I_3 =1/2 and I_3 =-1/2 respectively.

2.1.2 Bosons

Table 2.2 shows particle content of boson category. Bosons possess integer spin. The Gluon (g), photon (γ), W and Z are vector bosons with spin one. Higgs boson is a scalar boson with spin zero. The W boson having -1 charge has an antiparticle with equal and opposite charge. Vector bosons act as mediators of various forces and their masses decide the range of mediated force. Gluons and photon are massless and act as mediator for strong and electromagnetic force respectively, while $W^{\pm}$,

TABLE 2.1: Fermions

spin	class	generation	particles	mass (MeV)	charge (Q)
1/2	leptons	1^{st}	ν_e	$\leq 2 \times 10^{-6}$	0
			e	0.511	-1/2
		2^{nd}	ν_μ	≤ 0.19	0
			μ	105.7	-1/2
		3^{rd}	ν_τ	≤ 18.2	0
			τ	1777	-1/2
1/2	quarks	1^{st}	u	2.2	2/3
			d	4.7	-1/3
		2^{nd}	c	1.275×10^3	2/3
			s	95	-1/3
		3^{rd}	t	173.1×10^3	2/3
			b	4.18×10^3	-1/3

Z^0 are massive gauge bosons which mediate weak force. The scalar particle, Higgs boson is responsible for masses of fermions and bosons.

TABLE 2.2: Bosons

particle	mass (GeV)	charge (Q)	spin
g	0	0	1
γ	0	0	1
W	80.4	-1	1
Z	91.2	0	1
H	125.2	0	0

2.1.3 Fundamental interactions

The SM lagrangian is obtained by imposing three local gauge invariances on quark and lepton field vectors: $U(1)_{\mathrm{Y}} \otimes SU(2)_{\mathrm{L}} \otimes SU(3)_{\mathrm{C}}$. According to Noether's theorem, coupling of quark or lepton fields with gauge fields arises naturally as a result of imposed invariance. Photon is a gauge field associated to U(1) symmetry, $SU(2)_{\mathrm{L}}$ gauge symmetry gives rise to three weak vector bosons W^+, W^- and Z^0 and $SU(3)_{\mathrm{C}}$ symmetry implies existence of eight gluons. Here the generic property of $SU(n)$ group that it has $n^2 - 1$ independent generators, allows 3 and 8 gauge bosons (which are independent generators) for $SU(2)_{\mathrm{L}}$ isospin and $SU(3)_{\mathrm{C}}$ color symmetry respectively. In total, the SM thus has twelve force carrying gauge particles. Conserved quantities from $U(1)_{\mathrm{Y}}$, $SU(2)_{\mathrm{L}}$ and $SU(3)_{\mathrm{C}}$ symmetry are electric charge (Q), isospin and color respectively [26].

Strong interactions

The SU(3) symmetry of strong interaction forces a quantum number called "color" to be conserved in any such interaction. Quarks and gluons participate in strong interaction. Each of quarks, can have color quantum number R (called "red"), G (called

"green") or B (called “blue”). Eight gluons which are generators of this group, arise from unitary, traceless representations of 3×3 matrix and possess a linear combination of R, G, B colors. Eight independent generators are :

$$R\overline{G},\ R\overline{B},\ G\overline{B},\ G\overline{R},\ B\overline{R},\ B\overline{G},\ R\overline{R}-G\overline{G}/\sqrt{2},\ R\overline{R}+G\overline{G}-2B\overline{B}/\sqrt{6}$$

Here "bar" implies anticolor. Anti-quarks possess color opposite to that of quark. At every quark-gluon and gluon-gluon interaction, color exihibited at either of legs is so as to conserve color at the vertex. The strength of this interaction increases with distance. At very small distances, quark-quark coupling strength becomes so small that quarks within mesons and baryons become almost free (called asymptotic freedom) [27, 28]. Beyond $\sim$ 1 fm, the force among colored objects become so strong that no color object can have individual identity, but instead they are combined to form colorless objects, called "hadrons". This process of forming colorless objects via color exchanges, is called hadronization. The property that quarks and gluons can not be isolated beyond 1 fm and hence lead to hadronization, is called “color confinement”. Hadrons come in two categories. Mesons are quark-antiquark composition, while baryons are three quark states. For mesons and baryons, “Baryon number” are equal to zero and one respectively. These quark confined states, obey generalized Gell-Nishijima formula, which relates total charge (Q), third component of strong isospin (I_3), baryon number (B), strangeness (S), charm (C), bottomness (b) and topness (t):

$$Q = e[I_3 + \frac{B+S+C+t+b}{2}] \tag{2.1}$$

Proton is the lightest baryon with valence quark composition “uud” which possesses RGB color combination to make the proton color neutral. These valence quarks are embedded within a sea of gluons and non-valence quarks. Quarks and gluons within proton are together called as “partons”. During energetic proton-proton collision at LHC, at fundamental level, the interacting partons from two proton beams get enough energy to move apart from each other. But due to color confinement, these can not be isolated and combine with quarks and gluons spontaneously formed at the dispense of their potential energy to form a spray of colorless hadrons which gets boosted in the direction of net momentum carried by interacting partons. In experimental terms, this spray of boosted hadrons is called a “jet”. As LHC is a hadron machine (as protons are hadrons), strong interactions dominate over direct production of electroweak processes.

Electromagnetic interactions

All charged particles undergo electromagnetic interaction via photon exchange. As photon is massless, electromagnetic force has infinite range. Electric charge and parity conservation is respected at every photon-fermion interaction vertex. As photon does not possess electric charge, it can not have self interaction. As a result, triphoton vertex does not exist unlike trigluon vertex in case of strong interaction, where gluons themselves possess color charge. The popping up of electromagnetic component (electrons, photons) in hadron machine like LHC, starts with a fundamental process $\pi^0 \rightarrow 2\gamma$, where $\pi^0 s$ are copiously produced in hadronization of quarks and gluons. So even though, direct production of electroweak bosons is rarer compared to strong interactions, electromagnetic components (electrons and photons) are not that rarer at LHC because of $\pi^0 s$. The typical time scale of this process is $10^{-17} s$.

Weak interactions

The SU(2) isopin symmetry imparts left handed fermions with $I_3 = \pm 1/2$ paired up to form a doublet (ν_L, e_L). Particles in these doublet interact via W and Z bosons. Weak force is mediated by W and Z bosons and is experienced by particles carrying weak hypercharge Y. Left handed fermions have Y=-1 and weak force experienced is proportional to Y. The electric charge Q, I_3 and Y are connected by Gell-Mann-Nishijima formula

$$Q = e[I_3 + \frac{Y}{2}] \tag{2.2}$$

As an example, electron has I_3 =-1/2, Y=-1/2, hence Q=-e. In the SM, neutrinos are massless and do not have right handed counterparts. Right handed counterparts of fermions like electron, muon do not participate in weak interaction; and are represented as isospin singlets.

The heavy W and Z bosons, make a weak interaction a short range. As name suggests, this interaction is weaker (and slower) compared to strong and electromagnetic interaction. A typical time scale of weak decay $\pi^+ \rightarrow \mu^+\nu_\mu$ is $10^{-9}s$.

Electroweak symmetry breaking

At high energies, electromagnetic and weak force are unified and follow $U(1)_{\mathrm{Y}} \otimes SU(2)_{\mathrm{L}}$ symmetry. The γ, W^+, W^- and Z^0 are the mass states obtained from linear combination of four gauge fields, B^μ associated to U(1) symmetry and W_i^μ (i=1, 2, 3) associated to SU(2) symmetry.

$$W^+ = (W_1 - iW_2)/\sqrt{(2)} \tag{2.3}$$

$$W^- = (W_1 + iW_2)/\sqrt{(2)} \tag{2.4}$$

$$Z^0 = -Bsin\theta_{\mathrm{w}} + W_3 cos\theta_{\mathrm{w}} \tag{2.5}$$

$$\gamma = Bcos\theta_{\mathrm{w}} + W_3 sin\theta_{\mathrm{w}} \tag{2.6}$$

Here θ_{w} is called weak mixing angle. Photon is massless while $W^\pm$ and Z^0 are heavier. The origin of mass asymmetry between mediators of electromagnetic and weak force (so called "electroweak symmetry breaking") is explained by Higgs mechanism.

Higgs mechanism

In the SM, force mediators B^μ, W_i^μ and gluons are expected to be massless, as a naive introduction of mass term in SM lagrangian, breaks local gauge symmetry. But in reality, $W^\pm$ and Z^0 particles are found to be massive [29–31]. To explain the masses of SM heavy bosons (and fermions), the SM lagrangian is extended to include Higgs field with quartic potential [32, 33]. This potential has nonzero minimum (called vaccum expectation value). Expansion of SM lagrangian about minimum of potential leads to spontaneous symmetry breaking and gives rise to massive electroweak bosons.

Before spontaneous symmetry breaking, there is a massless B^μ gauge field associated to $U(1)_{\mathrm{Y}}$ symmetry (2 degrees of freedom (dof) associated to two transverse

polarizations), and three massless W_i^μ fields associated to $SU(2)_{\mathrm{L}}$ symmetry (6 dof). For spontaneous symmetry breaking, Higgs field takes a complex scalar doublet form (4 dof). Thus in total, there are 12 degree of freedom. Interaction of Higgs field with SM lagrangian, gives rise to a real scalar Higgs boson (dof 1), W_i^μ and B^μ mix to give three massive vector bosons ($W^\pm$ and Z^0) (dof 9; two transverse and one longitudinal polarization associated to each boson) and one massless photon (dof 2). Thus, total degrees of freedom is preserved after spontaneous symmetry breaking. Coupling of Higgs field with fermions is proportional to mass and hence strongest for the top quark, followed by weak gauge bosons.

2.2 Limitations of the SM

The SM is immensely successful in explaining fundamental nature of the world around. With Higgs discovery in 2012, the SM is complete in its elementary particle constituents. But some questions are still unanswered by SM.

- Out of four known fundamental forces, gravity, electromagnetic, weak and strong, SM has explanation only for electromagnetic, weak and strong, while gravity is not incorporated yet. It is postulated that gravity can become comparable to strengths of other three forces at energy as high as GUT scale (10^{15} GeV). The SM has no explanation for why and how this is the case.
- Based on astronomical observations, it is known that not all visible matter in universe can explain rotation curve of galaxy, instead there should be an extra matter (which is invisible by all means so far) distributed from galaxy centre to its halo to explain this curve [34]. This invisible matter (also called "dark matter") is supposed to have no interaction other than weak force. The SM has no candidate, which can explain this significant abundance of dark matter.
- Experimentally detected Higgs boson has mass 125 GeV [8, 9]. But considering quantum loop corrections to Higgs mass (from SM fermion coupling to Higgs field), Higgs mass can go as high as the Planck mass scale. For Higgs boson mass to be stable at 125 GeV, very high fine tuning in mass is needed. Thus SM does not satisfactorily explain Higgs boson mass at 125 GeV [35–38].

With these limitations in mind, many beyond SM theories are proposed to overcome drawbacks of SM. One of such theory is Supersymmetry or SUSY.

2.3 Supersymmetry

Supersymmetry is a theory which maps every fermion to boson and vice versa. Thus each SM particle has one superpartner. Superpartners of quarks and leptons, are called with prefix "s" to their SM counterparts (e.g. sup, selectron), while partners of gauge bosons are called with postfix "ino" to names of SM counterparts (e.g. guagino, wino). Each SUSY particle is denoted with "tilde" on symbol of their SM counterparts. In SUSY world, instead of single particle states, one has to deal with (super)multiplets of particle states.

Only chiral supermultiplets can contain fermions whose left handed parts transform differently than right handed parts. Hence all SM fermions should be member of chiral supermultiplet. As an example, partner of left and right handed Dirac electron, will be left and right handed components of selectron, $\widetilde{e_L}$ and $\widetilde{e_R}$. Similar is the case with other leptons and quarks.

It is observed that SM Higgs scalar boson should be a part of two Higgs chiral supermultiplets. This is because, a doublet with weak hypercharge Y=1/2 can only give mass to up-type quarks with charge +2/3 and a doublet with Y=-1/2 can give mass to down type quarks with charge -1/3. These doublets are denoted as (H_u^+, H_u^0) and (H_d^0, H_d^-). A physical SM Higgs boson is a linear combination of H_u^0 and H_d^0.

The vector bosons of the SM reside in gauge supermultiplets. Every gluon has associated gluino, to form spin 1/2 color-octet. The electroweak gauge symmetry $SU(2)_{\rm L} \otimes U(1)_{\rm Y}$ is associated to W_i^μ, B^μ and also to their superpartners called wino and bino. After electroweak symmetry breaking, bino, a neutral wino and two neutral Higgsinos combine to form four physical mass states, called neutralinos ($\widetilde{\chi}_i^0$ where i=1..4), while two charged winos and two charged Higgsinos combine to form four charged mass states, called charginos ($\widetilde{\chi}_i^\pm$ where i=1,2). This is called minimal supersymmetric extension of the SM spectrum (MSSM) where there are superpartners like higgsino, wino, zino, together with all squarks and sleptons [39–41].

TABLE 2.3: Chiral supermultiplets in MSSM

Name	spin 0	spin 1/2
squarks, quarks	($\widetilde{u_L}$, $\widetilde{d_L}$)	(u_L, d_L)
	$\widetilde{u_R}$	u_R
	$\widetilde{d_R}$	d_R
sleptons, leptons	($\widetilde{\nu}$, $\widetilde{e_L}$)	(ν, e_L)
	$\widetilde{e_R}$	e_R
Higgs, Higgsino	(H_u^+,H_u^0)	($\widetilde{H_u^+}$,$\widetilde{H_u^0}$)
	(H_d^-,H_d^0)	($\widetilde{H_d^-}$,$\widetilde{H_d^0}$)

TABLE 2.4: Gauge supermultiplets in MSSM

Name	spin 1/2	spin 1
gluino, gluon	$\widetilde{g}$	g
wino, W bosons	$\widetilde{W_i^\mu}$	W_i^μ
bino, B boson	$\widetilde{B^\mu}$	B^μ

The SM fermion (f) coupling $-\lambda_f H \overline{f} f$ induces one loop correction to Higgs mass:

$$\Delta m_H^2 = -2\lambda_f^2 \Lambda^2 \tag{2.7}$$

Similarly, a scalar (S) coupling $-\lambda_s |H|^2 |S|^2$ induces one loop correction to Higgs mass:

$$\Delta m_H^2 = \lambda_s \Lambda^2 \tag{2.8}$$

If the new physics is such that each quark and lepton of the SM were accompanied by two complex scalars having the same Higgs couplings of the quark and lepton ($\lambda_s = |\lambda_f|^2$), Λ^2 term will automatically cancel and the Higgs mass would be stabilized at its tree level.

At lower energies, coupling strength of interaction is maximum for strong, minimum for weak and intermediate for electromagnetic. With energy, coupling strengths of strong, electromagnetic and weak interactions vary. Extrapolation of strengths to higher energies show that, with a set of existing SM particles, the three gauge couplings become of comparable strength only approximately. In other words, unification of strong, electroweak forces will not be exact. But if one just allows for the MSSM, the extrapolation of three gauge couplings with energy meet exactly at one point, called GUT scale and also will let proton decay rate to be compatible with present experimental bounds.

In MSSM, gauge interactions of sleptons and squarks are expected to be similar to SM quarks and leptons, that is, left handed squarks couple to W bosons but right handed squarks do not. Chargino and neutralino are mass eigenstates formed after superposition of gauge eigenstates wino, bino and higgsino. Neutralinos which are the lightest supersymmetric particle, are considered to be stable and only weakly interacting, and hence is a favored dark matter candidate. These prospective outcomes of supersymmetry, make it attractive to search for at the colliders.

2.4 R parity

In the SM, baryon number (B) conservation does not allow decay of proton to any other lighter particle, as proton is the lightest baryon. But unlike stability of electron, which is ensured by local gauge symmetry associated with photon field, there is no fundamental symmetry to ensure baryon number conservation. Due to this, in the MSSM, baryon number and lepton number are not supposed to be conserved separately. As a result, decay of proton to (e^+,π^0) will be allowed in the MSSM. But experimental observation is that proton is stable and current limits on proton decay is about 10^{33} years [42]. To keep proton decay intact even in supersymmetric extension of SM, a new quantum number called " R parity" is defined [43, 44].

$$P_R = (-1)^{3(B-L)+2s} \tag{2.9}$$

Here B, L and s are respectively baryon number, lepton number and spin of a particle. All SM particles have P_R even. Their supersymmetric partners will have odd P_R.

Another consequences of R parity conservation will be:

- in proton-proton collision, SUSY particles should be produced in pairs.
- decay of each SUSY particle should have odd number of SUSY particles at each stage of decay chain.

2.5 Simplified models

In this thesis, SUSY models conserving R parity are investigated. For the interpretation of results, the models are considered in which either gluinos or squarks are produced in pairs in a proton-proton collision. Figure 2.1 shows representative strong SUSY production scenarios considered in this search [45–49]. The names of these simplified SUSY models are adopted in such a way that they define the sequence of production and decay of particles. Each of these models assumes that all other SUSY particles, except those appearing in diagrams, are very heavy and not accessible at the LHC. Also the decay chain as specified in the model is assumed to have

100% branching ratio. Looking into such topology featured simplified models allow results to be interpreted for many realistic scenarios. Following are the details of the models considered.

- T1tttt, T1bbbb, T1qqqq - T1 refers to scenarios with gluino pair production. For example, T1tttt represents a decay chain where gluino has decayed to $t\bar{t}$ and $\widetilde{\chi}_1^0$. Similarly, T1bbbb and T1qqqq respectively represent decay to $b\bar{b}$ and light flavoured $q\bar{q}$ in association with $\widetilde{\chi}_1^0$.
- T2tt, T2bb, T2qq - T2 refers to squark-antisquark pair production. In T2tt scenario, top squark-antisquark decay to their SM counterparts along with $\widetilde{\chi}_1^0$. In a similar way, T2bb and T2qq refer to production and decay of bottom and light flavoured squark-antisquark pair respectively.
- T5qqqqVV- T5 model again refers to gluino pair production, which decay to light flavoured $q\bar{q}$ but now in association with next-to-lightest neutralino $\widetilde{\chi}_2^0$ or chargino $\widetilde{\chi}_1^{\pm}$. The probability for decay to $\widetilde{\chi}_2^0$, $\widetilde{\chi}_1^+$ and $\widetilde{\chi}_1^-$ is assumed to be 1/3 each. The $\widetilde{\chi}_2^0$ and $\widetilde{\chi}_1^{\pm}$, further decay to Z or W boson and $\widetilde{\chi}_1^0$.

These models may or may not have leptons in the final state. The search described in this thesis only looks into zero lepton final state, i.e. the heavy quarks and bosons are only allowed to decay in hadronic final state. Depending on whether there is gluino or squark-antisquark pair production, these scenarios can populate in high or low jet multiplicity regions respectively. End product of each decay chain has neutralino as Lightest Supersymmetric Particle (LSP), which is stable and weakly interacting, hence can be a viable dark matter candidate. The presence of LSP in collision events leads to a distinctive signature of large missing transverse momentum ($p_{\mathrm{T}}^{\mathrm{miss}}$ or MET). As will be discussed in chapter 7, this analysis has achieved exclusion of gluino mass to 2180 GeV, 2310 GeV, 2030 GeV for T1tttt, T1bbbb and T5qqqqVV scenarios and squark mass exclusion to 1190 GeV, 1220 GeV and 1630 GeV for T2tt, T2bb and T2qq scenarios respectively.

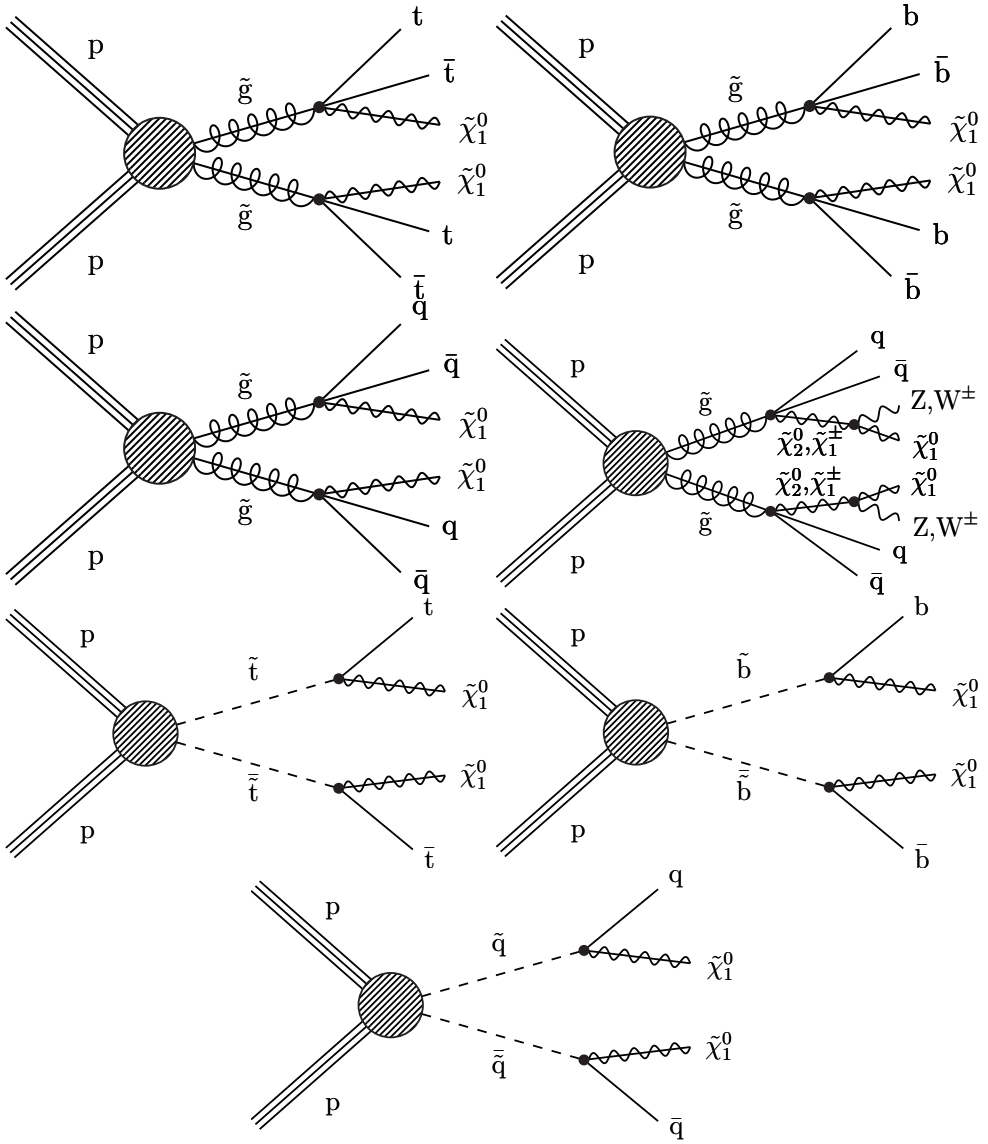

FIGURE 2.1: Representative diagrams of gluinos and squark pair production in rowwise left to right ordering: T1tttt, T1bbbb, T1qqqq, T5qqqqVV, T2tt, T2bb and T2qq

Chapter 3

Experimental set up and event reconstruction

This chapter gives an overview of the experimental set up used to create the pp collisions at the LHC, and to collect the collision data by the CMS detector. With each subdetector system optimized for its purpose, the chosen experimental set up is efficient in inferring physics processes which possibly took place as the proton bunches collided. Nevertheless, assessing the data quality is a crucial step before analyzing it for the physics motivation under consideration. Later part of this chapter tries to give glimpse of challenges with multijets + $p_{\mathrm{T}}^{\mathrm{miss}}$ search, and how detector related anomalies were identified and rejected in data taken during various phases of Run 2 operations.

3.1 The LHC as an accelerator and a collider

The LHC [50] is operated by European Organization for Nuclear Research (CERN). This underground experiment is running at 27 km diameter circular tunnel in the Geneva region near Switzerland and France borders. Two proton beams, each with a centre of mass (CM) energy 6.5 TeV are collided at a designed luminosity ($\mathcal{L}$) of $10^{34} cm^{-2} s^{-1}$. Each proton beam is like a train of proton bunches, where consecutive bunches are separated by about 25 ns in time. Each proton bunch has about 10^{11} protons and is flying almost close to speed of light.

3.1.1 Acceleration and bunch formation

The procedure of getting relativistic protons uses a chain of multiple accelerators each of which is designed to increase their energies in specific range before passing on the protons to the subsequent stage. At first, protons are extracted from hydrogen gas by applying very high voltage and sent to Radio Frequency (RF) cavity, where protons get focussed and assembled to form bunches. The RF cavity increases energy of protons from 90 keV to 750 keV. After that, energy of protons is increased successively through accelerators LINAC2 (50 MeV), Proton Synchrotron Booster (PSB) (1.4 GeV), Proton Synchrotron (PS) (25 GeV), Super Proton Synchrotron (SPS) (450 GeV), and finally the LHC (6.5 TeV). Fig. 3.1 shows a sketch of the CERN accelerator complex.

3.1.2 Parton-parton interaction at collision point

With a spacing of 25 ns or 7.5 meters, 3546 proton bunches are possible across the LHC ring of 27 km. Out of these only 2808 are filled while others are kept empty

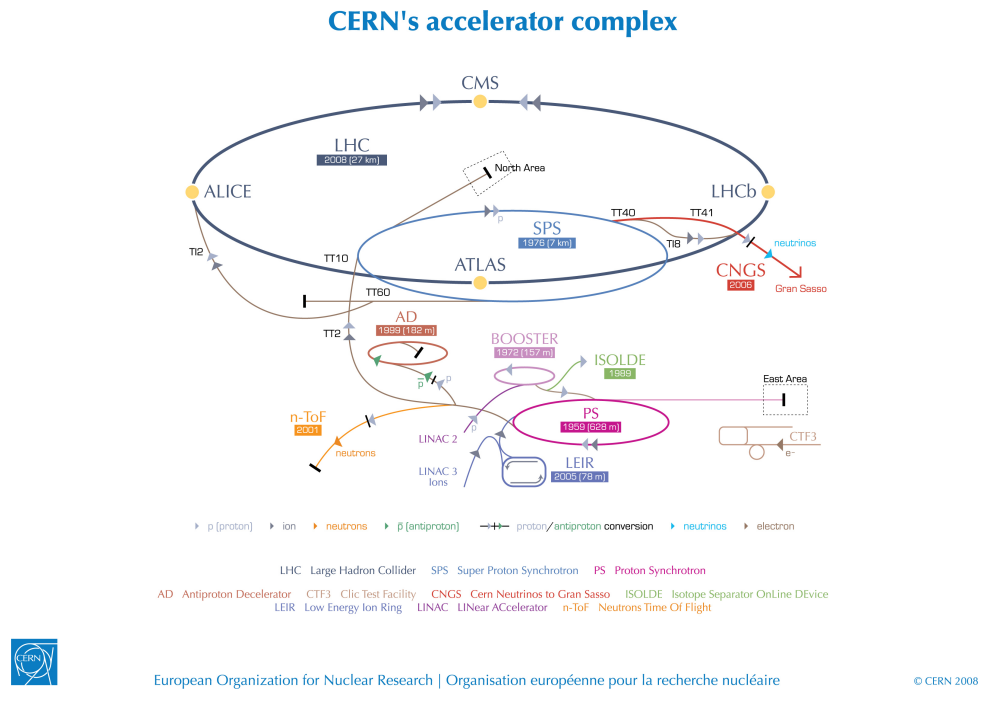

FIGURE 3.1: The CERN accelerator complex

for ejection of used proton beam once instantaneous luminosity has gone below certain threshold and also to guarantee safe dumping of beam in case of problems. At $\sqrt{s}$ =13 TeV, proton-proton inelastic collision cross section (σ_{inel}) is 80 mb (cross section for elastic and diffractive processes do not contribute as their output is hardly possible to fall within detector coverage). As a result, average number of inelastic proton-proton collisions (pp) per bunch crossing expected (after all 3546 bunches are crossed) are about 16.

$$\mathcal{L} \times \sigma_{\text{inel}} \times \text{bunch separation} \times \text{fraction of filled bunches} =$$
$$10^{34} cm^{-2} s^{-1} \times 80\, mb \times 25\, ns \times \frac{2808}{3546} = 16$$

This number is very close to the mean of average pp interactions per bunch crossing observed in year 2015 ($< \mu >$) as shown in fig. 3.2 (right). With dedicated work by experts on how to increase number of bunches in a single proton beam, how to squeeze the beam to have smaller area and what should be beam crossing configuration, the LHC by end of Run 2 could attain the instantaneous luminosity double of its nominal value, it had started with, as shown in fig. 3.2 (left).

As proton is a composite structure made up of quarks and gluons (collectively called partons), energy of interacting partons (Q^2) is different from centre of mass energy of the pp system. Parton Distribution Function (PDF) decides probability of fraction of proton energy carried by a parton, and hence the energy available in a parton-parton interaction to produce the SM or new particles. A hard collision event is the one where partons which interact are energetic enough to produce heavy particles. The probability to have hard collision increases with Q^2.

For a given bunch crossing, more than one pair of partons can undergo a hard interaction. The one with the maximum momentum transfer between the partons

is referred as "primary interaction vertex" or "primary vertex", while the others are called "pile up" interaction vertices. The particles produced in the pile up interactions result in undesired overlap with those produced in the primary interaction, and is called "in-time pile up". Another source of undesired overlap is "out-of-time pile up", where the outcomes of two consecutive bunch crossings get mixed up if signal capturing by detector is not fast and extends beyond 25 ns by when the next bunch crossing takes place. Both the "in-time" and the "out-of-time" pile up pose challenges at the time of acquisition of data (real time or online) and its analysis (offline). The LHC was designed to have on average 25 proton-proton interactions per crossing. But, as shown in fig. 3.2, average pp interactions during Run 2 were about 34 [51], with a large variation across years and also during data taking in single year.

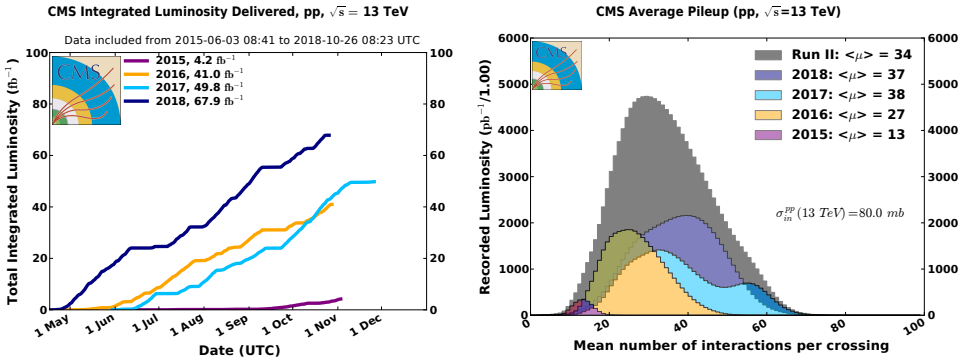

FIGURE 3.2: Integrated luminosity and average pile up during Run 2

The two proton beams, after being fully accelerated, are made to cross each other at four collision points across the LHC ring, and the outcome of collision is recorded by detectors surrounding these beam crossing points, namely the CMS, ATLAS, ALICE and LHCb. The CMS and the ATLAS are general purpose detectors, with maximum spatial coverage around the interaction points. A Large Ion Collider Experiment (ALICE) is designed to study heavy ion physics and the Large Hadron Collider beauty experiment (LHCb) uses a specialized detector to study flavor physics (mainly hadrons containing b-quarks).

3.2 The CMS detector

A perspective view of the CMS detector is shown in fig. 3.3. Silicon based pixel and strip tracker surrounds the pp interaction point and provides precise momentum measurement of charged tracks using bending of trajectories in high magnetic field. Tracker covers 5.8 m in length and 2.5 m in diameter. Outside tracker sits electromagnetic and hadronic calorimeters, which have a role of energy measurement by fully absorbing incoming energetic particles. Electromagnetic calorimeter is specialized for energy measurement of electrons and photons, while hadron calorimeter plays a crucial role in jets and $p_{\mathrm{T}}^{\mathrm{miss}}$ measurement. The assembly of tracker, electromagnetic and hadronic calorimeter sits inside a bore of superconducting solenoid, which provides the high magnetic field (3.8 T) necessary for bending of charged particles, especially muons. Muons do not shower inside calorimeter, hence only track measurement is possible for them. Also the highly energetic muons fly almost straight in the detector, hence a large bending power is needed for a good track measurement. Muon chamber is the outermost component of this detector and helps to

make muon momentum measurement precise in association with inner tracker measurement. With all these subdetectors in place, overall dimensions of CMS detector becomes, length 21.6 m, diameter 14.6 m and weight 12500 tonne.

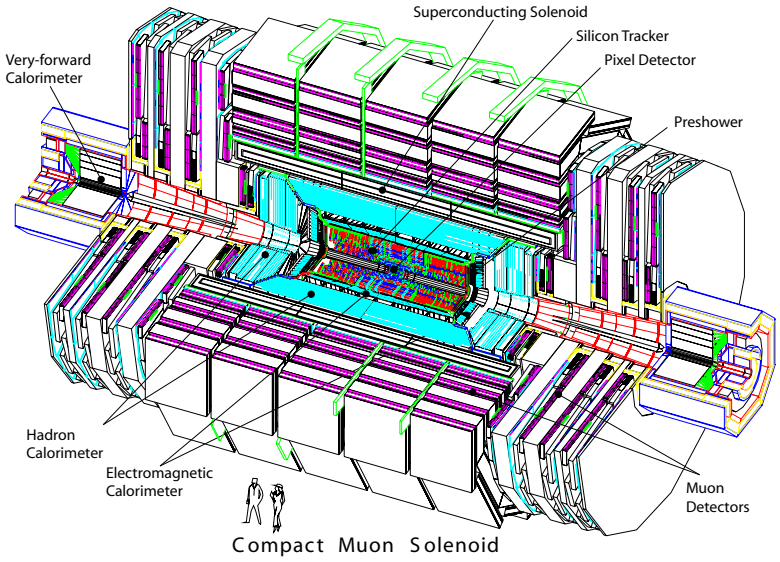

FIGURE 3.3: CMS detector perspective view

The common nomenclature associated with direction and four momentum of particles in CMS detector are as given in table 3.1

TABLE 3.1: CMS detector nomenclature

Nomenclature	Description
Origin	Nominal collision point
x-axis	radially inward pointing LHC centre
y-axis	vertically upward
z-axis	along beam direction
r	$\sqrt{x^2 + y^2}$
ϕ	in x-y plane with respect to x
θ	with respect to +z axis
Pseudorapidity η	-ln(tan(θ/2))
Energy and momentum in transverse direction	p_{T} and E_{T}
Imbalance of momentum in transverse direction	$p_{\mathrm{T}}^{\mathrm{miss}}$or MET

3.2.1 Solenoid

Superconducting magnet is an electromagnet made from coils of superconducting wire Niobium-Titanium (NbTi). They have zero electrical resistance (when cooled down to few kelvins), hence can carry extra high currents to generate very large magnetic field.

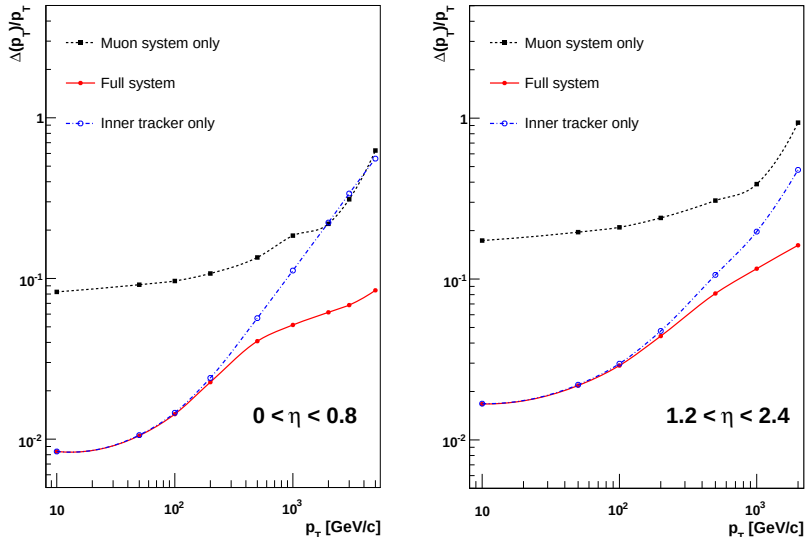

FIGURE 3.4: Muon momentum resolution in two different η rangeswith and without using muon system

Higher the momentum of particle, lesser is the bending of track (Lorentz force law: mv=qBr) which degrades the momentum resolution. Thus a large bending power is especially needed for precise momentum measurement of highly energetic charged particles, mostly produced in decay of heavy bosons. The individual hadrons within jets and electrons from heavy boson decays are not as challenging as muons for precise momentum measurement due to relatively low momentum of hadrons within jets while electomagnetic shower due to electrons and photons are short and get contained within electromagnetic calorimeter. Muons do not undergo shower to deposit their energy in calorimeter, hence relies only on track momentum measurement. Thus superconducting magnet best serves for the purpose of precise momentum measurement across MeV to TeV energy range.

The CMS superconducting solenoid is 13 m long, has 6 m inner diameter and provides 4T field. Bore of coil accomodates the inner tracker and the calorimetry inside. A 12 Tm bending power is achieved before muon is measured in muon system. Muons in final state provide a clean signature, as they do not undergo bremsstrahlung. Hence, it is advantageous to have a very good momentum resolution. The bending power of a solenoid is such that inner tracker along with muon chamber improves momentum resolution of tracker alone system from 60% to 8% in $0 < \eta < 0.8$ region and 50% to 16% in $1.2 < \eta < 2.4$ region as shown in Figure 3.4.

3.2.2 Tracker

The CMS tracking system is made up of silicon as active material. The foremost tracker encountered by particles produced at pp interaction is pixel detector. The pixel detector has three barrel pixel layers at radii of 4.4, 7.3 and 10.2 cm. On both sides of these pixel layers, there are disks of pixel modules on either η sides. The pixel detector delivers three high precision space points on each charged particle trajectory. The radial region between 20 cm and 116 cm is occupied by the silicon strip tracker. This region is formed using Inner Barrel and Disks (TIB/TID). The TIB/TID system delivers up to 4 r-ϕ measurements: a single point resolution of 23 μm and 35 μm for first two and a third layer respectively. The TIB/TID is surrounded by

Tracker Outer Barrel (TOB). It has an outer radius of 116 cm and consists of 6 barrel layers. They provide 6 r-ϕ measurements with single point resolution of 53 μm and 35 μm, respectively for first four and last two layers. The TOB extends in z between ±118 m. Endcap tracking detectors (TEC+ and TEC- where the sign indicates the location along the z axis) cover the region 124 cm < |z| < 282 cm and 22.5 cm < |r| < 113.5 cm. Average pitch of these strips is about 184 μm. Fig. 3.5 has schematic sketch of tracker.

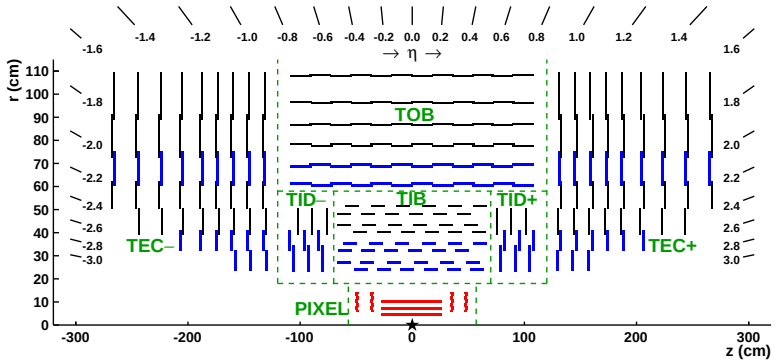

FIGURE 3.5: Tracker with different modules

A precise measurement of secondary vertices and impact parameters (section 3.3.4) is necessary for the efficient identification of heavy flavour particles (or quarks) which are produced in many of the interesting physics channels. This is where high granularity tracker helps. Tracking information is also heavily used in the high level trigger (HLT) of CMS, which has to do the job of reducing the event rate of L1 trigger output (100 kHz) to 0.5-1 kHz event rate to be written to the tape for permanent storage (section 3.4). The choice of detector technology is based on occupancy level. The pixelated detectors at radius<10 cm with dimension 100×150 μm^2 keep occupancy about 10^{-4} per pixel. Silicon micro strip detector at 20<r<55 cm of dimension 10 cm×80 μm has 2-3% occupancy per cell. The outer tracker has 500 μm thickness silicon as opposed to the 320 μm in the inner tracker. Thicker sensors would in principle have a higher depletion voltage. But since the radiation levels in the outer tracker are smaller, a higher initial resistivity can be chosen such that the initial depletion voltages of thick and thin sensors are in the same range of 100 V to 300 V.

3.2.3 Electromagnetic calorimeter (ECal)

For electrons and photons of energies of order of few GeV, bremsstrahlung and pair production are dominant mechanisms of energy loss inside material respectively. This leads to a cascade of secondary electron-positron pairs and photons, which is called an electromagnetic (EM) shower. The objective of ECal is to measure energy of incoming electron or photon by total energy deposition inside detector through electromagnetic shower. The characteristic lengths of electromagnetic shower are radiation length (X_0) and Moliere radius. The X_0 is the longitudinal distance along shower development in which incident electron losses 1/e of incident energy by bremsstrahlung and is equal to 7/9 of photon's free path of pair production. Moliere radius defines distance transverse to shower development, such that a cylinder of Moliere radius has 90% of energy of electromagnetic shower contained.

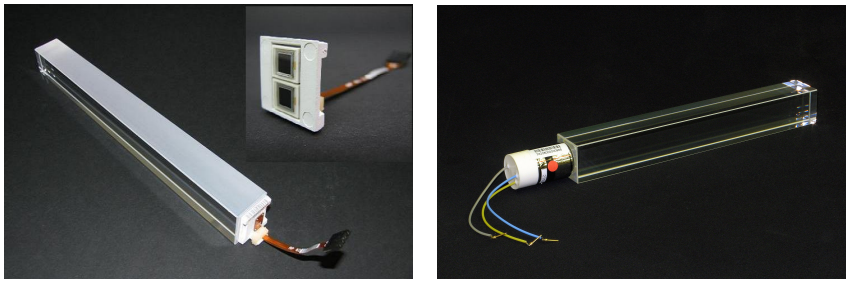

FIGURE 3.6: ECal crystals in EB and EE

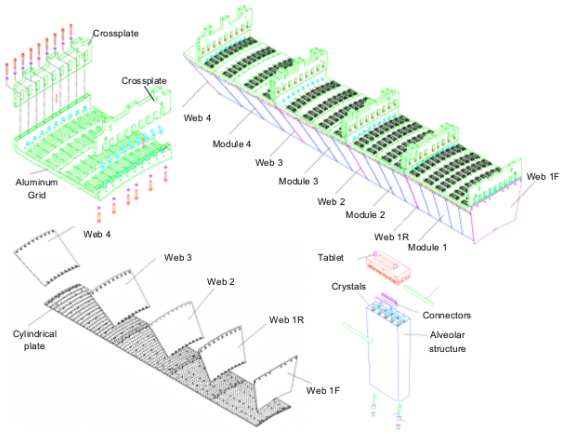

FIGURE 3.7: ECal barrel assembly in terms of crystals, modules and supermodules

In CMS, ECal is a homogeneous calorimeter made up of Lead tungstate ($PbWO_4$) scintillating crystals. Central barrel is made up of 61200 crystals, while each endcap is made up of 7324 crystals. High density (8.28 g/cm^3), short radiation length (0.89 cm) and small Molière radis (2.2 cm) of these crystals make detector fine granular and compact. Response time is such that 80% of the light is emitted in 25 ns; which is consistent with proton bunch crossing. Peak emission wavelength is at 420-430 nm. There is total internal reflection inside the crystal for optimum light collection on the photodetector. In case of barrel, non-uniform light collection by barrel crystals due to variation in length is corrected by making one side of crystal non-transparent.

ECal barrel

ECal barrel (EB) covers a range upto $|\eta| < 1.479$. Granularity in barrel is 360-fold in ϕ and (2×85) fold in η, making total of 61200 $PbWO_4$ crystals. Each crystal makes 3° angle with respect to the vector from the nominal interaction vertex, in both the ϕ and η projections, so as to avoid passing a particle directly through cracks between two crystals. Crystal front face cross section is 0.0174×0.0174 in η - ϕ plane (22×22 mm^2 at the front face of crystal, and 26×26 mm^2 at the rear face). Front face of ECal crystal (or ECal start) is at 1.29 m and crystal length is about 25.8 X_0. Assembly in ECal barrel is such that crystals are fitted into submodule and submodules are assembled into modules (400 to 500 crystals). Supermodules are formed by combining four such modules (1700 crystals). Eighteen supermodules, each with 20° in ϕ form a half barrel. Figure 3.7 shows different components in assembly of ECal barrel.

FIGURE 3.8: One "D" formed out of supercrystals which makes half of each ECal endcap

ECal endcap

The ECal endcap (EE) extended the barrel coverage to $1.479 \leq |\eta| \leq 3.0$. Crystal front face cross section is 0.0174×0.0174 in η - ϕ plane (28.62×28.62 mm^2 at the front face of crystal, and 30×30 mm^2 at the rear face). Crystals are arranged in x-y grid pointing off-interaction point, making angle 2-8°. Front face of ECal endcap envelope is at 315.4 cm from interaction point and crystal length is about 24.7 X_0. Assembly in endcap is such that 5× 5 crystals make one supercrystal. Such 138 full and 18 partial supercrystals make one "D" as shown in figure 3.8. Each endcap is made up of two such "D"s.

Photodetectors

The scintillation light produced by passage of particles through ECal crystals is converted to electronic signal by photodetectors located at the rear end of crystals. Requirements on photodetectors are they should be fast, radiation tolerant and insensitive to 4T magnetic field and particles passing through. Barrel uses avalanch photodiodes with active area of 5×5 mm^2. A pair of them is mounted on each crystal. Gain is 50 at maintained temperature of 18° celcius. ECal endcap uses vacuum phototriodes (VPT). Photomultipliers are having a single gain stage. An active area of approximately of each VPT is 280 mm^2. One VPT is glued to the back of each crystal. The lower quantum efficiency and internal gain of the vacuum phototriodes, compared to the avalanche photodiodes, make them appropriate for ECal endcap, which has higher expected level of radiation.

Electronic readout

The on-detector electronics has been designed to read a complete trigger tower (5×5 crystals in η-ϕ or a super-crystal for EB and EE respectively). The signal is shaped

by a multi-gain Pre-Amplifier and digitized by 40-MHz analog to digital converter (ADC). The digitized data are stored during the Level-1 trigger latency (the delay before a transfer of data begins) of about μs. The full ECal data for an event, if all channels are read out, exceeds the allocated target of 100kB/event for ECal by a factor of nearly 20. Reduction of the data volume is done by selective read-out based on energy with respect to various thresholds (E_{T}). Algorithm is based on categorization of trigger towers based on energy (if in barrel).

If a trigger tower belongs to the high interest class (E_{T} > 5 GeV) then the crystals of this trigger tower and of its neighbour trigger towers (225 crystals in total in the barrel case) are read with no zero suppression. If a trigger tower belongs to the medium interest class (E_{T} > 2.5 GeV), then crystals of this trigger tower (25 crystals in the barrel case) are read with no zero suppression. If a trigger tower belongs to the low interest class and it is not the neighbour of a high interest trigger tower, then the crystals in it are read with zero suppression at about 3σ.

3.2.4 Hadronic calorimeter (HCal)

The objective of HCal is measurement of hadron jets and which also lead to measurement of $p_{\mathrm{T}}^{\mathrm{miss}}$. HCal barrel is radially restricted by outer extent of the ECal (r = 1.77 m) and the inner extent of the magnet coil (r = 2.95 m). To compensate for the limited radial extent of HCal barrel, there is Outer HCal (HO) calorimeter (tail catcher) outside solenoid. Beyond $|\eta|$ = 3, the forward hadron calorimeters are placed at 11.2 m from the interaction point.

HCal barrel (HB)

HCal barrel plays significant role in new physics searches like SUSY (particularly searches with all hadronic final states), where large number of high p_{T} jets are expected mostly in central barrel region. As hadronic showers are longer in length, HCal barrel is designed to have a depth of more than 5λ at η=0. The depth in terms of λs is extended further using HO, which is discussed in following section. There are 36 (18+18) identical azimuthal wedges which form HB+ and HB- in $|\eta|$ = 1.3. Wedges are constructed out of flat brass absorber plates, where "layer 1" is on inside of LHC ring. Each wedge is divided into four ϕ divisions such that 18×4=72 cover $2\times\pi$ (Fig.3.9) . Table 3.2 shows various layers inside a wedge.

TABLE 3.2: Layers in HB wedge

layer	material	thickness
front plate	steel	40 mm
1-8	brass	50.5 mm
9-14	brass	56.5 mm
back plate	steel	75 mm

Scintillators used as an active medium, which are placed in between brass plates. Plastic scintillators are divided into 16 η sectors, resulting in a segmentation ($\Delta\eta$,$\Delta\phi$) = (0.087,0.087). The 16th η sector overlaps with HCal endcap. Scintillator tiles of a given ϕ layer (i.e. 16 η sectors) are grouped into a single mechanical scintillator tray unit, which corresponds to one readout.

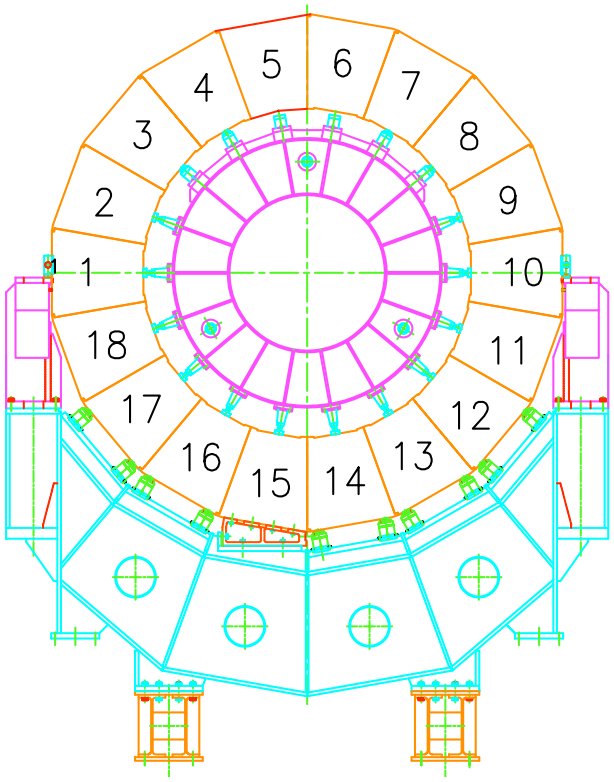

FIGURE 3.9: 18 wedges forming HB+ / HB-

Each wedge has 17 scintillator layers in depth; of which layer 0 and 16 have thickness of 9 mm, while layers 1-15 have 3.7 mm thickness. Layer zero is to sample hadronic showers developing in the inert material between the EB and HB, while layer 16 serves to correct for late developing showers leaking out the back of HB. Light from each tile is collected with a 0.94-mm-diameter green double-cladded wavelength-shifting fibre. After exiting the scintillator, the wavelength shifting fibres (WLS) are spliced and go to an optical connector at the end of the tray and then transferred to a hybrid photodiode (HPD).

HCal endcap (HE)

This covers $1.3 \leq |\eta| \leq 3.0$. The C26000 cartridge brass is used as absorber as it is nonmagnetic, high radiation tolerant and has good mechanical properties. Brass plates are 79-mm-thick with 9-mm gaps to accommodate the scintillators. Total length including ECal endcap become about 10λ. Trapezoidal shaped scintillators have grooves in which WLS fibres are inserted. The granularity of the calorimeters is ($\Delta\eta$,$\Delta\phi$)=(0.087,0.087) for $|\eta| < 1.6$ and ($\Delta\eta$,$\Delta\phi$)=(0.17,0.17) for $|\eta| \geq 1.6$. Figure 3.10 shows η-ϕ segmentation in HE.

Outer calorimeter (HO)

HO utilises the solenoid coil as an additional absorber and is used to identify late starting showers and to measure the shower energy deposited after HB. The magnetic field is returned through an iron yoke designed in the form of five 2.536 m wide (along z-axis) rings. The HO is placed as the first sensitive layer in each of these five

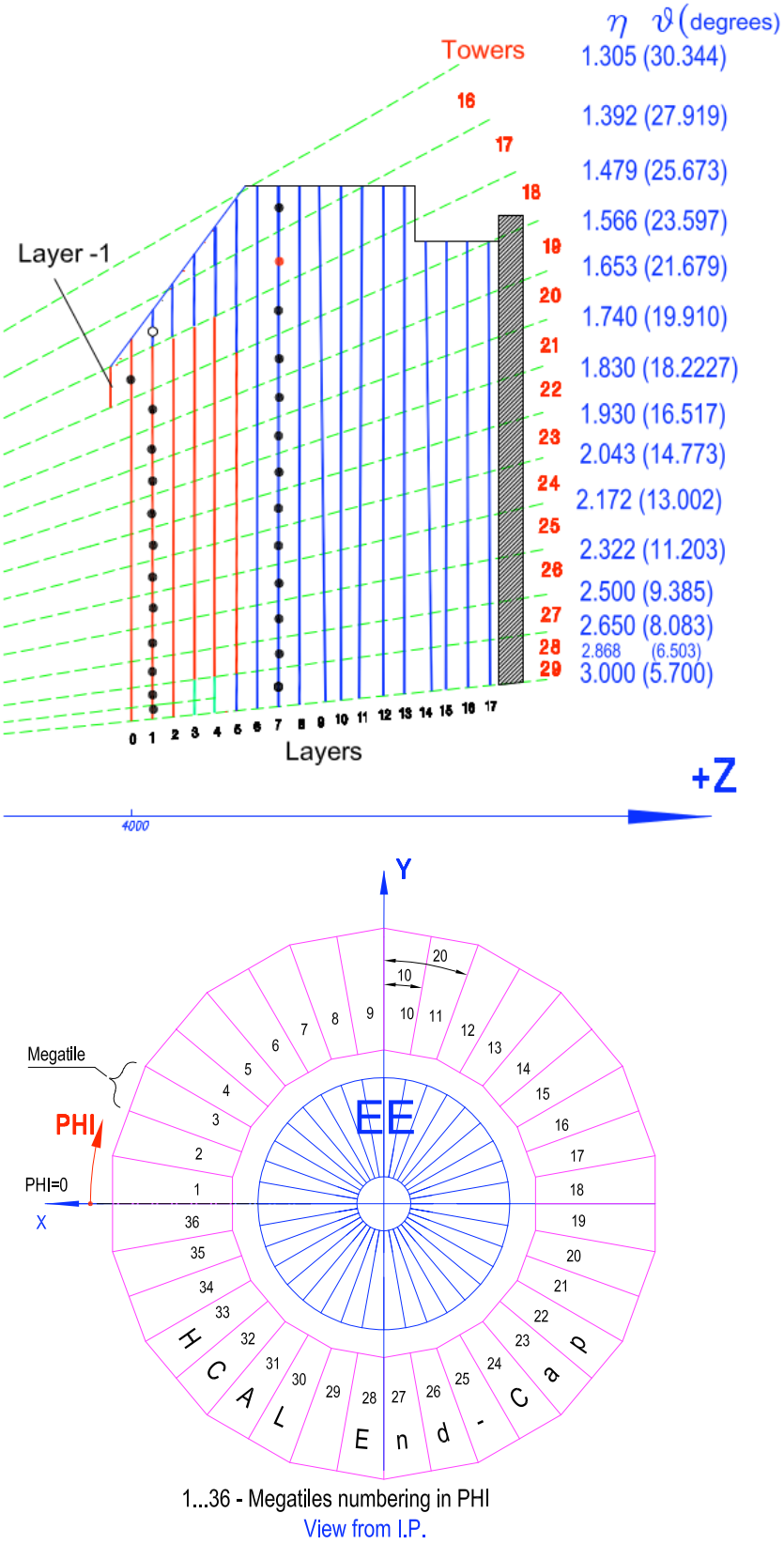

FIGURE 3.10: Segmentation of HE in (top) η and ϕ (bottom)

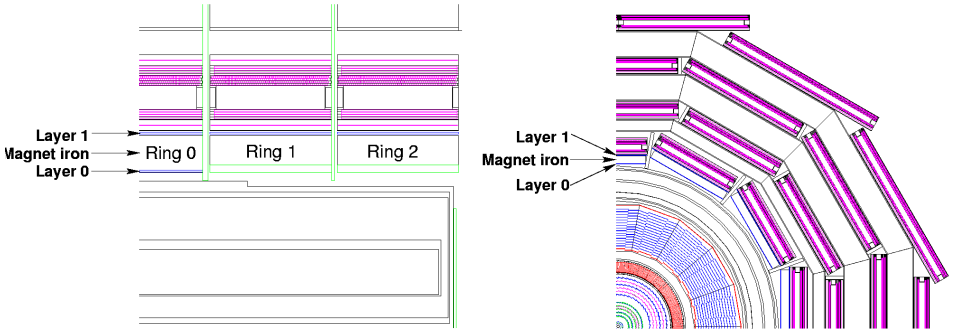

FIGURE 3.11: Longitudinal and transverse view of HO

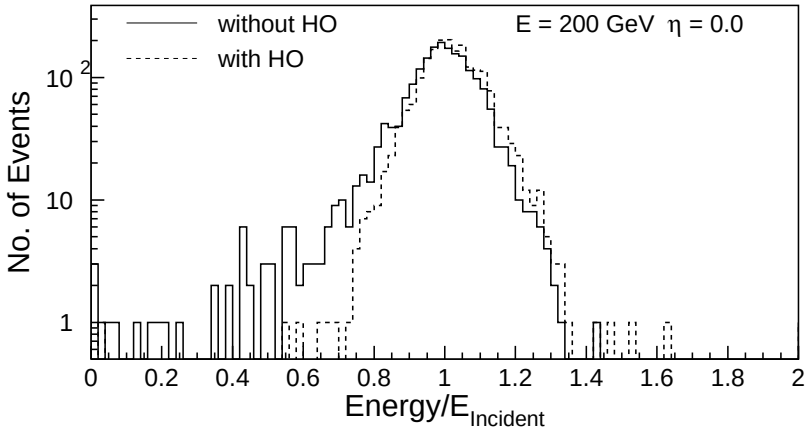

FIGURE 3.12: Improvement in pion energy resolution after using HO in addition to HB

rings. Location of rings in terms of z coordinates is -5.342 m, -2.686 m, 0, +2.686 m, +5.342 m (Fig. 3.11). As HB thickness is smaller at η=0, ring 0 has HO scintillators at radial distances 3.82 m and 4.07 m on either side of iron. For other rings, scintillator is only at a radial distance of 4.07 m. Tiles in HO are such that they map to towers in HB of granularity 0.087×0.087 in η and ϕ. HO helps to recover the effect of shower leakage as shown in Fig. 3.12.

Forward calorimeter (HF)

The HF covers eta range of $3.0 < |\eta| < 5.2$. Low p_{T} jets are mostly abundant in this region. As this eta region is beyond ECal and tracker coverage, jets and $p_{\mathrm{T}}^{\mathrm{miss}}$ measurement in forward region relies only on this calorimeter section.

Forward hadron calorimeter is designed with the requirement that both active and passive material should survive harsh radiation conditions. This led to choice of quartz fibre (fused-silica core and polymer hard-cladding) as active medium. Charged particles above the Cherenkov threshold (E $\geq$ 190 keV for electrons) shower and generate Cherenkov light. Only light that hits the core-cladding interface at an angle larger than the critical angle (71^0) contributes to the calorimeter signal. The fibres measure about 600 μm in diameter for fused silica core. Over 1000 km of fibres are used in the HF calorimeters. A steel absorber structure is composed of 5 mm

thick grooved plates. Fibres are inserted in these grooves. There are two sets of fibres; half run over the full depth of the absorber (165 cm), while the other half starts at a depth of 22 cm from the front of the detector. Electromagnetic showers are short in length, hence a large fraction of their energy is in the first 22 cm. Hadron showers are longer enough and deposit almost equal energy in both segments. This allows to measure electromagnetic and hadronic showers. The front face of the calorimeter is located at 11.2 m from the interaction point. Fibres run parallel to the beam line, and are bundled to form 0.175×0.175 towers in η, ϕ. Light is guided through air-core light guide, which has highly reflective (>90% at desired angle) lining to direct it inside of a hallow tube and then reach to photocathode. Photomultiplier tubes (PMTs) are used to convert the light signal to electronic measurement.

3.2.5 Muon chamber

Muon chambers have objective of muon identification, momentum measurement and triggering. Good muon momentum resolution and trigger capability is enabled by the high-field solenoidal magnet and its flux-return yoke. Barrel system has drift tubes which cover $|\eta| < 1.2$. Four muon stations are interspersed among the layers of the flux return plates. Out of four stations, one is inside the yoke, 2 additional layers embedded within the yoke and the other is outside. First three stations have 8 chambers; in two groups of four. Four chambers measure the muon coordinate in the r-ϕ bending plane, and other 4 chambers provide a measurement in the z direction. The fourth muon station does not contain the z-measuring planes. A drift-tube (DT) chamber is made of 3 (or 2) superlayers (SL), each made up of 4 layers of rectangular drift cells staggered by half a cell. The SL is the smallest independent unit of the design. Dead spots are avoided by such overlap of half a cell width. One SL, that is, a group of 4 consecutive layers of thin tubes in staggered geometry gives excellent time-tagging capability and provides a time resolution of a few nanoseconds. The constraints of mechanical stability, limited space, and the requirement of redundancy led to the choice of a tube cross section of 13×42 mm^2.

In endcaps, muon rates and background levels are high, and magnetic field is nonuniform. Due to these challenges, Cathode Strip Chambers (CSCs) are used in endcaps, which are multiwire proportional chambers comprised of 6 anode wire planes interleaved among 7 cathode panels. CSCs identify muons in range $0.9 < |\eta| < 2.4$. In the endcap-barrel overlap range, $0.9 < |\eta| < 1.2$, muons are detected by both the barrel drift tubes (DT) and endcap CSCs. A muon in the range $1.2 < |\eta| < 2.4$ crosses 3 or 4 CSCs. There are four stations of CSCs in each endcap. Chambers in each station are positioned perpendicular to the beam line and interspersed between the flux return plates. Cathode strips of each chamber run radially outward and provide a precision measurement in the r-ϕ bending plane. Offline reconstruction efficiency of simulated single-muon samples is typically 95-99% except in the regions around $|\eta|$=0.25 and 0.8 (between 2 DT wheels) and $|\eta|$=1.2 (transition between the DT and CSC systems), where the efficiency drops. Figure 3.13 shows one quarter view of muon chamber. A fast dedicated muon trigger device based on Resistive Plate Chamber (RPC) is used to identify the relevant bunch crossing to which a muon track is associated even in the presence of the high rate and background expected at the LHC. RPC is a basic double-gap module consisting of 2 gaps, operated in avalanche mode. The RPC is capable of tagging the time of an ionising event in a much shorter time than the 25 ns with adequate spatial resolution.

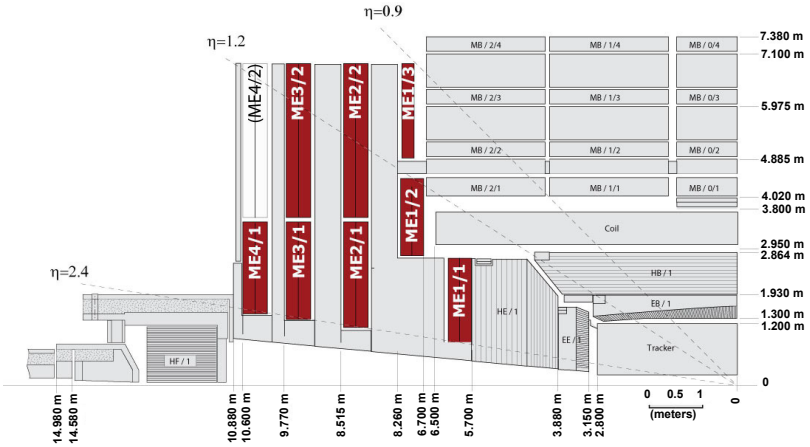

FIGURE 3.13: Muon chamber quarter view

3.3 Event reconstruction

As the various particles produced in a collision event traverses through the detector material, the energy deposited by them is converted into an electronic signature and recorded for further analysis. The next task is to convert these raw information into meaningful physics objects which can be further analyzed to understand the collision event in which these originated. The CMS uses a Particle Flow (PF) algorithm for event reconstruction [52], where information from all subdetectors is optimally combined and used for object reconstruction. The signature of various detector level objects can be simplistically summarized as follows:

- Electron: a track in tracker, energy deposit in ECal
- Photon: no track in tracker, energy deposit in ECal
- Muon: track in tracker, muon chamber
- Charged hadron: track in tracker, energy deposit in ECal and HCal
- Neutral hadron: no track in tracker, energy deposit in ECal and HCal
- Neutrino (or other weakly interacting particle): no signature in detector; imbalance in total transverse momenta

3.3.1 Track reconstruction

Every charged particle passing through a magnetic field gives rise to a bent trajectory in tracker in form of hits. Track reconstruction is done using iterative tracking [52]. It starts from a set of three or four hits as far from interaction point as possible to avoid confusion of nearby tracks, followed by extrapolation. If there are hits in chamber, within window of extrapolation error, extrapolation continues otherwise not. Once a track is reconstructed, the hits corresponding to that track are not considered for the next iteration. Curvature of a track can be used to get charge and momentum of a particle. As tracker is made up of light material (small atomic number), energy loss associated with production of hits in tracker is generally small. However, the support structure adds to the material budget which results in conversion of photons

to $e^{\pm}$ pair and bremsstrahlung radiation for electrons. These effects are accounted for in the momenta and position reconstruction of these particles.

3.3.2 Energy reconstruction in calorimeters

Electrons, photons and hadrons deposit most of their energies in ECal and HCal respectively, which are made up of heavy material like $PbWO_4$ and brass. Electrons and Photons loss energy by electromagnetic shower [53]. In electromagnetic showers (EM), electrons with energy above critical energy (where rate of energy loss by radiation and ionization become equal) travel one radiation length and give half of their energy to bremsstrahlung photon. Similarly a photon travels one radiation length and undergoes pair production with each created particle receiving half of its energy. Electrons below critical energy cease to radiate and ionization becomes the major energy loss mechanism. For photon, minimum energy required for pair production is $2m_ec^2$. Below that energy, compton scattering and photoelectric effect become dominant energy loss mechanisms. Electromagnetic showers are short and hence electrons and photons stop in ECal by depositing all of their energy. Hadron showers start with nuclear interaction. They have hadronic as well as purely electromagnetic components. Hadronic component is made up of protons, neutrons, charged pions etc. Electromagnetic component starts with π^0 decay to pair of photons, which further continues through pair production and bremsstrahlung of electrons. Hadronic showers are longer and deposit their energy both in ECal and HCal, before getting stopped. The difference in length scale of hadronic and EM shower provides ability to distinguish between particles.

Particle Flow (PF) reconstruction

The objective of calorimeter energy clustering is to detect energy from neutral particles and separate them from charged energy deposits. During Particle Flow (PF) reconstruction [52], separate clustering is done in each subdetector. No clustering is in HF (each cell gives one cluster). Clustering starts with local energy maximum, and grown such that at least one boundary of cell is shared with already existing cluster and has energy above threshold. For linking a cluster to track, a tangent is drawn from intersection point of charged track to tracker layers. This tangent is extrapolated to ECal. A cluster is linked to a track if extrapolated tangent position is within cluster boundaries. This way a bremsstrahlung photon energy cluster is linked to electron track. The output of PF algorithm is a list of PF candidates, namely, PF Electron, PF Muon, PF Charged Hadron, PF Neutral Hadron and PF Photon along with their position and four momenta coordinates.

3.3.3 Jet reconstruction

The quarks and gluons hadronize resulting in a spray or jet of particles. Exeprimentally, momenta of these particles can be inferred by clustering the final state particles. In CMS, we use Anti-kT algorithm with distance parameter R=0.4 (called AK4 jets) [54, 55] for clustering. For a Particle Flow (PF) jet, input to jet clustering is a list of PF candidates. The distance parameters used in this algorithm are as below:

$$d_{\mathrm{ij}} = min(p_{\mathrm{Ti}}^{-2}, p_{\mathrm{Tj}}^{-2}) \times \frac{\Delta_{\mathrm{ij}}^2}{R^2} \tag{3.1}$$

$$d_{\mathrm{iB}} = p_{\mathrm{Ti}}^{-2} \tag{3.2}$$

Here Δ^2_{ij} is spatial separation between two PF candidates (i,j):

$$\Delta^2_{ij} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2} \tag{3.3}$$

and R=0.4 decides the size of AK4 jet in (η,ϕ) plane. The p_T weighting of spatial distance ensures that clustering starts with highest p_T candidate. A closest PF candidate is merged with this highest p_T candidate only if the spatial distance between two is $\Delta^2_{ij} < 0.4$, else this highest p_T candidate is formed as an individual jet. For the next iteration of clustering, a merged system from previous iteration is treated like a new PF candidate. In this way, the process continues till all PF candidates are clustered into jets. Towards endcap and forward region, HCal cells get bigger, which results in lowered granularity and eventually resolution of jets. Jet reconstruction efficiency is function of p_T , it approaches unity at high p_T . For soft jets, it is possible that jets are discarded due to noise thresholds, hence efficiency is low there. Jet response is made uniform by applying Jet Energy Corrections (JEC) dependant on p_T and η of jets. A typical jet has 65% of jet energy as charged particles, 25% as photons, 10% as neutral hadrons.

3.3.4 b-jet reconstruction

With a superior CMS tracker, identification of jets originating from the hadronization of b-quarks is possible. The b-hadrons are hadrons containing b quark as constituent. These b-hadrons have sufficiently long lifetime (in rest frame 1.5 ps) that they decay to charmed hadron after traversing a finite distance. This leads to a displaced vertex with respect to the primary vertex of collision. This displaced vertex is called a "secondary vertex". Combined Secondary Vertex based on deep neural network (DeepCSV) is the algorithm used for b-jet tagging [56]. The csv based b-jet tagging relies on distinct properties of b-hadrons like large proper life time, large impact parameter (perpendicular distance between a back propagated track from secondary vertex and primary vertex), large mass, decays to final states with high charged track multiplicities (on average 5 charged tracks), large semileptonic branching ratio (20% times b hadrons decay to electrons or muons). Variables described above are combined into a single discriminating variable (CSV) using likelihood technique, which is used for tagging.

3.3.5 MET (p_T^{miss}) reconstruction

By conservation of momentum, total momentum before and after proton-proton collision should match both in longitudinal and transverse direction. Nonelementary nature of proton, makes it difficult to know longitudinal momentum transferred during collision on event by event basis. As net momentum in transverse plane before proton-proton collision is zero, in a balanced event net momentum in transverse plane after collision also should be zero. But, weakly interacting neutral particles leave undetected in the detector. In such scenario, vector sum of all visible particles in event does not add to zero. The p_T^{miss} is a measure of unbalanced p_T in a transverse plane to beam direction and is directed opposite to vector sum of p_T of all visible candidates in event [57]. However, any kind of malfunctioning of detector, anomalies in event reconstruction or resolution effects of jets can give rise to high p_T^{miss} , which can accidentally mimic signature of new physics or SUSY we are interested in. Precise p_T^{miss} measurement, is therefore, required in case of analyses having weakly interacting particles in final state (e.g. neutralino), which have true source

of $p_{\mathrm{T}}^{\mathrm{miss}}$. Section 3.5 discusses more on how handling the $p_{\mathrm{T}}^{\mathrm{miss}}$ becomes challenging and then how ensuring data quality becomes prime important.

3.4 Triggers

Trigger systems are indispensable components of the LHC experiments, and are aimed to select the pp collisions of physics interest for permanent storage for further detailed investigation [58]. With a collision rate of 40 MHz, and a typical collision event recorded by the CMS detector corresponding to 1-1.5 MB, approximately 40-60 TB data is produced per second while the permanent storage is limited to the rate of 1-5 GB/s to write to tape. Most of the collision events correspond to mundane multijet production, and production rate of new physics particles and electroweak SM processes is 5-6 orders of magnitude smaller. The concept of triggers allows to select the events of interest in real time and lead to selective storage of collision events for future investigations. The CMS trigger system does this event preselection online or in real time at two levels - hardware based Level 1 (L1) trigger and software based High Level Trigger (HLT).

3.4.1 L1 trigger

This is the first step of triggering where a preliminary event selection is done based on crude energy measurements in calorimeters and muon candidates in muon detectors. The CMS experiment does not yet use tracker information in L1-trigger. Instead of using a complete event reconstruction, decision of selecting an event is made based on presence of trigger primitive objects like electrons, photons, muons, jets above a certain threshold and high $p_{\mathrm{T}}^{\mathrm{miss}}$ in the event. The decision is conveyed to front end electronics, where information from all subdetectors at the most granularity and resolution is stored in buffers. If the decision of L1 is to store the event, the raw data is passed to the next level of trigger. Due to such crude selection based on reduced granularity, L1 manages to convey the decision within 3.2 μs and to reduce the event rate to about 100 kHz.

3.4.2 HLT trigger

The events selected by L1 trigger are passed to HLT, for the second level of data compression. Unlike L1 where decision is made at hardware level, HLT uses sophisticated algorithms to reconstruct the event using information from all sub detectors and employ a Particle Flow algorithm. These algorithms are run parallelly on a computing firm of around 5000 CPUs and the delivery of decision regarding whether to store an event takes around 50 ms. The advantage of being software based is that the thresholds can be adjusted even at the time of data taking depending on instantaneous luminosity. As independent trigger paths are checked simultaneously, sometimes the cross triggers with a cut on two or more independent physics objects are used for further data compression. Another approach used for data compression is "prescaling". The trigger paths with low p_{T} thresholds are maintained by setting prescale "n" to the path. This ensures that every n-th event triggered is only stored. This way, the event rate after passing through HLT is maintained to about 0.5-1 kHz, which can be stored on data storage devices.

3.5 Identification and rejection of events with large anomalous $p_{\mathrm{T}}^{\mathrm{miss}}$

3.5.1 Data and MC inputs

In 2018, analysis is carried with full Run2 data equivalent to 137 fb^{-1}. Data samples and CMSSW versions used for reconstruction are as follow:

- 2016 : 17Jul2018 94XReReco (35.9 fb^{-1})
- 2017 : 31Mar2018 94XReReco (41.5 fb^{-1})
- 2018 : 17Sep2018 102XReReco for runs A,B,C; PromptReco for run D (59.6 fb^{-1})

The details of various datasets along with corresponding luminosity are shown in appendix A.

SM background samples used for three years are as below:

- 2016 : Summer16 (94X MiniAODv3)
- 2017 : Fall17 (94X MiniAODv2)
- 2018 : Autumn18 (102X MiniAODv2)

More details on each sample are given in appendix B.

These MC samples are used to provide relevant inputs for background estimation strategies as will be discussed in following chapters. The samples are generated using MADGRAPH5_aMC@NLO [59, 60] and parton showering and hadronization is done using PYTHIA8 [61]. For 2016, CUETP8M1 tune [62] is used while for 2017 and 2018, CP5 tune [63] is used for underlying event modelling. The simulated background samples are normalized using most accurate available cross sections [59, 64–73]. The cross sections used correspond to next-to-next-to-leading-order (NNLO) calculation except for single top which used next-to-leading (NLO), QCD and γ + jet processes which used leading order (LO) cross sections. Detector response is modelled using GEANT4 [74] based package (also called "FullSim").

For signal MC, 94X Fast Simulation (FastSim) is used for 2016, 2017 and 102X Fast Simulation is used for 2018. Signal samples are generated using MADGRAPH5_aMC@NLO at leading order (LO) and normalized using NNLO plus next-to-next-to-leading logarithmic (NNLL) cross sections [75–79]. Parton showering and hadronization are done using PYTHIA8. For 2016, CUETP8M1 tune [62] is used while for 2017 and 2018, CP2 tune [63] is used. Detector response is described using CMS fast simulation programme [80, 81]. FastSim corrections are applied to account for FastSim against FullSim differences [82, 83].

3.5.2 Event cleaning to mitigate anomaouls large $p_{\mathrm{T}}^{\mathrm{miss}}$

Before starting any physics study using collected data, it is important to have a good understanding of data collected by detector to not get mislead by any false signal in detector. Data quality monitoring (DQM) phase before making data available for public usage, partially does this job; where data certification is done on weekly basis using a set of data quality histograms, which are further assessed by offline shifters and sub detector experts.

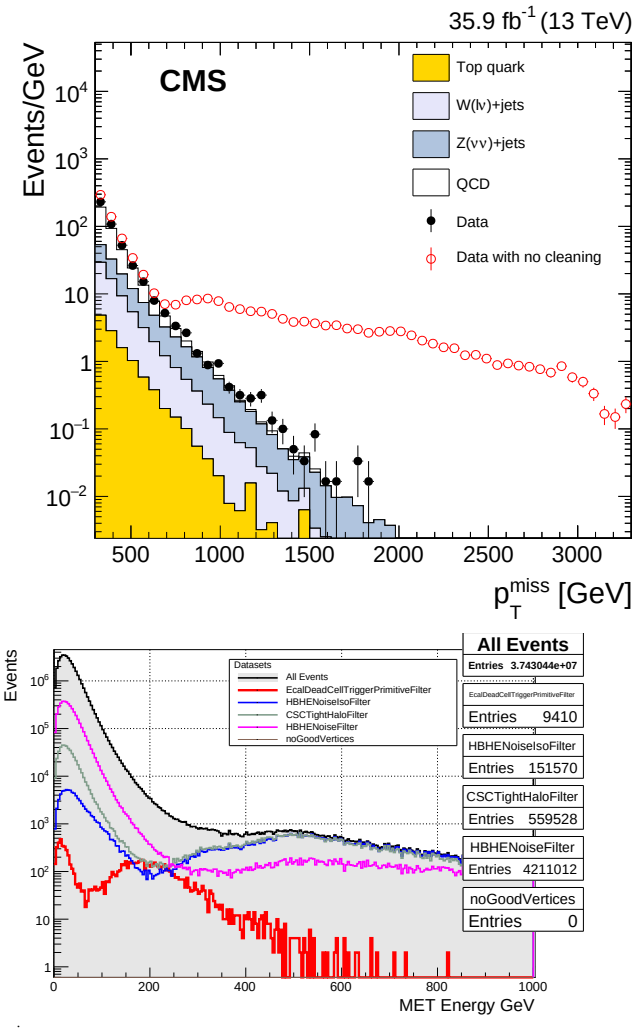

FIGURE 3.14: Anomalous MET event rejection during event cleaning [57]

But, even in a certified dataset, it is quite possible that certain anomalous MET events are there, where MET is not due to physics processes, but arising from detector malfunction. The fact that detector is constantly irradiated under such "busy environment" can lead to unforeseen detector efficiency loss or response delay in one or more channels causing spurious $p_{\mathrm{T}}^{\mathrm{miss}}$ in such cases. Such events are not useful for analysis purpose. Fig. 3.14 (top) shows a MET tail observed in data against MC simulation. In such cases, data is investigated to know origin of spurious $p_{\mathrm{T}}^{\mathrm{miss}}$ events. General strategy is to apply various event cleaning filters to take care of these issues. Fig. 3.14 (bottom) shows how MET tail events are rejected after applying various event cleaning filters. It also shows that high $p_{\mathrm{T}}^{\mathrm{miss}}$ region is affected more by the anomalous $p_{\mathrm{T}}^{\mathrm{miss}}$ issues compared to low $p_{\mathrm{T}}^{\mathrm{miss}}$ region.

Most of above filters are also used during Run 2 analysis. Following is the description of various filters [57].

- globalSuperTightHaloFilter

  The colliding proton beams get surrounded by halo of particles arising from interaction of incident protons with residual gases in beam pipe and beam pipe itself. Even after shielding used to protect from these particles, some high

energetic muons able to reach detector and can affect analysis at extreme cases. This filter is motivated towards tagging events with a halo muon.

- HBHENoiseFilter, HBHEIsoNoiseFilter

  To reject sporadic $p_{\mathrm{T}}^{\mathrm{miss}}$ events in HCal barrel and endcap, arising due to single-ion feedback noise, magnetic field dependant collective hybrid photodiode (HPD) noise and collective readout box (RBX) noise pulses. The filter uses geometric patterns of HPD or read out channels and pulse shape and timing information to tag anomalous $p_{\mathrm{T}}^{\mathrm{miss}}$ events, for example, an event is rejected if single HPD has more than 17 or 18 occupied cells.

- EcalDeadCellTriggerPrimitiveFilter

  In ECal, due to coarse resolution, it is not possible to get readout crystal by crystal. This filter uses trigger primitive information to identify fake $p_{\mathrm{T}}^{\mathrm{miss}}$ events. Trigger primitive value saturates above 127.5 GeV. If trigger primitive value is close to saturation value, in channels with no data link, this filter identifies and removes such events as they will be most probably populating $p_{\mathrm{T}}^{\mathrm{miss}}$ tail.

- ecalBadCalibReducedFilter

  Due to the intense dose of radiation, some of ECal endcap cells perform badly than their calibrated response, resulting into high p_{T} anomalous jets. The ecalBadCalibReducedFilter uses a list of bad cells in ECal as input and rejects an event with energy falling in such cell, which is most likely to be a noise event.

- BadPFMuonFilter

  To reject the scenarios where muon either fails the Particle Flow requirement or is a PF muon but has large uncertainty on its p_{T} .

- Good vertex filter

  To make sure event has originated from primary vertex. This filter requires that there should be atleast one reconstructed vertex which has $N_{\mathrm{dof}} > 4$ and $|z| < 24$ and $\rho < 2$.

In addition, some custom filters were used particular to analysis requirement.

- Muon jet filter

  Reject event with misreconstructed muon with $p_{\mathrm{T}} > 200$ GeV and muon energy fraction > 0.5 and

$$\Delta\Phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{jet}}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}) > \pi - 0.4 \tag{3.4}$$

- To reject anomalous jets in QCD CR

  Veto events if the leading jet (j_1) has neutral EM energy fraction < 0.03 and

$$\Delta\Phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{j_1}}, \vec{H}_{\mathrm{T}}^{\,\mathrm{miss}}) > (\pi - 0.4) \tag{3.5}$$

- PFCaloMET ratio filter

Uses MET calculated using Particle Flow objects (PFMET) and that before Particle Flow based reconstruction using calorimetry objects (CaloMET). To protect against PF failure reject event if

$$\frac{\text{PFMET}}{\text{CaloMET}} > 5 \tag{3.6}$$

- HT5/HT cut

  To reject anomalous jets in $3 < |\eta| < 3.1$ reject event if it fails below condition for a leading jet (j_1),

$$\frac{\text{HT5}}{\text{HT}} \leq max(1.2, 0.19 \times \Delta\Phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{j_1}}, \vec{H}_{\mathrm{T}}^{\,\mathrm{miss}}) + 0.90) \tag{3.7}$$

  Here HT5 represents scalar sum of p_{T} of jets with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 5.0$

As detector is continuously evolving with exposure to radiation during data taking, it may happen that existing set of cleaning filters are not enough to handle issues with data. In such cases, the data is further studied in analysis phase space, and criteria are developed to take care of various issues.

Run 2 round of analysis was more challenging after realization of various issues with data collected in different eras of 2016, 2017 and 2018. The 2016 and 2017 data was suffered by "Level 1 trigger primitive prefiring" issue. In addition ECal endcap noise issue badly affected later runs of 2017 data. Two disks in HE were found to have gone bad in later eras of 2018 data taking.

It was very important to understand the impact and having a recipe in hand to deal with these issues before starting with the real physics analysis.

3.5.3 Level-1 (L1) trigger primitive prefiring

During 2018 data commissioning, it was found that the gradual timing shift of ECal was not properly propagated to L1 trigger primitives in 2016 and 2017. This effect worsened with data taking as the transparency of crystals at high η gradually decresed over the course of time, and effective noise levels increased. As a result, by the end of 2017, a large fraction of L1 EGamma objects at $\eta > 2.0$ were associated to previous bunch crossing. This lead to “self veto” of events, as L1 trigger forbids consecutive bunch crossings to fire. The effect was corrected for 2018 data taking.

In analysis perspective, for 2016 and 2017 run eras, it was important to check how much is the impact of L1 prefiring issue on final state variables. To do this, in 2016 MC, analysis is repeated by removing events with jets having $p_{\mathrm{T}} > 100$ GeV and within $2.25 < |\eta| < 3.0$. As shown in figure 3.15, impact on variables of interest was 5-10% both in zero lepton (0L) and one lepton (1L) events, which was not insignificant. As a result, it was decided to apply central prefire maps to jets and electrons in 2016 and 2017 MC to make them consistent with data.

Impact of applying L1 prefiring maps

Data versus MC comparisons are made for leptons, H_{T}-$jets$ (jets within $|\eta| < 2.4$) and $H_{\mathrm{T}}^{\mathrm{miss}}$-$only$-$jets$ (jets within $2.4 < |\eta| < 5.0$) in respective η bins for objects having $p_{\mathrm{T}} > 100$ GeV as they were most affected by this issue. Figure 3.16 and 3.17 show how these comparisons look before and after applying prefire map on 2016 and 2017 MC respectively. Prefire corrections are more for 2017 compared to 2016. Data versus

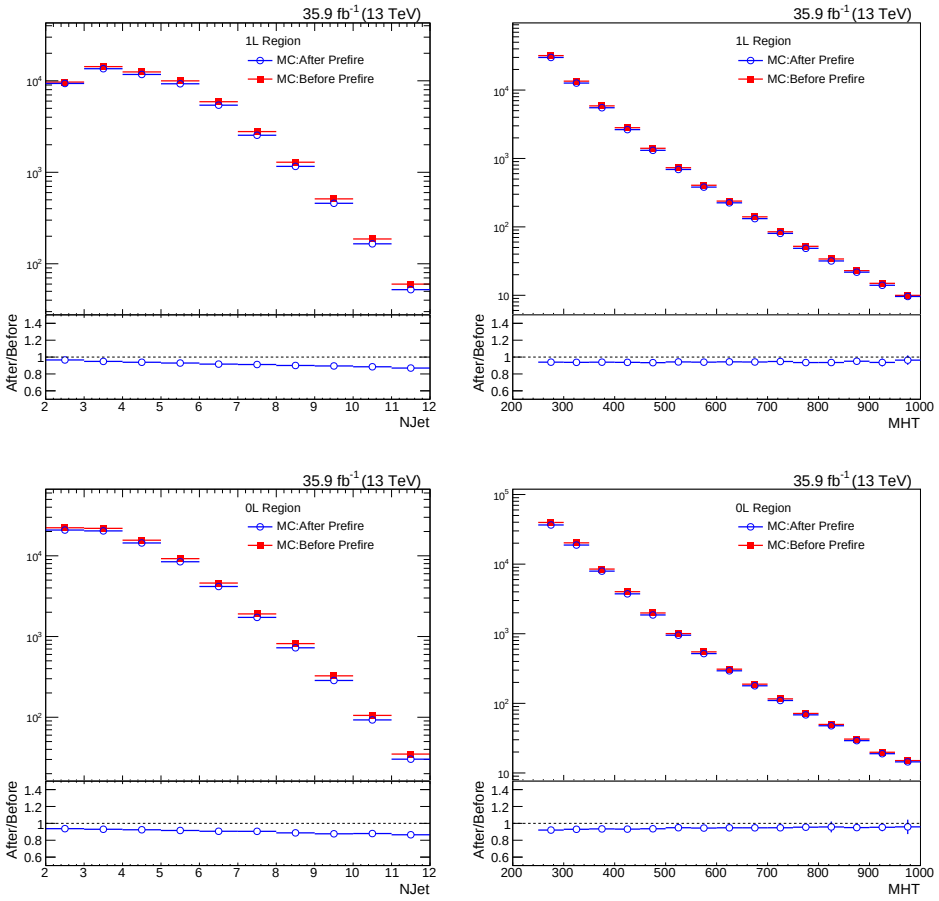

FIGURE 3.15: Impact on N_{jet}(left) and $H_{\text{T}}^{\text{miss}}$(right) distributions after skipping jets from prefire affected region in events containing an electron or a μ^- (1L) (top) or zero lepton (0L) (bottom)

MC agreement clearly shows improvement of about 10-15% after applying prefire map in most affected region.

3.5.4 ECal endcap (EE) noise

The CMS detector is continuously hit by energetic particles produced in pp collisions. As a result, detector performance degrades with time. By end of 2017, tansparency loss of ECal went up to 80-85%. This lead to increase in fake $p_{\text{T}}^{\text{miss}}$ (instrumental $p_{\text{T}}^{\text{miss}}$) more into later era of 2017 data (2017 F) than early eras (2017 B-E). To deal with this issue, centralized MET-v2 recipe was propagated at analysis level to 2017 dataset. MET-v2 recipe recalculated $p_{\text{T}}^{\text{miss}}$ after excluding jets with pT<75 GeV and unclustered candidates in 2.65< $|\eta|$ <3.139. Figure 3.18 shows that 2017B:E versus 2017F agreement improved by about 10% after applying MET-v2 recipe.

$H_{\text{T}}^{\text{miss}}$-*only-jets* : residual disagreements after MET-v2

Even after MET-v2 implementation, jets with 2.4< $|\eta|$ <5.0 (called $H_{\text{T}}^{\text{miss}}$-*only-jets* in following) showed prominent differences in F versus B:E comparison. In 2017 F, p_{T} vs η distribution showed asymmetry with respect to η =0 and p_{T} vs $\Delta\phi$ distribution showed a concentration of high p_{T} jets anti-aligned to $p_{\text{T}}^{\text{miss}}$ (Fig. 3.19 and Fig. 3.20).

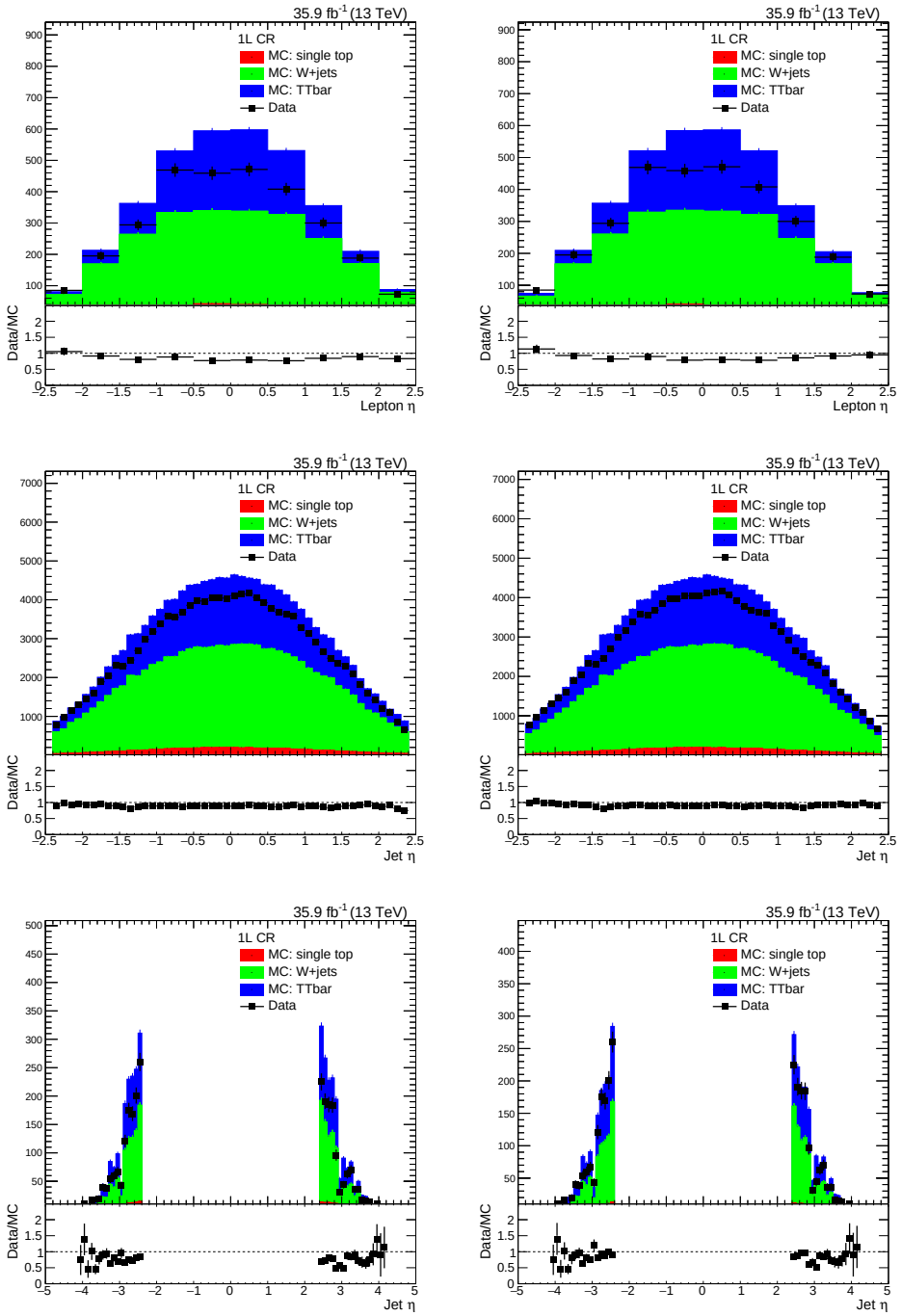

FIGURE 3.16: 2016 data versus MC comparison for lepton, H_{T}-*jet* and $H_{\mathrm{T}}^{\mathrm{miss}}$-*only-jet* with $p_{\mathrm{T}} > 100$ GeV after high $\Delta\Phi$ baseline selection in 1L control region before (left) and after (right) applying jet and photon prefire map to jets and leptons respectively.

EcalNoiseJet filter to reject residual EE noise effects

To reject high p_{T} noisy $H_{\mathrm{T}}^{\mathrm{miss}}$-*only-jets* appearing as bulge in their distribution of p_{T} versus $\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{jet}}, \vec{H}_{\mathrm{T}}^{\,\mathrm{miss}})$, a custom filter (called "EcalNoiseJet filter") was developed to reject the event from 2017F, if leading or subleading $H_{\mathrm{T}}^{\mathrm{miss}}$-*only-jet* with p_{T} >250 GeV has $\Delta\phi < 0.1$ or $\Delta\phi > 2.6$ with respect to $H_{\mathrm{T}}^{\mathrm{miss}}$. Figure 3.21 showed that

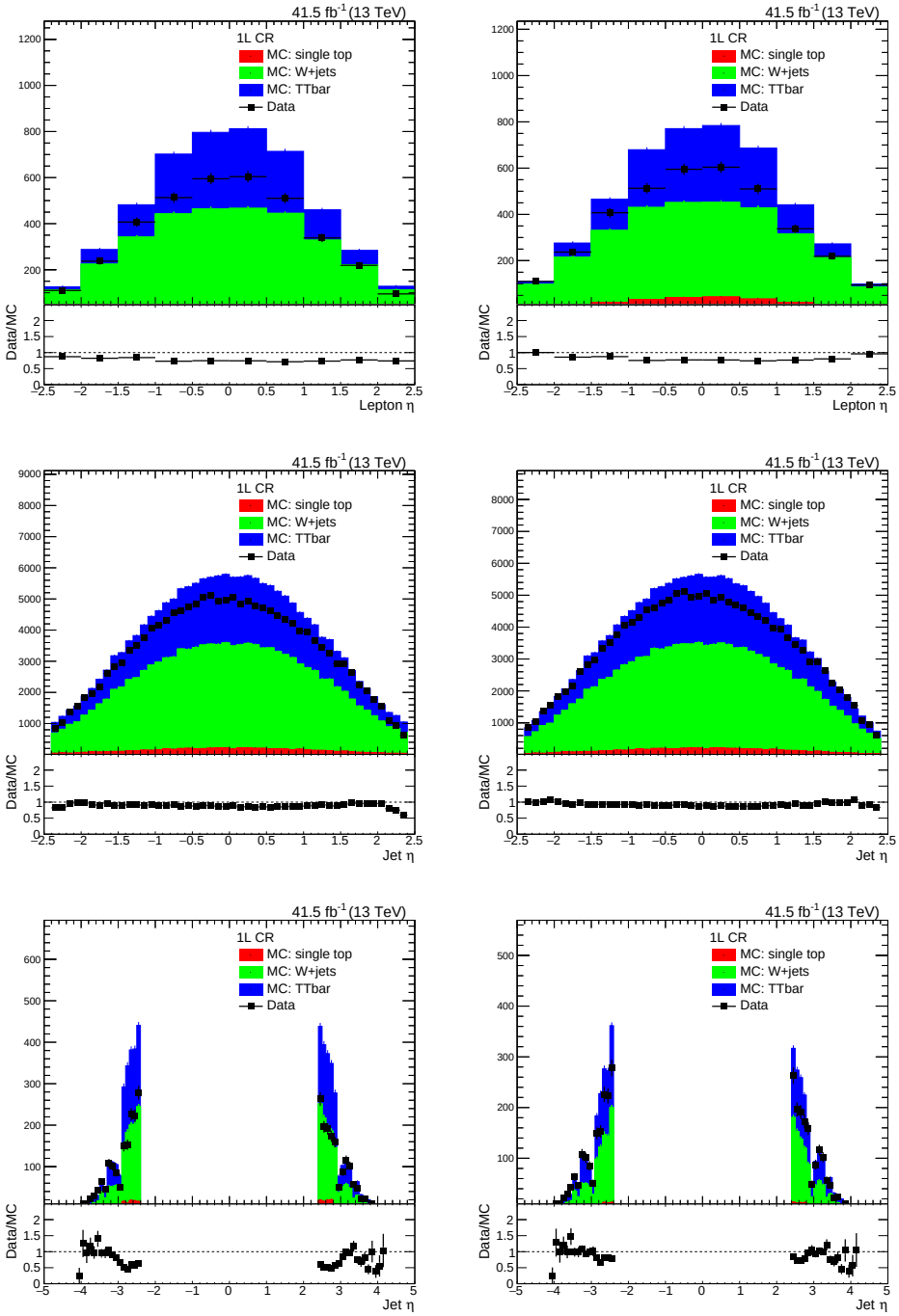

FIGURE 3.17: 2017 data versus MC comparison for lepton, H_{T}-*jet* and $H_{\mathrm{T}}^{\mathrm{miss}}$-*only-jet* with $p_{\mathrm{T}} > 100$ GeV after high $\Delta\Phi$ baseline selection in 1L control region before (left) and after (right) applying jet and photon prefire map to jets and leptons respectively.

after applying this filter, both p_{T} versus η and p_{T} versus $\Delta\phi$ distributions became symmetric in 2017 F. This filter also helped to make $H_{\mathrm{T}}^{\mathrm{miss}}$-*only-jet* p_{T} spectrum and neutral EM fraction comparison more consistent among early (2017 B:E) and later (2017 F) eras of 2017 data taking as shown in Fig. 3.22.

With the help of MET experts, the events rejected by EcalNoiseFilter were further

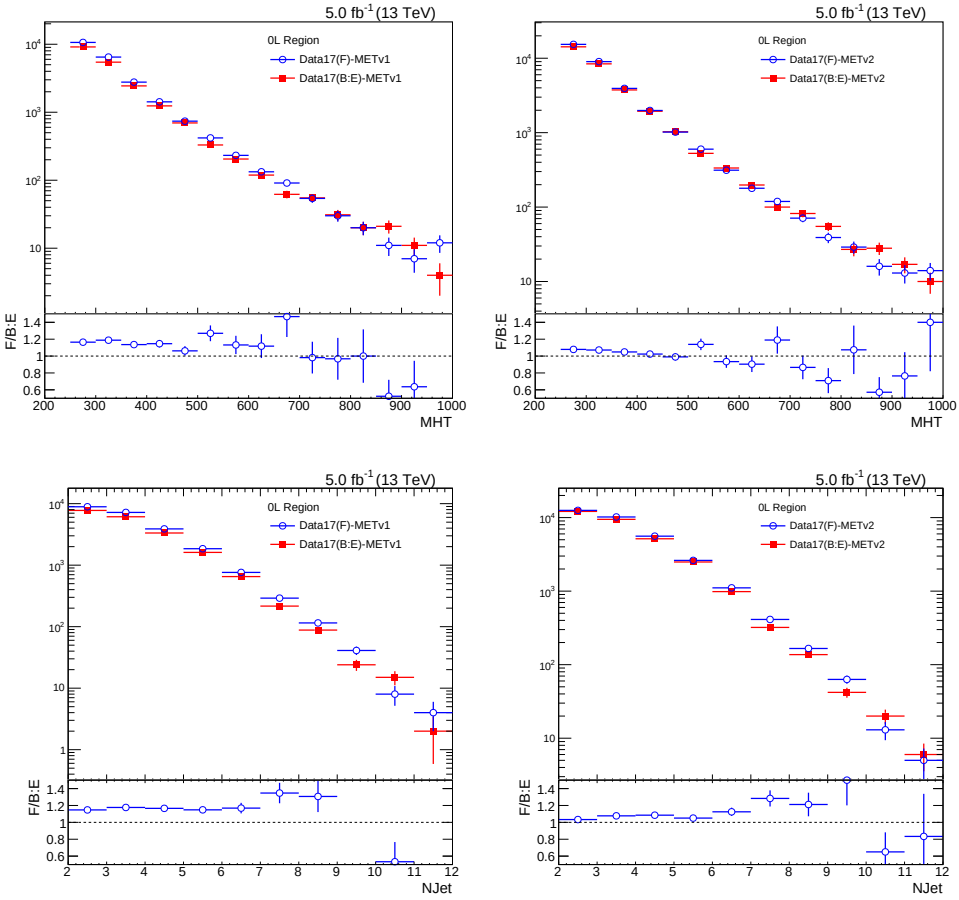

FIGURE 3.18: 2017F versus 2017 B:E data comparison in zero lepton high $\Delta\Phi$ baseline region before (left) and after (right) applying MET-v2 recipe ($p_{\mathrm{T}}^{\mathrm{miss}}$ after excluding jets with pT<75 GeV and unclustered candidates in 2.65< | η | <3.139)

scrutinized in more details. It was found that there are actually hot cells in Ecal may be arising due to calibration issue, which is causing an extra energetic jet to appear opposite to $H_{\mathrm{T}}^{\mathrm{miss}}$. This conclusion was consistent with the message extracted from above studies.

The understanding that extra bulge events are due to detector noise also got confirmed after seeing that EcalNoiseJet filter shows hardly any reduction in one lepton region dominated by true MET events as shown in Fig. 3.23 and Fig. 3.24. As this region is dominated by true $p_{\mathrm{T}}^{\mathrm{miss}}$ due to neutrinos accompanying lepton in the event, the effect of noisy jets is found to be very small or negligible.

Impact of EcalNoiseJet filter on signal samples

While the proposal had been to use EcalNoiseJet filter to get rid of $H_{\mathrm{T}}^{\mathrm{miss}}$-$only$-$jets$ coming from noise in forward region, it was important to check impact of this cut on signal efficiencies. Table 3.3 shows fractional events that pass when EcalNoiseJet filter is applied to inclusive set of events selected by applying baseline criteria (but as seen before impact is more on tails of jet p_{T}). Impact on signal and SM MC is less than 1% while in data, impact is 2-3% in high $\Delta\Phi$ region and 5-6% in low $\Delta\Phi$ region dominated by fake MET due to jet mismeasurement. As this cut is effectively

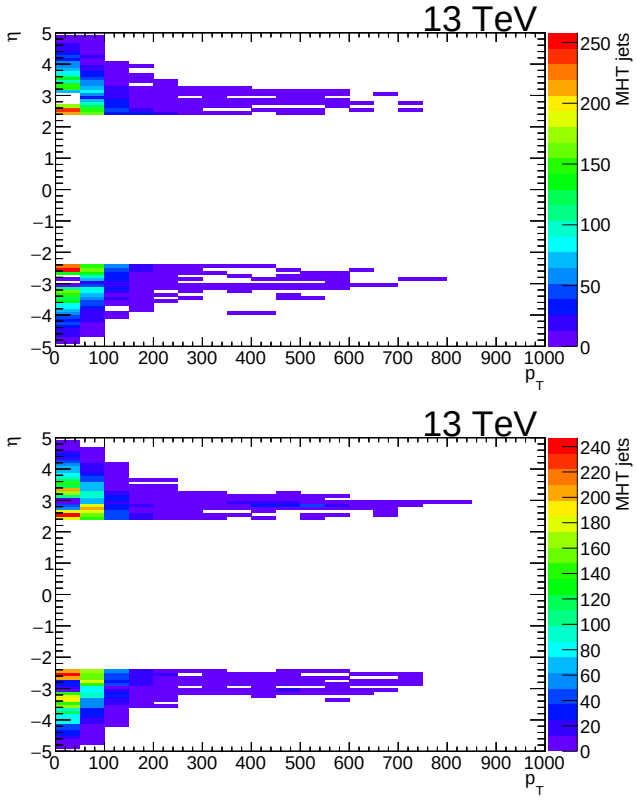

FIGURE 3.19: 2017F versus 2017 B:E data comparison: p_{T} vs η distribution for leading two $H_{\mathrm{T}}^{\mathrm{miss}}$-$only$-$jets$ for B:E (top) and F (bottom) in zero lepton high $\Delta\Phi$ baseline region after applying MET-v2 recipe ($p_{\mathrm{T}}^{\mathrm{miss}}$ after excluding jets with pT<75 GeV and unclustered candidates in 2.65<$|\eta|$<3.139)

rejecting events with fake MET without affecting true MET events, it was decided to apply it at analysis level.

TABLE 3.3: Impact of EcalNoiseJet filter

$\Delta\Phi$ baseline	SUSY model (mass point)	Passed filter (%)
high $\Delta\Phi$	T2qq: (1650,1)/ (1650,500)/ (1650,1400)	99.8/ 99.8/ 99.8
	T2tt: (1000,50)/ (1000,200)/ (1000,650)	99.8/ 99.7/ 99.7
	T1tttt: (1900,100)/ (1900,1000)/ (1900,1600)	99.9/ 99.8/ 99.6
	T1qqqq: (2000,1)/ (2000,500) /(2000,1400)	99.7/ 99.7/ 99.9
	Data: 2017B:E/ 2017F	98.4/ 97.8
	MC: $t\bar{t}$ +jets/ W+jets/ single top	99.4/ 99.7/ 99.7
low $\Delta\Phi$	Data: 2017B:E/ 2017F	94.3/ 93.1
	MC: $t\bar{t}$ +jets/ W+jets/ single top	99.0/ 99.6/ 99.3

H_{T}-$jets$: residual disagreements after MET-v2

Figure 3.18 (right), shows that differences in B:E vs F in N_{jet} have become severe after N_{jet}=6. To understand further, p_{T} distribution of H_{T}-$jets$ (p_{T} >30, $|\eta|$<2.4) is

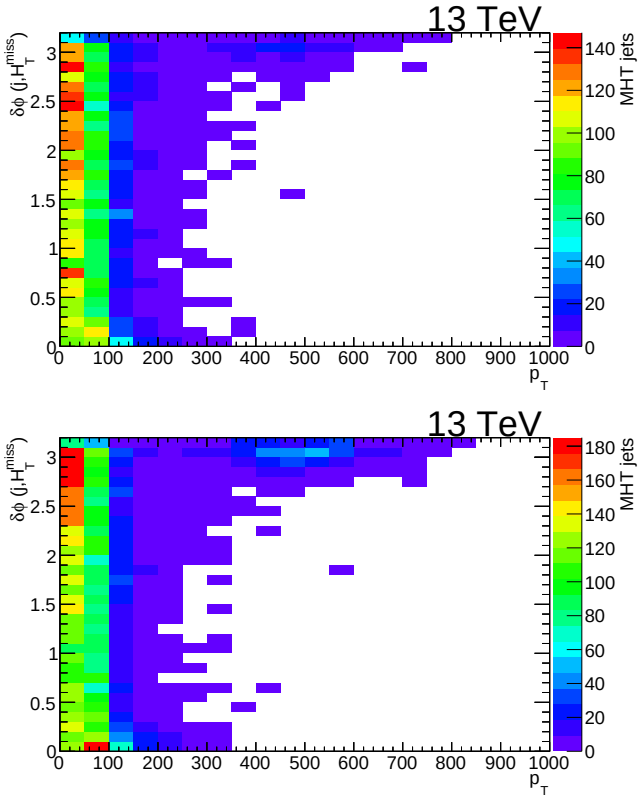

FIGURE 3.20: 2017F versus 2017 B:E data comparison: p_{T} vs $\Delta\Phi$ with respect to $H_{\mathrm{T}}^{\mathrm{miss}}$ for leading two $H_{\mathrm{T}}^{\mathrm{miss}}$-*only-jets* (2.4< $|\eta|$ <5.0) in B:E (top) and F (bottom) in zero lepton high $\Delta\Phi$ baseline region after applying MET-v2 recipe ($p_{\mathrm{T}}^{\mathrm{miss}}$ after excluding jets with pT<75 GeV and unclustered candidates in 2.65< $|\eta|$ <3.139)

plotted for $N_{\mathrm{jet}}<6$ and $N_{\mathrm{jet}}\geq 6$ cases (figure 3.25). For $N_{\mathrm{jet}}\geq 6$, flat 20% discrepancy is observed between B:E and F for jet p_{T} <500 GeV, which is not the case for $N_{\mathrm{jet}}<6$.

To investigate this discrepancy, jet p_{T} versus photon multiplicity is plotted for all H_{T}-*jets* in events passing high $\Delta\Phi$ baseline selection. Fig. 3.26 shows how the ratio 2017F/2017(B:E) looked. The bins where this ratio is ≥ 2 or ≤ 0.5 must be noise prone bins as energy cluster due to noise will not find associated track and will be categorized as photon in such cases. To check impact of these noise prone bins, an event passing high $\Delta\Phi$ baseline selection with at least one H_{T}-*jet* falling into any of noise prone bins is rejected and then inclusive distribution comparisons are rechecked (figure 3.27). Both distributions show improvement in F/B:E agreement about 5-7%. Considering total luminosity associated with full Run 2 dataset, this improvement will be of sub-percent order, hence it was decided not to put any hard cut based on photon multiplicity.

3.5.5 HCal endcap readout failure (HEM) in 2018 dataset

Two disks of HCal Endcap (15, 16) failed during 2018 data taking. The issue is known as HE-Missing or HEM failure. Missing HE-15/16 corresponds to -3 $<\eta<$ -1.4 and -1.57 $<\phi<$ -0.87 (40°). In this region it was expected to get fake leptons

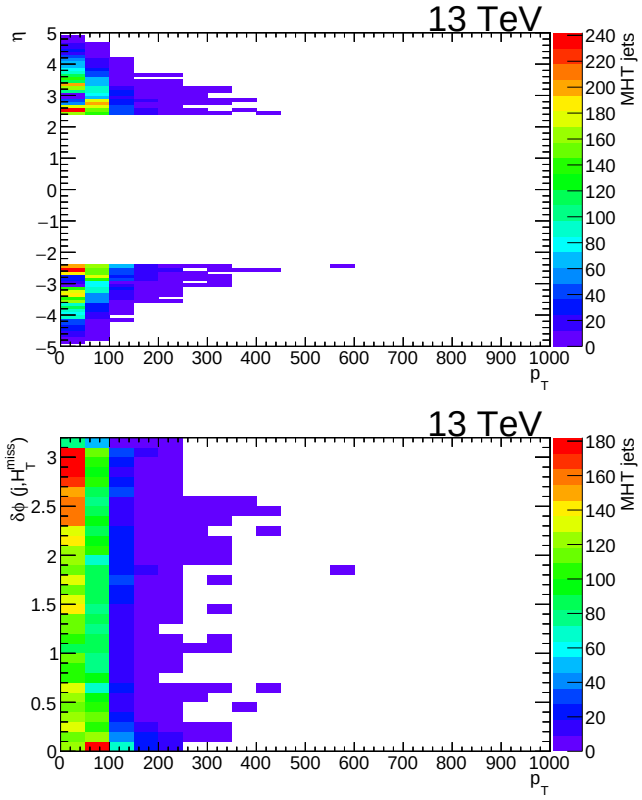

FIGURE 3.21: p_{T} vs η (top) and p_{T} vs $\Delta\Phi$ with respect to $H_{\mathrm{T}}^{\mathrm{miss}}$ (bottom) for leading two $H_{\mathrm{T}}^{\mathrm{miss}}$-*only-jets* (2.4< $|\eta|$ <5.0) in zero lepton high $\Delta\Phi$ baseline region after applying EcalNoiseJet filter along with MET-v2 recipe ($p_{\mathrm{T}}^{\mathrm{miss}}$ after excluding jets with pT<75 GeV and unclustered candidates in 2.65< $|\eta|$ <3.139) in 2017F

as there would not be any HCal deposit associated to ECal deposit hence accidentally categorizing an energy cluster falling into HE-15/16 region as EGamma object at PF level. Jets directed along HEM affected region will be undermeasured, causing fake MET in that direction, leading to overall extra MET measurement in the event.

Impact on one lepton region in data

For preliminary analysis of impact of HEM issue on one lepton region, Run2018B-RelVal MET dataset specially reconstructed with HEM (called HEM dataset) and without HEM issue (called nominal dataset) is used. As events in nominal and HEM datasets are exactly same (but reconstructed differently), a map of event number, lumiblock number and run number is used to achieve event by event matching of two datasets. Table 3.4 shows event yield in HEM dataset under condition of

A] Event has one electron [1e] or one muon [1μ]

B] Event has one electron or one muon and the same event is also tagged as one lepton [1L] event in nominal dataset

Thus HEM dataset has 91 extra one lepton events compared to nominal dataset. Using event map, these exact 91 events are picked from nominal dataset and cutflow is studied.

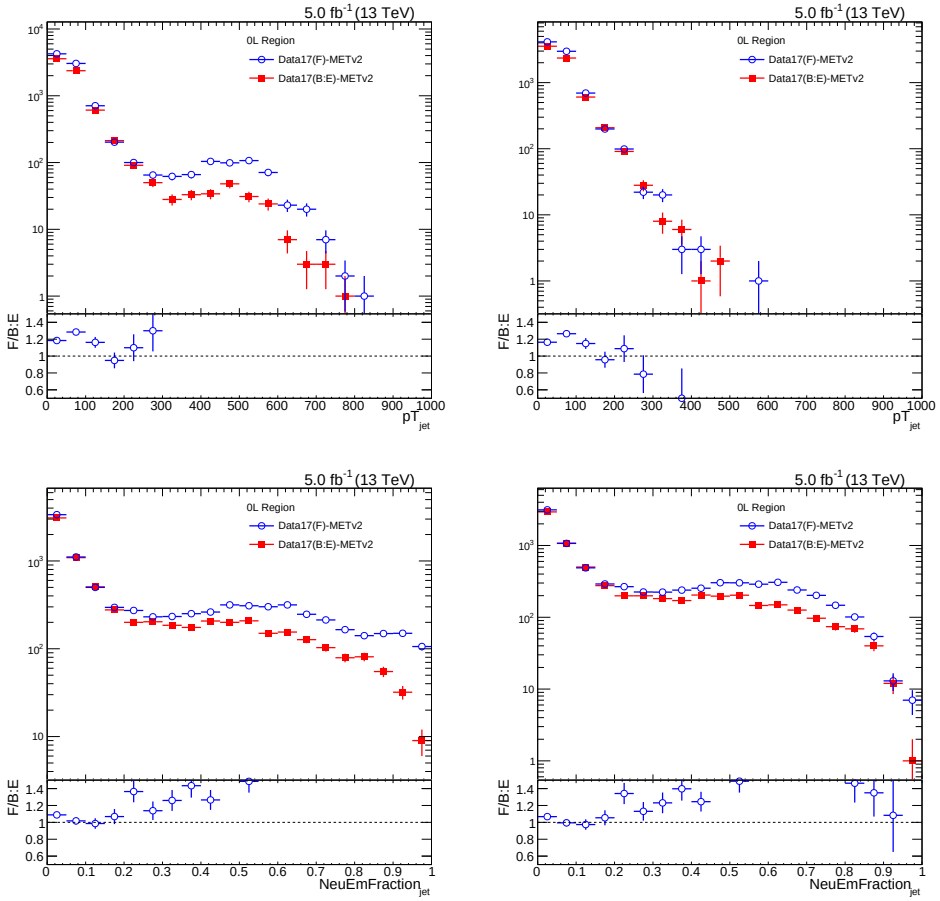

FIGURE 3.22: 2017F versus 2017 B:E data comparison for $H_{\mathrm{T}}^{\mathrm{miss}}$-*only-jets* (2.4< $|\eta|$ <5.0) : p_{T} and neutral EM fraction in zero lepton high $\Delta\Phi$ region (left) after MET-v2 recipe ($p_{\mathrm{T}}^{\mathrm{miss}}$ after excluding jets with pT<75 GeV and unclustered candidates in 2.65< $|\eta|$ <3.139) but before EcalNoiseJet filter (right) after MET-v2 recipe and after EcalNoiseJet filter

TABLE 3.4: 1L yield from HEM dataset before and after asking for matched 1L event in Nominal dataset

class	Yield after [A]	Yield after [B]
1L	3537	3446
1e	770	724
1μ	2767	2722

Table 3.5 shows that 46 extra 1e events coming in HEM are mostly due to lepton misreconstruction, while 45 extra 1μ events are either due to misreconstruction or accidental passing of some of the filters.

Due to low statistics of available sample, baseline selection was not imposed on the comparisons made with Run2018B-RelVal MET dataset. With full 2018 dataset, data versus MC comparison is made in one lepton region after baseline selections are applied.

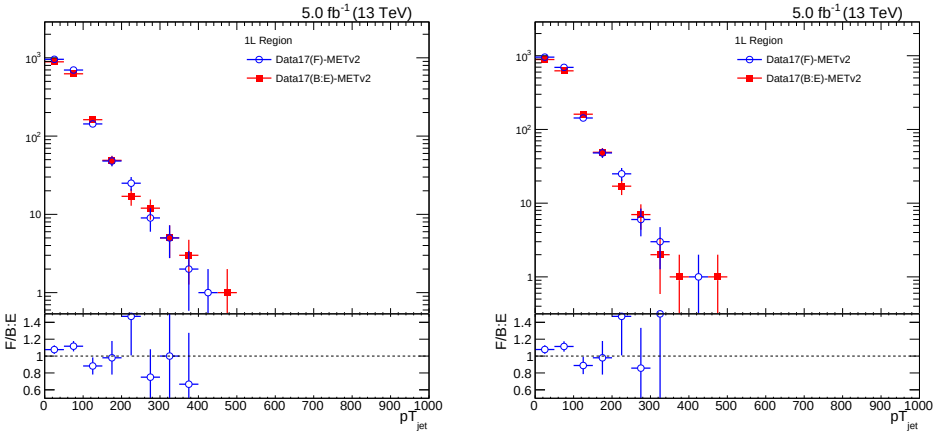

FIGURE 3.23: p_{T} spectrum of $H_{\mathrm{T}}^{\mathrm{miss}}$-*only-jets* in one lepton high $\Delta\Phi$ baseline region before (left) and after (right) applying EcalNoiseJet filter along with MET-v2 recipe ($p_{\mathrm{T}}^{\mathrm{miss}}$ after excluding jets with pT<75 GeV and unclustered candidates in 2.65< $|\eta|$ <3.139) in 2017F

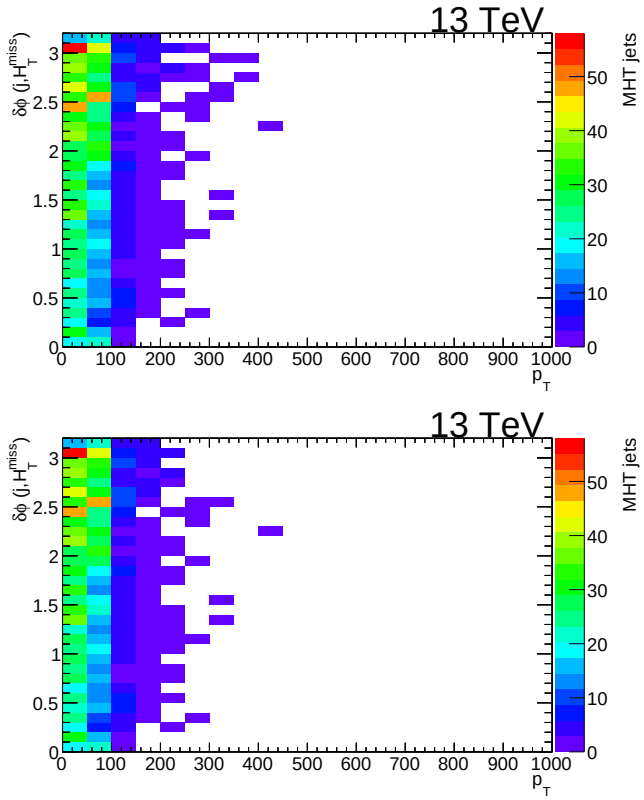

FIGURE 3.24: p_{T} vs $\Delta\Phi$ with respect to $H_{\mathrm{T}}^{\mathrm{miss}}$ for leading $H_{\mathrm{T}}^{\mathrm{miss}}$-*only-jet* in one lepton high $\Delta\Phi$ baseline region before (top) and after (bottom) applying EcalNoiseJet filter along with MET-v2 recipe ($p_{\mathrm{T}}^{\mathrm{miss}}$ after excluding jets with pT<75 GeV and unclustered candidates in 2.65< $|\eta|$ <3.139) in 2017F

Figure 3.28 shows that after imposing baseline selection, excess 1μ events seen in HEM dataset no longer persist further to contribute to one lepton region, but excess

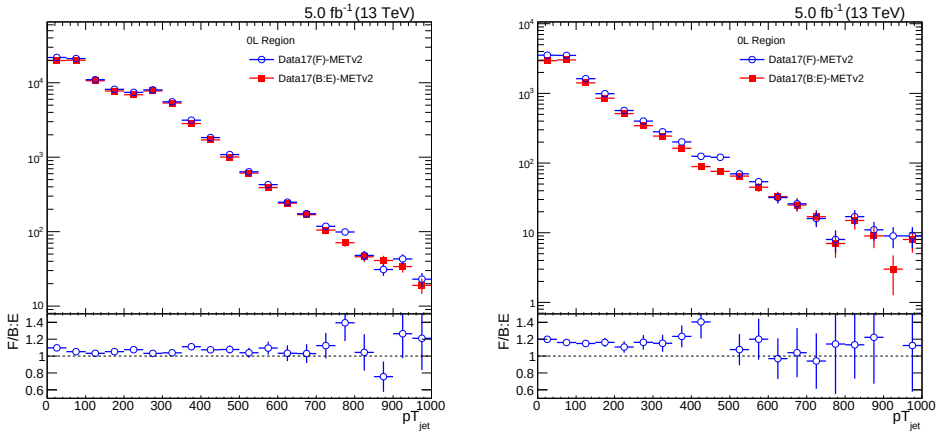

FIGURE 3.25: p_{T} spectrum of all H_{T}-$jets$ if event has total $N_{\mathrm{jet}} < 6$ (left) and $N_{\mathrm{jet}} \geq 6$ (right) for events passing high $\Delta\Phi$ baseline selection

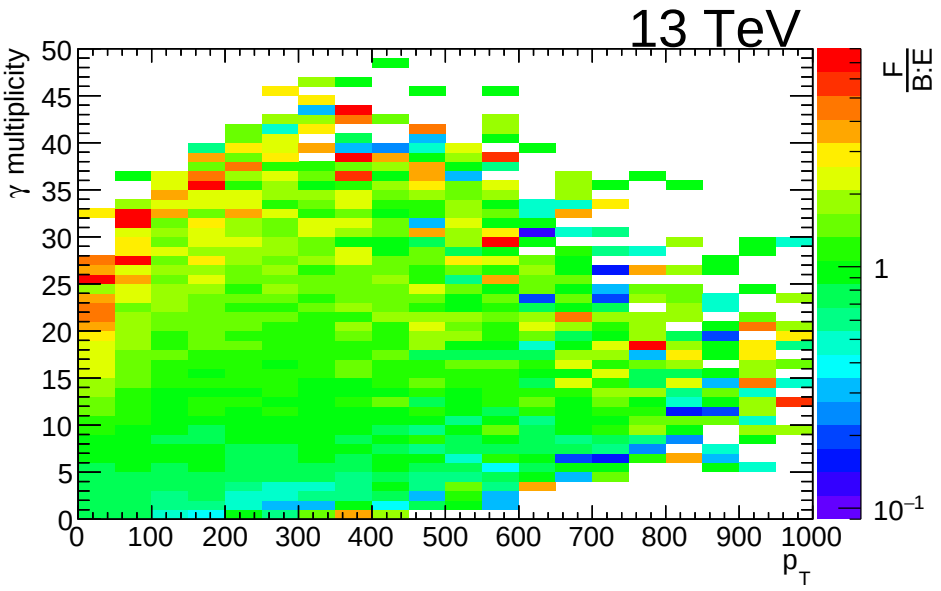

FIGURE 3.26: Ratio F/B:E obtained for p_{T} vs photon multiplicity distribution of H_{T}-$jets$ from events passing high $\Delta\Phi$ baseline selection

1e events show up as a bump in data versus MC comparison.

Nominal and extended HEM veto

As a conservative approach towards handling this issue, a first version for vetoing excess one electron events, called "Nominal HEM veto" was proposed as mentioned below:

Reject an event

- if there is an electron (passing identification and isolation requirement) falling in HEM affected region

  OR

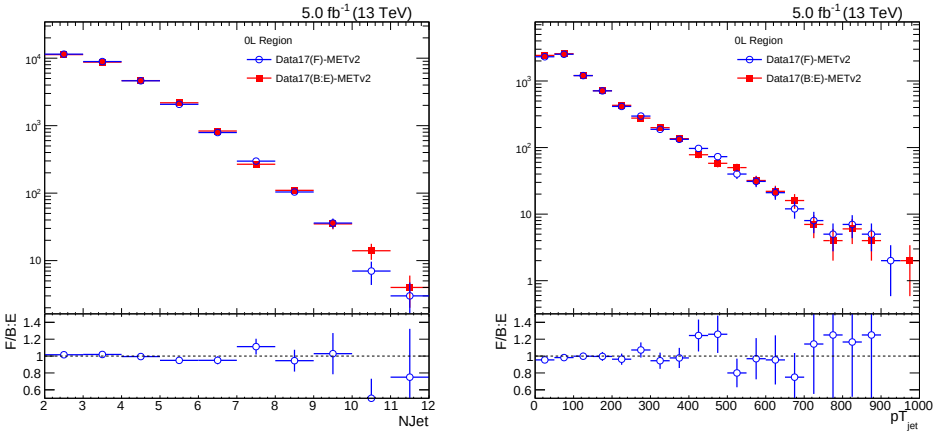

FIGURE 3.27: 2017F versus 2017 B:E data comparison in zero lepton high $\Delta\Phi$ baseline region after applying 2D photon cut with MET-v2 recipe ($p_{\mathrm{T}}^{\mathrm{miss}}$ after excluding jets with pT<75 GeV and unclustered candidates in 2.65< $|\eta|$ <3.139): (left) N_{jet} distribution, (right) p_{T} spectrum of all H_{T}-$jets$ if event has total $N_{\mathrm{jet}} \geq 6$

TABLE 3.5: Details of excess 1L events entering into HEM dataset due to issues like accidental passing of filters or lepton misreconstruction

Cut	HEM yield	Nominal yield	(Mis)identity of excess events in HEM [1e,1μ]
Total	91	91	0
PFCaloMETRatio	91	65	26 [0, 26]
noMuonJet	91	64	1 [0, 1]
JetID	91	63	1 [0, 1]
Trigger	91	63	0 [0, 0]
1L (without isolation)	91	25	38 [35, 3]
1L (isolated)	91	0	25 [11, 14]

- if there is a jet with $p_{\mathrm{T}} > 30$ GeV falling in HEM affected region

This improved data versus MC agreement in 1L control region considerably by rejecting fake lepton contribution as shown in figure 3.29.

But this veto does not seem to clear completely extra low pT jets counted in HEM affected region as shown in figure 3.30.

Figure 3.30 (left) shows the ratio of 2018D to 2018A datasets for p_{T} versus ϕ distribution plotted for $H_{\mathrm{T}}^{\mathrm{miss}}$-$only$-$jets$ after applying "Nominal HEM veto" to 2018D. Ratio distribution shows that there are remnant excess $H_{\mathrm{T}}^{\mathrm{miss}}$-$only$-$jets$ around HEM affected region. These must be the jets which partly fell into HEM affected region (hence undermeasured) but failed to get rejected by "Nominal HEM" veto. Figure 3.30 (right) shows ratio plot of p_{T} versus $\Delta\Phi$(jet,MHT) distribution of $H_{\mathrm{T}}^{\mathrm{miss}}$-$only$-$jets$ from 2018 RunD in a rectangle around HEM affected region (-3.2 < $|\eta|$ < -2.4, -1.8 < $|\phi|$ < -0.5) to that of all $H_{\mathrm{T}}^{\mathrm{miss}}$-$only$-$jets$ from 2018D in figure 3.30 (left). This ratio plot depicts that these excess $H_{\mathrm{T}}^{\mathrm{miss}}$-$only$-$jets$ are mostly low p_{T} jets aligned along direction of $H_{\mathrm{T}}^{\mathrm{miss}}$.

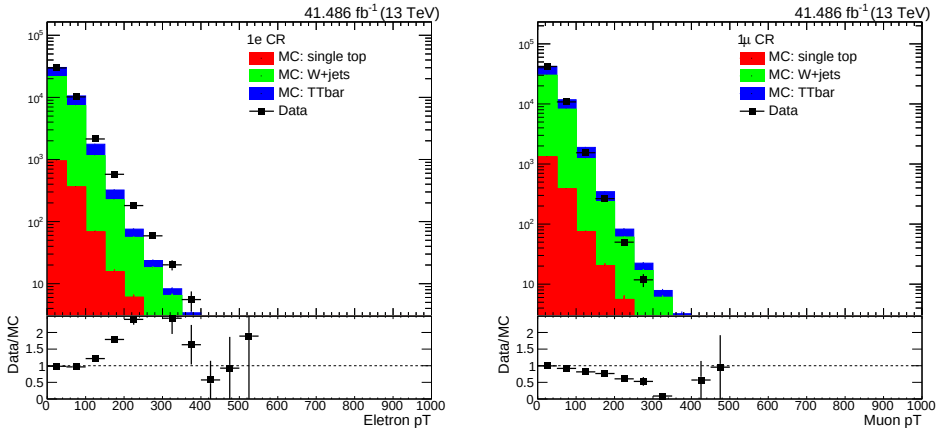

FIGURE 3.28: 2018 data versus 2017 MC comparison showing impact of HEM issue

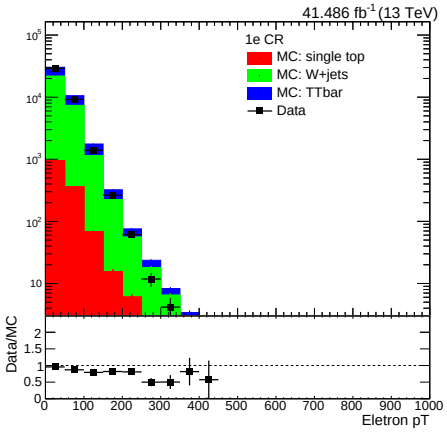

FIGURE 3.29: Impact on "Nominal HEM" veto on electron p_{T} distribution in 1L region as seen by 2018 data versus 2017 MC comparison

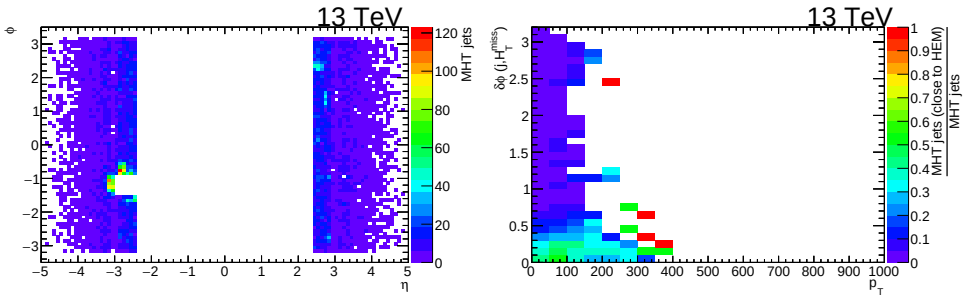

FIGURE 3.30: Remnant excess in leading two $H_{\mathrm{T}}^{\mathrm{miss}}$-*only-jets* seen around HEM affected region after "Nominal HEM veto" (left). Majority of these excess jets are found to be low p_{T} and aligned along direction of $H_{\mathrm{T}}^{\mathrm{miss}}$(right).

As a remedy, "Nominal HEM" veto was extended by 0.2 in η - ϕ directions, in order to reject undermeasured jets due to HEM. Also based on above studies, this

extended HEM veto is asked to check $\Delta\Phi$(jet,MHT) for a jet falling into the extended veto region and reject the event only in case $\Delta\Phi$(jet,MHT) < 0.5. The "Nominal HEM veto" on electrons is kept unchanged.

Figure 3.31 shows similar plots but now after extended HEM veto. Clearly low p_{T} excess jets entering in signal region due to undermeasurement are effectively rejected. This extended HEM veto with additional $\Delta\Phi$ requirement on jets is also found to be the best option as it's causing minimal signal efficiency loss of about 2%.

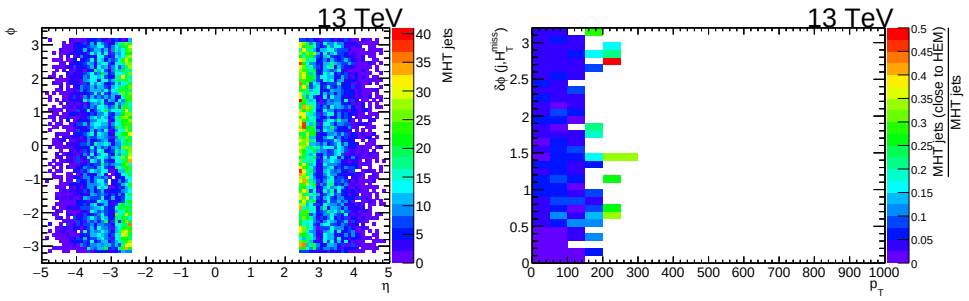

FIGURE 3.31: Distribution of $H_{\mathrm{T}}^{\mathrm{miss}}$-*only-jets* after extended HEM veto (left). Majority excess jets seen with "Nominal veto" in low p_{T} region are rejected with extended veto (right)

Chapter 4

Fast Simulation of hadron showers in CMS calorimeters

4.1 Introduction

In High Energy Physics (HEP) research, the role of Monte Carlo (MC) simulation is indispensable. Every generated event has to account for detector response to make it equivalent to the real event produced in data. The data delivered in any HEP experiment, is "nature given" and hardly possible (rather not intended) to be controlled manually. As an example, SM Higgs boson production cross section at 13 TeV is of order of few pb. This is many orders of magnitude smaller than the QCD multijet production ($\sim mb$) or electroweak boson production ($\sim nb$). The cross section just gives relative probability of production, but in data, there is no intuitive way to predict which SM process would happen in a given pp interaction. As a result, the number of events produced in data for various detailed measurements is limited by the production cross section of the process and total integrated luminosity. On other hand, in simulations one can generate events corresponding to any process, however rare, as high as one desires and then scale down the total number to match expected number for given luminosity in data. Due to the manual control involved, the MC simulations have played vital role in many Standard Model (SM) measurements, detector performance studies and calibration work.

In CMS, currently GEANT4 based full simulation (called FullSim) is most commonly used. The FullSim involves detail modelling of how CMS detector responds to various particles coming out of proton-proton collision. This modelling is done based on the information of how particle-matter interactions happen, saved as a database in Geant4 package. FullSim response is validated extensively using test beam experiments, which makes it equivalent to data. Along with FullSim, a parameterized mode of simulation using a simplified geometry, called FastSim is also employed by CMS. As an example, the energy loss by a particle passing through a detector block of length "L" (ΔE), will be modelled in FastSim as a particle passing through a thin layer of material, with associated parameterized energy response to this layer, such that particle-trace through this layer will loss ΔE amount of energy in case of a physical detector. Currently, FastSim is extensively used for "new physics sample" generations. This chapter discusses various studies done to check and improve performance of FastSim with reference to FullSim.

4.2 Motivation to study FastSim performance

Currently, with the availability of a well validated FullSim, the MC simulation of most of the Standard Model processes uses FullSim. But, because of detailed simulation of Geant4 based material interactions and actual detector geometry, FullSim response becomes really slow. As an example, for a single $t\bar{t}$ event simulation, the time taken is around 50 sec even in clean environment (without pile up). This time consumption becomes a real concern when there is a need to generate large number of events of particular rare process. The processes predicted by any supersymmetry (SUSY) model have very low cross-section compared to SM processes, implying very less number of events are expected compared to SM process for a given luminosity. As a result, any SUSY signal simulated for purpose of data analysis, need to have large number of events to get enough statistics. Also the SUSY parameters are not known. This requires that, during a SUSY search, such large statistics sample needs to be produced not just for a single SUSY mother-daughter mass pair, but for multiple mass points in SUSY mother-daughter mass plane, which are scanned simultaneously to ensure that the phase space covered in the search is as maximum as possible. Because of lengthy time demand, FullSim no longer remains an optimal choice in such cases and FastSim is preferred over FullSim in such cases. At present, SUSY and Exotica analysis groups rely extensively on FastSim for sample generation of various models. The need of FastSim is going to increase in future in High Luminosity LHC era (HL-LHC), when even simulating a single SM process sample with FullSim corresponding to such high luminosity can be cumbersome because of being computationally intensive. This underlines the fact that "FastSim" can be a potential substitute for "FullSim", if its performance is in very good agreement with FullSim.

4.3 Performance of FastSim

When FastSim was developed, the objective was to be as close to FullSim as possible but to reduce the time consumption of simulation. To achieve faster and still comparable response to FullSim, FastSim uses simplified detector geometry and parameterization for modelling particle interactions. For example, in FastSim, one detector block will be approximated by a thin material layer and instead of using Geant4 to predict how much energy, particle would have lost while traversing that detector block, a parameterized mathematical form will be used as a plug in to find the lost energy fraction. Due to the approximations involved, FastSim response is not as perfect as FullSim but the time consumption is reduced considerably.

While FastSim was devised as an alternative to FullSim, it was designed such that the measurement of high level physics objects, like leptons, jets, MET is consistent with FullSim description. But after looking at some more detailed quantities, it is clearly seen that FastSim response deviates from that of FullSim, at many places. As an example figure 4.1 shows that MET Significance in $t\bar{t}$ simulated events is consistent with data in FullSim but not in FastSim. To understand the origin of such discrepancies, it was important to revisit FastSim starting from simulation of basic quantities.

4.3.1 Hadron showers in calorimeters

The jets, formed by hadronization of quarks and gluons are ubiquitous at the LHC & are indispensable tools for new discoveries at the CMS experiment. Hadrons within

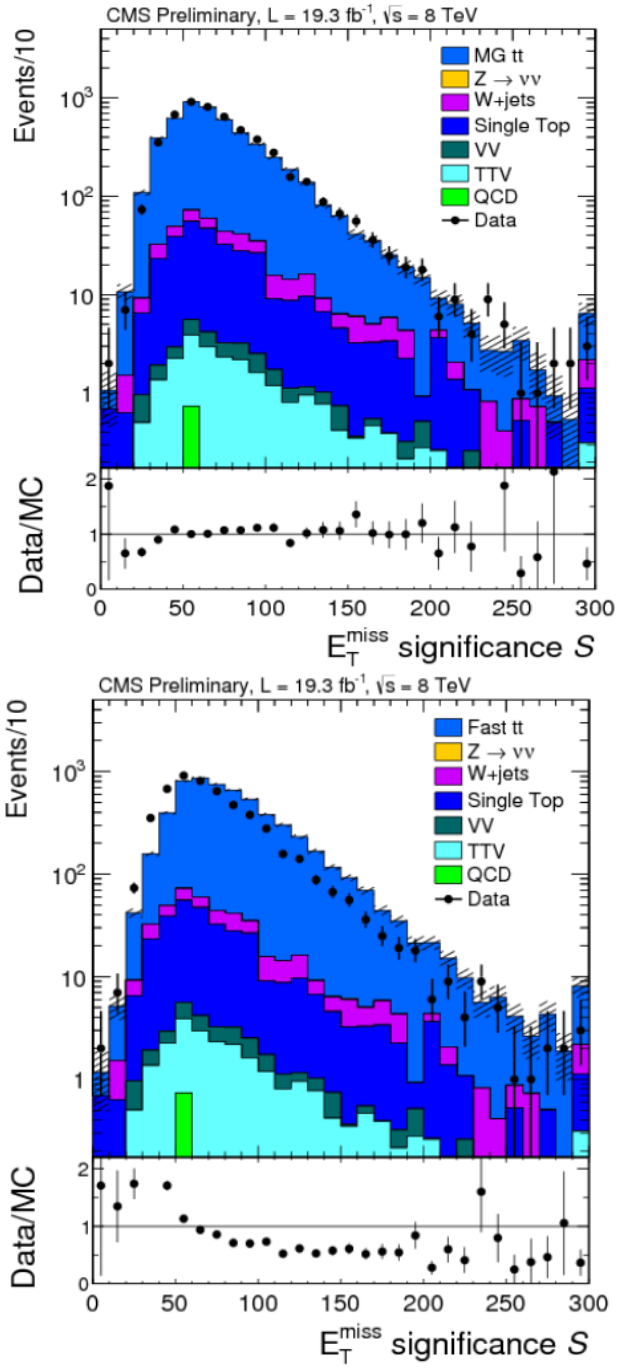

FIGURE 4.1: MET significance in simulated $t\bar{t}$ events in FullSim (top) and FastSim (bottom) [74X CMSSW version]

jets, undergo nuclear interaction with detector medium and produce what is called as "hadronic showers". In simple words, hadronic shower is a multiparticle production, which takes place at cost of disposal of energy of incoming hadron. Hadronic shower is generally characterized in terms of interaction length λ, which defines the distance along trajectory, in which 63%, i.e $(1\text{- }\exp(-1)\ )$ of incident hadrons are lost due to nuclear interaction. Figure 4.2 shows schematic of hadron shower [84]. Shower has essentially two basic components, hadronic component, dominated by charged (and some neutral) hadrons and electromagnetic component initiated by π^0

decay into 2γ, followed by electron-positron pair production by each γ, and subsequent development of an electromagnetic shower.

The pions are lightest among hadron category. As a result, they form major constituent of hadronization products of quarks and gluons- the jets. Hence, correct modelling of pion response is crucial in simulation to get the modelling of jets correct. This lead to prioritizing pion response understanding to analyze observed differences in FastSim versus FullSim.

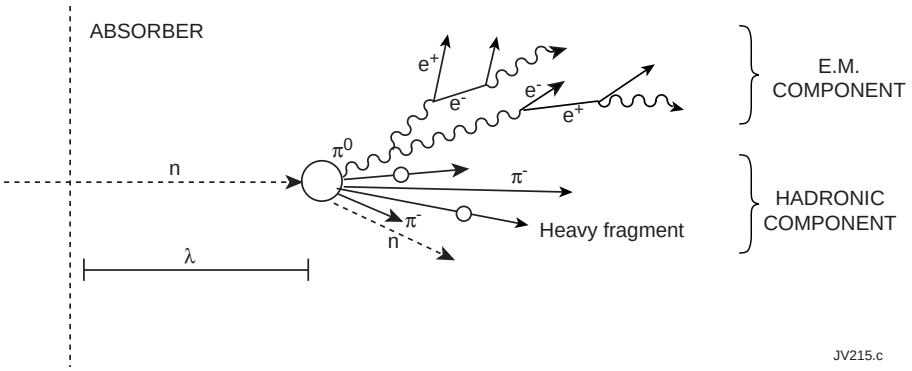

FIGURE 4.2: Sketch of hadronic shower in detector medium

4.4 Modeling of charged pion energy response

An incident pion will interact with detector medium to generate a hadron shower, which completely deposits energy of incoming pion into the detector in form of energy hits. To study modelling of a pion response, a pion sample was created by following a chain of generation, simulation and reconstruction (GEN-SIM-DIGI-RECO). The response of detector is generally dependent on incoming particle energy. The pions within jets, mostly have energies of order of few GeV or even lesser. Hence to start with, it was decided to study response of a pion sample with single flat energy in above range. The charged pions in the energy range 5 GeV to 100 GeV were considered, and distributions for 9 GeV energy are presented here for illustration purpose. When a sample is generated by following a chain GEN-SIM-DIGI-RECO, energy can be retrieved at various stages.

4.4.1 Energy retrieval at various stages of simulation

Here events were generated using Pythia flat energy particle gun [61]. This is called "GEN" step, and the energy of particle at generation step is termed as "gen level particle energy" in the following. During detector simulation, tracker is not considered (as energy lost there will be negligible) and magnetic field is set to 3.8T. At end of this step, energy deposition is recorded for further analysis in terms of hits (called simulated hits or simhits) in the detector. The energy corresponding to simhits will be called as "simhit energy". This is called "SIM" step. Now, the simhits only within the active part of detector are recordable in the actual experiment, and hence relevant for this study too. But, major energy loss happens in passive material of detector. Based on Geant4 simulation, for every pair of (active+passive) layer, a "sampling fraction" is known a priori, which represents the ratio of energy deposited in active

medium to energy deposited in (active+passive) medium. To retrieve energy loss in passive medium and get back total simhit energy corresponding to incident particle, simhit energy in active medium is corrected by applying "sampling fraction". After simulation step, energy hits in terms of analog pulse are digitized (called "DIGI" step). Electronic signal pulse shape as a function of time and electronic noise are taken into account at this step. Next step of reconstruction chain called "RECO" is the one where the digitized information obtained from the detector is converted to momenta or energy measurements, by converting digitized energy pulse to equivalent reconstructed hits which will appear at detector level (called "rechits"). Energy of "rechits" is called reconstructed energy (also called "reco" energy) which should be equivalent to energy of incoming particle. After proper calibration, total reconstructed energy should peak at generator level energy, with a distribution around the peak contributed by the detector resolution effects.

4.4.2 Comparisons of simulated and reconstructed energies

In the CMS calorimeter set up, the ECal corresponds to $\sim 1\lambda$ material in front of HCal. Hence, almost 63% of charged pions starts showering in ECal. However, pion showers are deep and they deposit most of their energy in HCal. To get total energy of incoming pion, the hits in the detector are reconstructed and their energies are combined. At "simhit" level there is no detector noise, hence total simhit energy should match incident pion energy. As HCal is a sampling detector, at simhit level only the energy fraction deposited in active layer is recorded. To correct for the energy deposited in absorber (passive layer), energy deposit in every active layer of HCal is multiplied by a "sampling fraction" associated to that layer to account for energy lost in corresponding passive layer. At "rechit" level, total energy after combining all rechits may exceed the true energy of incoming particle due to detector noise and resolution effects already accounted. For this reason, to make energy measurement at rechit level legitimate, only rechits above certain energy threshold are used to get total energy by suppressing detector noise contribution. To make an educational decision on what threshold to be put, detector response to neutrino sample is studied. Neutrino being a weakly interacting particle, does not undergo any interaction in the detector and just flies off. Hence, for an incoming neutrino, expected energy deposit in detector is zero. Figure 4.3 shows the individual rechit energy distribution in ECal Barrel (EB) and HCal Barrel (HB) for 9 GeV incident neutrino generated using Pythia particle gun similar to pions. Thus there is non-zero energy recorded in detector even in absence of detector-particle interaction.

Looking at the mean (μ) and RMS (σ) of above distributions, a noise rejection cut off of $\mu \pm 2\sigma$ was applied in both EB and HB. This came out to be 0.8 GeV for HB rechit and 0.118 GeV for EB rechit. Figure 4.4 shows total simhit energy distribution (left) and total rechit energy distribution after noise threshold (right) for 9 GeV energy incident pion.

Looking at Figure 4.4, it looks that FastSim and FullSim are fairly consistent. But, when looked at ECal and HCal energy deposits separately, FastSim showed a considerable disagreement with respect to FullSim. Figure 4.5 shows total simhit energy distribution in ECal and HCal barrel separately for 9 GeV incident pion with η=[0,0.1].

This fundamental disagreement at simhit level was important to be addressed as it posed question on how well and consistently hadron shower development is modelled in FastSim as compared to FullSim. To understand that, first place to check

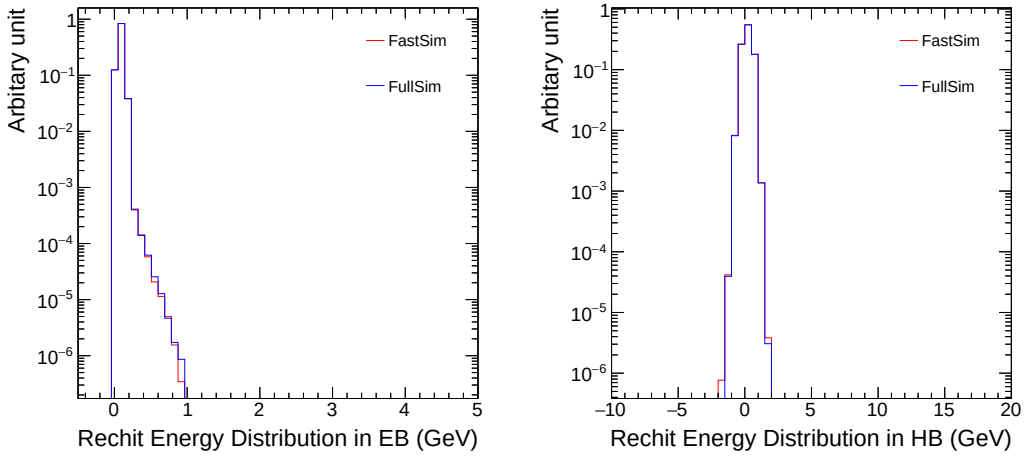

FIGURE 4.3: Rechit energy distribution in ECal Barrel (left) and HCal Barrel (right) for 9GeV neutrino.

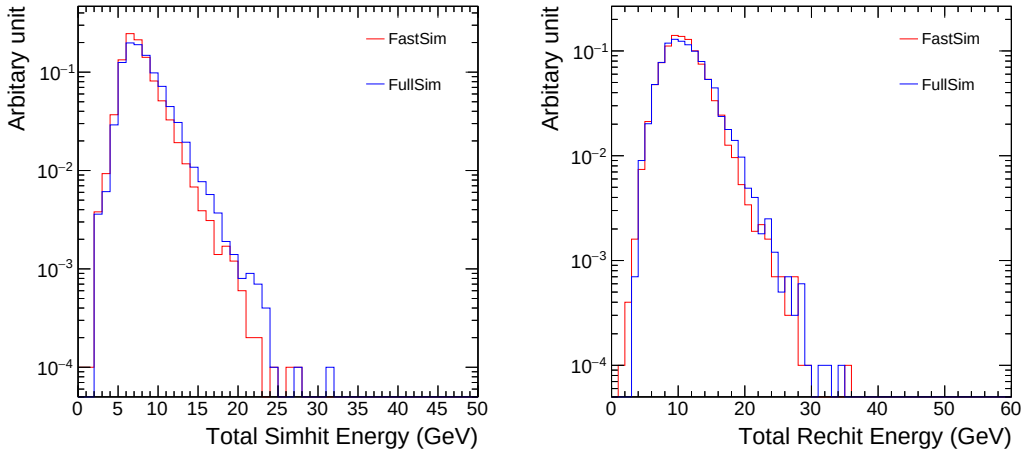

FIGURE 4.4: Total simhit energy (left) and total rechit energy after noise threshold (right) for 9 GeV incident pion

consistency was the shower start position, as it decides energy sharing between ECal and HCal.

4.5 Modelling of shower start location

After entering into calorimeter, a particle may travel a finite distance before it undergoes its first hadronic interaction with the detector material, and hence start shower development. The position corresponding to this first inelastic interaction is called a shower start position. A shower start position can be in ECal or HCal based on probability of first nuclear interaction.

In a coordinate system aligned along the incoming particle direction, the position of shower start position will be a function of path length travelled by a particle. Figure 4.6 shows the dependence of path length on incident angle θ. As an example,

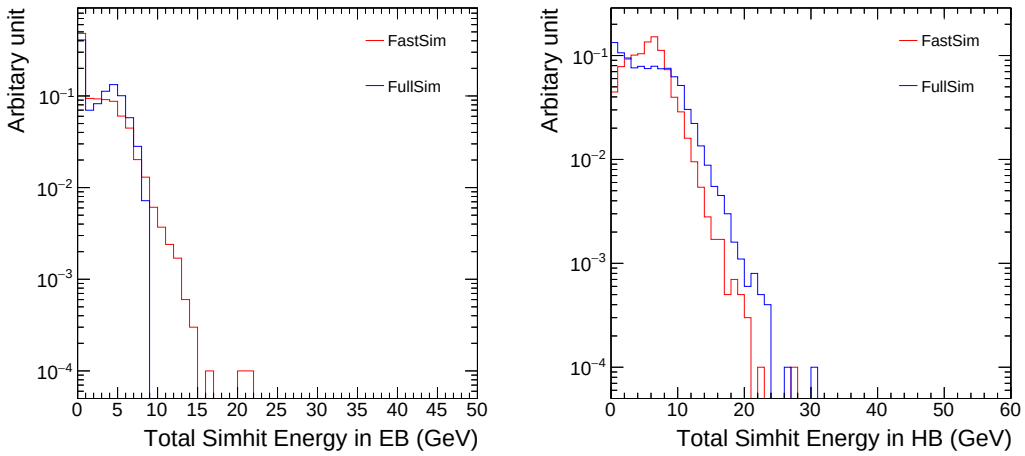

FIGURE 4.5: Total simhit energy distribution for 9 GeV pion in (left) ECal barrel and (right) HCal barrel

a pion directed at η =0 crosses ECal entrance at distance 129 cm from the collision point while that directed along η =0.5, crosses ECal entrance at $\sim$ 145 cm. Thus during shower start distribution comparison it is needed to decouple the effect of θ dependent path length on shower start distribution.

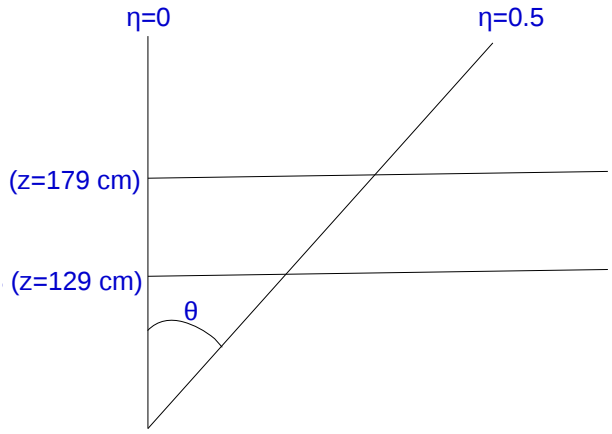

FIGURE 4.6: Path length as a function of incident angle(θ)

Figure 4.7 (left) shows the distribution of a calorimeter entrance point (also called starting point) and first inelastic interaction point in calorimeter obtained using the default G4SimWatcher functionality provided by Geant4 simulation package as compared against the same distributions obtained using the FullSim for an incident pion with $0.0 < |\eta| < 0.04$.

FastSim gives a peak at 129 cm for a starting point in calorimeter which is the ECal entrance position for η = 0. It was important to understand why FullSim does not give a sharp peak at 129 cm but shows a spread between 120-135 cm. The difference was understood after realizing how the incident particle trajectories are traced. Fastsim uses the parameterization and the incident pions are recorded at the front

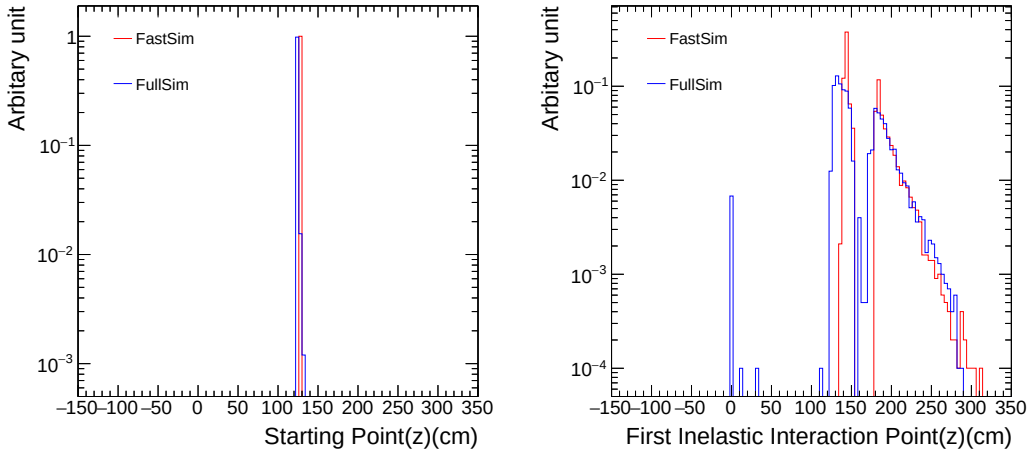

FIGURE 4.7: The position of calorimeter entrance point (left) and distribution of shower start point (right) compared between FullSim and FastSim (with default SimWatcher class) for a pion shooted at $0.0 < |\eta| < 0.04$

face of the ECal. But the Fullsim traces the incident particle trajectory in steps, hence also goes through a lot of dead material that is in front of ECal. Hence for proper comparison of a starting point (calorimeter entrance point) and first inelastic interaction point, it was needed to only select the detector volumes corresponding to ECal and HCal. After this proper selection of detector volumes and also restricting to exact η =0.04 (as there is physical gap at η =0.0), comparisons look as shown in figure 4.8.

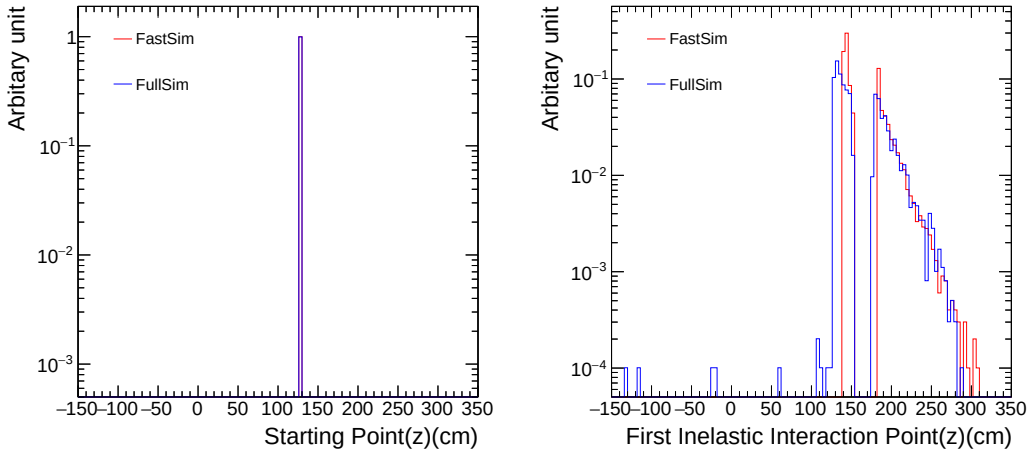

FIGURE 4.8: Calorimeter entrance point (left) and shower start point (right) distribution comparison between FullSim and FastSim (with default SimWatcher class) for a pion shooted at η =0.04 after proper selection of physical volumes

Calorimeter entrance point in FastSim matches with FullSim after selection of proper physical volumes. Both peak at 129 cm which corresponds to ECal front face at central tower. Two issues in FastSim shower start distribution are still noticeable.

One is, in ECal, shower start distribution is not starting exactly at ECal front face, but almost from middle of ECal longitudinal length, at around 140 cm. Also the FastSim distribution in ECal is not exponential as expected. The FastSim model was carefully revisited, to understand these features in details.

According to this model:

- ECal is treated as a single step while HCal is considered as combination of steps of fixed size. Also the model says that shower start position can lie only beyond half a step size. As a result, for ECal, shower start distribution only starts after half a ECal length.
- Instead of exponential distribution, a flat distribution is sampled to get shower start position in ECal, as a result there is no exponential behaviour seen in ECal.

Due to the above observed issues with current FastSim model, an alternative FastSim model available (GFlash) but not in use is also checked doing similar comparisons. This model is based on Grindhammer parameterization [85]. The shower start distribution by this model matches with FullSim in ECal but in HCal slope of distribution seems to disagree at considerable extent as shown in Figure 4.9.

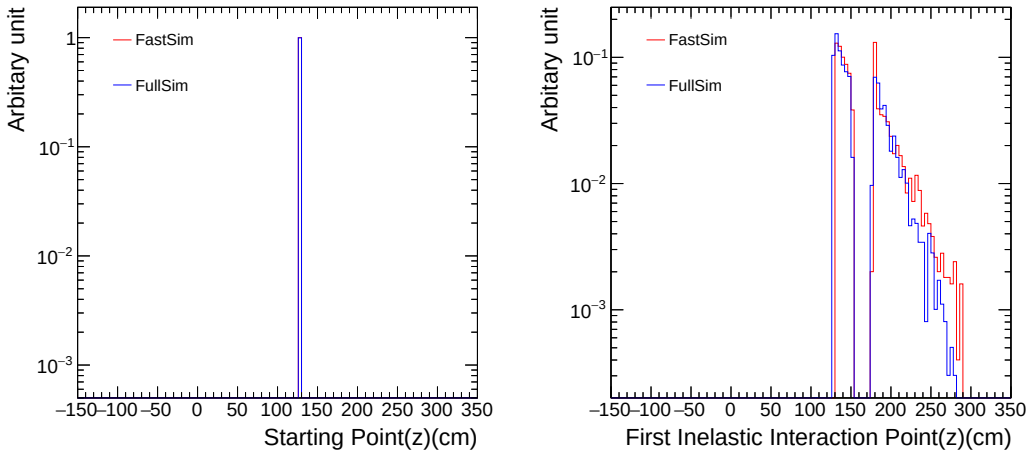

FIGURE 4.9: Calorimeter entrance point (left) and shower start point (right) distribution comparison between FullSim and FastSim (GFlash) for a pion shooted at η =0.04

4.5.1 Proposing new hadronic shower model for FastSim

Above studies marked that FastSim does show deviations with respect to FullSim when we start looking into details. On the other side, FastSim is gaining more and more importance due to the vital role it plays in new physics searches and also for upcoming High Luminosity LHC (HL-LHC) era. With this in mind, it was decided to write a new hadronic shower model for FastSim from scratch. It was beneficial to keep a modular form for this model for future refinements. Different modules in this model are going to be - shower start module, transverse parameterization and longitudinal parameterization module.

4.6 Updated model for shower start location

4.6.1 Shower start and interaction length(λ)

As mentioned in previous section shower start defines the position where an incident particle undergoes inelastic interaction to produce a cascade of particles. Shower start distribution is generally characterized in terms of interaction length (λ).

Suppose N(0) denotes number of primary particles incident at detector. Then interaction length(λ) is the distance from detector entrance point over which around 63% primaries undergo inelastic interaction to generate shower. Then number of primaries (particles which are not yet showered) survived at distance(x) from entrance point are-

$$N(x) = N(0) \times exp(-x/\lambda) \tag{4.1}$$

As a result, shower start distribution is expected to be exponential in terms of distance travelled in λs.

4.6.2 Shower start distributions using pion sample

For developing a model, a single pion sample is generated using Geant4 based simulation of CMS detector but with simplified calorimeter geometry for various energies. The simplified geometry had configuration ECal+HCal, with gap between two filled by an air. Each calorimeter layer was represented by a flat square shaped layer. Samples were generated with fixed η=0 and uniform in ϕ direction. Energies of samples were ranging from 0-400 GeV. For each event in sample, a shower start position is saved by coordinates (x,y,z) defined in particle's local coordinate system, such that z is defined along the particle trajectory. A shower start distribution is obtained, one for each energy, by extracting z coordinate of shower start position from all events corresponding to that energy.

It was observed that at higher energies shower start distributions follow exponential behaviour but not at very low energies as shown in figures 4.10.

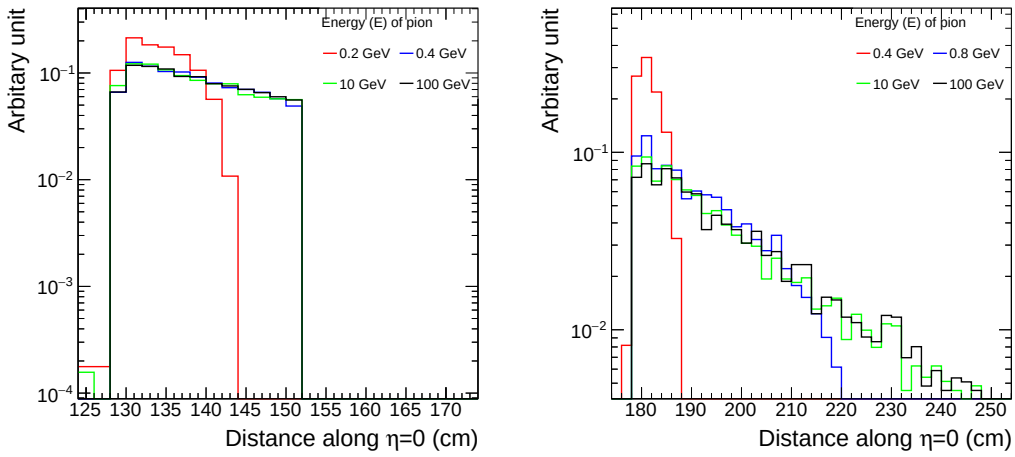

FIGURE 4.10: Shower start distributions for various energies in ECal (left) and HCal (right)

The reason for deviation of shower start distribution from exponential behaviour at low energies can be attributed to modulation of shower start probability due to energy lost by ionization processes.

4.6.3 Ionization loss at low energies

At low energies (E<1 GeV or so), ionization is a major process contributing to energy loss.

The rate of energy loss (-dE/dz) for a particle with charge q and velocity v due to ionization in a given medium is given by a Bethe Bloch formula given below [53, 84].

$$-dE/dz = Dq^2 n_e/\beta^2[ln(2m_e c^2\beta^2\gamma^2/I) - \beta^2 - \delta(\gamma)/2] \tag{4.2}$$

where z - the distance travelled through medium
$D \propto$ (Z/A), where Z is atomic number and A is mass number of the medium
m_e - mass of electron
n_e - electron density of medium
I - mean ionization potential of medium
The important features of energy loss by ionization are summarized in the following -

- Energy is lost particularly to the electrons from atoms of a given medium
- Slower the particle, faster is the energy loss. Hence ionization loss becomes significant at low energies.
- Energy loss is proportional to atomic number (Z) of a medium
- Energy loss rate becomes always minimum at around $\beta\gamma \approx$4, where particle is called as a minimum ionizing particle (mip). Until the particle undergoes nuclear interaction , it deposits energy via ionization as it traverses the detector material, hence behaves like a mip if it is sufficiently energetic.
- Beyond $\beta\gamma \approx$4 (i.e for relativistic particles) rate of energy loss again increases as relativistic expansion of particle's electric field allow further and further electrons of the medium to interact with a particle.

As shower start distribution follows exponential fall off, for any given incident particle, its shower start position will be a random position sampled from the exponential fall. If an incident particle has lost all of its energy before reaching this random shower start position (which is probabilistic), a particle is lost and there will be no shower. More the fraction of primaries lost by ionization, more will be the deviation of shower start distribution from exponential fall.

4.6.4 Shower start distribution including ionization losses

Based on Bethe Bloch form, for a charged particle of given energy (E) in given medium, an average position (μ) where particle will stop due to complete energy loss by ionization, can be given by equation:

$$\mu = \frac{E}{\int(dE/dz)} \tag{4.3}$$

To a good approximation, the distribution of stopping position (G(z)) just by ionization for a particle of energy E, will be a gaussian with mean μ.

$$G(z) = \frac{1}{\sqrt{(2\pi\sigma^2)}} \times exp(\frac{-(z-\mu)^2}{2\sigma^2}) \tag{4.4}$$

Thus the probability that an incident primary of energy E will be stopped by ionization energy loss (D(z)) before position z can be obtained by integrating G(z).

$$D(z) = \int_{-\infty}^{z} G(z')dz' = \frac{1}{2} \times (1 + erf(\frac{z-\mu}{\sqrt{2}\sigma})) \tag{4.5}$$

Thus only the fraction (1-D(z)) of incident particles will be available to undergo shower at position z. As a result, the probability that a particle will actually shower at position z in presence of ionization losses will be

$$P(z) = exp(-z/\lambda) \times (1 - D(z)). \tag{4.6}$$

4.6.5 Parameterizing shower start distribution

To define a generic model for shower start distribution, it was necessary to parameterize Equation 4.6 in terms of incident pion energy and distance travelled by it along trajectory in CMS detector. For that a tuning function is defined with following 10 input parameters:

Energy of particle (E), ECalEntrance, ECalExit, HCalEntrance, HCalExit, ECal interaction length (λ_{E}), HCal interaction length (λ_{H}), mean and standard deviation of a gaussian about average stopping position due to ionization loss (μ and σ respectively) and Normalization.

As there is no material in between ECal and HCal and hence no material interactions possible, the gap between the two is ignored while deriving parameterization. Thus for the given position z, the distance travelled in terms of interaction lengths (z_λ) for ECal and HCal will be as follows.

Within ECal,

$$z_\lambda = \frac{z - \mathrm{ECalEntrance}}{\lambda_{\mathrm{E}}} \tag{4.7}$$

Within HCal,

$$z_\lambda = \frac{\mathrm{ECalExit} - \mathrm{ECalEntrance}}{\lambda_{\mathrm{E}}} + \frac{z - \mathrm{HCalEntrance}}{\lambda_{\mathrm{H}}} \tag{4.8}$$

With z_λ defining distance travelled by particle trajectory, parameterized form of Equation 4.6 becomes:

$$P(z) = \mathrm{Normalization} \times exp(-z_\lambda) \times 0.5 \times (1 - erf(\frac{z-\mu}{\sqrt{2}\sigma})) \tag{4.9}$$

Pion shower start distributions for various energies are fitted with Equation 4.9 to extract μ, σ, λ_{E} and λ_{H}. It is observed that, mean (μ) and standard deviation (σ) of a gaussian, signifying ionization stop position have characteristic energy dependence as given by equations 4.10 and 4.11. Figure 4.11 shows these energy dependencies.

$$\mu = -0.6464 + 88.4 \times E \tag{4.10}$$

$$\sigma = 3.502 + 1.13 \times ln(E) \tag{4.11}$$

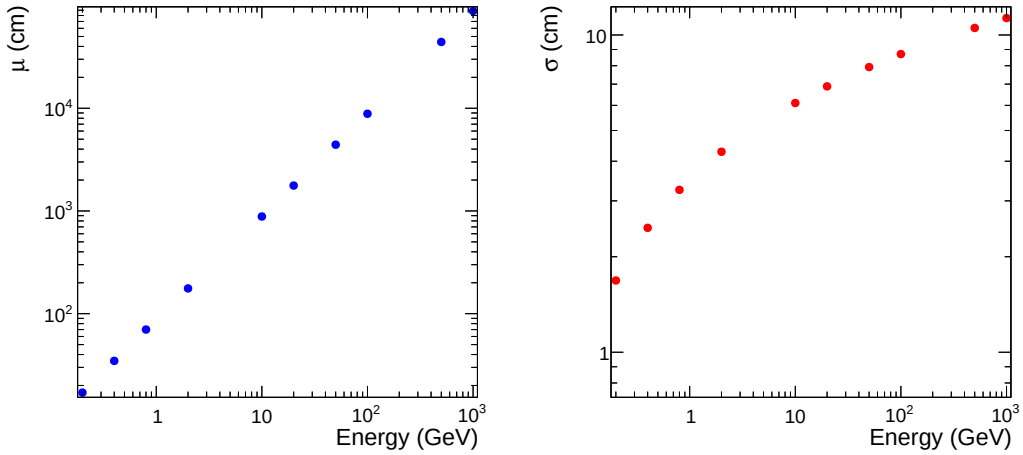

FIGURE 4.11: Mean(μ) and standard deviation(σ) of ionization stop position versus energy(E)

For λ_E and λ_H, energy dependence is found to be very subtle and only observed for the very low energies.

TABLE 4.1: Energy dependence of λ_E and λ_H

Energy (GeV)	λ_E (cm)	Energy (GeV)	λ_H (cm)
E≤2.6	19	E≤1.8	16
2.6 <E≤4.2	20	1.8 <E≤3.8	17.5
4.2 <E≤10	20.5	3.8 <E≤4.8	18
10 <E	22.5	4.8 <E≤5.6	19.5
		5.6 <E≤10	20
		10 <E	20.5

4.6.6 Validating shower start parameterization

For validation, parameterized shower start distributions are first compared against the same Geant 4 based shower start distributions which were used for extracting parameterization. Remember that those Geant 4 based samples had used simplified detector geometry and no magnetic field. To get parameterized distributions, for every energy, μ, σ, λ_E and λ_H values are obtained based on Equations 4.10, 4.11 and Table 4.1. An ionization stop position ($IonizationStop$) (in cm) due to complete energy loss is obtained by random sampling of a gaussian defined with (μ,σ).

$$IonizationStop = gRandom \rightarrow Gauss(\mu, \sigma) \tag{4.12}$$

A shower start position ($showerstart$) (in units of λ) is defined by random sampling an exponential with slope -1.

$$showerstart = gRandom \rightarrow Exp(-1) \tag{4.13}$$

IonizationStop and shower start positions are transformed in proper way so that they represent real positions in CMS ECal or HCal. Given these two positions, an incoming particle will shower only when shower start position is before IonizationStop position and can contribute to shower start distribution in such case. The parameterized shower start distribution is obtained by iteratively filling random shower start position in it whenever above condition is satisfied. Parameterized shower start distributions are compared against corresponding Geant4 distributions for three different angle of incidence of pion as $\cos(\theta)$ =1, $\cos(\theta)$ =0.5 and $\cos(\theta)$ =0.33. Here $\cos(\theta)$ =0.5 and $\cos(\theta)$ =0.33 correspond to cases where path length as seen by a particle is stretched to twice or thrice the length corresponding to θ =0 geometry. Hence parameters ECalEntrance, ECalExit, HCalEntrance, HCalExit for these three cases respectively take values (129, 152, 179, 270), (258, 304, 358, 540), (387, 456, 537, 810). The shower start distribution comparisons of parameterized versus Geant 4 distributions are shown in figures 4.12, 4.13, 4.14.

The agreement between Geant4 based and parameterized shower start distributions is within 10-20% in most of the cases for both standard and extended simplified CMS geometry. This gives the confidence that the generalized parameterization will hold for any given geometry and any angle of incidence and also in presence of magnetic field, where curvature of trajectory can be seen as particle is viewing an extended geometry, similar to the cases we have already discussed before.

4.6.7 Performance validation using standard CMS geometry samples

As a last step of validation, instead of Geant 4 samples with simplified geometry, the standard CMS geometry samples were used for this purpose. A single pion samples of around 10,000 events are generated for various energies across energy range of 0.2 GeV to 1000 GeV. A uniform distribution in η=[-0.05,0.05] and ϕ=[-π,π] is used for each of these energy samples, along with non-empty gap between ECal and HCal and 3.8 T magnetic field considered during simulation, to represent realistic event scenario in CMS detector.

Figures 4.15 show that even with real geometry and magnetic field on, parameterized distributions are consistent with Geant 4 distributions within 20-40% over energy range of 5 GeV-1000 GeV. Some activity is observed in a gap between Ecal and Hcal in case of real CMS geometry, which is expected as there exists some dead material associated with electronics and structural support. For energies below 5 GeV, a little lateral shift in distribution is observed with respect to that expected from a parameterized distribution. At low energies, curvature of trajectory due to magnetic field is pronounced compared to high energy. As a result, the effect of magnetic field on path length is more for low energies. Thus, if there is any inconsistency in orientation of the coordinate system in simplified and real geometry samples, it can lead to such lateral shift. The orientation of coordinate system in real CMS geometry samples to understand the observed shift will be revisited and accounted for when adopting the new parametrization for central simulation. Apart from this small discrepancy for very low energy pions, our newly proposed parameterization faithfully models the distributions predicted by GEANT4 over a wide energy range.

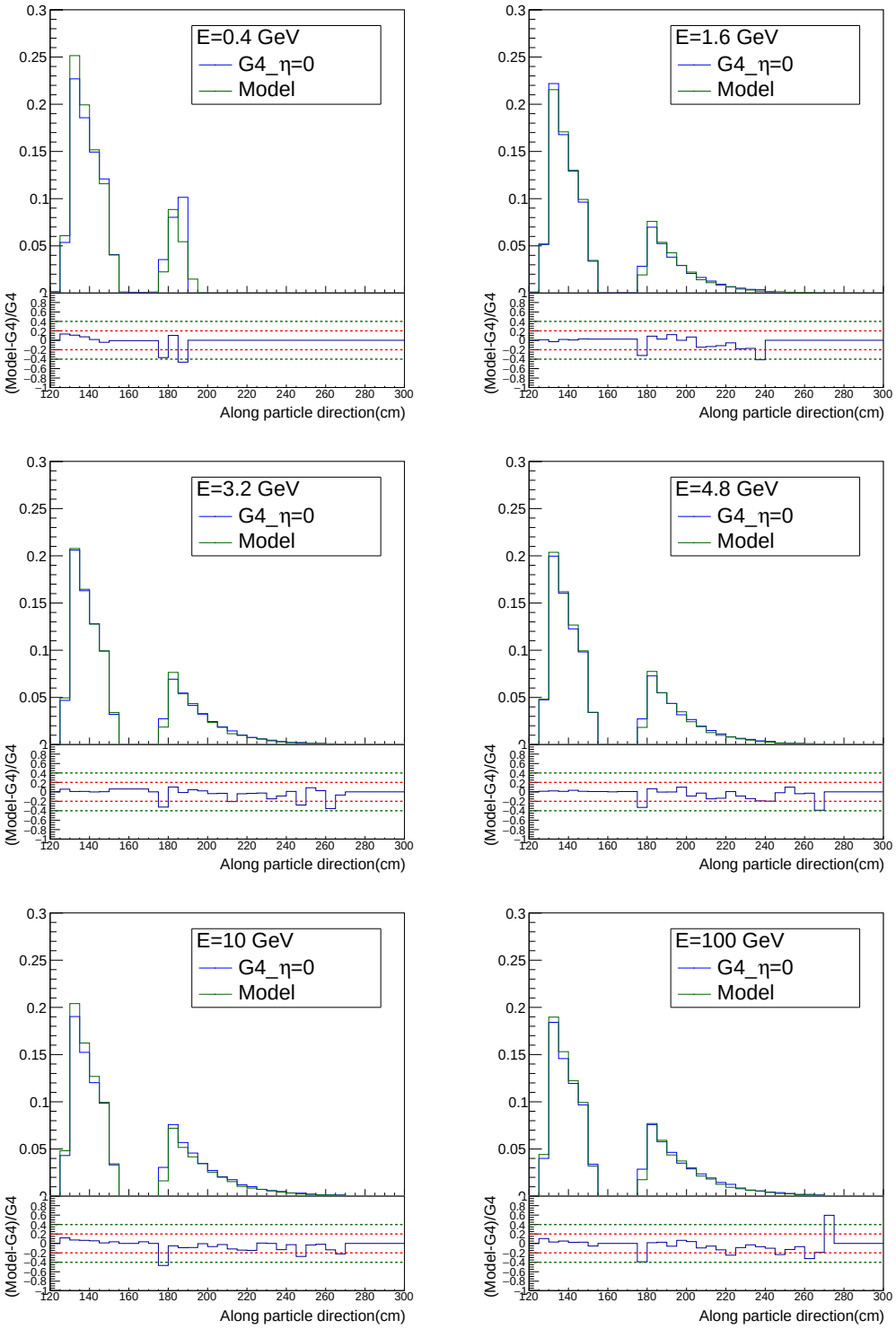

FIGURE 4.12: Validation of modified shower start parameterization using a pion sample where each pion is shooted at angle θ such that $\cos(\theta)$=1

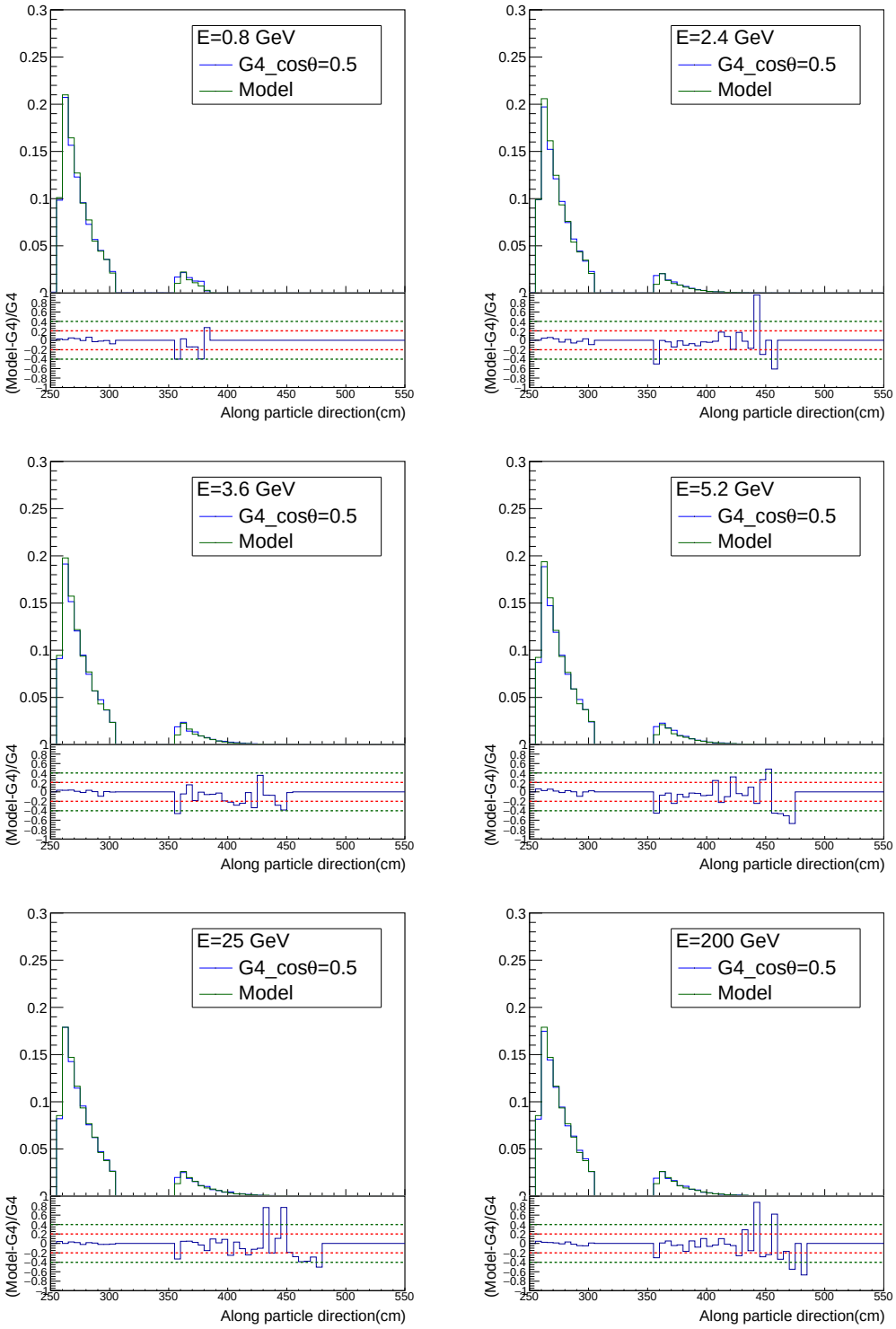

FIGURE 4.13: Validation of modified shower start parameterization using a pion sample where each pion is shooted at angle θ such that $\cos(\theta)$=0.5

4.7 New transverse shower model

4.7.1 Transverse shower models in FastSim

Hadronic showers are relatively complex as compared to electromagnetic showers as they have purely hadronic and also electromagnetic component (arising from

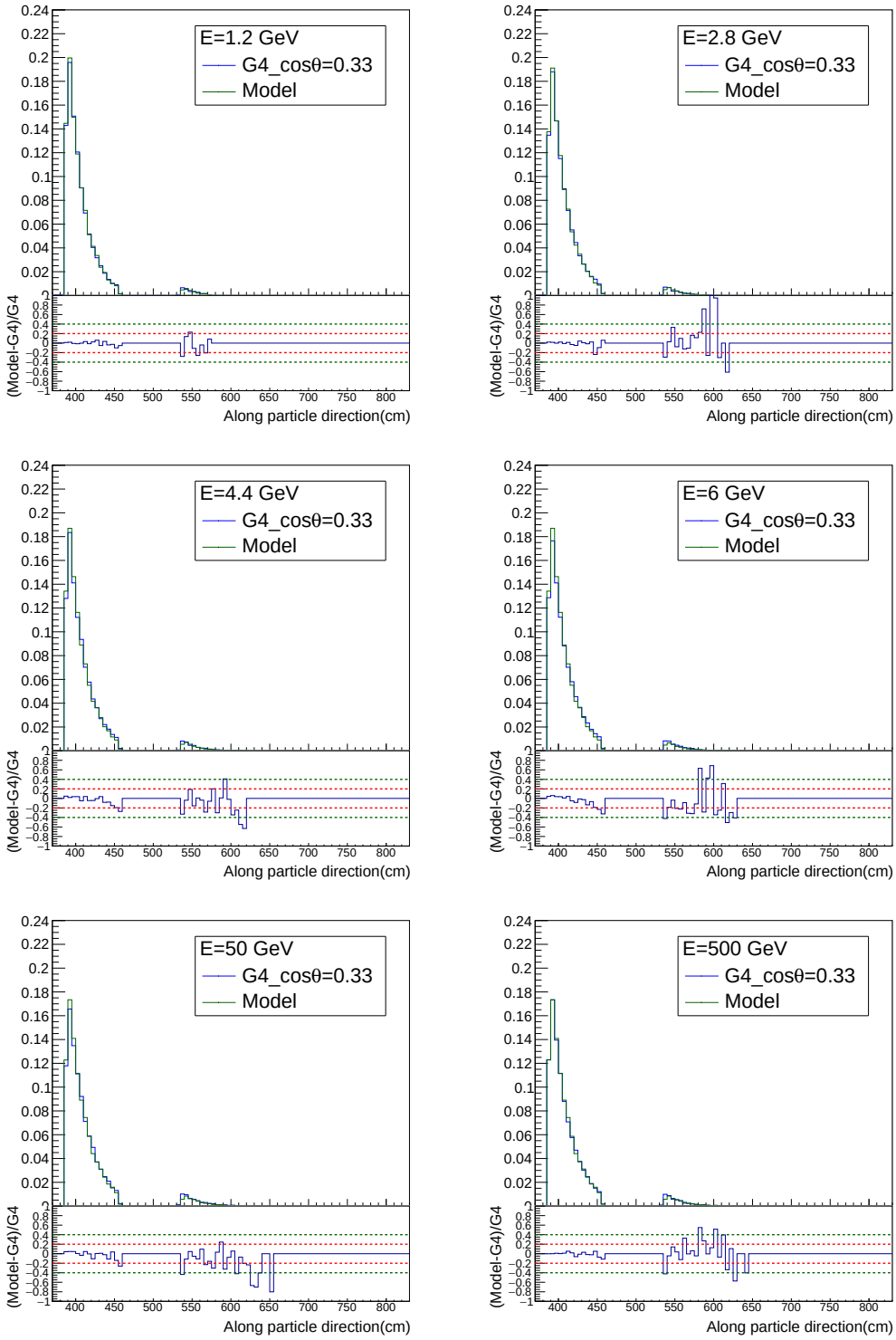

FIGURE 4.14: Validation of modified shower start parameterization using a pion sample where each pion is shooted at angle θ such that $\cos(\theta)$=0.33

π^0). Relative fraction of π^0 and hadronic component fluctuates a lot from shower to shower. Hence, simulation of hadronic shower has to take into account:

1. Energy dependence of fraction of π^0 and its fluctuations

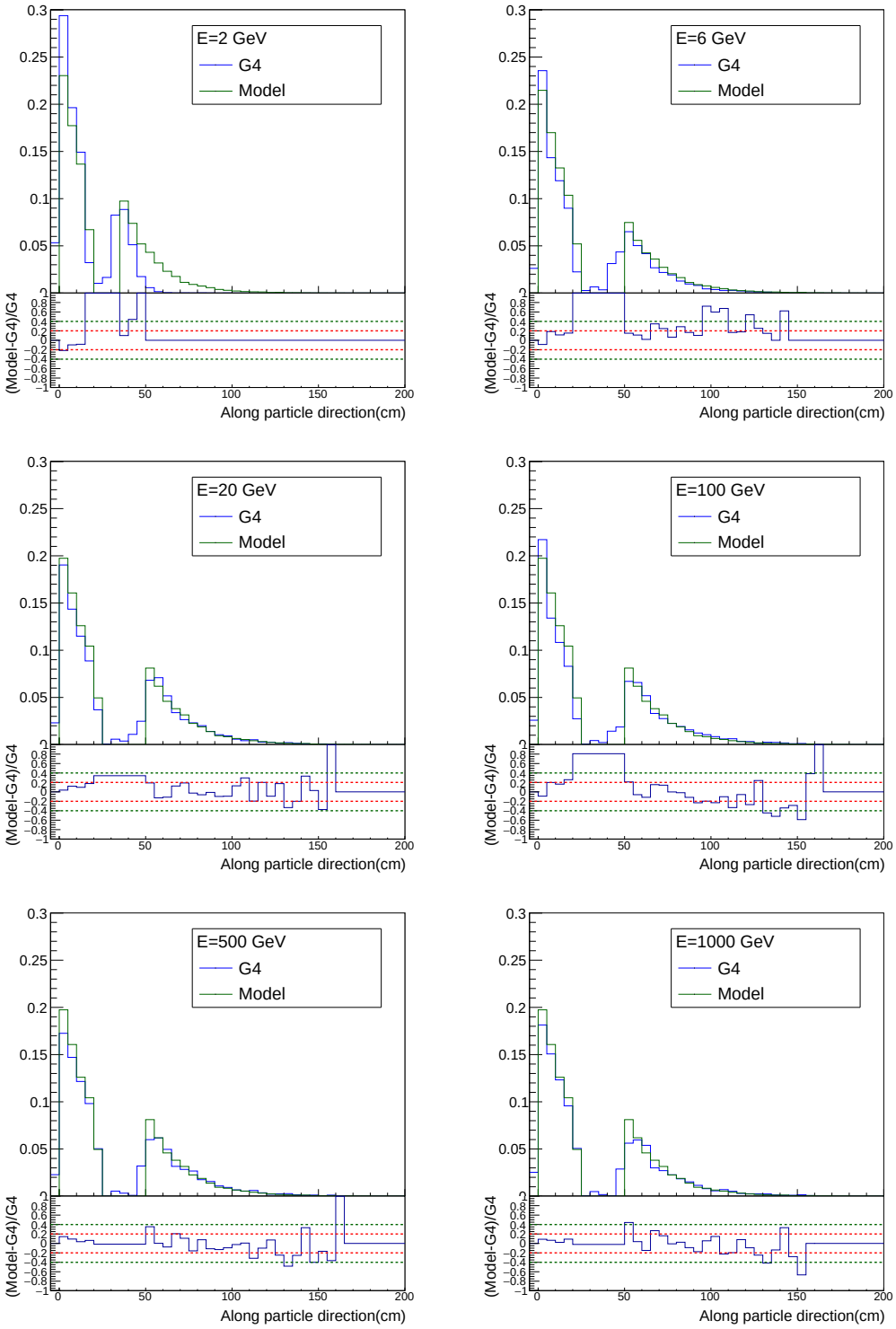

FIGURE 4.15: Validation of modified shower start parameterization using a pion sample with real CMS geometry and magnetic field on. Here pion sample has used uniform distribution in η=[-0.05,0.05] and ϕ=[-π,π]

2. Response of calorimeter differs for π^0 and purely hadronic component.

3. Propagation scale of π^0 is radiation length (X_0) while that for hadronic component is interaction length (λ_0).

In current FastSim model, a lot many adjustments are done on the fly without any reasoning given. Adjustments like defining "40" transverse steps for each longitudinal step (where energy is to be deposited), randomly varying energy and number of energy spots to be deposited at given step make this model difficult to follow.

An alternative FastSim model, which is available but not used, "GFlash", uses Grindhammer parameterization [85] to define transverse shower as shown in equation 4.14 below.

$$f(r) = 2 \times Norm \times r[\lambda] \times \frac{R_{50}[\lambda]^2}{(r[\lambda]^2 + R_{50}[\lambda]^2)^2} \tag{4.14}$$

Here the $Norm$ represents normalization factor, $r[\lambda]$ and R_{50} [λ] respectively represent radial distance in transverse plane from shower axis and median of a shower in given transverse plane in terms of interaction length λ. A sample of various incident pion energies 0.4-1000 GeV is used to check how well Grindhammer parameterization fits transverse hadron shower profile. For current study, only showers starting in HCal are considered. It is found that above parameterization fits transverse profile in various longitudinal depths reasonably well for low energy pions but towards high energy the fit starts deviating from true transverse profile (Figure 4.16). In these comparisons, "E" corresponds to energy of particle and "L" corresponds to longitudinal depth (e.g. L=0 represents average transverse profile over first interaction length).

As a solution over inaccessible nature of existing FastSim model and discrepancies with respect to Grindhammer parameterization, it was decided to tune Grindhammer parameterization further to make it more general to be applicable over wide energy and longitudinal depth range.

4.7.2 Modification to Grindhammer parameterization

For modification, different additional terms to Grindhammer form as a correction factor were tried. One realization was parameter R_{50} is very sensitive to fitting range. Hence, to derive parameterization, fitting range is kept fixed to be 120 cm for all layers and all energies, and then tried which can be a generic form to fit.

Following model, as given by equation 4.15, seem to work for transverse profile of all energies and for various longitudinal depths.

$$f(r) = (A + \frac{B}{r^2[\lambda]} + \frac{C}{r^4[\lambda]}) \times \frac{R_{50}^2[\lambda]}{(r^2[\lambda] + R_{50}^2[\lambda])^2} \tag{4.15}$$

Here A, B, C are normalization constants. One longitudinal depth is considered at a time and transverse profile from that depth is fitted with above form to get values for A, B, C and R_{50} [λ]. After repeating the procedure for various energies, approximate energy based parameterization is derived for above parameters (Table 4.2).

Figures 4.17, 4.18, 4.19, 4.20, 4.21 show transverse shower profile of various energies fitted with new parameterization for various longitudinal depths (L=0 to L=4). The performance of new parameterized model is better compared to Grindhammer parametrization, particularly towards high energies.

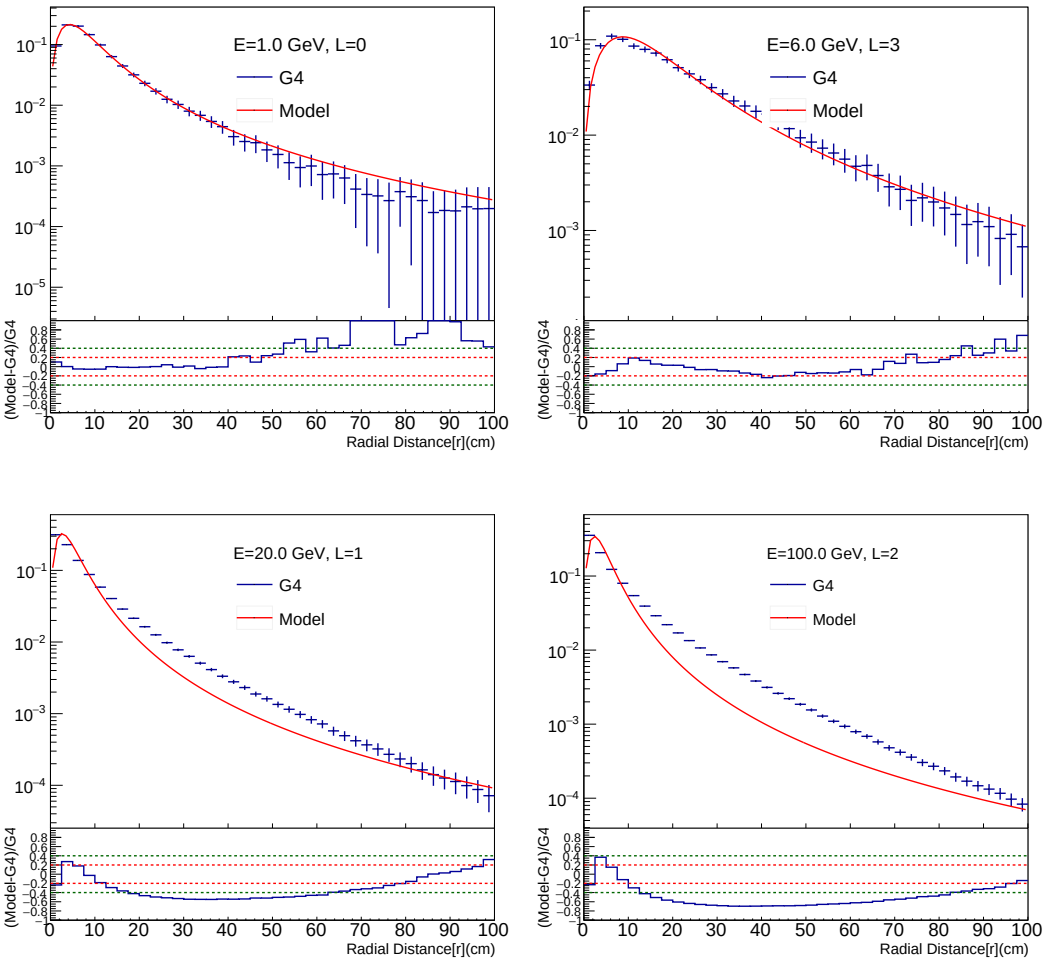

FIGURE 4.16: Comparison of transverse shower profile of pion against fit using Grindhammer parameterization

4.8 Modelling of sampling fluctuations

In existing model, sampling fraction is determined for each pair of (active+passive) layer and is taken as a constant for that particular pair. But, in reality hadronic showers are complex and as a result sampling fraction for a given pair also fluctuates from shower to shower. In new model for hadronic shower in FastSim, these fluctuations in sampling factor were supposed to be included.

For that at first using Geant4 showers, a distribution of sampling fraction is obtained for each (active+passive) layer. This is done for all 36 identical azimuthal wedges in HCal, for various energies and also for various particle types (hadrons, mips, π^0) and different shower starting point. It was found that Landau distribution fits irrespective of different shower types. The mean (μ) and standard deviation (σ) of Landau fit is extracted for each of the case mentioned above. Once known, Landau distribution given by (μ,σ) can be used to model fluctuations in sampling fraction. Figure 4.22 shows a Landau fit to one of the layer.

TABLE 4.2: Energy dependence of various parameters in modified Grindhammer parameterization

λ depth	Energy dependence
L0	R_{50} [λ]=31.78 - 1466/6×E + 1496/(0.6×E^2 + 6×E) A=0.00577 + 0.2048/E -0.1672/(E^2 + E) B=0.00631 - 0.05404/E + 0.02575/E^2 C=-10^{-5} (E<=4 GeV), C=10^{-5} (E>4 GeV)
L1	R_{50} [λ]=16.3 + 12.29/E - 0.00805×E A=0.01617 + 1.787/E -16.9/(E^2 + 10×E) B=0.004714 - 0.01305/E + 0.002404/E^2 C=-10^{-5} (E<=500 GeV), C=10^{-5} (E>500 GeV)
L2	R_{50} [λ]=19 + 29.41/E - 5.022/E^2 A=0.04181 + 2.691/E -26.16/(E^2 + 10×E) B=0.004883 - 0.03741/E + 0.03398/(0.5×E^2 + E) C=-10^{-5}
L3	R_{50} [λ]=20.25 + 52.45/E - 13.75/E^2 A=0.08721 + 1.105/E -1.383/(E^2 + E) B=0.004326 - 0.0311/E + 0.0096/E^2 C=10^{-5}
L4	R_{50} [λ]=23.88 + 12.67/E + 377/(E^2 + 10×E) A=0.0828 + 2.687/E -24.72/(E^2 + 10×E) B=0.007382 - 0.08527/E + 0.08958/(0.5×E^2 + E) C=10^{-5}

4.9 Longitudinal shower parameterization

In current FastSim, longitudinal shower profile is based on Grindhammer parameterization [85]. This parameterization used two different gamma functions, one for electromagnetic component and other for hadronic component. But, in fact electromagnetic component fluctuates a lot based on whether π^0 is created in early part of shower or in later part, as π^0 energy profile will be different in these two cases.

In new model, instead of two components, parameterization is developed considering three different components of shower, hadronic, early and late π^0.

Figure 4.23 gives a rough sketch of modelling longitudinal shower profile for new FastSim model. To model different parameters, single pion sample is generated for various energies and simulated using simplified CMS detector geometry. At first, fractions of three components are determined. Fig. 4.24 (top) shows fraction of total deposited energy (f_{dp}) in calorimeter, fraction of f_{dp} representing electromagnetic component (f_{π^0}) and fraction of f_{π^0} showing late π^0 component ($f_{\pi^0_l}$). Once these fractions are found, α, β for different components determined by fitting rate of energy loss along longitudinal direction (dE/dx) by gamma function and corresponding α, β are retrieved. This is done on shower by shower basis to get distribution of α, β. Figure 4.24 (bottom) shows that average of gamma distributions fitted to FullSim distributions (green) and that by picking random α, β values from their distributions (red) almost match. But, they deviate slightly from average FullSim distribution (blue). This was because few showers could not be described well with gamma distribution. But, on average, gamma distribution is a reasonably good approximation to describe longitudinal shower profile.

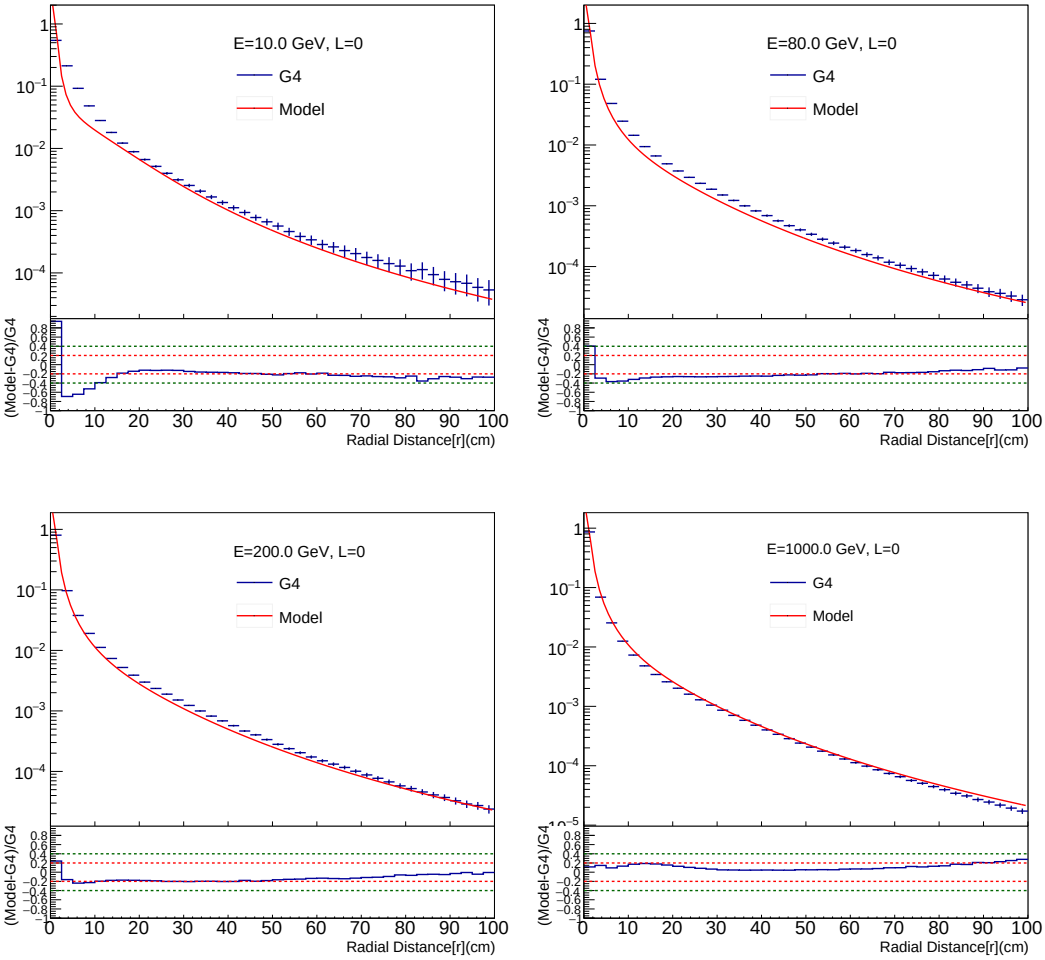

FIGURE 4.17: Comparison of transverse shower profile of pion against fit using modified parameterization for longitudinal depth first longitudinal depth (L=0)

4.10 Summary

A new parameterization is developed for different components of hadronic shower and performance is checked against standard Geant4 based simulation. New parameterization for different components of hadron shower is performing better compared to existing FastSim model and may help to resolve some of FastSim versus FullSim discrepancies. These components can be used as different modules in new hadron shower model. Because of the modular nature it will also be easier to do any modification if needed in future. As various components of modeling have been updated and validated against GEANT4, it has been proposed to integrate all modules into the CMS nominal FastSimulation.

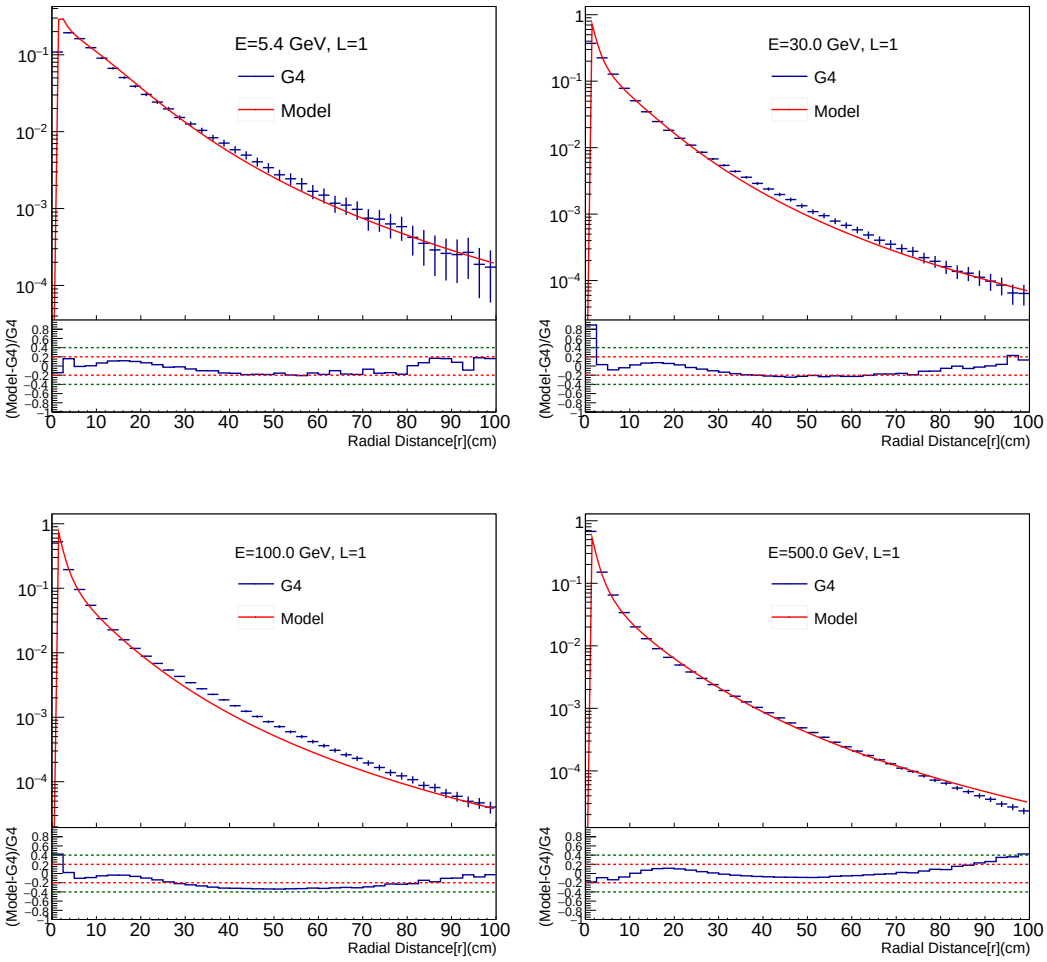

FIGURE 4.18: Comparison of transverse shower profile of pion against fit using modified parameterization for longitudinal depth second longitudinal depth (L=1)

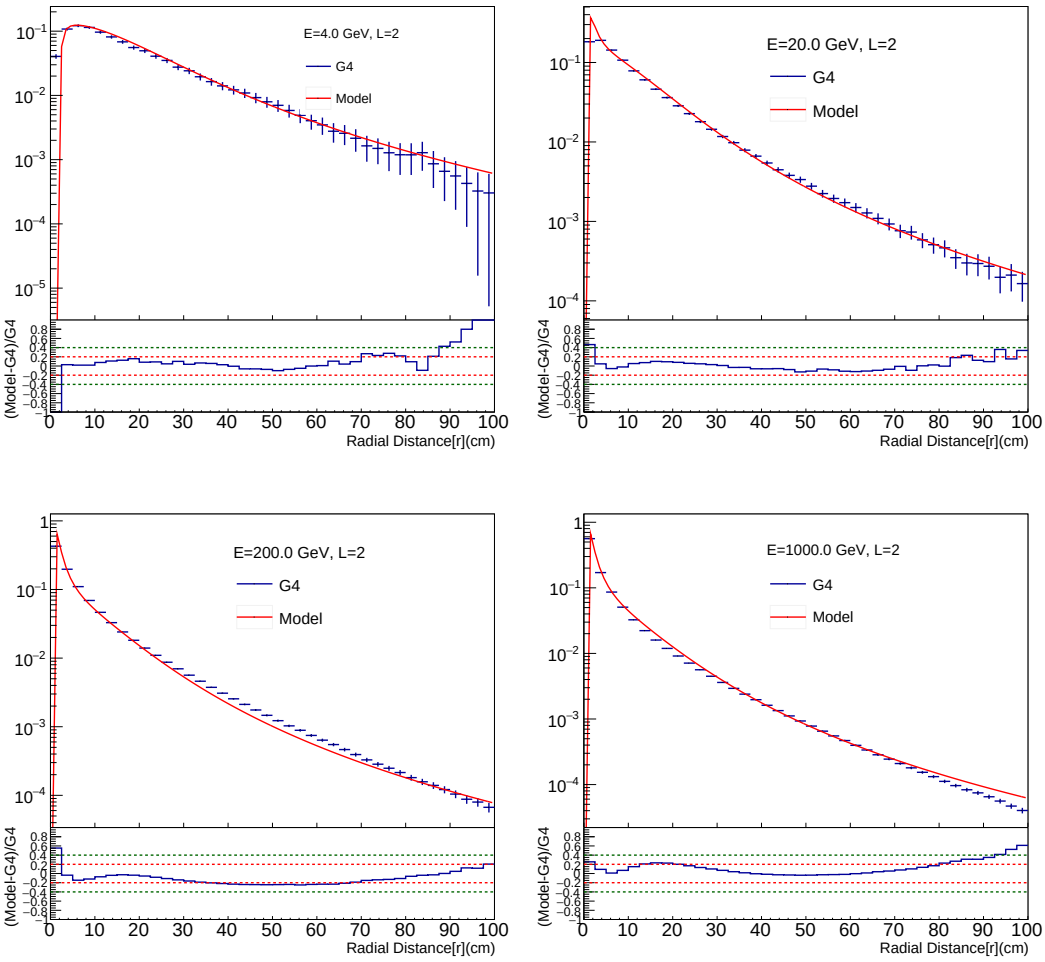

FIGURE 4.19: Comparison of transverse shower profile of pion against fit using modified parameterization for longitudinal depth third longitudinal depth (L=2)

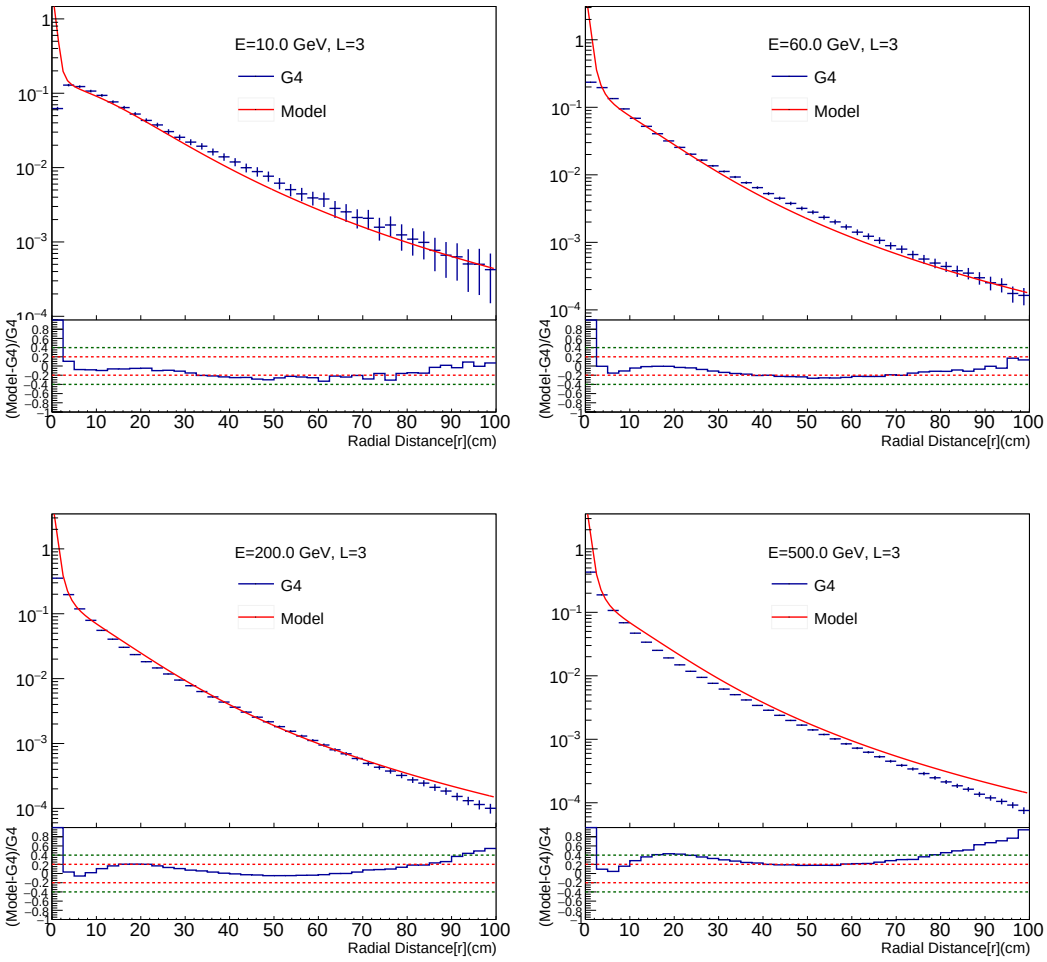

FIGURE 4.20: Comparison of transverse shower profile of pion against fit using modified parameterization for longitudinal depth fourth longitudinal depth (L=3)

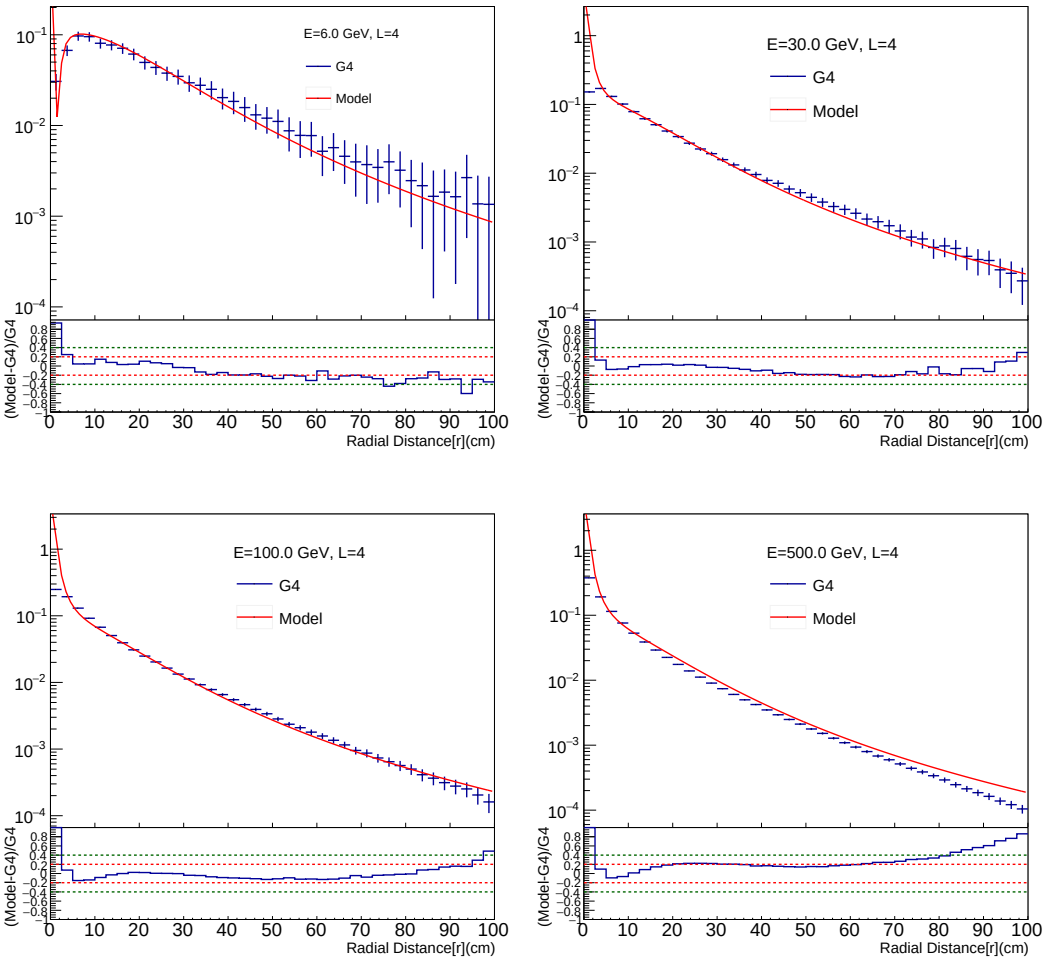

FIGURE 4.21: Comparison of transverse shower profile of pion against fit using modified parameterization for longitudinal depth fifth longitudinal depth (L=4)

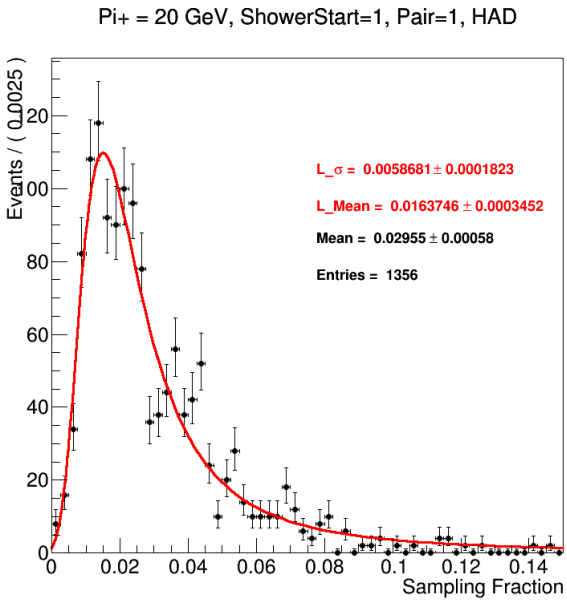

FIGURE 4.22: (μ,σ) derived by fitting Landau distribution to sampling fraction distribution

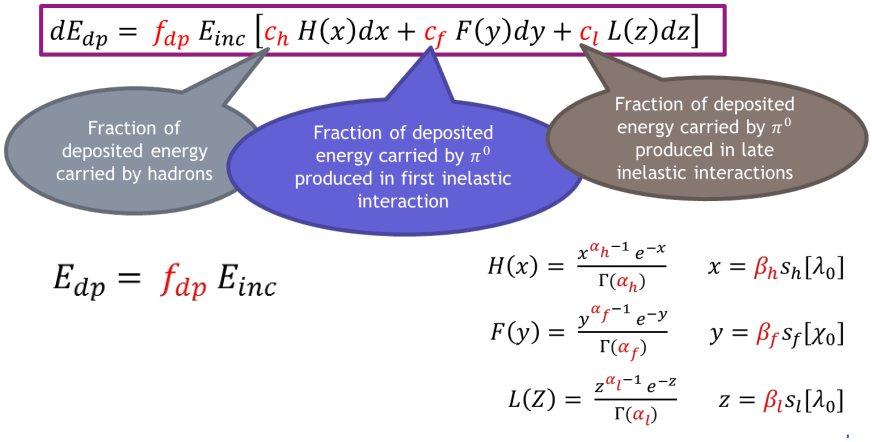

FIGURE 4.23: Parameterization for new longitudinal shower modelling

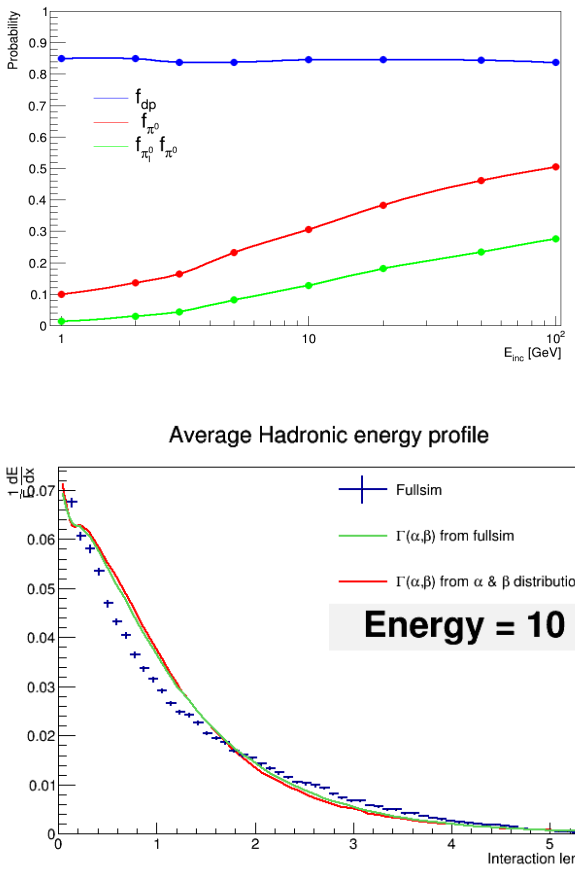

FIGURE 4.24: (top) fractions corresponding to total deposited energy in calorimeter (f_{dp}), total electromagnetic component (f_{π^0}) and late electromagnetic component (f_{π^0} $f_{\pi_l^0}$) obtained by fitting Geant 4 distributions. (bottom) Consistency of gamma distributions obtained using mean of α and β distributions (green), randomly chosen α, β from respective distributions (red) with FullSim distribution (blue).

Chapter 5

Search for supersymmetry with jets & $p_{\mathrm{T}}^{\mathrm{miss}}$ final states

This chapter starts with a brief discussion on what is the importance of SUSY searches in jets + $p_{\mathrm{T}}^{\mathrm{miss}}$ final state and where such searches benefit against SUSY search in final states containing leptons or photons. After this brief introduction, the goal is to give the broad idea of the SUSY search carried as a major part of this thesis work. The chapter discusses what can be the major SM backgrounds for the SUSY search with jets + $p_{\mathrm{T}}^{\mathrm{miss}}$ final state, what are the event selections used and their impact on SM backgrounds and potential SUSY signal and the set of triggers applied on data collected of years 2016, 2017 and 2018. Later part of the chapter includes details of various SM background estimation strategies used in this SUSY search.

Presence of SUSY particles in proton proton collision events can be manifested in a large variety of final states [86]. The fig. 5.1 shows production cross sections for various SUSY processes as predicted by theory. Gluino and squark pair production via strong interaction dominates over electroweak production of sleptons and electroweakinos [14].

Section 2.5 discussed strong SUSY production scenarios considered in this search. The Lightest Supersymmetric Particle (LSP), $\widetilde{\chi_1^0}$ is assumed to be stable and weakly interacting and hence gives large $p_{\mathrm{T}}^{\mathrm{miss}}$ in the event. Thus, this final state will be enriched with jets coming from light flavor (or bottom) quarks and gluons and $p_{\mathrm{T}}^{\mathrm{miss}}$ predominantly coming from $\widetilde{\chi_1^0}$ [11–13, 19, 87–89]. Jets and $p_{\mathrm{T}}^{\mathrm{miss}}$ can be accompanied with a lepton from leptonic W boson decay. Lepton can also originate from semileptonic decay of b-hadrons. In the inclusive analysis described here, we target fully hadronic final state with no lepton in the final state. Events with isolated lepton or an isolated track are vetoed to reject physics processes resulting in high $p_{\mathrm{T}}^{\mathrm{miss}}$ due to associated neutrinos (e.g. W+jets and $t\bar{t}$ production). More details on event selection will be discussed in section 5.2.

5.1 The Standard Model backgrounds

The Standard Model (SM) processes giving jets and $p_{\mathrm{T}}^{\mathrm{miss}}$ in the final state are the background (noise) for this search, as events originated from such processes are indistinguishable from potential signal events in data. The dominant backgrounds are Z+jets where Z decays to pair of neutrino and antineutrino, W+jets or $t\bar{t}$ +jets where electron or muon from W decay is lost or W decays to hadronic τ and multijet events originating from Quantum Chromodynamics (QCD) processes with significant fake $p_{\mathrm{T}}^{\mathrm{miss}}$.

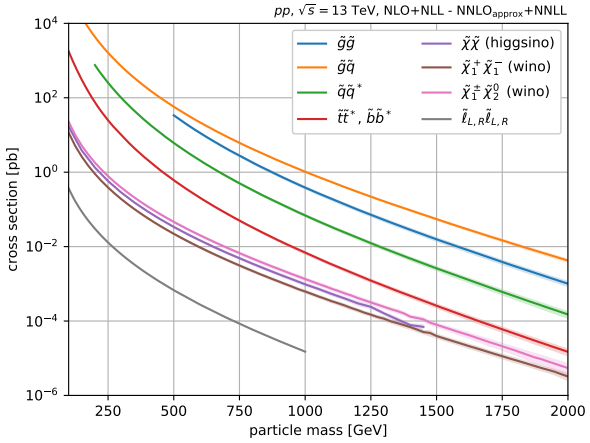

FIGURE 5.1: SUSY production cross sections for various processes at CM energy of 13 TeV [90]

TABLE 5.1: Decay modes of Z, W and τ

Particle	Mode	Branching Fraction (%)
Z	e^+e^-	3.36
	$\mu^+\mu^-$	3.36
	$\tau^+\tau^-$	3.37
	$\nu\overline{\nu}$	20
	$q\overline{q}$	70
$W^+(W^-)$	$e^+\nu_e(e^-\nu_{\overline{e}})$	10.71
	$\mu^+\nu\mu(\mu^-\nu\mu)$	10.63
	$\tau^+\nu\tau(\tau^-\nu\tau)$	11.38
	$q\overline{q'}$	67.41
$\tau^+(\tau^-)$	$e^+\nu_e\nu\tau(e^-\nu_{\overline{e}}\nu\tau)$	17.82
	$\mu^+\nu_\mu\nu_\tau(\mu^-\nu_{\overline{\mu}}\nu_{\overline{\tau}})$	17.39
	$q\overline{q'}\ \nu_\tau(\nu_{\overline{\tau}})$	65

5.1.1 Z+jets

Zero lepton requirement rejects events where Z boson decays leptonically. Also $p_{\mathrm{T}}^{\mathrm{miss}}$ requirement does not allow fully hadronic decay mode of Z to be a background for this search. Only possible decay modes causing background for the search are to pair of neutrinos or both τs from Z decaying hadronically. But as can be seen from Table 5.1 later one has a very small probability (3.37% $\times$ 65% $\times$ 65% = 1.4%). On similar argument, possibility that both e or μ from leptonic decay mode fail to be identified or reconstructed or fall out of detector acceptance and appear as jets+$p_{\mathrm{T}}^{\mathrm{miss}}$ event is highly improbable. Then only dominant mode that becomes background is the invisible decay (figure 5.2 (left)). This is an irreducible background for the search, which means it can not be further reduced by imposing any selection on event properties.

5.1.2 W+jets and $t\bar{t}$ +jets processes

The W+jets process has contribution from direct W production at the interaction vertex as well as from those produced in top quark decays. Similar to Z, leptonic and fully hadronic decay modes of W do not enter into signal region due to zero lepton and high $p_{\mathrm{T}}^{\mathrm{miss}}$ requirements. Leptonic decay mode of W is more probable compared to that of Z. When a W decays leptonically, the lepton can be lost in two ways; e or μ^- can be lost if it failed identification or isolation or falls outside acceptance, or W decays to hadronically decaying tau giving true jet+$p_{\mathrm{T}}^{\mathrm{miss}}$ signature in final state, which again becomes irreducible background for the search. Both these cases contribute to "lost lepton" background (Figure 5.2 (center)).

5.1.3 Quantum Chromodynamics (QCD) processes

The QCD basically has contribution from light flavour quark jets and gluon jets. These processes appear like jets+MET signal events, if either there is a fake $H_{\mathrm{T}}^{\mathrm{miss}}$ arising from jet mismeasurement or true $H_{\mathrm{T}}^{\mathrm{miss}}$ coming from semileptonic decay of bottom or charm hadrons (Figure 5.2 (right)). For events with $N_{\mathrm{b-jet}}$=0, primary $H_{\mathrm{T}}^{\mathrm{miss}}$ source is the jet mismeasurement and $H_{\mathrm{T}}^{\mathrm{miss}}$ will be directed along the mismeasured jet. For events with $N_{\mathrm{b-jet}}$> 0, semileptonic b-jet decay is a prime source of $H_{\mathrm{T}}^{\mathrm{miss}}$, and in a boosted b-jet topology, neutrino from semileptonic decay will be aligned with a final state jet as there is no other true source of $H_{\mathrm{T}}^{\mathrm{miss}}$ in the event. Thus $H_{\mathrm{T}}^{\mathrm{miss}}$ in both cases is aligned along the direction of jet and hence the characteristic of such events is low $\Delta\Phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{jet}},\vec{H}_{\mathrm{T}}^{\,\mathrm{miss}})$ angle as shown in figure 5.3.

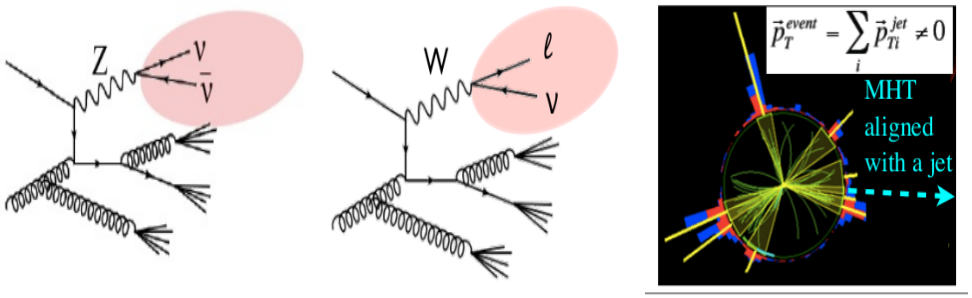

FIGURE 5.2: Schematic sketch of SM backgrounds: (left) Z + jets, (center) W + jets, and (right) QCD multijet.

Based on major SM background composition to search, event selection is made in order to achieve good signal sensitivity by efficient rejection of SM backgrounds.

5.2 Event selection

For this inclusive search to cover a wide variety of SUSY topologies, event selection is done using four orthogonal search variables.

- N_{jet} - The number of jets with p_{T} >30GeV and $|\eta|$ <2.4
- $N_{\mathrm{b-jet}}$ - The number of b-tagged jets with p_{T} > 30 GeV and $|\eta|$ < 2.4
- H_{T} (or HT)- Scalar sum of p_{T} of jets in N_{jet} collection

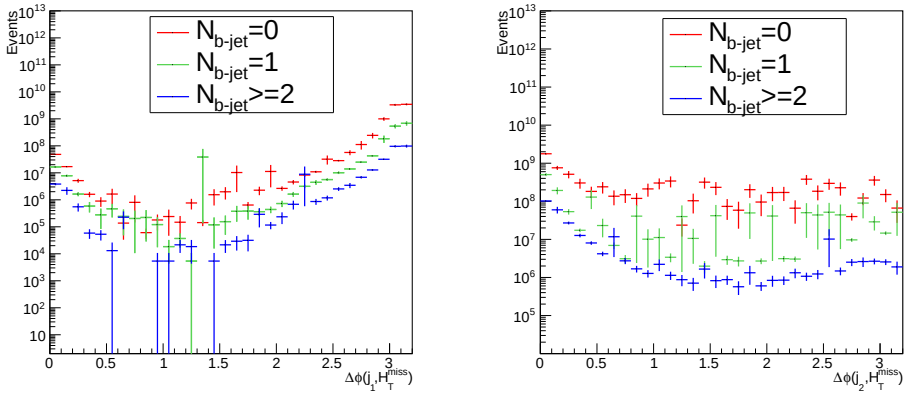

FIGURE 5.3: $\Delta\Phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{jet}},\vec{H}_{\mathrm{T}}^{\,\mathrm{miss}})$ distribution in QCD events plotted in zero lepton signal region for events with zero, one or atleast two $N_{\mathrm{b-jet}}$s: (left) leading jet, (right) subleading jet within $|\eta|$ <2.4 after applying baseline selection of H_{T}> 300 GeV, $H_{\mathrm{T}}^{\mathrm{miss}}$ > 300 GeV, $N_{\mathrm{jet}}\geq 2$

- $H_{\mathrm{T}}^{\mathrm{miss}}$ (or MHT)- Magnitude of negative vector sum of p_{T} of jets with p_{T} > 30 GeV and $|\eta| < 5.0$

For substantial SM background rejection and minimal impact on signal models following baseline selections are applied:

- $N_{\mathrm{jet}} \geq 2$, H_{T} >300 GeV, $H_{\mathrm{T}}^{\mathrm{miss}}$ >300 GeV

  This defines the baseline selection for this search. In addition, a region with H_{T} >300 GeV and 250 GeV < $H_{\mathrm{T}}^{\mathrm{miss}}$ < 300 GeV is used as a low-$H_{\mathrm{T}}^{\mathrm{miss}}$ sideband to derive normalization and validation of R+S strategy.

  On top of above selections, to avoid contribution from leptonic decay modes of SM backgrounds, events with isolated lepton and/or isolated track are vetoed.

- Muon veto:

  Muon candidates selected are those which satisfy "Medium ID" requirement. Medium ID is optimized towards prompt muon selection from Z and W boson decay with efficiency of about $\geq$ 99.5% [91]. In addition, impact Parameter (IP) with respect to primary vertex (PV) should satisfy condition

$$d_{xy}(\mu, PV) \leq 0.2\ cm \tag{5.1}$$

$$d_{z}(\mu, PV) \leq 0.5\ cm \tag{5.2}$$

  Medium Muon candidates should have p_{T} >10 GeV and $|\eta|$ <2.4. To distinguish between prompt muons and muons from b hadron decays, it should also satisfy mini-isolation requirement of I_{mini} <0.2. An event satisfying above criteria is rejected.

- Electron veto:

Electron candidates have to satisfy "Cut based veto" selection. Due to "looser selection" imposed, "veto ID" has maximum efficiency to select real electrons but can also have more fakes. Candidates should have p_{T} >10 GeV and $|\eta|$ <2.5. Isolation requirement is I_{mini} <0.1. Any event satisfying above conditions is rejected.

- Isolated track veto:

  This is applied on top of isolated lepton veto, to reject events where any leptonic event has failed lepton veto due to lepton misidentification or isolation requirement not being satisfied. It also helps to reduce hadronic tau background by vetoing events with isolated charged pion track.

  Lepton track needs to satisfy requirement of p_{T} >5 GeV and I_{tk} <0.2, while hadronic track (mostly pion) should have p_{T} >10 GeV and I_{tk} <0.1, where track isolation I_{tk} is defined as the ratio of scalar p_{T} sum of charged tracks within ΔR<0.3 of primary track to p_{T} of primary track.

- $\Delta\Phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{jet}},\vec{H}_{\mathrm{T}}^{\,\mathrm{miss}})$

  As enormous QCD background has the feature that $p_{\mathrm{T}}^{\mathrm{miss}}$ due to jet mismeasurement is aligned to mismeasured jet, events in signal region are needed to satisfy $\Delta\Phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{jet}},\vec{H}_{\mathrm{T}}^{\,\mathrm{miss}}) > 0.3,0.3,0.5,0.5$ for four leading H_{T} jets, to reject this background. QCD dominated control region is selected by inverting above ΔΦ requirement (at least one of the leading jet should fail either of $\Delta\Phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{jet}},\vec{H}_{\mathrm{T}}^{\,\mathrm{miss}})$ > 0.3,0.3,0.5,0.5 cuts) and asking for $N_{\mathrm{b-jet}}$ =0 and 250 GeV < $H_{\mathrm{T}}^{\mathrm{miss}}$ < 300 GeV (called low ΔΦ QCD validation region). A data-driven strategy to estimate, called "Rebalance and Smear (R+S)" explained later, is validated separately in low-$H_{\mathrm{T}}^{\mathrm{miss}}$ and low ΔΦ validation regions.

- MET filters for rejecting events with spurious high $p_{\mathrm{T}}^{\mathrm{miss}}$

  Section 3.5 has discussed various event cleaning filters used during legacy analysis. An event of interest is expected to pass all imposed filters.

In various SUSY topologies, a rich variety of final states are possible. One can design a SUSY search exclusive to either of possible topologies or it can be an inclusive search covering maximum phase space to allow simultaneous scan of various SUSY topologies. The analysis described here uses inclusive approach for SUSY search.

Events survived after above cuts are considered for this search and are distributed across 174 search bins. Figure 5.4 shows $N_{\mathrm{jet}} - N_{\mathrm{b-jet}}$ and $H_{\mathrm{T}} - H_{\mathrm{T}}^{\mathrm{miss}}$ binning used during 2016 [19] and full Run 2 analysis. Each $N_{\mathrm{jet}} - N_{\mathrm{b-jet}}$ bin is further divided into $H_{\mathrm{T}} - H_{\mathrm{T}}^{\mathrm{miss}}$ bins as shown in fig.5.5. To get advantage of higher luminosity, bins are made tighter for full Run 2 analysis.

5.3 Signal and SM background composition

The figures 5.6 and 5.7 respectively show distribution of various search variables only with H_{T} > 200 GeV, $H_{\mathrm{T}}^{\mathrm{miss}}$ > 200 GeV requirement and after full baseline selections applied. Here event weights of both background and signal processes are normalized to their respective cross-sections and lumi of 35.9 fb^{-1}, so as to match total number of events (observed for SM and theoretical for signal) expected in 35.9 fb^{-1}data. Kinematic distributions from signal mass points, which were buried under huge SM background yield started becoming visible after putting baseline cuts, as observed in fig. 5.7.

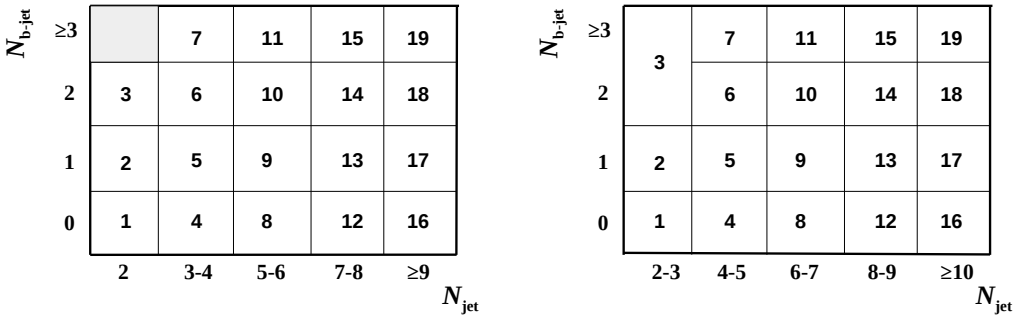

FIGURE 5.4: N_{jet} - $N_{\mathrm{b-jet}}$ binning used for 2016 (left) and full Run 2 analysis (right) with 35.9 fb^{-1}and 137 fb^{-1}data respectively

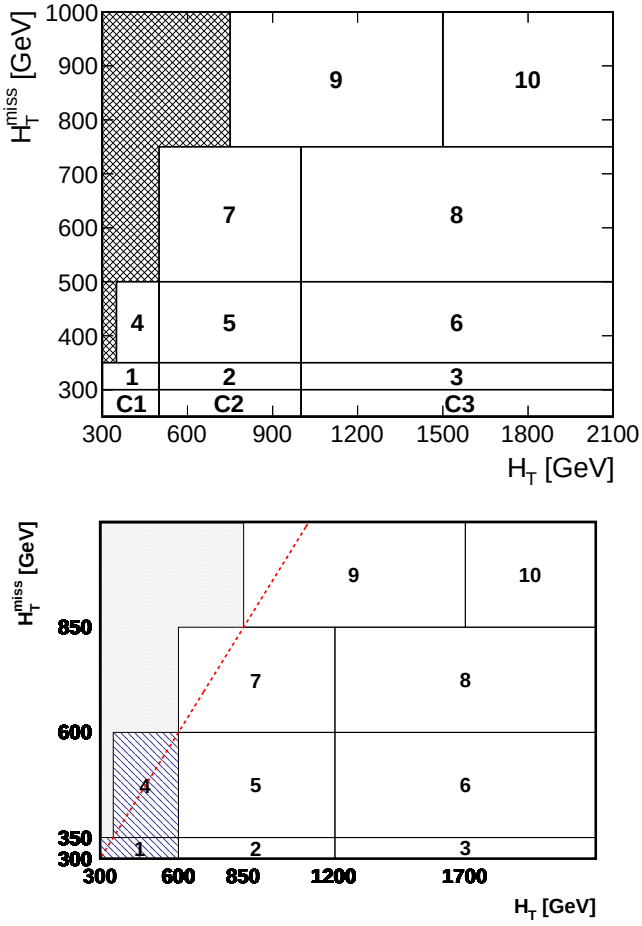

FIGURE 5.5: H_{T} - $H_{\mathrm{T}}^{\mathrm{miss}}$ binning used for 2016 (top) and full Run 2 analysis (bottom) with 35.9 fb^{-1}and 137 fb^{-1}data respectively

By starting with a set of events satisfying $H_{\mathrm{T}} > 200$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV requirement, how various cuts help in rejection of SM background processes and final survived fraction contributing as SM background in signal region are shown in Fig. 5.8. The H_{T} , $H_{\mathrm{T}}^{\mathrm{miss}}$, N_{jet} baseline selection brings potential SM background down to few percent. Relative rejection by $\Delta\Phi$(jets, $H_{\mathrm{T}}^{\mathrm{miss}}$) cut is maximum for QCD

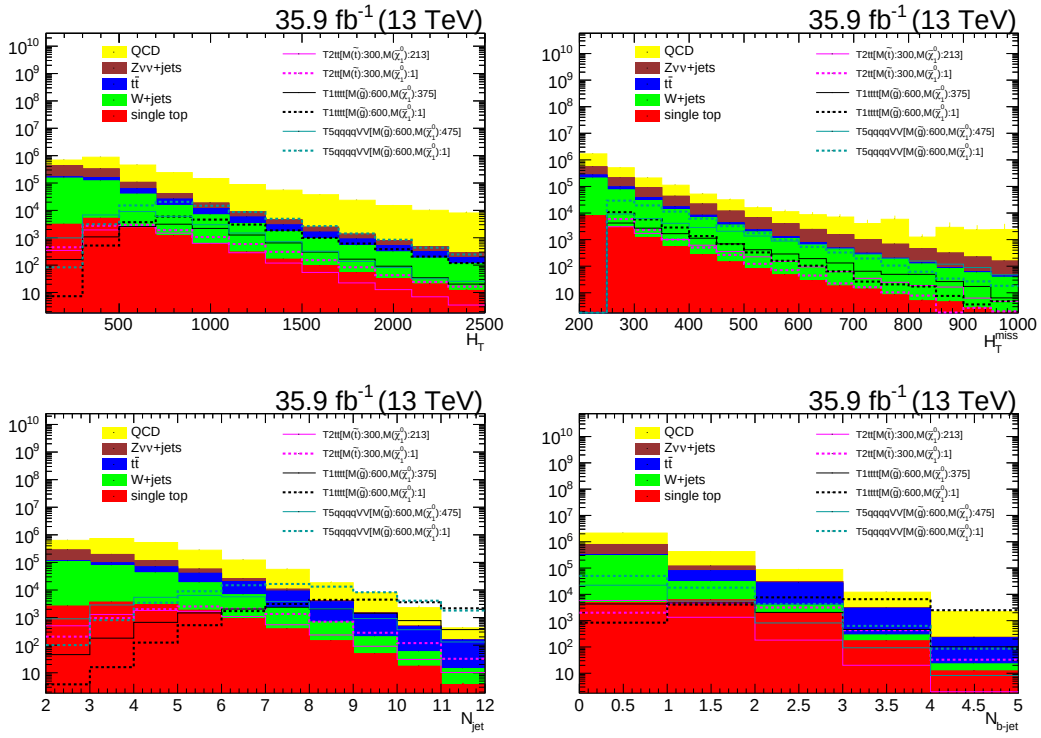

FIGURE 5.6: Distribution of kinematic variables for SM backgrounds and various SUSY signal topologies before applying baseline selections

as expected. In a true physics event, $H_{\mathrm{T}}^{\mathrm{miss}}$ is expected to be less than H_{T} , hence rejection is minimum with $H_{\mathrm{T}}^{\mathrm{miss}} < H_{\mathrm{T}}$ requirement (red line in figure 5.5 (bottom)). Lepton veto and isotrack veto rejection is few percent or less than that. The $p_{\mathrm{T}}^{\mathrm{miss}}$ filters are supposed to reject anomalous $p_{\mathrm{T}}^{\mathrm{miss}}$ events in data due to detector issues as discussed in section 3.5 but does not have significant impact in MC as expected. Survived background fraction is maximum for Z+jets (about 4%), followed by $t\bar{t}$, W+jets, single top and QCD in descending order.

Background composition changes depending on kinematic phase space. Figure 5.9 shows how Z+jet, $t\bar{t}$/W + jets (Lost Lepton) and QCD contribution changes as a function of one search variable at a time and inclusive in other three search variables. The Z+jet appears as largest background both at low and high $H_{\mathrm{T}}^{\mathrm{miss}}$ ends. But as a function of N_{jet} , $N_{\mathrm{b-jet}}$ multiplicity, lost lepton contribution increases due to dominating $t\bar{t}$ process at higher end. Figure 5.10 shows a composition in two dimensional [N_{jet} ,$N_{\mathrm{b-jet}}$] phase space but inclusive in H_{T} >300 GeV and $H_{\mathrm{T}}^{\mathrm{miss}}$ >300 GeV for top plot and H_{T} >600 GeV and $H_{\mathrm{T}}^{\mathrm{miss}}$ >600 GeV for bottom plot. With higher H_{T} and $H_{\mathrm{T}}^{\mathrm{miss}}$ requirement, both QCD and lost lepton background from $t\bar{t}$/W +jets drop, except at $N_{\mathrm{b-jet}}$ =2+ for highest two N_{jet} bins, where lost lepton remains a dominant background even at high H_{T} , $H_{\mathrm{T}}^{\mathrm{miss}}$ requirement due to $t\bar{t}$ contribution. In case of invisible Z+jet event, two neutrinos are most likely to fall in same hemisphere giving high $H_{\mathrm{T}}^{\mathrm{miss}}$ as compared to $t\bar{t}$, where neutrino, antineutrino from top and antitop fall mostly in opposite hemispheres leading to small values of $H_{\mathrm{T}}^{\mathrm{miss}}$ due to mutual cancellation. As a result, Z+jet dominates in $H_{\mathrm{T}}^{\mathrm{miss}}$ spectrum.

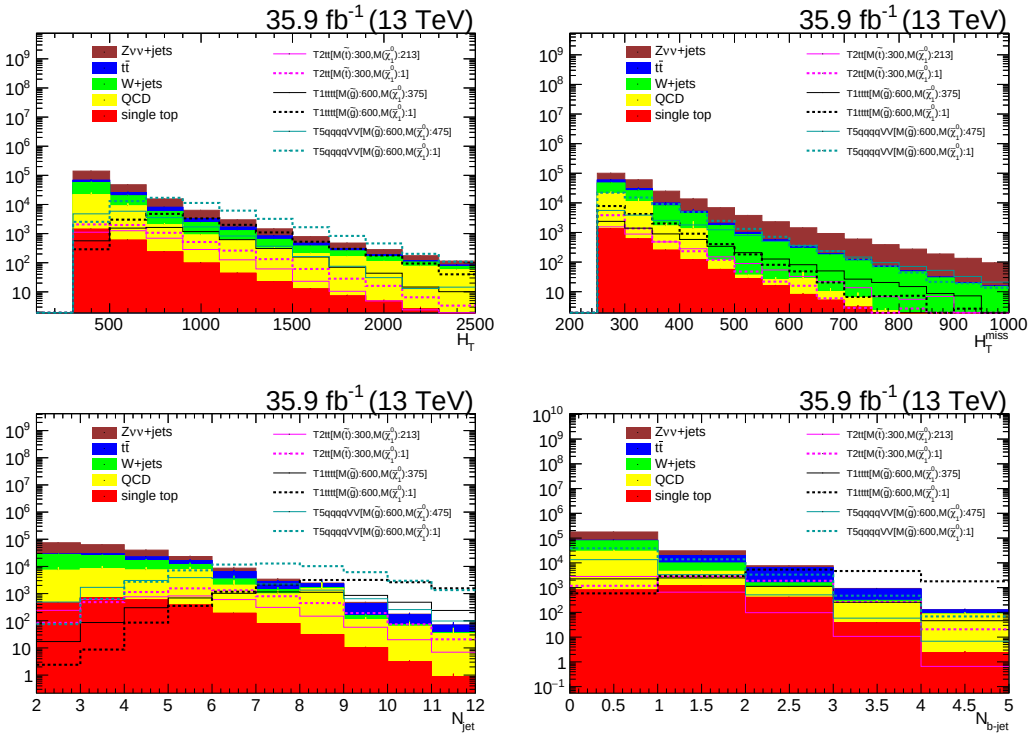

FIGURE 5.7: Distribution of kinematic variables for SM backgrounds and various SUSY signal topologies after applying baseline selections

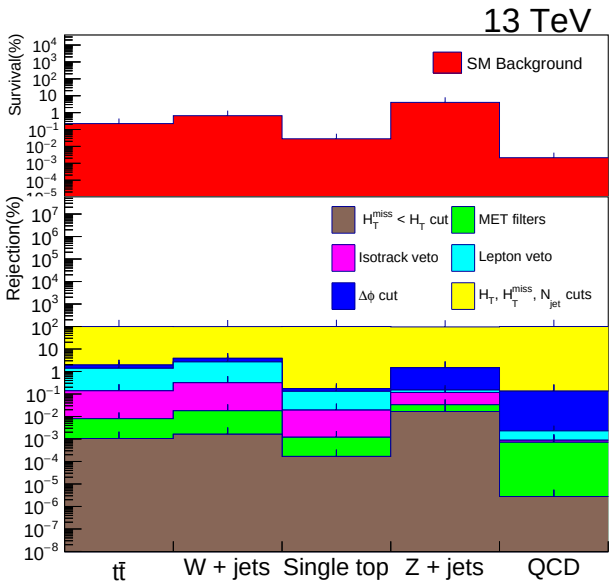

FIGURE 5.8: Starting with a set of events satisfying $H_{\mathrm{T}} > 200$ GeV, $H_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV requirement, relative percentage rejection by various selections are shown in bottom panel, while final events survived after all selections are shown as a percentage in top panel separately for $t\bar{t}$, W+jet, ST, Z+jet and QCD processes.

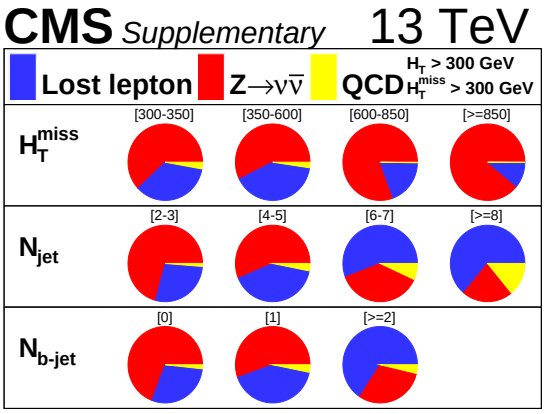

FIGURE 5.9: Background composition as a function of different search variable bins in 137 fb^{-1}data (see table F.1 for numbers)

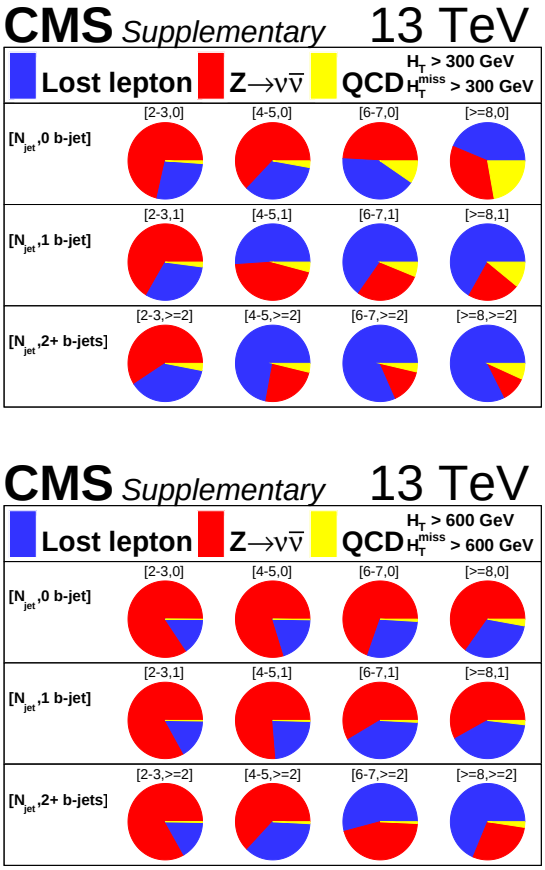

FIGURE 5.10: Background composition against [N_{jet} , $N_{\text{b-jet}}$] inclusive in H_{T} and $H_{\text{T}}^{\text{miss}}$ with selection of $H_{\text{T}} > 300$ GeV, $H_{\text{T}}^{\text{miss}} > 300$ GeV for top plot and $H_{\text{T}} > 600$ GeV, $H_{\text{T}}^{\text{miss}} > 600$ GeV for bottom plot in 137 fb^{-1}data (see tables F.2 and F.3 for numbers)

5.4 Triggers

Triggers are used to store only a subset of all data events, which are useful for physics analysis purpose [58]. A set of MET $\times$ MHT triggers are used for event selection from data in signal region.

- HLT_PFMETx_PFMHTx_IDTight_v* (x=90,100,110,120,130,140)
- HLT_PFMETNoMux_PFMHTNoMux_IDTight_v* (x=90,100,110,120,130,140)
- HLT_PFMETx_PFMHTx_IDTight_PFHT60_v* (x=100,110,120,130,140)
- HLT_PFMETNoMux_PFMHTNoMux_IDTight_PFHT60_v* (x=100,110,120,130,140)

Here the variables and their values refer to trigger level reconstruction and calibration. The high threshold triggers act as a back-up to compensate efficiency loss when low threshold triggers are prescaled at higher instantaneous luminosity. The same set of triggers was used also for lost lepton control region (CR) and QCD low $\Delta\Phi$ validation region in 2018. The efficiency of this set of triggers is determined both in true and fake $p_{\mathrm{T}}^{\mathrm{miss}}$ region. Efficiency in true $p_{\mathrm{T}}^{\mathrm{miss}}$ region is calculated using single electron dataset as reference while efficiency in fake $p_{\mathrm{T}}^{\mathrm{miss}}$ region is calculated using single photon dataset as reference.

$$\epsilon_{true-MET} = \frac{\text{Events passing MET and single electron trigger } (H_{\mathrm{T}} > 300,\ N_{\mathrm{jet}} > 1)}{\text{Events passing single electron trigger } (H_{\mathrm{T}} > 300,\ N_{\mathrm{jet}} > 1)} \tag{5.3}$$

$$\epsilon_{fake-MET} = \frac{\text{Events passing MET and single photon trigger } (H_{\mathrm{T}} > 300,\ N_{\mathrm{jet}} > 1)}{\text{Events passing single photon trigger } (H_{\mathrm{T}} > 300,\ N_{\mathrm{jet}} > 1)} \tag{5.4}$$

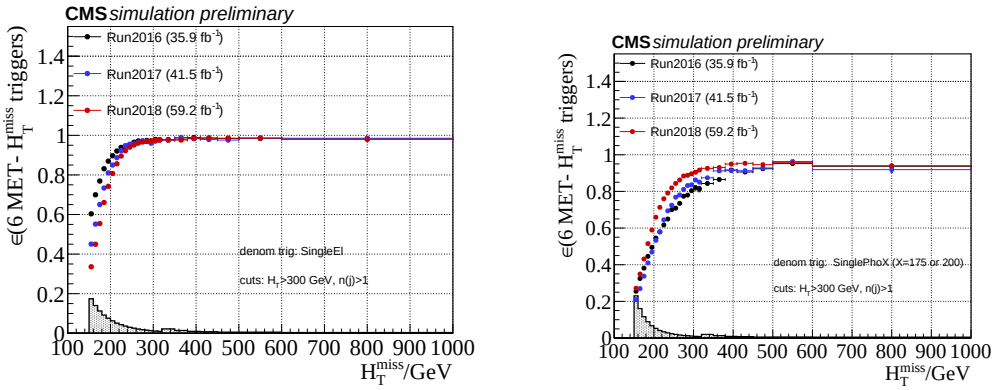

FIGURE 5.11: Signal region trigger efficiencies in True and Fake MET region

Efficiencies are shown in Figure 5.11 and are used to correct simulated signal yield and events used to get Z+jet and R+S QCD prediction.

Following are the triggers used for various CR selections.

- Single electron or muon triggers:

  These are used to collect events in di-electron and di- muon CR for Z+jet prediction. Single-muon triggers are found to be fully efficient in di-muon CR. Single-electron trigger efficiency is fully recovered for HEM failure by vetoing electrons in HEM failure region.

- A set of single photon triggers:

  These are used for Z+jet CR selection.

  - HLT_Photon175_v* — for 2016
  - HLT_Photon200_v* — for 2017 and 2018

  They are found to be 99% efficient in high $\Delta\Phi$ baseline region ($p_{\mathrm{T}}^{\gamma} > 250$ GeV).

- Suite of H_{T} triggers

  These are used for R+S QCD seed selection.

  HLT_PFHTx_v* (x=200, 250, 300, —— 900)

  Many of them are prescaled, hence corrections are applied to account for prescaling.

5.5 Estimation of SM backgrounds

Signal region (SR) events recorded in data can be admixture of contribution from SM processes and SUSY signal if it exists. In this admixture, there is no way to know whether a particular data event has originated from SM process or potential SUSY event. But from MC simulation it is possible to get an idea about estimate of expected SM background and strength of potential SUSY signal to populate in a particular search bin. The general idea of getting background prediction is to start with a kinematic phase space which is dominated by SM background and is almost free from signal, called a " control region" (CR). Based on best known understanding of CR and SR, a method is developed which is applied on CR event yield to get expected SM background yield in SR. This method which maps from CR to SR takes into account the physics based difference between CR and SR. The SM background yields predicted by this way can be compared directly to expected MC yields in SR. If expectation and prediction agrees within uncertainties, it provides confidence on estimation strategy. Such well validated estimation strategy from MC, is applied to data CR to get data driven background prediction. A necessity of "data driven" approach comes from the fact that MC generation restricted to next to next leading order (and not including higher order diagrams) lets spectra of various search variables to differ in MC with respect to data both in normalization and shape. In addition, with changing collider and detector conditions across data taking, it is not always possible to simulate all detector level effects. In data driven approach used for this analysis, CR for background estimation is taken from data and MC based inputs used in estimation strategy (e.g. lepton efficiencies) are corrected for data versus MC differences by applying associated data/MC scale factors as correction.

Following sections will describe Z+jet, QCD and lost lepton background estimation strategies used in 2016 [19]. For Z+jet and QCD, the strategies were unchanged during full Run 2 analysis [20].

5.5.1 Z($\nu\overline{\nu}$) + jets background estimation

The Z($\nu\overline{\nu}$) + jets background uses γ + jets as a control region (CR) for estimation in 46 $N_{\mathrm{b-jet}}$ =0 bins while estimation in higher $N_{\mathrm{b-jet}}$ bins use $Z \rightarrow \mathrm{l^+l^-}$ as CR [19]. Using $Z \rightarrow \mathrm{l^+l^-}$ as CR for overall background estimation would have been an option. But, $Z \rightarrow \mathrm{l^+l^-}$ suffers by low statistics (branching fraction $\approx$ 10%). The production of prompt γ + jets events is similar to Z + jet process, and production rates become of similar order for high boson p_{T} with respect to Z mass. Also γ + jets is benefitted by higher statistics compared to $Z \rightarrow \mathrm{l^+l^-}$ in $N_{\mathrm{b-jet}}$ =0 bins. For these reasons, γ + jets is used as CR, for Z($\nu\overline{\nu}$) + jets estimation in $N_{\mathrm{b-jet}}$ =0 bins, while extrapolation to higher $N_{\mathrm{b-jet}}$ bins is done using $Z \rightarrow \mathrm{l^+l^-}$ spectrum directly from data. Both γ + jets and $Z \rightarrow \mathrm{l^+l^-}$ CRs do not have true $p_{\mathrm{T}}^{\mathrm{miss}}$. Jet clustering is redone by removing γ or Z($\mathrm{l^+l^-}$) from γ + jets and Z($\mathrm{l^+l^-}$) + jets events, to mimic $p_{\mathrm{T}}^{\mathrm{miss}}$ from Z($\nu\overline{\nu}$) + jets process. Both γ and dilepton pair from two CRs should have $p_{\mathrm{T}} > 200$ GeV.

Only direct prompt production of γ + jets resembles production of Z at vertex. To avoid contamination from indirect (from π^0 decay) and fragmentation photons (from gluon radiation), γ + jets sample is corrected by applying "direct photon fraction" (F_{dir}) and photon purity (β_γ) derived from the MC events. Normalization difference between γ + jets and Z + jets is accounted by applying transfer factor $R_{\mathrm{Z_{ll}}/\gamma}$ derived from MC events as a function of 46 $N_{\mathrm{b-jet}}$ =0 bins. Double ratio ($R^{data}_{\mathrm{Z_{ll}}/\gamma}/R^{sim}_{\mathrm{Z_{ll}}/\gamma}$) derived as a function of $H_{\mathrm{T}}^{\mathrm{miss}}$ is used to correct transfer factor for trigger efficiency and

data versus simulation differences in photon reconstruction efficiency [92]. Equation 5.5 shows estimated $Z \to \nu\overline{\nu}$ background as a function of γ + jets yield.

$$N^{pred}_{Z\to\nu\overline{\nu}} = (R^{data}_{\mathrm{Z_{ll}}/\gamma} / R^{sim}_{\mathrm{Z_{ll}}/\gamma}) \times R_{\mathrm{Z_{ll}}/\gamma} \times F_{\mathrm{dir}} \times \beta_{\gamma}(H_{\mathrm{T}}^{\mathrm{miss}}) \times N^{obs}_{\gamma} \tag{5.5}$$

Due to lack of statistics, $R_{\mathrm{Z_{ll}}/\gamma}$ can not be trusted for high $N_{\mathrm{b-jet}}$ case. For this reason, Z($\mathrm{l^+l^-}$) + jets events are used for extrapolation to higher $N_{\mathrm{b-jet}}$. Extrapolation factor F_{jb} is measured for each N_{jet} , $N_{\mathrm{b-jet}}$ using Z($\mathrm{l^+l^-}$) + jets data corrected for its purity (equation 5.6).

$$F_{jb} = N_{j,b}/N_{j,0} \tag{5.6}$$

This extrapolation factor is used as a multiplying factor to get event yield in given [N_{jet} ,$N_{\mathrm{b-jet}}$,H_{T} ,$H_{\mathrm{T}}^{\mathrm{miss}}$] bin from yield in [N_{jet} ,0,H_{T} ,$H_{\mathrm{T}}^{\mathrm{miss}}$] bin (equation 5.7).

$$N(Z \to \nu\overline{\nu})^{k}_{j,b} = N(Z \to \nu\overline{\nu})^{k}_{j,0} \times F_{jb} \tag{5.7}$$

In equations 5.6 and 5.7, index j, b correspond to N_{jet} , $N_{\mathrm{b-jet}}$ respectively while index k corresponds to a kinematic bin in $H_{\mathrm{T}} - H_{\mathrm{T}}^{\mathrm{miss}}$ plane.

Figure 5.12 (top) shows $R_{\mathrm{Z_{ll}}/\gamma}$ ratio and (bottom) shows F_{jb} extrapolation factors in bottom panel.

Figure 5.13 shows MC closure for Z invisible background expected and predicted after scaling to 137 $\mathrm{fb^{-1}}$. By definition, there is a perfect closure observed in $N_{\mathrm{b-jet}}$ = 0 bin. In other bins also expectation and prediction agree within statistical uncertainties.

5.5.2 QCD multijet background estimation

The QCD multijet background estimation uses Rebalance and Smear (R+S) method [19, 93, 94]. A set of low $\Delta\Phi$ control region (CR) events are selected using a set of H_{T} triggers (with no condition on $H_{\mathrm{T}}^{\mathrm{miss}}$) with thresholds of 200 to 1050 GeV.

In first step of rebalancing an event, p_{T} of each measured jet in CR event is varied in order to maximize the probability:

$$P(\overline{J}_{true}|\overline{J}_{meas}) = P(\overline{J}_{meas}|\overline{J}_{true})/\pi(H_{\mathrm{T}}^{\mathrm{miss}}, \delta\phi_{j\,1\,(b)}) \tag{5.8}$$

The $P(\overline{J}_{true}|\overline{J}_{meas})$ represents posterior probability density of assigning true jet momenta $\overline{J}_{true}$ to a configuration of jets with measured momenta $\overline{J}_{meas}$. The likelihood to observe a configuration of measured jets $\overline{J}_{meas}$ given a configuration of jets with true momenta $\overline{J}_{true}$ is represented by $P(\overline{J}_{meas}|\overline{J}_{true})$. The $\pi(H_{\mathrm{T}}^{\mathrm{miss}}, \delta\phi_{j\,1\,(b)})$ represents magnitude and direction (with respect to highest p_{T} or leading b-tagged jet) of $H_{\mathrm{T}}^{\mathrm{miss}}$ in QCD events.

In second step called smearing, every individual jet in rebalanced event is smeared by random sampling of jet response templates which accounts for jet resolution in detector. Each rebalanced event is sampled numerous times and is weighted by number of times it is smeared. This R+S procedure results in generating event sample which resembles original CR event sample but after electroweak background contamination subtracted. Figure 5.14 shows validation of R+S method in QCD CR. Observed data yield in CR agrees with total predicted background within (statistical + systematic) uncertainties.

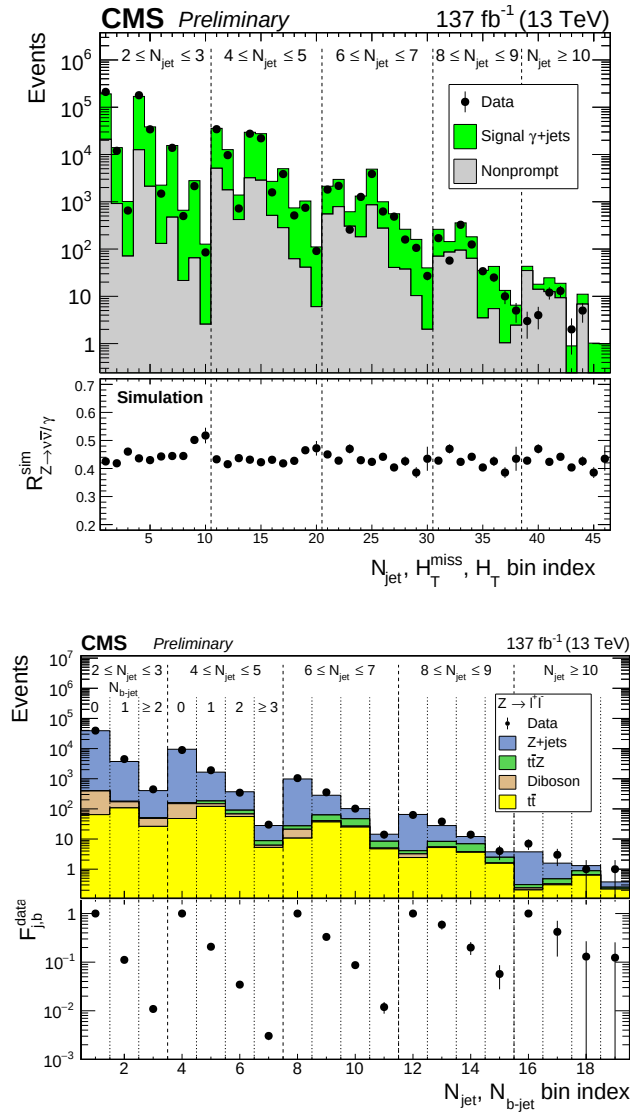

FIGURE 5.12: (top)Top panel shows $\gamma+\mathrm{jets}$ events in data overlaid on decomposition as prompt and non-prompt $\gamma+\mathrm{jets}$ events in MC. Bottom panel shows $R_{\mathrm{Z_{ll}}/\gamma}$ ratio in 46 0-$N_{\mathrm{b-jet}}$ bins.(bottom) Top panel shows $Z \rightarrow \mathrm{l^+l^-}$ data overlaid with different backgrounds. Bottom panel shows F_{jb} extrapolation factors derived for various N_{jet}-$N_{\mathrm{b-jet}}$ bins.

5.5.3 Lost lepton background estimate

As discussed before, lost lepton background will have contribution from lost electron and muon ($lost^{\mathrm{e},\mu}$) and from hadronically decaying tau ($lost^{\tau_{\mathrm{h}}}$).

The $lost^{\mathrm{e},\mu}$ background estimation during 2016 analysis used event by event approach [19]. A SM event with either an electron or muon and satisfying baseline selection, can enter into a signal region if lepton fails at any of the following three steps:

1] kinematic acceptance
2] reconstruction
3] isolation

Kinematic acceptance here is defined as a lepton should have p_{T} >20 GeV and $|\eta|$

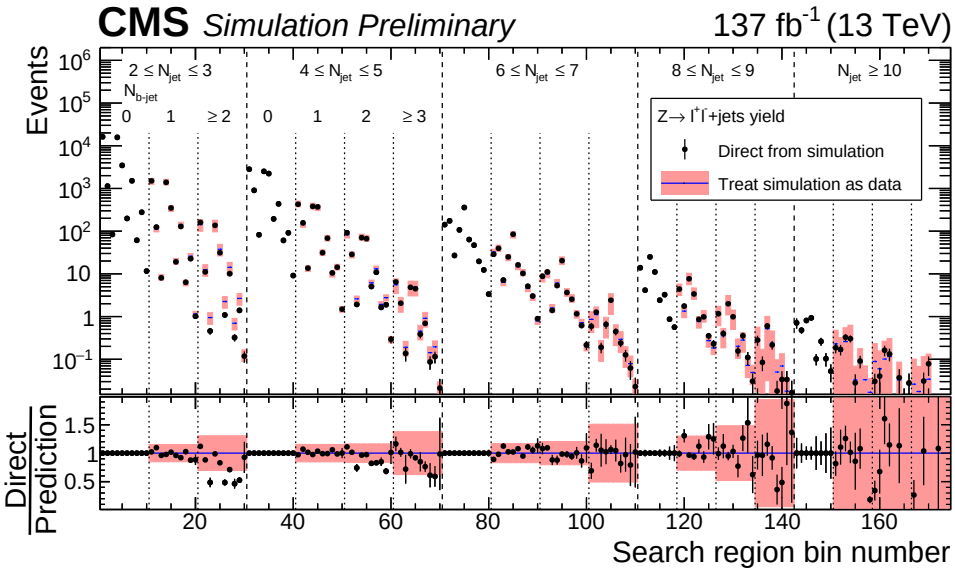

FIGURE 5.13: Comparison of expected and predicted Z invisible background

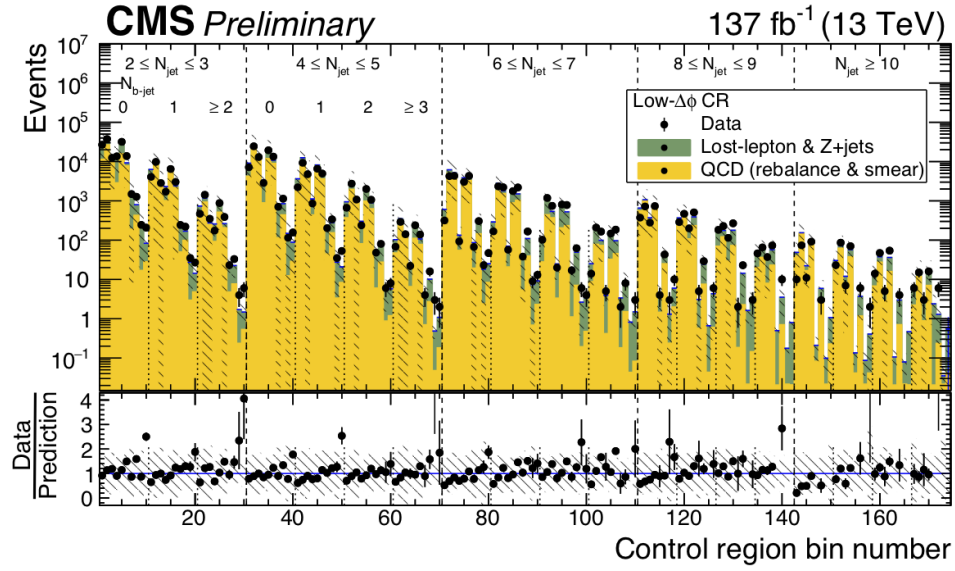

FIGURE 5.14: Comparison of data versus predicted backgrounds in QCD low $\Delta\Phi$ CR

<2.1. Identification and isolation requirements for electron and muon are as specified in section 5.2. The "found lepton" events populating one lepton control region (CR) are those after successively passing all above three requirements. To get "lost lepton" events either because of failing at acceptance or reconstruction or isolation step, each CR event weight is scaled appropriately to account for efficiency associated to that step and all following steps. This scaling by efficiency to get "lost lepton" count from "found lepton" event counts is done on event by event basis, by using efficiencies as a function of [p_T , η] of "found lepton" event in CR.

Method was validated in MC by comparing predicted background from one lepton CR yield to expected yield directly from MC (called MC closure test) as shown in Fig. 5.15. In such validation, both single lepton CR and zero lepton signal region

(SR) are both derived from the same MC sample.

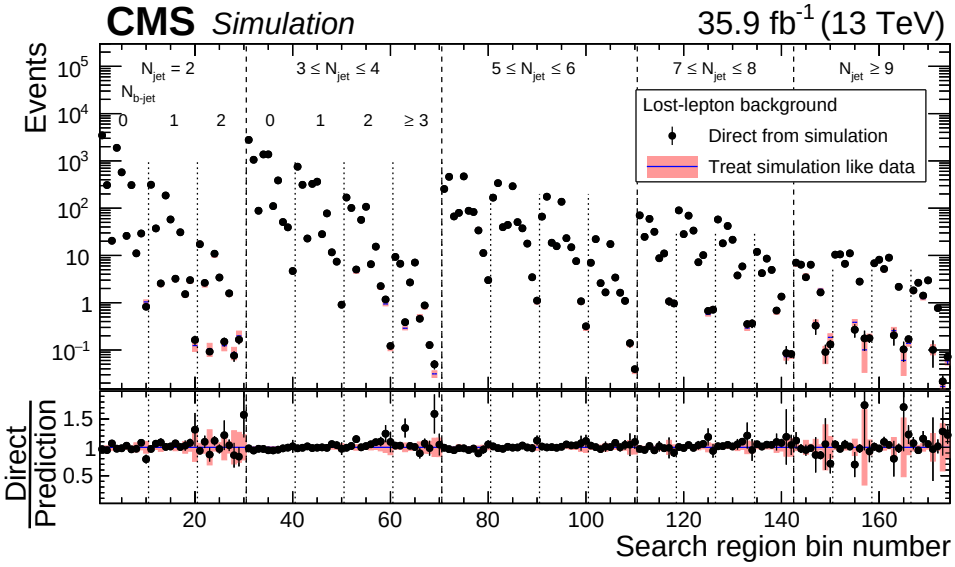

FIGURE 5.15: $lost^{\mathrm{e},\mu}$ closure in 174 search bins

The $lost^{\tau_{\mathrm{h}}}$ estimation in 2016 was done using template method. For full Run 2 analysis, $lost^{\tau_{\mathrm{h}}}$ component was estimated together with $lost^{\mathrm{e},\mu}$ using Average Transfer Factor (TF) method. The $lost^{\tau_{\mathrm{h}}}$ estimation for 2016 analysis and estimation of $lost^{\mathrm{e},\mu}$ + $lost^{\tau_{\mathrm{h}}}$ using TF method for full Run2 analysis are major parts of the thesis work. Next chapter discusses details of template method and average TF method.

Chapter 6

Estimation of $t\bar{t}$ +jets & W+jets backgrounds

This chapter discusses estimation of hadronically decaying tau ($lost^{\tau_h}$) using template method used in 2016 analysis and estimation of $lost^{\tau_h}$ together with lost-electron or muon ($lost^{e,\mu}$) using average TF method used during full Run 2 analysis. The chapter discusses details of template method and then motivation for choosing average TF method over template method for Run 2 analysis. Later part discusses details of average TF method and also justifies why it is called "average" TF.

6.1 Estimation of $lost^{\tau_h}$ background using template method

The τ from W decay has 65% probability to decay hadronically (τ_h) and 35% probability to decay to electron or muon. At the generator level, W boson decaying to muon or tau are kinematically equivalent (as W boson much heavier compared to muon or tau). Figure 6.1 shows that p_T distribution of generator level tau (gen-τ) and generator level and reconstructed level muon (gen-μ and reco-μ) match very well after kinematic cuts of p_T >20 GeV and $|\eta|$ <2.1.

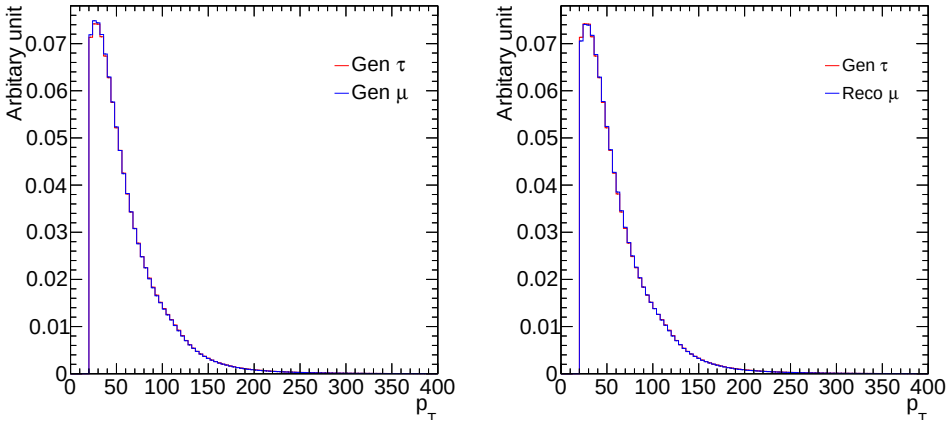

FIGURE 6.1: Comparison of gen-τ and generator level muon p_T (left) and that of gen-τ and reco-μ p_T (right) after kinematic cuts of p_T >20 GeV and $|\eta|$ <2.1

The difference between muon and tau appears only at the reconstruction level. Muon is reconstructed with very good momentum resolution in detector (average p_T resolution of global muon in barrel is about 1.5% for up to p_T < 600-700 GeV). The τ_h produces real multijet+p_T^{miss} signature at detector level. The p_T of gen-τ gets divided

between τ_h-jet and associated tau neutrino as soon as τ decays. In MC simulation, for a given p_T of gen-τ, it is possible to get probability distribution of fractional p_T carried by τ_h-jet.

Tau template method used for background estimation in 2016 [19] relies on following two facts:

- As muon and tau are equivalent at generator level, μ + jet events can be used as control region (CR) for τ_h background estimation
- Probability distribution of fractional p_T carried by τ_h-jet can be used as a template to simulate τ_h + jet event from μ + jet CR event.

The τ_h background estimation is done in high $\Delta\Phi$ and low $\Delta\Phi$ zero lepton region. High $\Delta\Phi$ zero lepton region defines signal region (SR) while low $\Delta\Phi$ zero lepton region is used to estimate electroweak contamination (here τ_h) to the QCD CR.

6.1.1 The τ_h template construction

To get the p_T spectrum of τ_h-jet as a function of gen-τ p_T , W + jet, $t\bar{t}$ and single top MC simulated samples are used. The events having at least one gen-τ are selected. The gen-τ carrying maximum p_T is used for template construction.

In order to find a τ_h-jet associated with gen-τ, a matching function is used. This matching function compares position of gen-τ against position of each of the reconstructed level jet in η-ϕ space. The jet which is closest to gen-τ within a cone of ΔR=0.2 if gen-τ p_T < 50 GeV and ΔR=0.1 otherwise, is determined as associated τ_h-jet. Here relaxation of ΔR for low gen-τ p_T is done to ensure reasonable efficiency for matching with τ_h-jet. The ratio of p_T of a matching τ_h-jet to p_T of gen-τ defines "τ_h-jet response". For every such τ_h-jet event, τ_h-jet response is filled with proper eventweight. This is done on event by event basis for full MC sample to generate probability distribution of fractional p_T of gen-τ carried by τ_h-jet (called τ_h template). Four different τ_h templates are constructed corresponding to gen-τ p_T ranges of 20-30 GeV, 30-50 GeV, 50-100 GeV and above 100 GeV, each with 50 bins of jet response as shown in figure 6.2. Probability distribution is normalized such that integral over all 50 bins equals unity for each of the templates. Detector resolution is worse at low p_T compared to high p_T . As a result fig. 6.2 shows that, a tail in probability distribution beyond one is prominent for low gen-τ p_T template compared to high p_T ones.

While simulating τ_h + jet event from muon CR event, muon p_T is supposed to be replaced by τ_h template. For this reason, τ_h template should depict the response of a τ_h-jet p_T as a function of gen-τ p_T without any contamination from surrounding activity in event and would be expected to be less than one ideally. But the templates derived using SM MC ($t\bar{t}$ +jet and W+jet MC) can be contaminated by some contribution from nearby activity making it "impure". This impurity as well as detector resolution will show up as template response shifted to higher fractional p_T values than one.

To avoid such contaminated response, two alternative ways of tau template construction were considered.

- Use Pythia8 [61] particle gun to produce tau and force it to decay hadronically. Use the closest matching jet to gen-τ for template construction (called "τ-gun" templates).

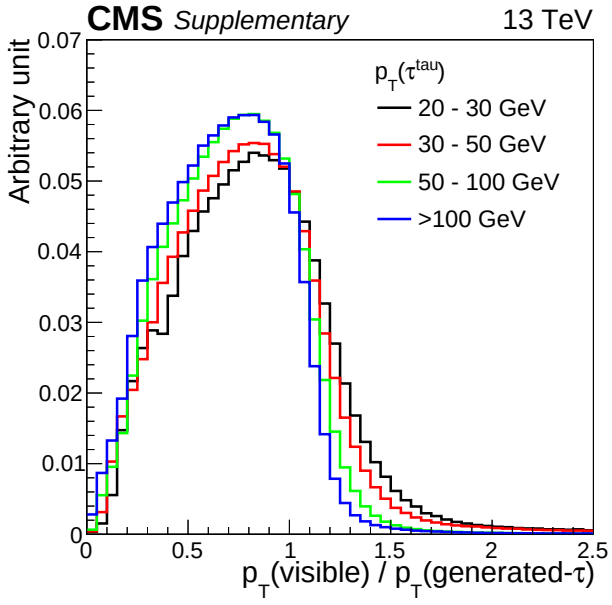

FIGURE 6.2: τ_h templates obtained as a function of various gen-τ p_T ranges

- Use Pythia8 particle gun to produce W and force it to decay to tau, and further decay of tau in hadronic channel. Use the closest matching jet to gen-τ for template construction (called "W-gun" templates).

In both cases, the generated event is passed through nominal CMS event reconstruction steps.

For a consistency check, τ-gun and W-gun template responses are compared at generator level, i.e instead of reconstructed jet, a closest matching generator level jet to gen-τ is used for template construction. For W-gun template, W is produced using Pythia gun within $|\eta|$ <4.0 without any p_T requirement. For τ-gun template, τ is allowed a range of η = [-2.4,2.4], ϕ = [-3.14, 3.14] and p_T = [0.0,160 GeV].

Figure 6.3 shows comparison of response templates obtained from τ-gun and W-gun for different p_T ranges. These comparisons clearly show that τ-gun templates failed to recover a shape of hadtau p_T distribution from W decay. The discrepancy is arising because decay products of W are polarized and possess certain angular correlations. This polarization effect is accounted properly in W-gun templates but not in τ-gun templates. This lead to the conclusion that τ-gun templates will not serve the purpose and W-gun templates will be a better choice for simulating τ_h-jet response over SM MC templates and τ-gun templates.

To make reconstructed level W-gun template response look closer to true response, additional adjustments are done as follow:

- To minimize “no matching jet” cases while template construction, the minimum jet p_T threshold in PF jet parameters is moved from 5 GeV to 1 GeV. As low p_T jets are more prone to detector mismeasurement compared to high p_T ones, this lowering of threshold p_T can significantly improve low gen-τ p_T templates.

- To account for the rare events where matching jet has $p_T < 1$ GeV and hence inevitably falling under “no matching jet” case, such events are forced to fill zero

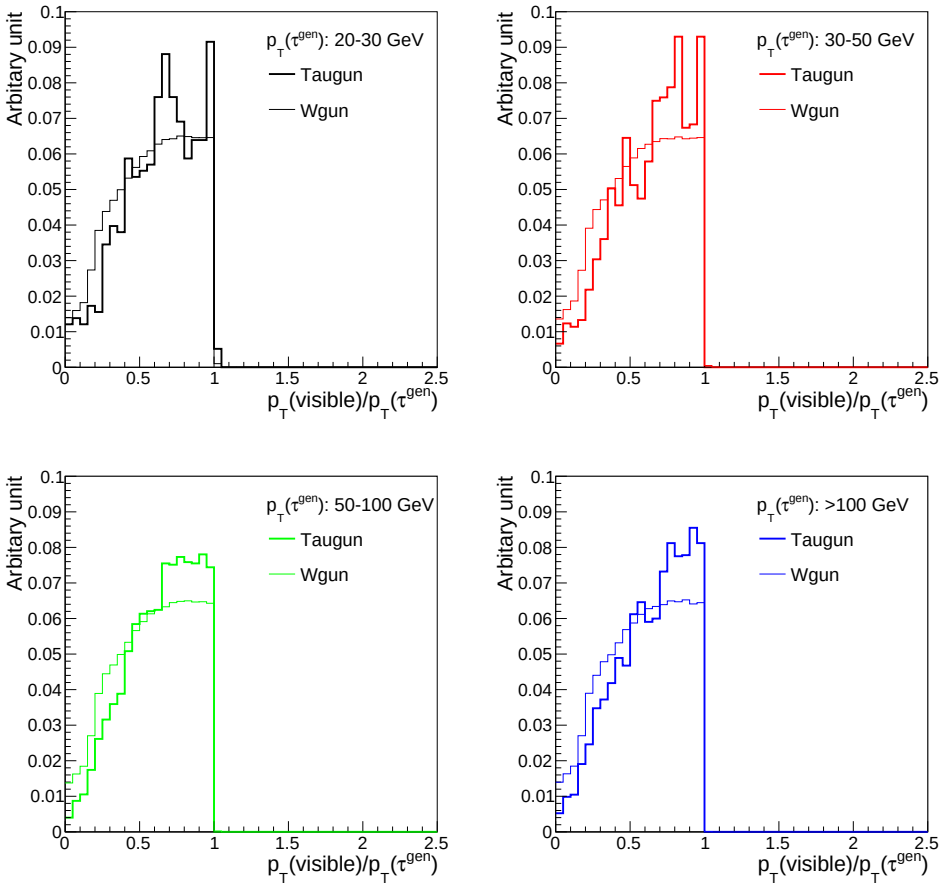

FIGURE 6.3: Comparison of τ_h templates obtained with tau gun and W-gun for p_T ranges (top left) 20-30 GeV, (top right) 30-50 GeV, (bottom left) 50-100 GeV and (bottom right) >100 GeV.

response template bin instead of just ignoring them. This condition ensures that each matching jet event contributes to either of the template bin.

Figure 6.4 shows comparison of templates obtained using SM MC samples and W-gun sample at reconstruction level. It can be seen that W-gun is removing tails in p_T distribution, significantly for low gen-τ p_T template compared to high gen-τ p_T template, which is expected as low p_T tau generally has stronger activity going around, and hence more contaminated. Figure 6.5 shows τ_h-jet response templates produced using W-gun events.

Impact of using W-gun templates instead of templates obtained from W+jets and $t\bar{t}$ +jets MC events will be discussed in section 6.1.4.

6.1.2 Emulation of events containing τ_h +jet event

The τ_h +jet events are simulated from μ+jets events by bootstrapping over 50 bins of τ_h template. That is, for every muon+jet event, there are 50 simulated τ_h + jet events, such that their probabilities add up to unity.

The muon from control region (CR) is selected by applying following conditions:

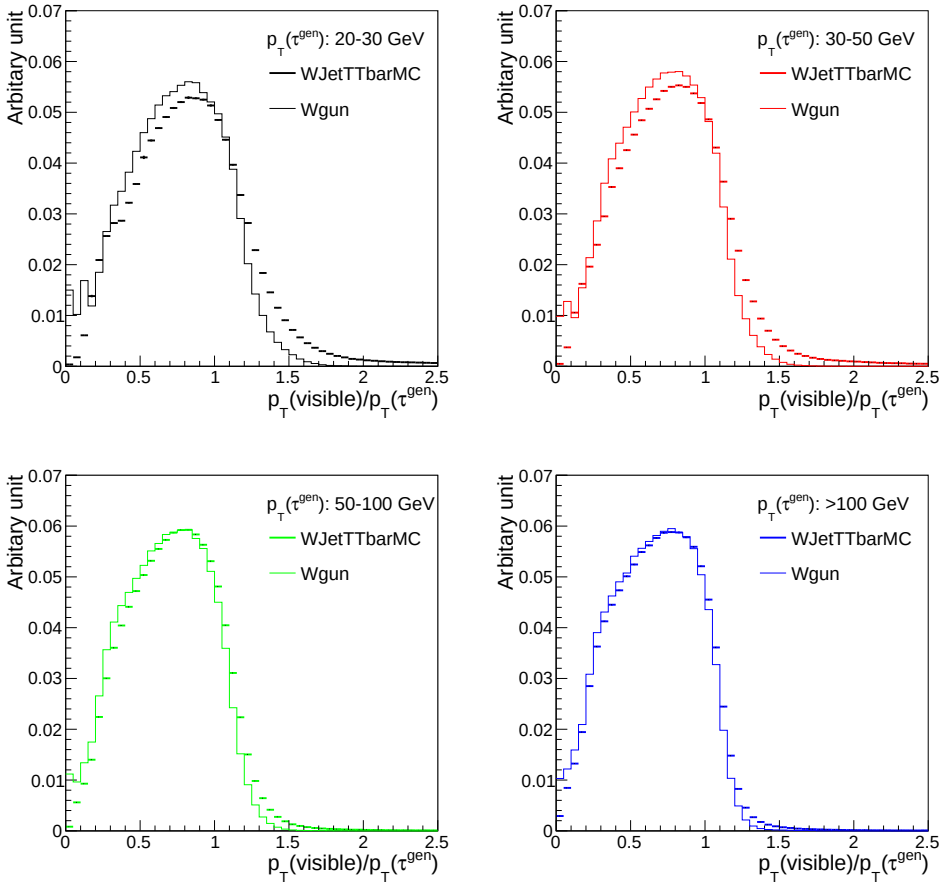

FIGURE 6.4: Comparison of τ_h templates obtained with W+jet and $t\bar{t}$ +jet MC or W-gun for p_T ranges (top left) 20-30 GeV, (top right) 30-50 GeV, (bottom left) 50-100 GeV and (bottom right) >100 GeV.

- exactly one isolated muon with $p_T > 20$ GeV and $\eta < 2.1$ ($p_T > 25$ GeV for 300 GeV $< H_T <$ 500 GeV due to trigger requirement). The muon should satisfy identification and isolation requirement as given under section 5.2
- no isolated electron candidate as defined under section 5.2.
- $m_T < 100$ requirement to reduce potential signal event contribution.

In a muon+jets events, a jet matching to a reco-μ within $\Delta R < 0.4$ is found. For every iteration of bootstrapping, a muon from this matched jet is to be replaced with a τ_h-jet response given by x-axis value corresponding to the template bin picked for that iteration. At first, raw p_T of a matched jet is obtained by undoing jet energy corrections (JEC). Then simulated τ_h-jet p_T using τ_h-jet response becomes:

$$p_{T\tau_h-jet} = \text{JEC} \times (\frac{p_T(\text{jet})}{\text{JEC}} - p_T(\mu)) + p_T(\mu) \times \tau_h \text{ response} \quad (6.1)$$

Probability of this simulated tau+jet event, is given by y-axis value of the bin selected for iteration and is called "bootstrap-weight" ($w_{bootstrap}$). Hence the weight of a simulated τ_h-jet event during each bootstrapping iteration becomes:

$$weight[\tau + jet] = weight[\mu + jet] \times w_{bootstrap} \quad (6.2)$$

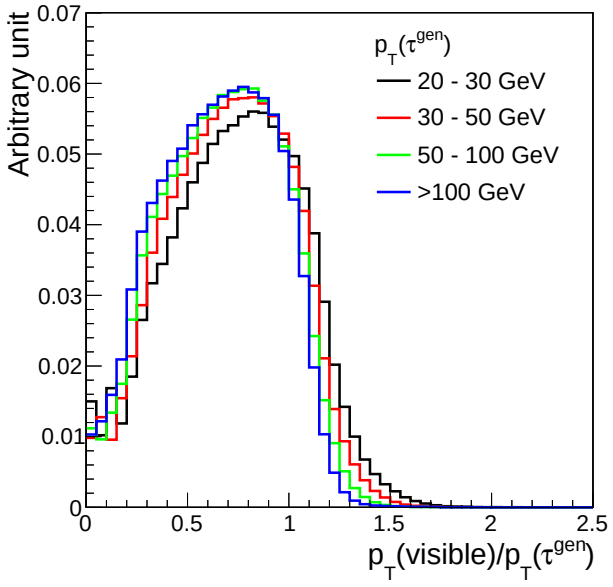

FIGURE 6.5: The $\tau_{\rm h}$ templates constructed using W-gun sample.

Here $weight[\mu + jet]$ represents weight based on cross section times luminosity. This scaling of $weight[\mu + jet]$ ensures that weights of 50 simulated $\tau_{\rm h}$-jet events add to unity for each of CR event.

As simulated $\tau_{\rm h}$-jet has different $p_{\rm T}$ than muon jet $p_{\rm T}$, it is possible that simulated $\tau_{\rm h}$-jet either pass or fail kinematic cuts on jet during baseline selection. For this reason, it becomes necessary to recalculate all search variables for each of simulated $\tau_{\rm h}$-jet event.

- $N_{\rm jet}$ - if both muon jet (before smearing) and simulated $\tau_{\rm h}$-jet pass kinematic cut of $p_{\rm T}$ >30 GeV and $|\eta|$ <2.4, then $N_{\rm jet}$ remains unchanged. If muon jet was not counted under $N_{\rm jet}$, but simulated $\tau_{\rm h}$-jet passes kinematic cuts, then $N_{\rm jet}$ increases by one. If muon jet was counted under $N_{\rm jet}$, but simulated $\tau_{\rm h}$-jet does not pass kinematic cuts, then $N_{\rm jet}$ decreases by one.
- $H_{\rm T}$: use new $N_{\rm jet}$ collection obtained above to get $H_{\rm T}$
- $H_{\rm T}^{\rm miss}$: recalculate $H_{\rm T}^{\rm miss}$ by subtracting muon jet $p_{\rm T}$ and adding simulated $\tau_{\rm h}$-jet $p_{\rm T}$ vectorially to original $H_{\rm T}^{\rm miss}$ only if both satisfy $p_{\rm T}$ >30 GeV and $|\eta|$ <5.0 requirement
- $\Delta\Phi$: use new $H_{\rm T}^{\rm miss}$ to get $H_{\rm T}^{\rm miss}$ -ϕ. For leading four jets (based on $p_{\rm T}$) in new $N_{\rm jet}$ collection, find $\Delta\Phi$ with respect to new $H_{\rm T}^{\rm miss}$ -ϕ
- $N_{\rm b-jet}$: determined by accounting for the fact that $\tau_{\rm h}$-jet has significant probability of getting b-mistagged. More details of this strategy are discussed in subsection 6.1.3.

These recalculated search variables are used to assign a search bin associated to simulated $\tau_{\rm h}$-jet event.

6.1.3 Efficiencies associated with muon control sample

The simulated τ_h +jet events need to be corrected for inefficiencies associated with muon identification, purity of CR and for residual differences in topology of muon + jet and τ_h + jet event. These efficiencies are determined from MC as discussed below.

Lepton acceptance

Muon CR is selected based on condition that muon should have $p_T > 20$ GeV and $|\eta|$ <2.1. Requirement on muon η is based on tracker coverage being upto $|\eta|$ <2.5 and jet cone radius of ΔR =0.4.

Along with $|\eta|$ requirement, $p_T > 20$ GeV requirement comes from the fact that at very low p_T , momentum measurement in forward direction becomes challenging due to the dense environment about lepton track. To avoid these criticalities, muon CR is selected based on kinematic conditions on muon as mentioned above (called "muon acceptance"). To correct muon CR yield for efficiencies associated to these kinematic cuts, muon CR event is scaled by reciprocal of "muon acceptance" on event by event basis. If μ_{all} represents total single muon events passing baseline cuts and falling in [N_{jet} , H_T , H_T^{miss}] bin and μ_{pass} represents muon CR events passing baseline and $p_T > 20$GeV and $|\eta|$ <2.1 cuts and fall into [N_{jet} , H_T , H_T^{miss}] , then mean "Muon acceptance" parameterized in terms of [N_{jet} , H_T , H_T^{miss}] is as given below:

$$\epsilon^{\mu}_{Acc}[N_{jet}, H_T, H_T^{miss}] = \frac{\mu_{pass}[N_{jet}, H_T, H_T^{miss}]}{\mu_{all}[N_{jet}, H_T, H_T^{miss}]} \tag{6.3}$$

Figure 6.6 shows how muon acceptance looks in high $\Delta\Phi$ region. Towards high jet multiplicity, $t\bar{t}$ contribution dominates over W+jets. As a result, leptons coming from W decay are more boosted and hence fall in central eta region of detector, which lead to improvement in lepton kinematic acceptance.

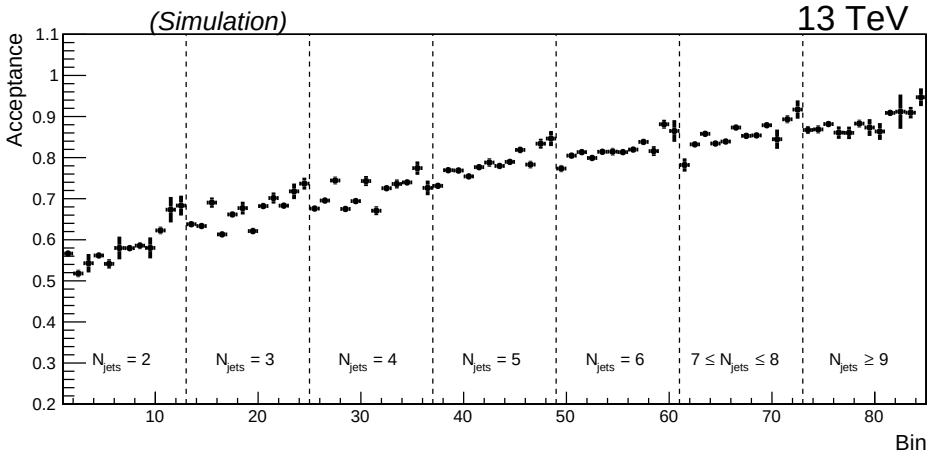

FIGURE 6.6: Lepton acceptance in 84 bins of [N_{jet} , H_T , H_T^{miss}] in high $\Delta\Phi$ region. The [H_T , H_T^{miss}] bins for each N_{jet} are similar to search region [H_T , H_T^{miss}] binning except two H_T bins corresponding to highest H_T^{miss} bin are combined

The N_{b-jet} dependence of muon acceptance is also considered on top of [N_{jet} , H_T , H_T^{miss}] dependence. If μ_{b-all} represents total single muon events passing baseline and falling in [N_{jet} , N_{b-jet} , H_T , H_T^{miss}] bin and μ_{b-pass} represents muon CR

events passing baseline and also p_{T} >20 GeV and $|\eta|$ <2.1 cuts and fall into [N_{jet} , $N_{\mathrm{b-jet}}$, H_{T} , $H_{\mathrm{T}}^{\mathrm{miss}}$] , $N_{\mathrm{b-jet}}$ dependence of muon acceptance becomes:

$$\epsilon^{\mu}_{\mathrm{Acc}}[N_{\mathrm{b-jet}}] = \frac{\mu_{b-pass}[N_{\mathrm{jet}}, N_{\mathrm{b-jet}}, H_{\mathrm{T}}, H_{\mathrm{T}}^{\mathrm{miss}}]}{\mu_{pass}[N_{\mathrm{jet}}, H_{\mathrm{T}}, H_{\mathrm{T}}^{\mathrm{miss}}]} \times \frac{\mu_{all}[N_{\mathrm{jet}}, H_{\mathrm{T}}, H_{\mathrm{T}}^{\mathrm{miss}}]}{\mu_{b-all}[N_{\mathrm{jet}}, N_{\mathrm{b-jet}}, H_{\mathrm{T}}, H_{\mathrm{T}}^{\mathrm{miss}}]} \quad (6.4)$$

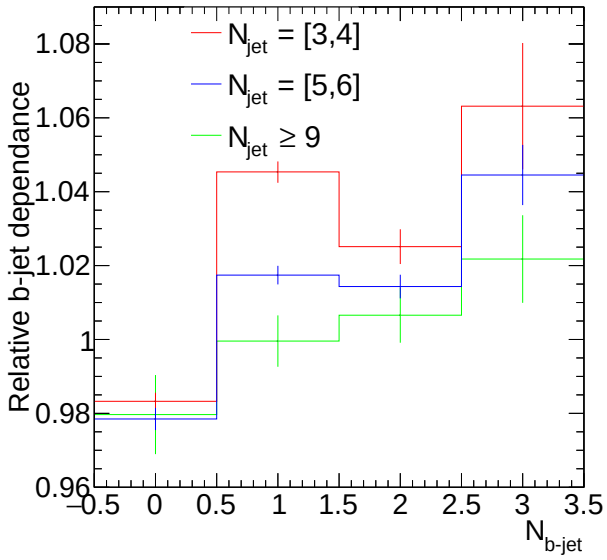

FIGURE 6.7: The $N_{\mathrm{b-jet}}$ dependence with respect to central lepton acceptance for N_{jet} bins [3,4], [5,6] and ≥ 9

The $N_{\mathrm{b-jet}}$ dependence observed for various N_{jet} bins is shown in fig. 6.7.

Isotrack veto efficiency

One prong and three prong decay of τ_{h} have characteristic charged track coming from a charged pion. Hence to suppress τ_{h} background in signal region, the events with isolated track are rejected in signal region. Similar to lepton acceptance, isotrack veto efficiency is first derived in bins of search variables N_{jet} , H_{T} and $H_{\mathrm{T}}^{\mathrm{miss}}$. If τ_{h-all} represents number of events having τ_{h} at generator level and $\tau_{h-noTrack}$ represents events having τ_{h} at generator level but no isolated track at reconstruction level, then isolated track veto efficiency as function of [N_{jet} , H_{T} , $H_{\mathrm{T}}^{\mathrm{miss}}$] becomes:

$$\epsilon_{\mathrm{isotrk}}[N_{\mathrm{jet}}, H_{\mathrm{T}}, H_{\mathrm{T}}^{\mathrm{miss}}] = \frac{\tau_{h-noTrack}[N_{\mathrm{jet}}, H_{\mathrm{T}}, H_{\mathrm{T}}^{\mathrm{miss}}]}{\tau_{h-all}[N_{\mathrm{jet}}, H_{\mathrm{T}}, H_{\mathrm{T}}^{\mathrm{miss}}]} \quad (6.5)$$

Figure 6.8 shows isotrack veto efficiency as a function of [N_{jet} , H_{T}, $H_{\mathrm{T}}^{\mathrm{miss}}$]. Towards high jet multiplicity, due to surrounding activity, it becomes difficult to find isolated track and hence efficiency drops down.

Similar to lepton acceptance, $N_{\mathrm{b-jet}}$ dependence on top of mean efficiency is determined. If $\tau_{h-b-all}$ represents number of events having τ_{h} at generator level and falling in [N_{jet} , $N_{\mathrm{b-jet}}$, H_{T} , $H_{\mathrm{T}}^{\mathrm{miss}}$] bin and $\tau_{h-b-noTrack}$ represents events having τ_{h} at generator level but no isolated track at reconstruction level and fall into [N_{jet} , $N_{\mathrm{b-jet}}$, H_{T} , $H_{\mathrm{T}}^{\mathrm{miss}}$] , $N_{\mathrm{b-jet}}$ dependence of isotrack veto efficiency becomes:

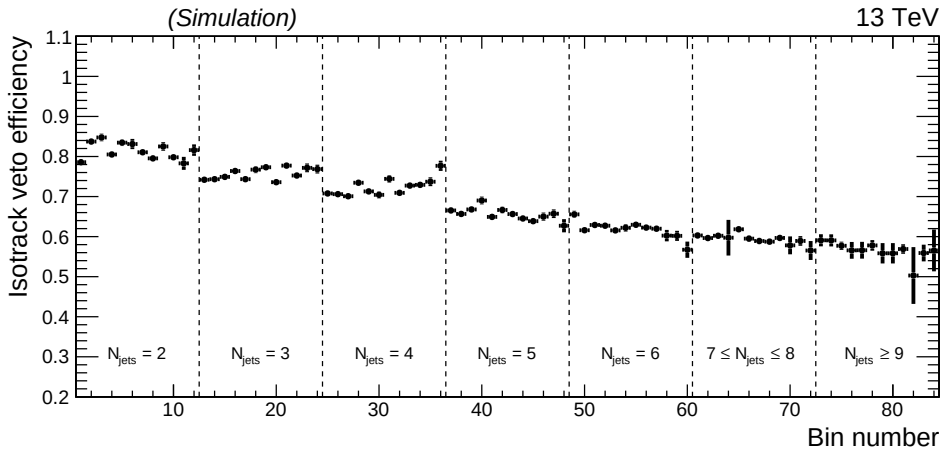

FIGURE 6.8: Isotrack veto efficiency in 84 bins of [$N_{\rm jet}$, $H_{\rm T}$, $H_{\rm T}^{\rm miss}$] in high $\Delta\Phi$ region. [$H_{\rm T}$, $H_{\rm T}^{\rm miss}$] bins are as in Fig. 6.6

$$\epsilon_{\rm isotrk}[N_{\rm b-jet}] = \frac{\tau_{h-b-noTrack}[N_{\rm jet}, N_{\rm b-jet}, H_{\rm T}, H_{\rm T}^{\rm miss}]}{\tau_{h-noTrack}[N_{\rm jet}, H_{\rm T}, H_{\rm T}^{\rm miss}]} \times \frac{\tau_{h-all}[N_{\rm jet}, H_{\rm T}, H_{\rm T}^{\rm miss}]}{\tau_{h-b-all}[N_{\rm jet}, N_{\rm b-jet}, H_{\rm T}, H_{\rm T}^{\rm miss}]} \tag{6.6}$$

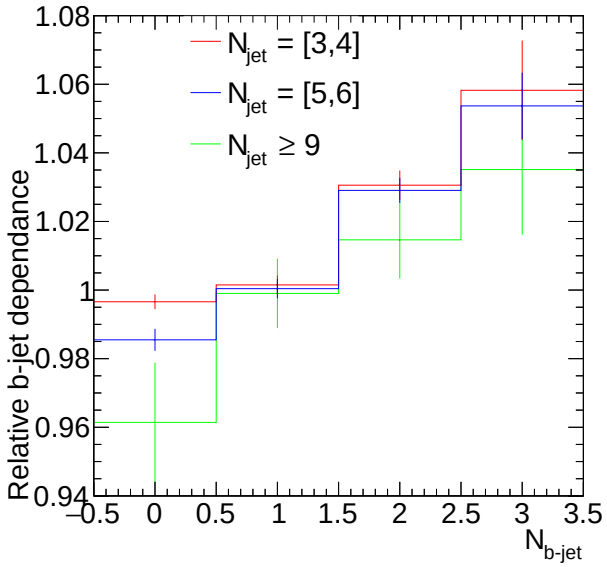

FIGURE 6.9: The $N_{\rm b-jet}$ dependence with respect to central Isotrack veto efficiency for $N_{\rm jet}$ bins [3,4], [5,6] and ≥ 9

The dependance of isolated track veto efficiency on $N_{\rm b-jet}$in different $N_{\rm jet}$bins is shown in fig. 6.9. Variation against central isolated track veto efficiency is maximum for highest $N_{\rm b-jet}$.

$\tau \to \mu$ probability

The $\tau \to \mu$ probability is used to correct for contamination in muon CR arising from τ decay to muon. The muons in the CR are expected to be coming directly from W decay for treating them equivalent to gen-τ. But, CR can have contamination from taus coming from W, which decay leptonically to muon. In order to correct for this contamination, $\tau \to \mu$ probability is determined as a function of [$N_{\rm jet}$, $H_{\rm T}$, $H_{\rm T}^{\rm miss}$] bins. If μ_{all} represents total muon CR events passing baseline cuts and falling in [$N_{\rm jet}$, $H_{\rm T}$, $H_{\rm T}^{\rm miss}$] bin and μ_{no-gen} represents muon CR events passing baseline, falling in [$N_{\rm jet}$, $H_{\rm T}$, $H_{\rm T}^{\rm miss}$] bin but either there is no matching generator level muon found or muon is identified as coming from tau decay, then $\tau \to \mu$ probability becomes:

$$\mathrm{f}_{\tau\to\mu}[\mathrm{N_{jet}}, \mathrm{H_T}, \mathrm{H_T^{miss}}] = \frac{\mu_{\rm no-gen}[\mathrm{N_{jet}}, \mathrm{H_T}, \mathrm{H_T^{miss}}]}{\mu_{\rm all}[\mathrm{N_{jet}}, \mathrm{H_T}, \mathrm{H_T^{miss}}]} \tag{6.7}$$

Figure 6.10 shows $\tau \to \mu$ probability as a function of [$N_{\rm jet}$, $H_{\rm T}$, $H_{\rm T}^{\rm miss}$]. The probability is found to be almost flat and hence no $N_{\rm b-jet}$ dependence is derived for this component.

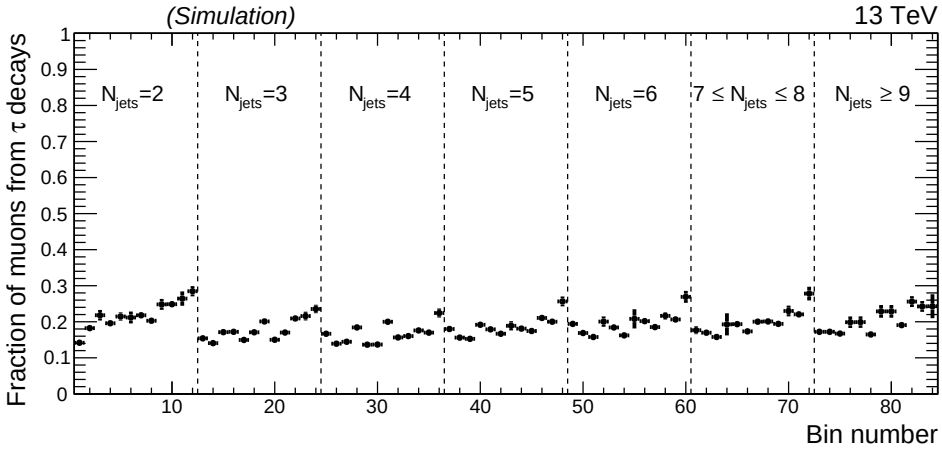

FIGURE 6.10: Probablity of muons coming from taus in 84 bins of [$N_{\rm jet}$, $H_{\rm T}$, $H_{\rm T}^{\rm miss}$] in high $\Delta\Phi$ region. [$H_{\rm T}$, $H_{\rm T}^{\rm miss}$] bins are as in Fig. 6.6

$m_{\rm T}$ efficiency

The selection of muon CR is made with additional requirement on its transverse mass as given by equation 6.8, where tk represents the particle trajectory and $\Delta\phi$ represents the angle between the trajectory and $p_{\rm T}^{\rm miss}$. If muon has originated from SM W+jet process, where neutrino is the only source of $p_{\rm T}^{\rm miss}$, and $m_{\rm T}$ distribution has sharp cut at W boson mass. If SUSY exists and muon is originated from any such SUSY event, then along with neutrino, $p_{\rm T}^{\rm miss}$ will have additional contribution from the LSP in the event, which will remain undetected in the event leaving extra $p_{\rm T}^{\rm miss}$. In those cases $m_{\rm T}$ will not be bounded by W mass and will not satisfy equation 6.8. Thus the condition $m_{\rm T} < 100$ GeV cut helps to reduce CR contamination due to non SM processes.

$$m_{\rm T}(tk, p_{\rm T}^{\rm miss}) = \sqrt{2p_{\rm T}^{\rm tk}p_{\rm T}^{\rm miss}(1-\cos(\Delta\phi))} < 100GeV \tag{6.8}$$

To correct for the reduction in muon CR yield due to m_{T}, efficiency of selection is determined as a function of [N_{jet} , H_{T} , $H_{\mathrm{T}}^{\mathrm{miss}}$] bins. If μ_{all} represents total single muon events passing baseline cuts and falling in [N_{jet} , H_{T} , $H_{\mathrm{T}}^{\mathrm{miss}}$] bin and μ_{pass-m_T} represents muon CR events, falling in [N_{jet} , H_{T} , $H_{\mathrm{T}}^{\mathrm{miss}}$] which actually satisfy m_{T} requirement given by equation 6.8, then m_{T} efficiency becomes:

$$\epsilon^{\mu}_{\mathrm{m_T}}[N_{\mathrm{jet}}, H_{\mathrm{T}}, H_{\mathrm{T}}^{\mathrm{miss}}] = \frac{\mu_{pass-m_T}[N_{\mathrm{jet}}, H_{\mathrm{T}}, H_{\mathrm{T}}^{\mathrm{miss}}]}{\mu_{all}[N_{\mathrm{jet}}, H_{\mathrm{T}}, H_{\mathrm{T}}^{\mathrm{miss}}]} \tag{6.9}$$

Figure 6.11 shows m_{T} efficiency in 84 bins of [N_{jet} , H_{T} , $H_{\mathrm{T}}^{\mathrm{miss}}$].

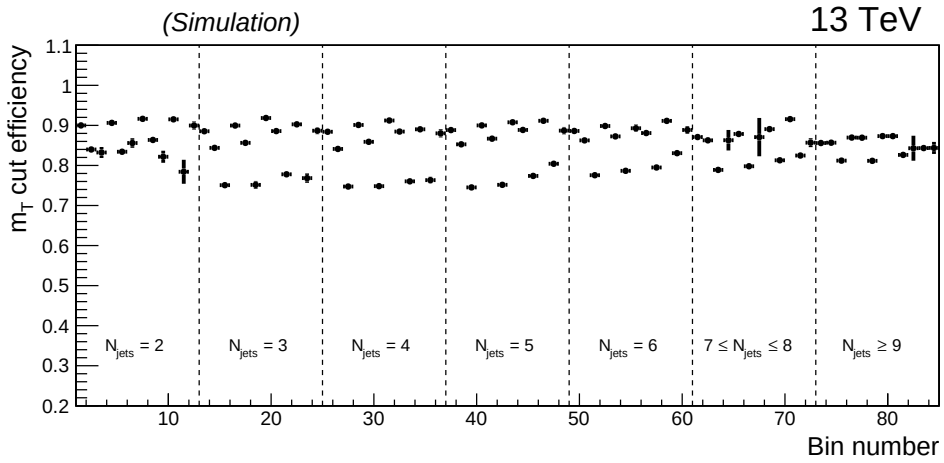

FIGURE 6.11: m_{T} efficiency in 84 bins of [N_{jet} , H_{T} , $H_{\mathrm{T}}^{\mathrm{miss}}$] in high $\Delta\Phi$ region. [H_{T} , $H_{\mathrm{T}}^{\mathrm{miss}}$] bins are as in Fig. 6.6

Figure 6.11 shows that m_{T} efficiency is almost flat and does not really show up the true feature that m_{T} efficiency is proportional to p_{T} of a lepton. Figure 6.12 shows m_{T} efficiency as a function of p_{T} and η of a muon in CR. As expected, there is no η dependence, but towards high p_{T} , m_{T} efficiency reduces.

As m_{T} efficiency is p_{T} dependent, p_{T} spectrum of muon from CR and gen-τ are compared before and after imposing m_{T} cut.

Figure 6.13 shows that after m_{T} cut, reco-μ and gen-τ p_{T} spectra no longer remain equivalent. The impact of m_{T} cut on reco-μ p_{T} spectrum is checked in various $N_{\mathrm{b-jet}}$ bins as shown in figure 6.14. To account for this residual m_{T} cut impact, it was decided to reweight muon p_{T} spectrum in CR to make it equivalent to gen-τ. If ratio of reco-μ p_{T} to gen-τ p_{T} as a function of $N_{\mathrm{b-jet}}$ represents "p_{T} reweighting factor" then, corrected τ_{h}-jet response after accounting for "p_{T} reweighting factor" for given bootstrapping iteration becomes,

$$\text{corrected } \tau_{\mathrm{h}} \text{ response} = \frac{p_{\mathrm{T}} \text{ of } \tau \text{ jet}}{p_{\mathrm{T}} \text{ of gen } \tau} = \frac{p_{\mathrm{T}} \text{ of } \tau \text{ jet}}{p_{\mathrm{T}} \text{ of reco } \mu} \times \frac{p_{\mathrm{T}} \text{ of reco } \mu}{p_{\mathrm{T}} \text{ of gen } \tau} \tag{6.10}$$

$$\text{corrected } \tau_{\mathrm{h}} \text{ response} = \tau_{\mathrm{h}} \text{ response} \times p_{\mathrm{T}} \text{ reweighting factor} \tag{6.11}$$

This recipe to account for differences in p_{T} spectra of gen-τ and reco-μ in $N_{\mathrm{b-jet}}$ bins and then correcting for m_{T} efficiency as a function of [p_{T} , η] showed better performance over using average m_{T} efficiency as a function of [N_{jet} , H_{T} , $H_{\mathrm{T}}^{\mathrm{miss}}$]. Results will be discussed in section 6.1.4.

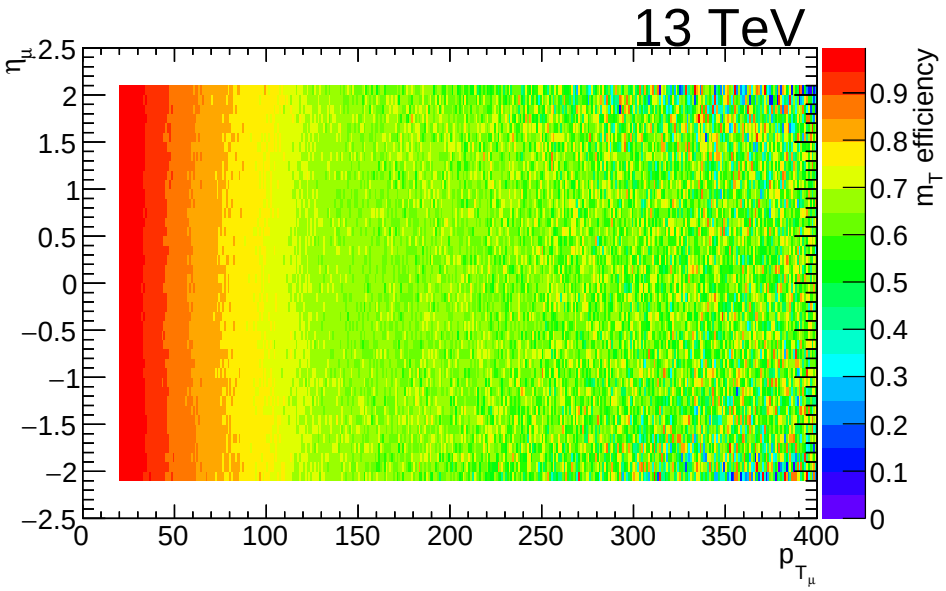

FIGURE 6.12: m_{T} efficiency as a function of p_{T} , η of muon in CR

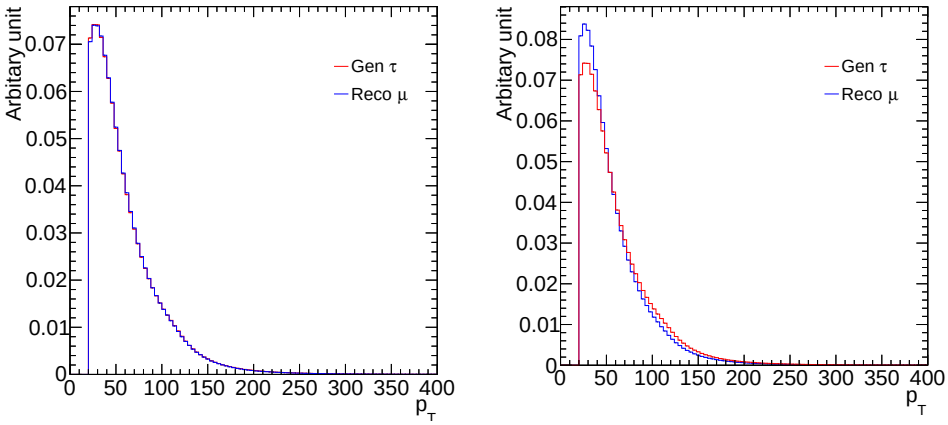

FIGURE 6.13: Comparison of p_{T} of gen-τ and reco-μ before and after m_{T} cut

Accounting for b-mistag rate of a τ_{h}-jet

Tau lepton has a life time of about 10^{-13}s in its rest frame before it decays to a jet. On the other hand B hadrons (hadrons with b quark content) also have a life time of the order 10^{-12}s before decaying to light quark jets (called b-jet) due to a suppressed couplings of bottom quark to up or charm quarks. As a result b-jets have a characteristic displaced vertex. As tau lepton lifetime is of similar order as for B hadrons, there is a significant probability that τ_{h}-jet is mistagged as a b-jet. If this mistagged jet satisfies the kinematic jet requirements, an event will be wrongly placed in higher $N_{\mathrm{b-jet}}$ search bin. To account for this effect, b-mistag probability of τ_{h}-jet is determined, and used to correct the final τ_{h} + jet prediction.

The b-mistag probability of τ_{h}-jet is determined as a function of gen-τ p_{T} . A fraction of events with gen-τ, where a matched reconstructed jet to gen-τ exceeds

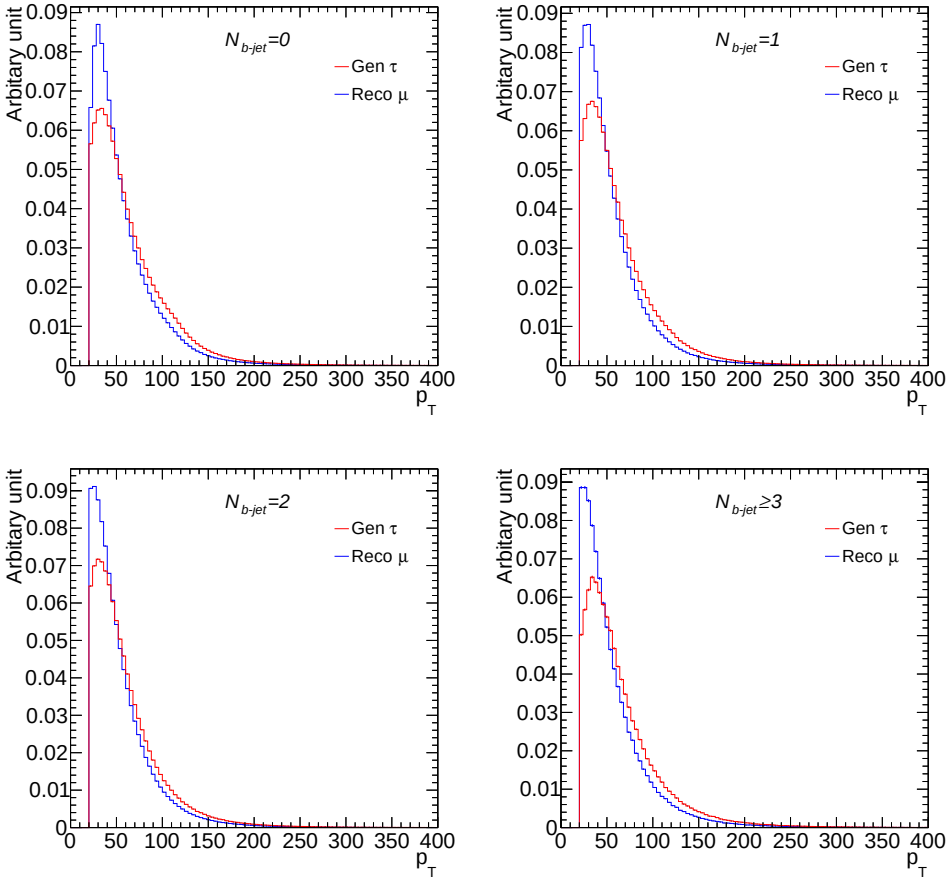

FIGURE 6.14: Comparison of p_{T} of gen-τ and reco-μ after m_{T} cut in $N_{\mathrm{b-jet}}$ bins 0, 1 ,2 and ≥ 3 respectively

the csv threshold for b-tagging (csv >0.8) defines b-mistag probability. Figure 6.15 gives the b-mistag probabilities obtained for a matching jet to gen-τ and reco-μ in different samples. It is observed that in $t\bar{t}$ + jet events mistag rate increases with increasing p_{T} due to the increasing proximity of real b-jet in the event. In the W+jet event, b-mistag rate is almost flat both for muon jet and τ_{h}-jet. We used the b-mistag rate of τ_{h}-jet from W+jet sample as an intrinsic b-mistag rate of τ_{h}-jet.

As b-mistag rate is nonzero across p_{T} range, for every simulated τ_{h}-jet event there are two possibilities; either τ_{h}-jet is b-mistagged or not and depending on that the simulated event can fall in different $N_{\mathrm{b-jet}}$ bins. To account for this, two iterations ("k") are considered for every simulated τ_{h}-jet event.

If muon jet from CR event is already b-tagged based on the csv value, then it is highly unlikely that simulated τ_{h}-jet will not be b-tagged, as Fig. 6.15 shows that b-mistag rate of τ_{h}-jet is always higher than muon jet. Hence simulated event with new values of N_{jet} , H_{T} and $H_{\mathrm{T}}^{\mathrm{miss}}$ in such case will belong to same $N_{\mathrm{b-jet}}$ bin as CR event during both iterations making total probability of that $N_{\mathrm{b-jet}}$ to be one. But if muon jet is not b-tagged, then in first iteration, simulated τ_{h}-jet event is filled in $N_{\mathrm{b-jet}}$ bin of CR event with eventweight scaled by (1-b-mistag rate). For second iteration, simulated τ_{h}-jet event is filled in $N_{\mathrm{b-jet}}$ +1 bin with eventweight scaled by b-mistag rate. This way, probability of two $N_{\mathrm{b-jet}}$ bins associated to simulated τ_{h}-jet event sum up to one.

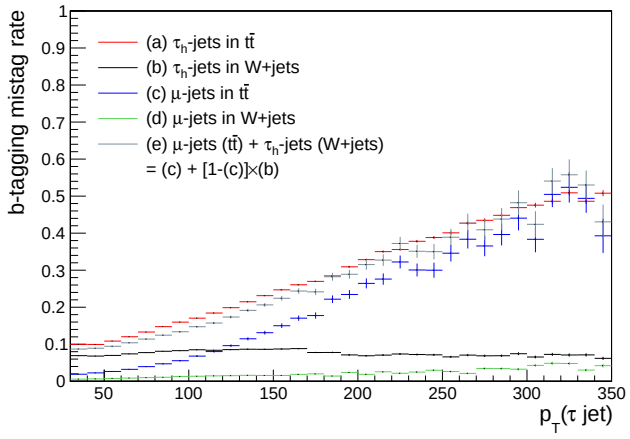

FIGURE 6.15: b-mistag probability of muon and τ_h-jet as a function of p_T in W+jet and TTbar+jet MC samples

6.1.4 Validation of method in MC event sample

With all efficiency corrections applied, τ_h background estimate can be given as:

$$N_{\tau_h} = \sum_i^{N^{\mu}_{CS}} \left(\sum_j^{\text{Template bins}} \left(p^{\text{resp}}_{\tau_h} \sum_k w^{\tau_h}_{b-\text{mistag}} \right) \frac{1}{\epsilon^{\mu}_{\text{Reco}} \epsilon^{\mu}_{\text{ISO}} \epsilon^{\mu}_{\text{Acc}} \epsilon^{\mu}_{m_T}} (1\text{-}f_{\tau\to\mu})(1\text{-}f_{ll}) \frac{\text{Br}(W\to\tau_h\nu)}{\text{Br}(W\to\mu\nu)} \epsilon_{\text{isotrk}} \right) \quad (6.12)$$

Figure 6.16 and 6.17 show the comparison of the predicted background from single muon control sample against expected τ_h background in high $\Delta\Phi$ search region and low $\Delta\Phi$ region including QCD sidebands using MC after scaling to 35.9 fb^{-1}. The expected τ_h background in search (low $\Delta\Phi$) region is obtained by selecting events with only one τ_h at generator level which satisfy baseline selection under high (low) $\Delta\Phi$ conditions and a jet is found matched to the gen-τ. The predicted yield from muon CR has used W-gun templates, average m_T efficiency against [N_{jet} , H_T , H_T^{miss}] and average [N_{jet} , N_{b-jet}] corrections developed to correct for residual nonclosure seen against [N_{jet} , N_{b-jet}]. The error bars on both expectation and prediction are statistical only. The results agree within statistical uncertainty in all search bins for both high and low $\Delta\Phi$ regions. This validation is referred as a closure test since the various efficiencies derived from MC are applied back on MC CR itself to check consistency of prediction against expectation.

The remaining lack of closure is corrected using residual correction factors. The average residual [N_{jet} , N_{b-jet}] corrections are derived as a ratio of MC expectation to MC prediction by integrating yields in corresponding H_T -H_T^{miss} bins. Figure 6.18 and 6.19 (left) show these correction factors while using a template derived from $t\bar{t}$ +jet and W+jet MC. The average corrections found for closure while using W-gun template but m_T efficiency versus [N_{jet} , H_T , H_T^{miss}] and W-gun template and m_T efficiency versus [p_T , η] after reweighting muon p_T spectrum for prediction are shown in figures 6.18 and 6.19 (right) respectively. The gradual reduction in [N_{jet} , N_{b-jet}] average corrections imply the improvement in closure after these changes.

Closures are also checked for inclusive 1D distribution in different search variables as shown in figure 6.20. These closures again have used W-gun templates, average m_T efficiency as function of [N_{jet} , H_T , H_T^{miss}] and are after applying average [N_{jet} , N_{b-jet}] corrections developed to correct for residual nonclosure seen against [N_{jet} , N_{b-jet}].

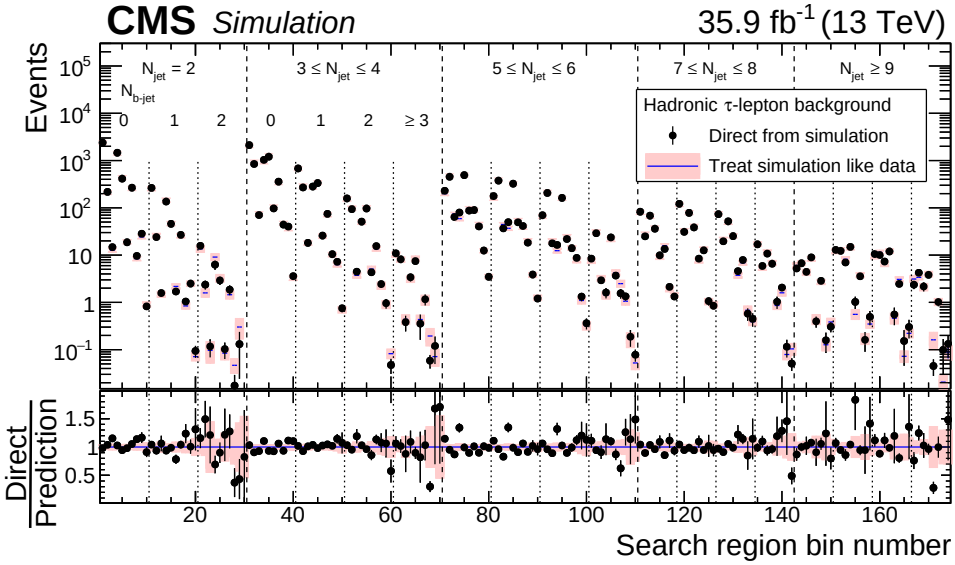

FIGURE 6.16: The comparison of MC τ_h background expected in 174 search region (black solid circles) against the predicted τ_h background in search region using single muon control sample (pink shaded region).

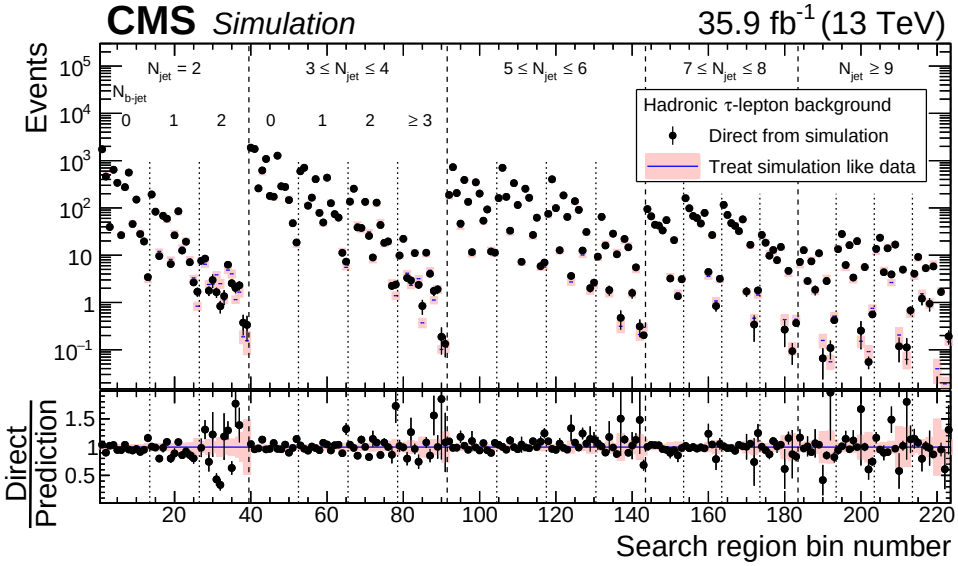

FIGURE 6.17: The comparison of MC τ_h background expected in 174 search region (black solid circles) against the predicted τ_h background in low $\Delta\Phi$ region using single muon control sample (pink shaded region).

6.1.5 Estimation of τ_h background in 2016 dataset

After validation, the same estimation strategy is applied on single muon control region (CR) from data to get τ_h prediction. Prediction is corrected for various muon CR efficiencies as discussed before. The NjNb residual correction factors obtained from MC closure as shown in figure 6.18 (left) are also applied on top to data prediction. The only correction factors which are additionally applied to data prediction but not

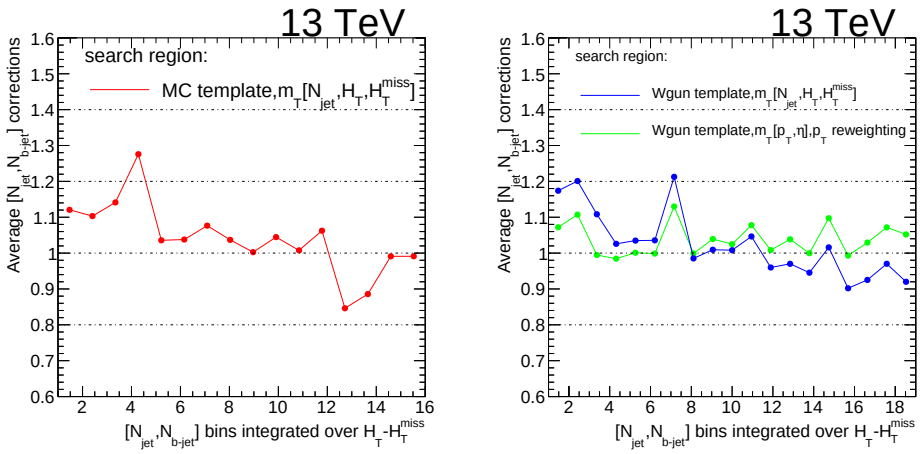

FIGURE 6.18: Average [N_{jet} , $N_{\mathrm{b-jet}}$] correction factors in (left) using $t\bar{t}$ +jet and W+jet templates, (right) using W-gun template and m_{T} efficiency versus [N_{jet} , H_{T} , $H_{\mathrm{T}}^{\mathrm{miss}}$] and using W-gun template and m_{T} efficiency versus [p_{T}, η] of muon from CR after residual correction for p_{T} differences in high $\Delta\Phi$ search region.

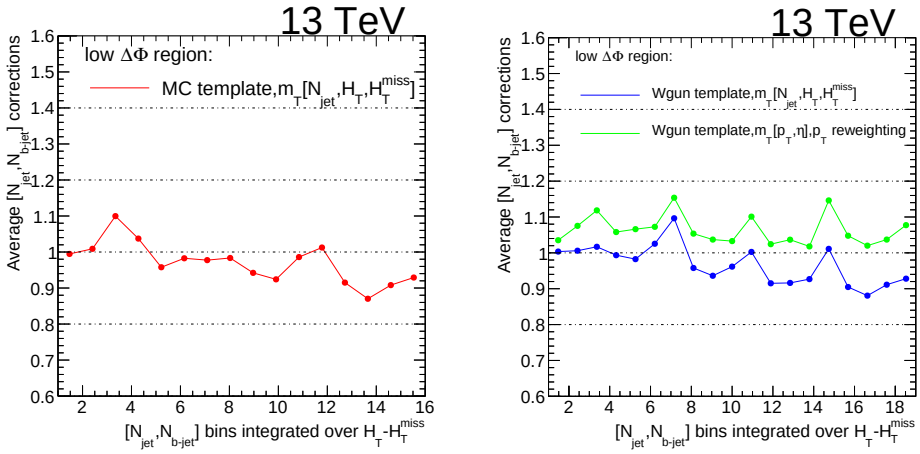

FIGURE 6.19: Average [N_{jet} , $N_{\mathrm{b-jet}}$] correction factors in (left) using $t\bar{t}$ +jet and W+jet templates, (right) using W-gun template and m_{T} efficiency versus [N_{jet} , H_{T} , $H_{\mathrm{T}}^{\mathrm{miss}}$] and using W-gun template and m_{T} efficiency versus [p_{T}, η] of muon from CR after residual correction for p_{T} differences in low $\Delta\Phi$ region.

to MC prediction comes from various trigger efficiencies involved. As CR is selected using single muon triggers, while search region uses MET triggers, background prediction from data CR is first scaled up by reciprocal of single muon trigger efficiency and then scaled down by MET trigger efficiency factor to get final prediction.

In addition, for low H_{T} region to remain above trigger turn on, the muon p_{T} threshold was moved to 24 GeV from 20 GeV. The effect of this change on prediction is estimated in MC. The ratio of MC prediction with muon p_{T} threshold of 24 GeV and that with 20 GeV, obtained bin by bin is applied as a correction factor to data prediction in low H_{T} region.

Figure 6.21 and 6.22 give the final data prediction obtained in signal region after applying all the efficiency correction factors in 174 search bins and in 1D search

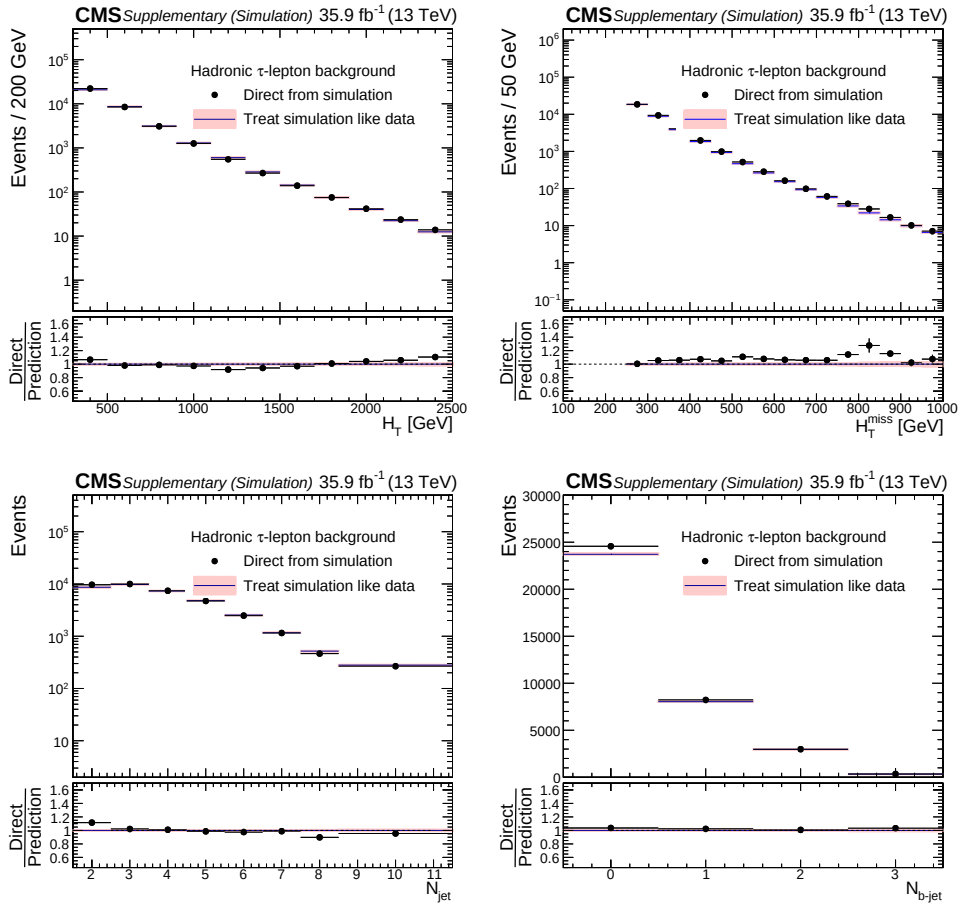

FIGURE 6.20: The comparison of MC τ_h background expected in search variables H_T, H_T^{miss}, N_{jet} and N_{b-jet}(black solid circles) against the predicted τ_h background using single muon control sample (pink shaded region).

variable projections respectively. Above comparison of data prediction against MC expectation is for a trivial sanity check of method but not used explicitly in any of final results.

Systematics associated to τ_h estimate

Major systematics considered associated with $lost^{\tau_h}$ background estimate using template method are as described below. They are used as nuisance parameters for final statistical interpretations.

- JEC systematics:

  The τ_h template is rederived by varying jet p_T by JEC uncertainties. MC Prediction obtained with these new templates is compared against MC prediction obtained without considering any uncertainty. The fractional change in prediction per bin is used as JEC systematics.

- Lepton SF systematics:

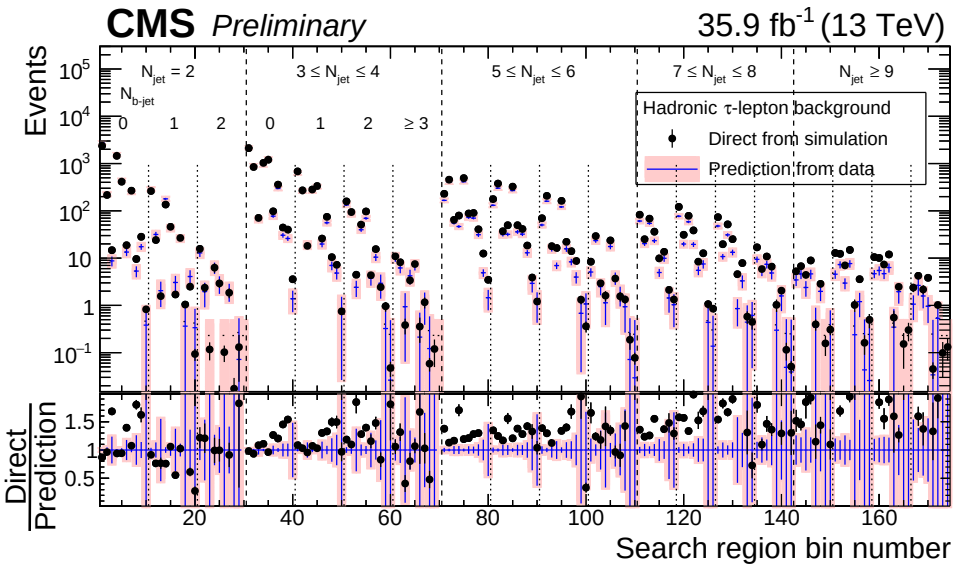

FIGURE 6.21: The τ_h prediction in data (pink shaded region) against τ_h background expected from MC (solid dots).

Data/MC SF associated to muon identification, isolation and reconstruction obtained from SUS lepton POG are varied by associated uncertainty and prediction is reobtained. The change in prediction is assigned as systematics.

- Btag systematics:

  b-mistag probability of τ_h-jet is varied by 50% and prediction is obtained again. Fractional change in prediction due to this, is considered as systematics.

- Acceptance systematics

  Acceptance systematics is the variation in lepton acceptance due to change in Parton Density Function (PDF) or Renormalization Scale used in MC simulation within uncertainty [95]. To get PDF systematics, eventweight is scaled independently by a set of 100 variations in PDF weights, to get 100 variations in lepton acceptance. The maximum variation in lepton acceptance with respect to nominal acceptance is considered as a PDF systematics.

  In a similar way, to get Scale systematics, eventweight is scaled independently by a set of 9 variations in Renormalization Scale weights, to get 9 variations in lepton acceptance. The maximum variation in lepton acceptance with respect to nominal acceptance is considered as a Scale systematics.

Final data prediction in 174 search bins and associated systematics are used as inputs to get statistical interpretations based on final comparison of data against all SM background predictions. These results have been published in [19]. The interpretation of results in the context of strong SUSY production are shown in appendix C.

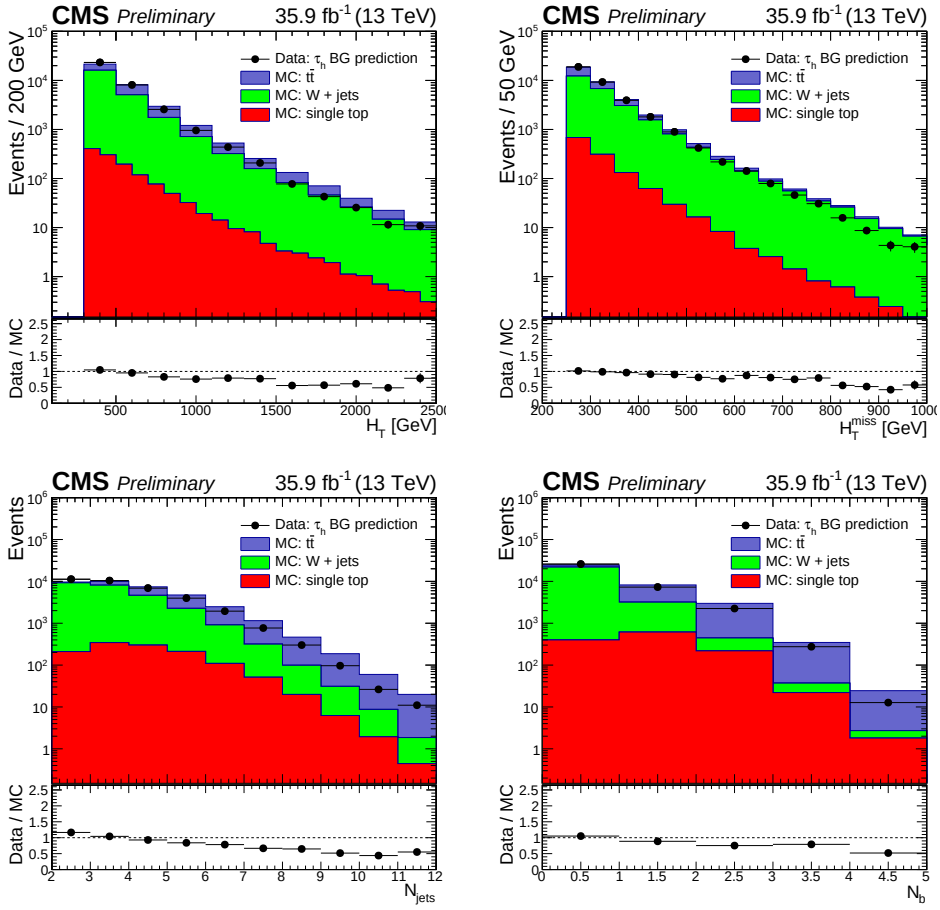

FIGURE 6.22: The τ_h prediction in data (solid dots) against τ_h background expected from MC (stacked histograms) in H_T ,H_T^{miss} , N_{jet} and N_{b-jet} bins.

6.2 Estimation of $lost^{e,\mu}$ +$lost^{\tau_h}$ background using average Transfer Factor (TF) method

In 2016 analysis, $lost^{\tau_h}$ prediction was derived using template method. The method relied on analogy of τ+jet event to μ+jet event at generator level and then smearing of μ+jet event based on probability distribution of fractional p_T carried by τ jet from generator τ. This approach although robust, did invite extra complexities like correcting single muon control sample on event by event basis for efficiency associated to detector acceptance of muon, $\tau \rightarrow \mu$ contamination, isotrack veto, mT cut and dilepton contamination. Also background strategy was computation intensive (at least 50 and at most 100 iterations for each control region (CR) event).

As at the time of start of this analysis, when it was an exploring stage of how detector behaves, how data looks in comparison to simulation, it was decided to choose a conservative event by event approach. This allowed to extract kinematic shape of each estimated background event as a function of each CR event. For this reason, template method was implemented during 2016 analysis.

But now with time as understanding of detector has become comprehensive and have gained enough confidence on how data looks in comparison with MC and

undersanding of how to deal with residual data versus MC differences, it was possible to think of a simpler approach like average TF method for prediction of $lost^{e,\mu}$ +$lost^{\tau_h}$ event from one lepton CR event.

The combined prediction of $lost^{e,\mu}$ +$lost^{\tau_h}$ is expected to reduce systematics significantly helping constraining the limits on SUSY scenarios. Figure 6.23 shows $lost^{e,\mu}$ +$lost^{\tau_h}$ expected yield contribution from $t\bar{t}$ +jets, W+jets and single top MC events. The $t\bar{t}$ +jets and W+jets form the major background components. The relative fraction of $t\bar{t}$ +jets contribution increase towards high H_T , N_{jet} , N_{b-jet} multiplicity, while for H_T^{miss} , $t\bar{t}$ +jets contribution decreases relative to W+jets towards high values due to presence of two neutrinos in the $t\bar{t}$ event.

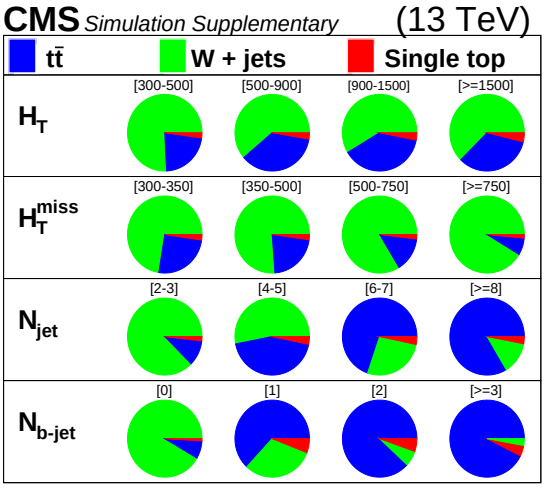

FIGURE 6.23: Combined $lost^{e,\mu}$ +$lost^{\tau_h}$ yield expected in zero lepton signal region as obtained from $t\bar{t}$ +jets, W+jets or single top MC (see table G.1 for numbers).

6.2.1 Average Transfer Factor (TF)

Transfer Factor will represent unique transition factor to be used to get signal region background prediction ($lost^{e,\mu}$ + $lost^{\tau_h}$) from control region yield (one lepton). In MC event sample, with help of generator level information available, it is possible to tag an event as either found lepton (lepton present both at generator and reconstructed level) or $lost^{e,\mu}$ or $lost^{\tau_h}$ event (lepton present at generator level but not at reconstructed level) and hence is straightforward to obtain ratio of $lost^{e,\mu}$ +$lost^{\tau_h}$ event yield to found lepton event yield representing Transfer Factor in MC. In data, such categorization of events is not possible as no generator level information available. Hence, the crucial point of this method becomes to correct Transfer Factor derived from MC for data versus MC differences in order to make Transfer Factor applicable on one lepton CR in data to derive $lost^{e,\mu}$ +$lost^{\tau_h}$ background prediction in data signal region.

In method opted here, Transfer Factor is derived as a function of 174 search bins. The word "Average" dictates the fact that Transfer Factor obtained for every bin represents average fraction of $lost^{e,\mu}$ +$lost^{\tau_h}$ to found lepton events in that bin and does not care for kinematics of lepton on event by event basis (as was done in template method). As a consequence of this averaging, it is impossible to extract the full kinematics of a CR or SR event.

6.2.2 Strategy

Average Transfer Factor implementation involves four main steps:

- step 1: Get Data/MC scale factor (SF) for one lepton CR as a function of search bins.
- step 2: Use SF derived in step 1 to get Data/ MC SF for zero lepton SR as a function of search bins.
- step 3: Use SFs derived in steps 1 and 2, and derive zero lepton to one lepton TF in data
- step 4: Apply TF derived in step 3 on one lepton CR from data to get $lost^{e,\mu}$ +$lost^{\tau_h}$ background prediction in data signal region.

In steps [1] and [2], product of POG recommended SUSY lepton SFs and Btag SFs are used to get final data/MC correction factor per event. Estimation relies on method "event reweighting by ignoring true b jets in the event".

6.2.3 Event reweighting by ignoring true b jets

A jet, reconstructed as a b jet, may or may not have associated generator level b jet. This efficiency of tagging a jet as b jet is different in data and SM MC. To account for the differences in efficiency, the b tag scale factors are provided, which are function of jet flavor, jet p_T and η.

The event reweighting method used here has the objective to correctly predict the number of b jets in data from MC, but by ignoring b tagged status of individual jets in MC simulated events [96]. In this method, for every reconstructed jet [j] in event, a non-zero b-tagging probability (b[j]) is assigned. Here b[j] is obtained by multiplying MC b-tagging efficiency by b-tag SF. Once b[j] is determined for all jets, combinatorial probability (P_b) of event being tagged as 0-btag, 1-btag or 2-btag is obtained. As an example, for an event with N_{jet} =3, probability to have 0-bjets and 1-bjets in event will be as given by equations 6.13 and 6.14.

$$P_0 = (1 - b[0]) \times (1 - b[1]) \times (1 - b[2]) \tag{6.13}$$

$$P_1 = b[0]\times(1-b[1])\times(1-b[2])+b[1]\times(1-b[0])\times(1-b[2])+b[2]\times(1-b[0])\times(1-b[1]) \tag{6.14}$$

and similarly for probability of higher b-tagged jets. The MC eventweight is scaled by P_b to assign probability to different b-tagged events. Thus same MC event contributes to different N_{b-jet} bins after accounting for data versus MC differences in such a way that their total probabilities add to one.

6.2.4 Scale factors for one lepton CR (SFCR)

Eventweight of every one lepton event (N^{MC}_{found}) falling within detector acceptance (p_T > 5 GeV, $|\eta|$<2.5) is scaled with Lepton Correction Factor (LCF) on event by event basis which is given by product of SFs associated with lepton identification, isolation and reconstruction as given by equation 6.15.

$$\text{Lepton Correction Factor (LCF)} = idSF \times isoSF \times recoSF \tag{6.15}$$

Here $idSF$, $isoSF$ and $recoSF$ are SUSY electron (muon) scale factors associated to electron (muon) identification, isolation and reconstruction given as a function of

its p_{T} and η. Using $t\bar{t}$, W+jets and single top MC samples and distributing the events in 174 search bins, represents corrected MC yield in one lepton CR ($N_{\mathrm{found}}^{MC-corrected}$) for data versus MC differences. Then average electron SF for CR becomes:

$$SFCR_{\mathrm{e}} = \frac{N_{\mathrm{found}}^{e,MC-corrected}}{N_{\mathrm{found}}^{e,MC}} \tag{6.16}$$

In similar manner, SFCR can be defined for one muon events as,

$$SFCR_{\mu} = \frac{N_{\mathrm{found}}^{\mu,MC-corrected}}{N_{\mathrm{found}}^{\mu,MC}} \tag{6.17}$$

6.2.5 Scale factors for zero lepton SR (SFSR)

To overcome the difficulty that in "lepton lost" case, it is not possible to extract associated SFs, the fact used is total of lepton lost (N_{lost}) and lepton (or track) found (N_{found}) events should be same in both data and MC and should match prompt one lepton events at generator level ($N_{\mathrm{prompt}}^{MC,truth}$).

$$N_{\mathrm{prompt}}^{MC,truth} = N_{\mathrm{found}}^{MC} + N_{\mathrm{lost}}^{MC} = N_{\mathrm{found}}^{data} + N_{\mathrm{lost}}^{data} \tag{6.18}$$

As data can be equivalently represented by corrected MC yield,

$$N_{\mathrm{prompt}}^{MC,truth} = N_{\mathrm{found}}^{MC} + N_{\mathrm{lost}}^{MC} = N_{\mathrm{found}}^{MC-corrected} + N_{\mathrm{lost}}^{MC-corrected} \tag{6.19}$$

$$N_{\mathrm{lost}}^{MC-corrected} = N_{\mathrm{prompt}}^{MC,truth} - N_{\mathrm{found}}^{MC-corrected} \tag{6.20}$$

Given this, average lepton SF in zero lepton signal region (SFSR), can be expressed as

$$SFSR = \frac{N_{\mathrm{lost}}^{MC-corrected}}{N_{\mathrm{lost}}^{MC}} = \frac{N_{\mathrm{prompt}}^{MC,truth} - N_{\mathrm{found}}^{MC-corrected}}{N_{\mathrm{lost}}^{MC}} \tag{6.21}$$

Here $N_{\mathrm{prompt}}^{MC,truth}$ represents MC event yield having an electron (muon) falling within detector acceptance ($p_{\mathrm{T}} > 5$ GeV, $|\eta|<2.5$) at truth level and N_{lost}^{MC} represents susbset of $N_{\mathrm{prompt}}^{MC,truth}$ where no matching prompt lepton is found in the event at reco level. The $N_{\mathrm{found}}^{MC-corrected}$ is a complementary set of events where there is either a matching lepton or track found at reco level.

- If a well identified and isolated lepton, matching to a gen lepton is found , the lepton correction factor (LCF) to get corrected MC yield ($N_{\mathrm{found}}^{MC-corrected}$) becomes:

$$\text{Lepton Correction Factor (LCF)} = trackSF \times isoSF \times recoSF \tag{6.22}$$

- If no well identified and isolated matching lepton is found but there is an isolated matching lepton track found, then lepton correction factor in such cases becomes:

$$\text{Lepton Correction Factor (LCF)} = trackSF \tag{6.23}$$

Here $trackSF$ represents SUSY scale factors associated to isolated tracks.

Similar to SFCR, SF for SR is obtained separately for electrons ($SFSR_e$) and muons ($SFSR_\mu$).

6.2.6 TF in bins of signal region using MC

Nominal $t\bar{t}$ +jets, W+jets and single top MC simulated samples are used here to derive TF in bins of signal region. In one lepton CR, to make sure that an electron or a muon selected is a prompt lepton coming from W decay, a reconstructed level lepton should match to generator level lepton and satisfy an additional requirement of $m_T < 100$ GeV. Here matching is defined as angle between generator level ($p_T > 5$ GeV and $|\eta| < 2.5$) and reconstructed level lepton should be within $\Delta R < 0.3$ and relative p_T difference between reco and gen lepton ($\Delta p_T/p_T$) should be less than 10%. One lepton events selected in this way are distributed across 174 search bins by following recommended method of "event reweighting by ignoring true b jets in the event"and they define one lepton MC yield in control region (N^{MC}_{found}). To get "$lost^{e,\mu}$ +$lost^{\tau_h}$" yield in signal region (N^{MC}_{lost}), from the events passing baseline selection, a set of events with no isolated lepton and no isolated track are distributed across search bins by following similar method. Once determined these, TF in MC in bins search region (i) is given as:

$$TF_i^{\text{MC}} = \frac{N^{MC}_{\text{lost,i}}}{N^{MC}_{\text{found,i}}} \tag{6.24}$$

To apply on data, TF_i^{MC} is corrected using $N^{MC-corrected}_{\text{lost}}$ and $N^{MC-corrected}_{\text{found}}$. To get $N^{MC-corrected}_{\text{found}}$ from N^{MC}_{found} following four cases are considered:

- If there is a reconstructed level electron and also a matching generator level electron ($p_T > 5$ GeV, $|\eta|$<2.5) in the event, use $SFCR_e$ [i] as correction factor
- If there is a reconstructed level muon and also a matching generator level muon ($p_T > 5$ GeV, $|\eta|$<2.5) in the event, use $SFCR_\mu$ [i] as correction factor
- If there is a reconstructed level electron (muon) but no generator level electron (muon) ($p_T > 5$ GeV, $|\eta|$<2.5) in event, assign correction factor to be one
- If there is a reconstructed level electron (muon) and also generator level electron (muon) ($p_T > 5$ GeV, $|\eta|$<2.5) in event, but it fails matching condition, assign correction factor to be one

To get $N^{MC-corrected}_{\text{lost}}$ from N^{MC}_{lost} following four cases are considered:

- If there is no reconstructed level electron but has generator level electron ($p_T > 5$ GeV, $|\eta|$<2.5) in the event, use $SFSR_e$ [i] as correction factor
- If there is no reconstructed level muon but has generator level muon ($p_T > 5$ GeV, $|\eta|$<2.5) in the event, use $SFSR_\mu$ [i] as correction factor
- If there is no reconstructed level electron (muon) and also no generator level electron (muon) ($p_T > 5$ GeV, $|\eta|$<2.5) in the event, assign correction factor to be one
- If there is no reconstructed level electron (muon) and also no generator level electron (muon) ($p_T > 5$ GeV, $|\eta|$<2.5) but there is generator level tau in the event, assign correction factor to be one

6.2.7 TF dependence on search variables

As a simpler way to look at TF, the variation with respect to [N_{jet} , $N_{\mathrm{b-jet}}$] is checked by integrating yields in CR (SR) across [H_{T} , $H_{\mathrm{T}}^{\mathrm{miss}}$] bins as shown in figure 6.24.

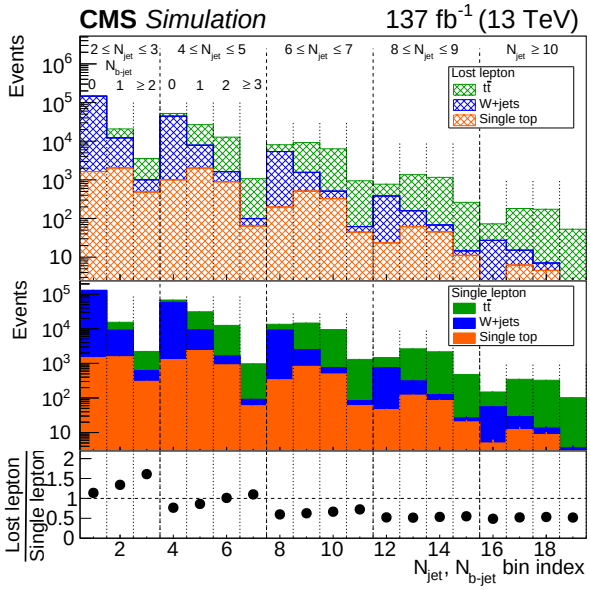

FIGURE 6.24: N_{jet} - $N_{\mathrm{b-jet}}$ dependence of TF from MC simulated events

To understand the observed trend in TF, similar distribution checked separately for $t\bar{t}$ +jets and W+jets by dividing zero lepton contribution in categories like lost due to failing acceptance or identification or isolation ($lost^{e,\mu}$) or $lost^{\tau_h}$ as shown in figure 6.25.

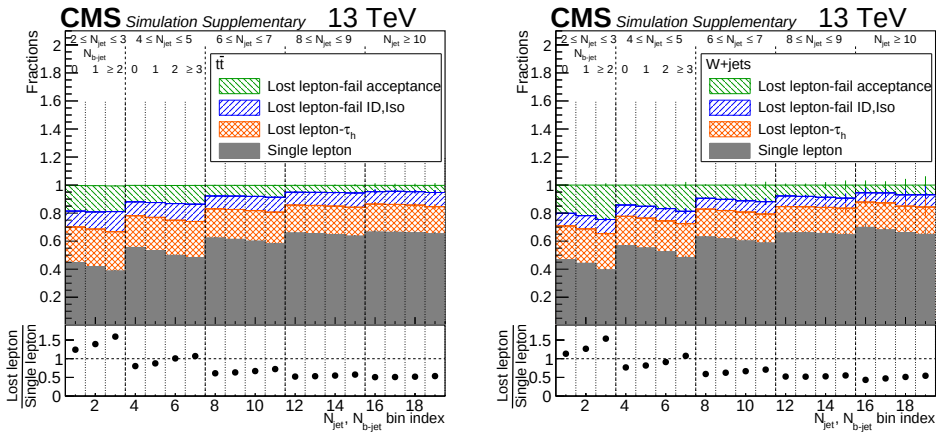

FIGURE 6.25: TF variation for (left) $t\bar{t}$ +jets and (right) W+jets from MC simulated events

Figure 6.25 reveals following facts:

- TF is independent of whether the process is $t\bar{t}$ +jets or W+jets

- Going from low to high jet multiplicity, $lost^{e,\mu}$ +$lost^{\tau_h}$ contribution due to failing acceptance has significantly reduced by about 10-15 %.

  This can be understood as while going towards high jet multiplicity, leptons coming from W decay are highly boosted and more central in η, which allow them to pass lepton acceptance requirement more often. This is the major reason causing fall in TF towards high jet multiplicity.

- At low jet multiplicity, all three components of zero lepton contribution show dependence on $N_{\mathrm{b-jet}}$

To understand $N_{\mathrm{b-jet}}$ dependence further, $lost^{e,\mu}$ +$lost^{\tau_h}$ contribution is divided in two categories based on proximity of generator level lepton ($gen^{e,\mu}$) in event to b-jet as given by equation 6.25.

$$\Delta R_{min}(gen^{e,\mu}, \text{b-jet}) < 0.4 \tag{6.25}$$

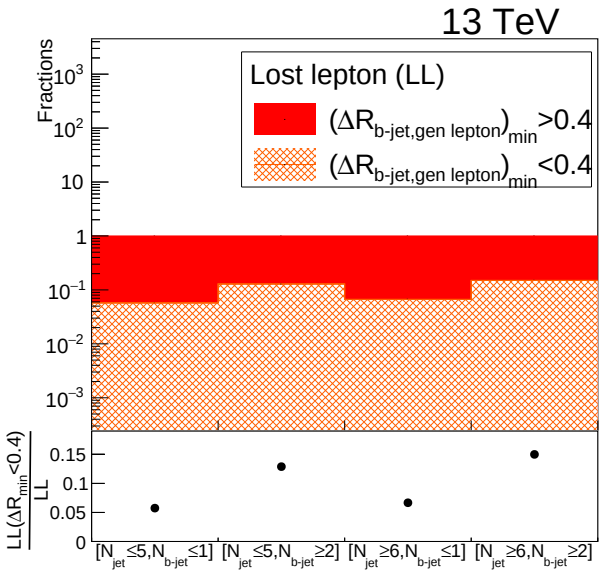

FIGURE 6.26: The $lost^{e,\mu}$ +$lost^{\tau_h}$ fraction falling into proximity of b-jet

Figure 6.26 shows that both for low jet and high jet multiplicity, higher the b-jet multiplicity in event, it is more probable that a lepton falls within any of the b-jet or alternatively because of proximity of lepton a jet gets mistagged as b-jet. In such cases, lepton fails isolation requirement, which shows up as increasing TF trend versus $N_{\mathrm{b-jet}}$ at low jet multiplicity. At high jet multiplicity, single lepton contribution dominates over $lost^{e,\mu}$ +$lost^{\tau_h}$. As a result although this effect is there, it does not show up that prominently for higher $N_{\mathrm{b-jet}}$.

6.2.8 Validation of method using MC simulated sample

Figures 6.27 and 6.28 show validation of method by comparing expected $lost^{e,\mu}$ +$lost^{\tau_h}$ yield against predicted $lost^{e,\mu}$ +$lost^{\tau_h}$ yield by applying TF_{MC} to 1L CR from MC. The 174 bin closure is perfectly at one as expected by method. As TF is derived as a function of 174 search bins, depending on bin boundaries in inclusive 1D distributions from Fig. 6.28, direct to prediction ratio fluctuates about one. A detailed

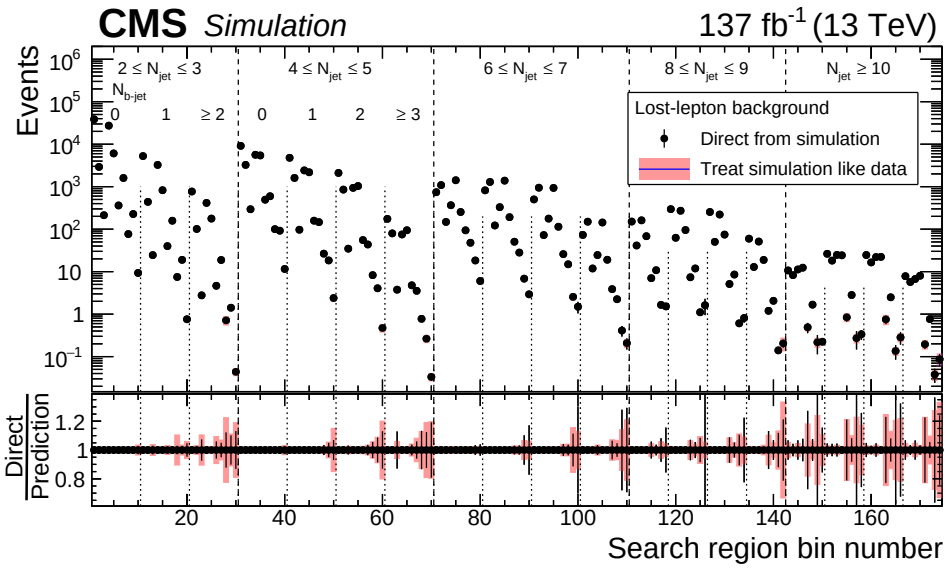

FIGURE 6.27: 174 bin closure with 2016+2017+2018 MC

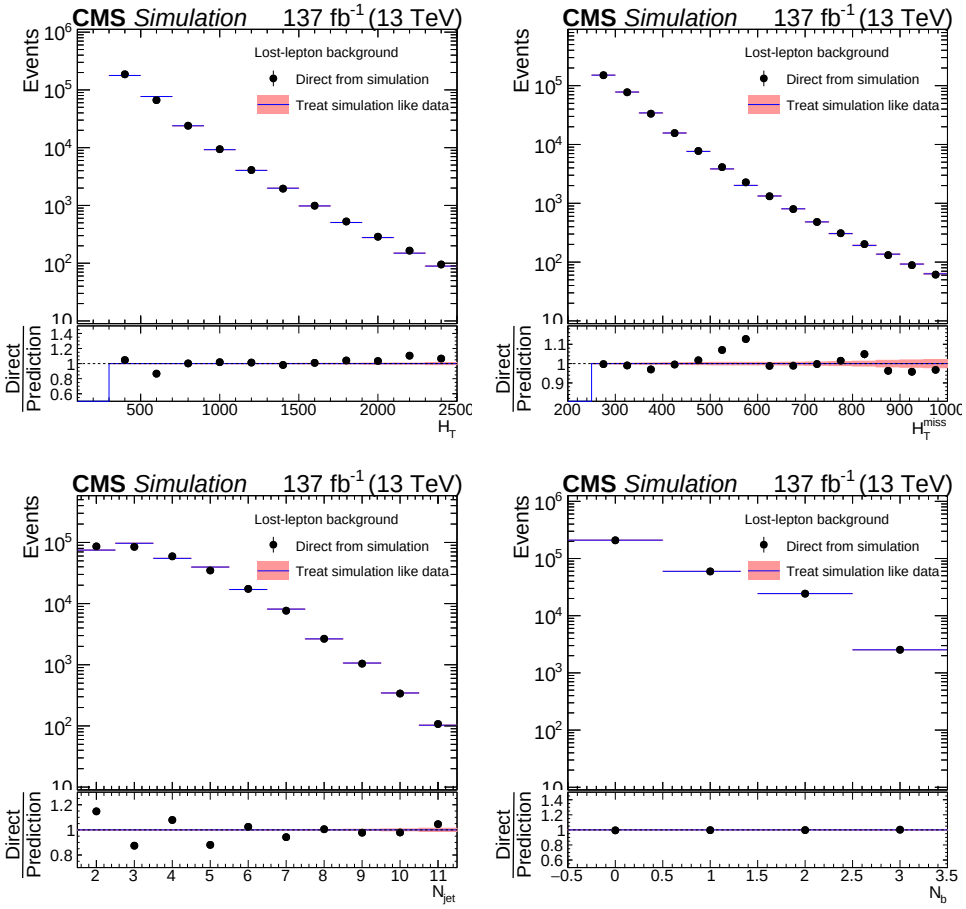

FIGURE 6.28: Closure in search variables with 2016+2017+2018 MC

validation of this method is done with respect to 2016 published results using event by event approach for $lost^{e,\mu}$ +$lost^{\tau_h}$ estimates. The studies are summarized in App.

D.

6.2.9 Estimation of $lost^{e,\mu}$ +$lost^{\tau_h}$ in data and systematic uncertainties

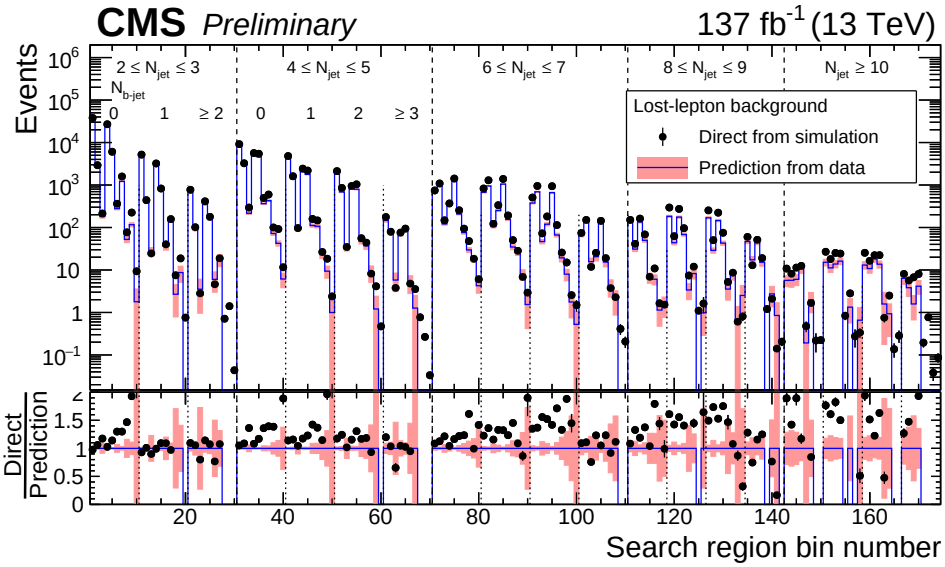

FIGURE 6.29: $lost^{e,\mu}$ +$lost^{\tau_h}$ prediction from data in 174 bins against MC expectation

Figures 6.29 and 6.30 show comparison of $lost^{e,\mu}$ +$lost^{\tau_h}$ prediction using 137 fb^{-1}data against MC expectation from 2016+2017+1018 MC. Similar to τ_h estimation with template method, these data versus MC comparisons are used to check that no abrupt features shown by data prediction but are not used in final results used for statistical interpretations.

As TF is applied as a simple scaling factor to one lepton CR yield, various systematics are derived by checking impact of propagation of various uncertainties on TF.

- JEC systematics:

  JECs applied to jets in event are scaled down by associated JEC uncertainties and jet p_T are rederived. Using new jet collection, all search variables are rederived and TF is obtained again. Variation of new TF with respect to nominal TF represents JEC systematics.

- Muon SF systematics:

  For muons, tracking SF is close to one. Hence only muon SFs associated with identification and isolation are varied by associated uncertainties independently and TF is rederived in both cases. Variation with respect to nominal TF represent muon id and iso SF systematics respectively.

- Electron SF systematics:

  For electrons, track reconstruction SF, idSF and isoSF are defined. Systematics derived by varying these three SF independently. Variation with respect to nominal TF represent electron id, iso and reco systematics respectively.

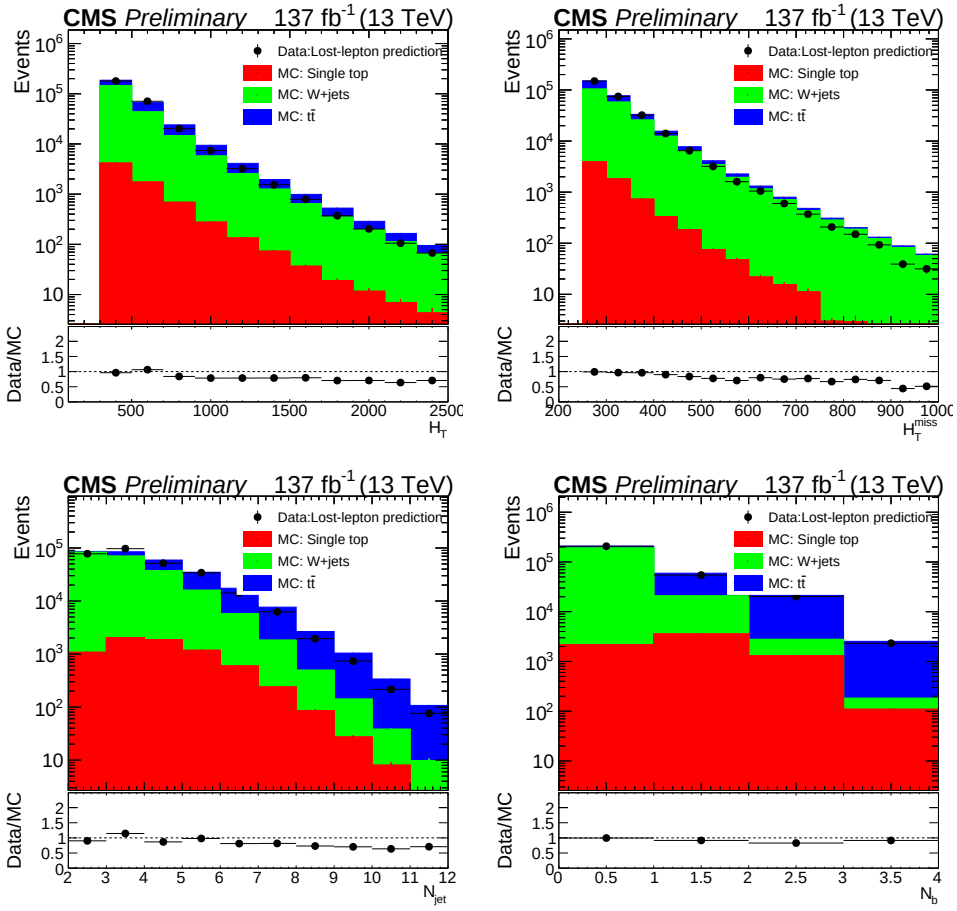

FIGURE 6.30: Comparison of $lost^{e,\mu}$ +$lost^{\tau_h}$ prediction from data against MC expectation in search variables

- b-tag systematics:

  Btag SFs are varied by associated uncertainties and variation in TF is considered as b-tag systematics.

- m_T systematics:

  The p_T^{miss} is varied by associated uncertainty and is then propagated to m_T. With this new m_T, TF is rederived. Variation with respect to nominal TF represent mT systematics.

- Acceptance systematics:

  Event weight is scaled by a set of 100 PDFweights or 9 Renormalization Scaleweights available and TF is rederived independently for each of them. Maximum variation with respect to nominal TF given by either of PDF weights / scale weights, is considered as PDF or scale systematics.

In addition there are statistical uncertainties from MC used to get TF and data statistical uncertainty from data CR. Statistical uncertainty on data CR dominates over other systematic sources of uncertainty.

6.2.10 Merging of systematics across years

As final $lost^{e,\mu}$ +$lost^{\tau_h}$ prediction is obtained by combining predictions from individual years, systematics for three years are added properly to account for correlation among them.

- Correlated systematics:

  All systematic uncertainties mentioned above (except TF statistical error and Data CR statistical error) use luminosity weighted average across three years.

  If $S_{y[i]}$ are values of particular systematics across different years and $L_{y[i]}$ represent luminosity for these years respectively, then final systematics after combining three years become:

$$S = \frac{\sum_{i=1}^{3}(S_{y[i]} \times L_{y[i]})}{\sum_{i=1}^{3} L_{y[i]}} \tag{6.26}$$

  Here index "i" corresponds to years 2016, 2017 and 2018 respectively.

- Uncorrelated systematics associated to statistical error on TF:

  If $0L_{y[i]}$ and $1L_{y[i]}$ represent "$lost^{e,\mu}$ +$lost^{\tau_h}$ " and "one lepton" yields for different years, then final MC based TF derived is:

$$TF_{\text{final}} = \frac{\sum_{i=1}^{3} 0L_{y[i]}}{\sum_{i=1}^{3} 1L_{y[i]}} \tag{6.27}$$

  Statistical error on TF_{final} represents final MC based statistical error on TF.

- Data statistical error systematics:

  Here data CR statistics for three years is simply added and statistical error on it is considered as Data statistical error systematics.

Table 6.1 gives order of various systematics after merging across three years. Data and MC statistical error uncertainties dominate over other systematics.

TABLE 6.1: $lost^{e,\mu}$ +$lost^{\tau_h}$ systematics after merging three years

systematics source	% uncertainty
B mistag, m_T, PDF, scale, MuID, MuIso, EleID, EleIso	0-1
EleReco	0 - 2
JEC	0 - 5
MC statistical	1-70
Data statistical	1-100

The final results along with the predicted backgrounds and systematic uncertainties in each search region are presented in chapter 7 and are published as [20]. The results are presented graphically in 7.1 including the data observed in various search regions.

Chapter 7

Results and summary

This chapter summarizes the results from a SUSY search carried in multijets+$p_{\mathrm{T}}^{\mathrm{miss}}$ final state using data equivalent to 137 fb^{-1}integrated luminosity collected by CMS experiment over years 2016, 2017 and 2018. The event yield in zero lepton signal region (SR) from data (observed event yield) is compared against the stacked pre-fit predictions for the SM backgrounds as shown in fig. 7.1. The stacked SM background predictions are obtained using the methods discussed under chapters 5 and 6. The corresponding numerical values are given in appendix E. The similar comparison is also made in one dimensional projection in different search variables as shown in fig. 7.2 and for 12 aggregate search bins, each representing a potentially interesting SUSY topology (Table 7.2) as shown in figure 7.3. The observed event yield agrees with predicted SM background yield in all search bins within uncertainties. Thus we did not observe evidence for production of SUSY particles in this data. Upper limits are evaluated for the production cross sections of the signal scenarios using a likelihood fit. The signal strength and its associated systematics, estimated backgrounds, their uncertainties and control region (CR) yield are input to the fit.

Likelihood function for each search bin is defined as the product of probability density functions; poisson distribution considered for control region yield and log-normal distribution for each of the nuisance parameter coming from various sources of systematic uncertainties in signal and background yields. The major sources of signal systematics and corresponding values are as given below.

TABLE 7.1: Uncorrelated signal systematics percentage

Source	Value (%)
Renormalization/Factorization: uncertainty in renormalization and factorization scales	0-5.7
Jet energy scale: impact of JEC uncertainty on jet p_{T} and η	0-14
Jet energy resolution: impact of JER uncertainty on jet p_{T} and η	0-10
Pile-up reweighting: uncertainty in total inelastic cross section	2.4
Isolated lepton and track veto: uncertainty in lepton and track reconstruction efficiency	2
Integrated luminosity: uncertainty in luminosity determination	2.3-2.5

Other than these, systematics due to uncertainty in b-jet tagging efficiency and misidentification move the event across bin boundaries hence lead to shape change

and are considered as correlated systematics across bins.

The observed data counts in signal region, central background predictions, total statistical and systematic uncertainties on each background prediction are shown in App. E. They are used as inputs to likelihood fit along with expected signal strength for the particular bin. The correlation among bins is taken into account. While setting limits, $CL_{(s)}$ method [97, 98] is used where a model with signal+background is compared against background only (null) hypothesis. Here "model" implies a cross section prediction with certain free parameters. Before fitting, signal yields are corrected for any contamination to control region. The approximate NNLO+NNLL cross section is used to determine corresponding exclusion curves. If λ_s and λ_b represent mean expected signal strength and mean background prediction and "n" is the observed data count in particular search bin, then probability of observing "n" data events under "signal+background" hypothesis will be

$$P_{(b+s)} = P(n, \lambda_s + \lambda_b) = e^{[-(\lambda_s+\lambda_b)]} \frac{(\lambda_s + \lambda_b)^n}{n!} \tag{7.1}$$

Under background only hypothesis, probability of observing "n" data events becomes

$$P_b = P(n, \lambda_b) = e^{[-(\lambda_b)]} \frac{(\lambda_b)^n}{n!} \tag{7.2}$$

Once, these probability density functions in hand, a test statistic is defined as:

$$-2ln(Q) = -2 \times ln(\frac{P(n, \lambda_s + \lambda_b)}{P(n, \lambda_b)}) \tag{7.3}$$

This test statistic (also called likelihood ratio) is optimal to discriminate signal+background hypothesis from background only hypothesis. Thus, for every "n" there is associated unique Q value.

- If P(n,λ_s +λ_b) = P(n,λ_b), then -2 ln(Q) = 0
- If P(n,λ_s +λ_b) > P(n,λ_b), then -2 ln(Q) < 0
- If P(n,λ_s +λ_b) < P(n,λ_b), then -2 ln(Q) > 0

For n=n_{obs}, observed data count in a search bin, corresponding Q value is called Q_{obs}.

With known Q_{obs}, one defines two quantities; $CL_{(b)}$ and $CL_{(b+s)}$ as probability of getting Q $\leq Q_{obs}$ under background only and signal+background hypothesis respectively.

$$CL_{(b)} = P_b(Q \leq Q_{obs}) \tag{7.4}$$

$$CL_{(b+s)} = P_{(b+s)}(Q \leq Q_{obs}) \tag{7.5}$$

The $CL_{(b)}$ and $CL_{(b+s)}$ can be determined easily by fraction of area of background only and signal+background probability distribution (with [-2 ln(Q)] on x-axis) to the right of -2 ln(Q_{obs}) value. Then confidence level of signal+background hypothesis normalized to background only hypothesis becomes:

$$CL_{(s)} = \frac{CL_{(b+s)}}{CL_{(b)}} \tag{7.6}$$

To find 95% confidence limit, a maximum signal strength $\lambda_{\mathrm{s}}^{max}$ is determined such that observing n_{obs} or fewer data events reaches 5% assuming a Poisson statistics with a mean of $\lambda_{\mathrm{s}}^{max}$ +λ_{b}. In other words, to get 95% confidence limit, $CL_{(\mathrm{s})} <$ 0.05. The signal with $\lambda_s > \lambda_{\mathrm{s}}^{max}$ is then excluded at 95% confidence limit (CL). For observed limit, n_{obs} is taken to be observed data count, while for expected limit n_{obs} is taken to be equal to total background estimate. Expected limit is a measure of how (in)compatible the background estimate is to the model with mean (λ_s +λ_{b}), hence it defines experimental sensitivity to rule out signal+background hypothesis.

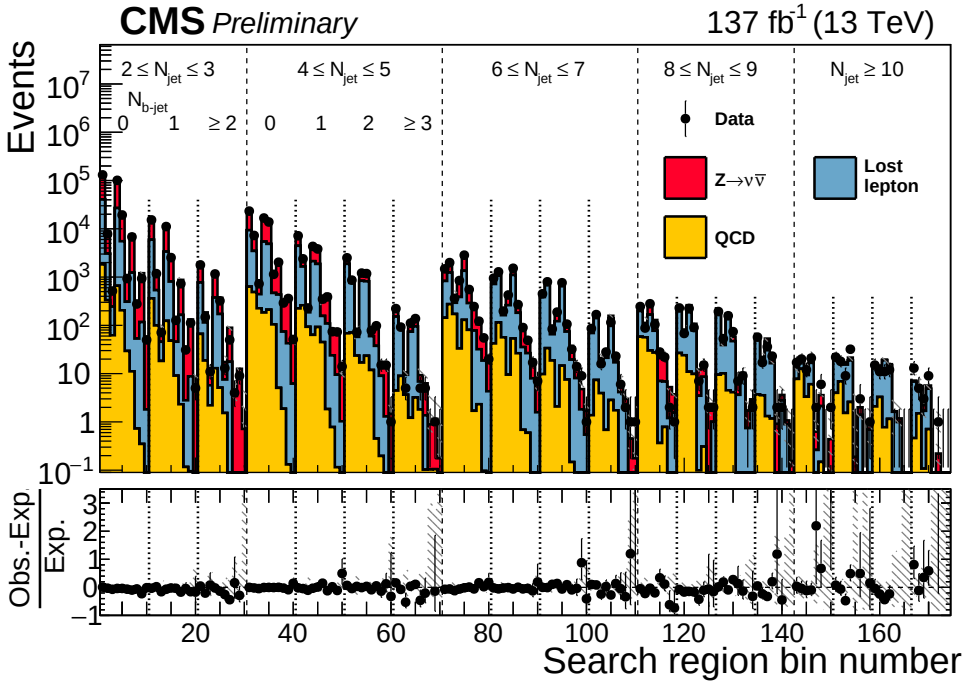

FIGURE 7.1: Data versus background prediction across 174 search bins using 2016+2017+2018 data

The exclusion is expressed in terms of gluino (squark)-neutralino masses for gluino (squark) pair production models. Gluinos with masses as large as 2180, 2310, 2000, and 2030 GeV, are excluded based on T1tttt, T1bbbb, T1qqqq, T5qqqqVV model scenarios shown in fig. 2.1. Figure 7.5 show similar results for squark pair production scenarios. Squarks with masses up to 1190, 1220, and 1630 GeV, are excluded respectively based on T2tt, T2bb and T2qq scenarios. For T2tt model, upper limit is not set in case of lower diagonal region, where the model becomes very similar to SM $t\bar{t}$ process and hence CR has significant contamination. For T2qq model, along with results with four mass-degenerate squark flavors (up, down, strange, and charm), each arising from two different quark spin states, exclusion is also shown in case only one of the eight states is accessible at LHC. In such case, upper limit on squark is reduced to 1130 GeV. Table 7.3 show comparison of expected exclusion obtained with 137 fb^{-1}Run 2 data against 35.9 fb^{-1}data in 2016 [19]. On average exclusion is extended by 200-300 GeV for both gluino and squark models.

TABLE 7.2: Aggregate search bins

Bin	Parton multiplicity	Heavy flavour	δm
1	Low	No	Small
2	Low	No	Large
3	Medium	No	Small
4	Medium	No	Large
5	High	No	All
6	Low	Yes	Small
7	Low	Yes	Large
8	Medium	Yes	Small
9	Medium	Yes	Large
10	High	Yes	Small
11	High	Yes	Large
12	High	Yes	All

TABLE 7.3: Exclusion comparison with 137 fb^{-1}data (Run 2) and 35.9 fb^{-1}(2016)

Model	Exclusion with 137 fb^{-1}data [GeV]	Exclusion with 35.9 fb^{-1}data [GeV]
T1tttt	2180	1960
T1bbbb	2310	1950
T5qqqqVV	2030	1800
T2tt	1190	960
T2bb	1220	990
T2qq	1630	1390

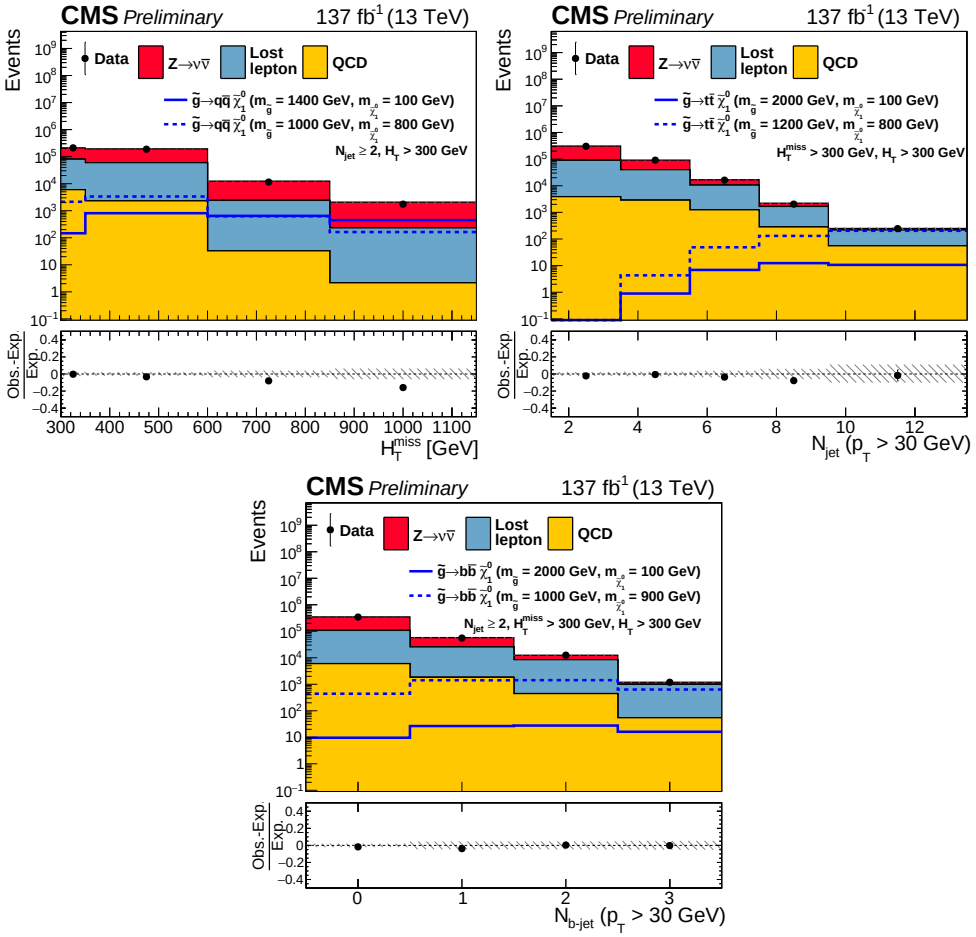

FIGURE 7.2: Data versus background prediction in one dimensional projection in search variables N_{jet}, $H_{\mathrm{T}}$and $H_{\mathrm{T}}^{\mathrm{miss}}$using 2016+2017+2018 data

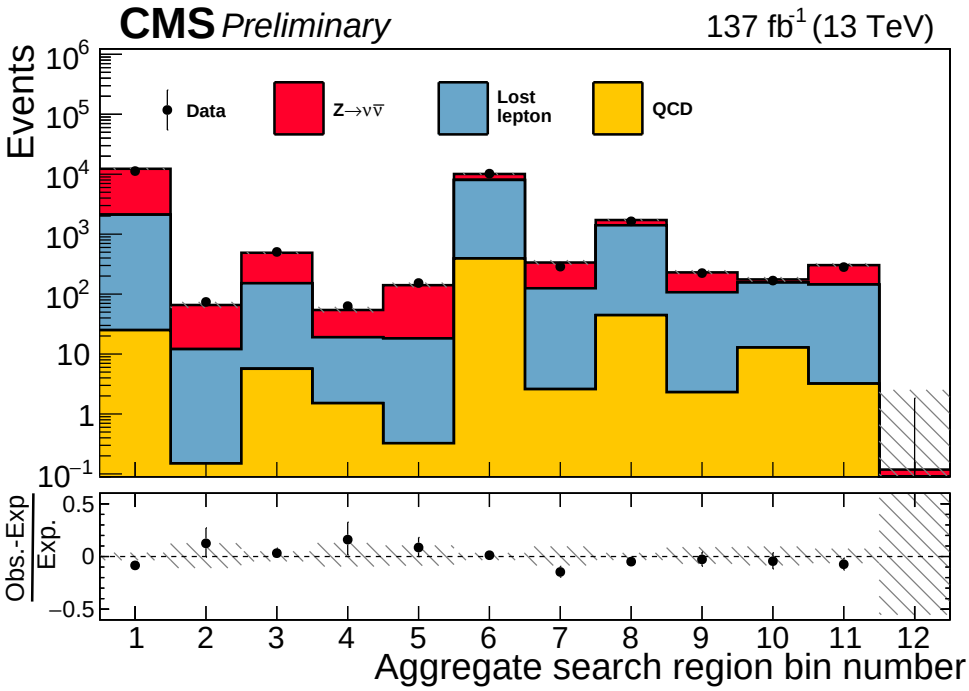

FIGURE 7.3: Data versus background prediction in twelve aggregate search bins corresponding to various topologies using 2016+2017+2018 data

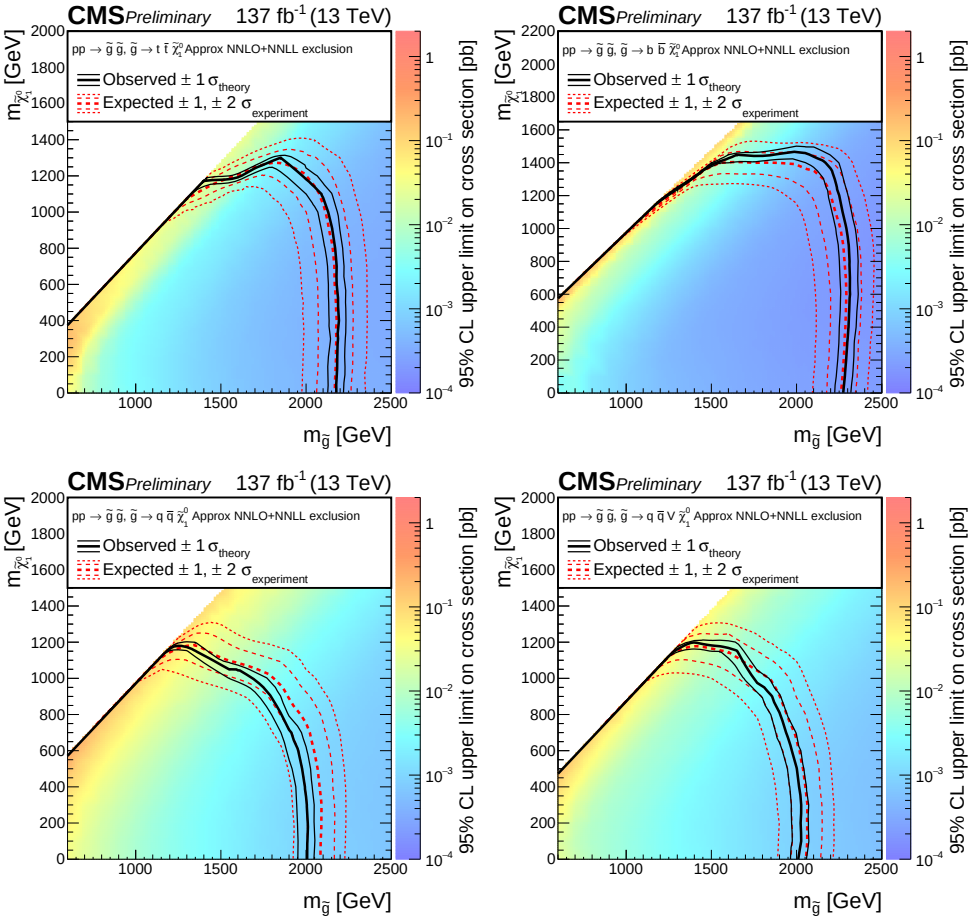

FIGURE 7.4: 95% confidence level upper limit on crosssection for gluino pair production models T1tttt, T1bbbb, T1qqqq and T5qqqqVV. Thick solid black line shows observed exclusion while dotted black line shows limits after varying theoretical crosssection by associated uncertainty. Thick red dashed line shows expected exclusion under background only hypothesis while thin red lines show region containing 68% and 95% distribution in limit under this hypothesis.

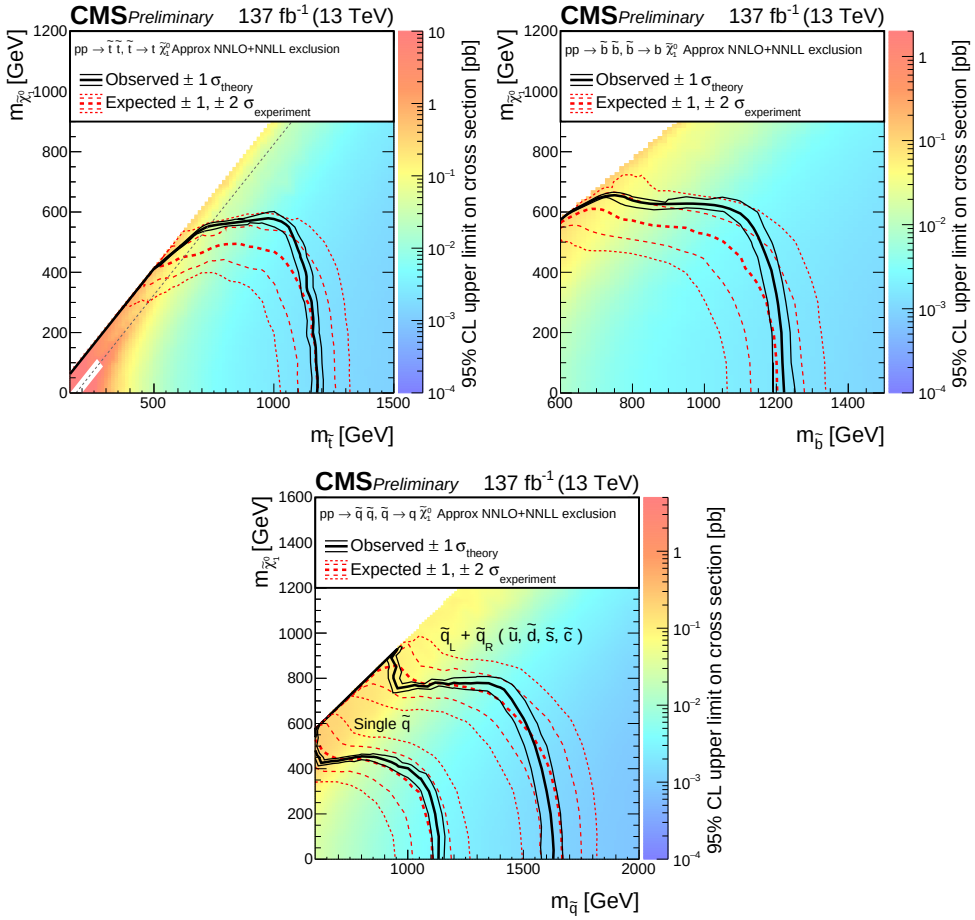

FIGURE 7.5: 95% confidence level upper limit on crosssection for squark pair production models T2tt, T2bb, T2qq. Thick solid black line shows observed exclusion while dotted black line shows limits after varying theoretical crosssection by associated uncertainty. Thick red dashed line shows expected exclusion under background only hypothesis while thin red lines show region containing 68% and 95% distribution in limit under this hypothesis.

Appendix A

Datasets used for full Run 2 analysis

The following table shows datasets collected from three years of data taking. The number for each dataset corresponds to integrated luminosity across three years listed in fb^{-1}.

Year	Primary Dataset	A	B	C	D	E	F	G	H	Total
2016	HTMHT	–	5746.365	2572.903	4242.289	4024.754	3104.509	7574.961	8650.622	35916.403
	JetHT	–	5750.126	2572.903	4242.292	4024.754	3104.509	7575.824	8650.628	35921.036
	MET	–	5746.370	2572.903	4242.287	3924.254	3104.508	7575.824	8649.019	35815.165
	SingleElectron	–	5746.183	2572.813	4242.201	4025.019	3104.288	7575.483	8650.155	35916.142
	SingleMuon	–	5746.010	2572.903	4242.292	4025.228	3104.509	7575.579	8650.628	35917.149
	SinglePhoton	–	5746.364	2572.903	4242.286	4025.226	3104.509	7575.824	8650.626	35917.738
2017	HTMHT	–	4793.951	9631.154	4247.680	9313.949	13534.500	–	–	41521.234
	JetHT	–	4793.961	9631.214	4247.682	9313.949	13534.525	–	–	41521.331
	MET	–	4793.348	9632.741	4247.682	9313.950	13498.415	–	–	41486.136
	SingleElectron	–	4793.904	9630.900	4247.670	9313.642	13539.222	–	–	41525.338
	SingleMuon	–	4793.961	9631.215	4247.682	9313.642	13538.559	–	–	41525.059
	SinglePhoton	–	4793.961	9631.210	4247.680	9313.642	13539.211	–	–	41525.704
2018	EGamma	13950.619	7060.617	6890.713	31741.739	–	–	–	–	59643.739
	JetHT	14027.047	7060.785	6894.782	31710.692	–	–	–	–	59692.692
	MET	14024.176	6907.892	6894.782	31719.531	–	–	–	–	59546.381
	SingleMuon	14027.047	7060.622	6894.771	31742.979	–	–	–	–	59725.419

FIGURE A.1: Details of datasets for years 2016, 2017 and 2018

Appendix B

SM MC simulated samples used for full Run 2 analysis

The following table shows SM MC simulated samples used for full Run 2 analysis.

Year	Dataset	σ (pb)	$\int \mathcal{L}\,dt$ (fb^{-1})
2016	TTJets_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	831.76	12.26
	TTJets_SingleLeptFromT_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	182.72	337.24
	TTJets_SingleLeptFromTbar_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	182.72	330.25
	TTJets_DiLept_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	88.34	349.06
	TTJets_SingleLeptFromT_genMET-150_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	9.684	1792.22
	TTJets_SingleLeptFromTbar_genMET-150_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	9.658	1760.63
	TTJets_DiLept_genMET-150_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	5.919	1647.82
	TTJets_HT-600to800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2.685	5343.28
	TTJets_HT-800to1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1.096	9607.90
	TTJets_HT-1200to2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.194	15097.94
	TTJets_HT-2500toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.002	646450.58
2017	TTJets_TuneCP5_13TeV-madgraphMLM-pythia8	831.76	9.63
	TTJets_SingleLeptFromT_TuneCP5_13TeV-madgraphMLM-pythia8	182.72	337.26
	TTJets_SingleLeptFromTbar_TuneCP5_13TeV-madgraphMLM-pythia8	182.72	309.66
	TTJets_DiLept_TuneCP5_13TeV-madgraphMLM-pythia8	88.34	320.56
	TTJets_SingleLeptFromT_genMET-150_TuneCP5_13TeV-madgraphMLM-pythia8	9.684	1476.65
	TTJets_SingleLeptFromTbar_genMET-150_TuneCP5_13TeV-madgraphMLM-pythia8	9.658	848.46
	TTJets_DiLept_genMET-150_TuneCP5_13TeV-madgraphMLM-pythia8	5.919	1463.03
	TTJets_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8	2.685	29849.58
	TTJets_HT-800to1200_TuneCP5_13TeV-madgraphMLM-pythia8	1.096	35778.74
	TTJets_HT-1200to2500_TuneCP5_13TeV-madgraphMLM-pythia8	0.194	65027.68
	TTJets_HT-2500toInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.002	1899830.47
2018	TTJets_TuneCP5_13TeV-madgraphMLM-pythia8	831.76	12.29
	TTJets_SingleLeptFromT_TuneCP5_13TeV-madgraphMLM-pythia8	182.72	312.78
	TTJets_SingleLeptFromTbar_TuneCP5_13TeV-madgraphMLM-pythia8	182.72	327.36
	TTJets_DiLept_TuneCP5_13TeV-madgraphMLM-pythia8	88.34	324.29
	TTJets_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8	2.685	5208.14
	TTJets_HT-800to1200_TuneCP5_13TeV-madgraphMLM-pythia8	1.096	9301.11
	TTJets_HT-1200to2500_TuneCP5_13TeV-madgraphMLM-pythia8	0.194	13862.50
	TTJets_HT-2500toInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.002	555382.49

FIGURE B.1: SM MC $t\bar{t}$ +jets samples used in analysis. The cross sections are calculated to NNLO.

Year	Dataset	σ (pb)	$\int \mathcal{L}\,dt$ (fb^{-1})
2016	QCD_HT200to300_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1717000.00	0.03
	QCD_HT300to500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	351300.00	0.15
	QCD_HT500to700_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	31630.00	1.98
	QCD_HT700to1000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	6802.00	2.30
	QCD_HT1000to1500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1206.00	12.61
	QCD_HT1500to2000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	120.40	98.33
	QCD_HT2000toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	25.24	238.49
2017	QCD_HT200to300_TuneCP5_13TeV-madgraph-pythia8	1717000.00	0.03
	QCD_HT300to500_TuneCP5_13TeV-madgraph-pythia8	351300.00	0.06
	QCD_HT500to700_TuneCP5_13TeV-madgraph-pythia8	31630.00	1.77
	QCD_HT700to1000_TuneCP5_13TeV-madgraph-pythia8	6802.00	2.92
	QCD_HT1000to1500_TuneCP5_13TeV-madgraph-pythia8	1206.00	13.81
	QCD_HT1500to2000_TuneCP5_13TeV-madgraph-pythia8	120.40	94.55
	QCD_HT2000toInf_TuneCP5_13TeV-madgraph-pythia8	25.24	226.31
2018	QCD_HT200to300_TuneCP5_13TeV-madgraphMLM-pythia8	1717000.00	0.03
	QCD_HT300to500_TuneCP5_13TeV-madgraphMLM-pythia8	351300.00	0.15
	QCD_HT500to700_TuneCP5_13TeV-madgraphMLM-pythia8	31630.00	1.74
	QCD_HT700to1000_TuneCP5_13TeV-madgraphMLM-pythia8	6802.00	7.04
	QCD_HT1000to1500_TuneCP5_13TeV-madgraphMLM-pythia8	1206.00	12.73
	QCD_HT1500to2000_TuneCP5_13TeV-madgraphMLM-pythia8	120.40	89.87
	QCD_HT2000toInf_TuneCP5_13TeV-madgraphMLM-pythia8	25.24	212.13

FIGURE B.2: SM MC QCD samples used in analysis. The cross sections are calculated to LO.

Year	Dataset	σ (pb)	$\int \mathcal{L}\,dt$ (fb^{-1})
2016	WJetsToLNu_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1627.45	6.11
	WJetsToLNu_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	435.24	89.57
	WJetsToLNu_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	59.18	131.12
	WJetsToLNu_HT-600To800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	14.58	1281.72
	WJetsToLNu_HT-800To1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	6.660	1175.76
	WJetsToLNu_HT-1200To2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1.608	4273.91
	WJetsToLNu_HT-2500ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.039	67792.88
2017	WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	1627.45	21.96
	WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	435.24	48.56
	WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	59.18	239.73
	WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8	14.58	1471.62
	WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8	6.660	3020.20
	WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8	1.608	12269.07
	WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.039	508831.27
2018	WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	1627.45	18.10
	WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	435.24	58.31
	WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.039	79961.87
	WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	59.18	99.68
	WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8	14.58	1346.26
	WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8	6.660	1249.54
	WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8	1.608	4670.23

FIGURE B.3: SM MC W+jets samples used in analysis. The cross sections are calculated to NNLO.

Year	Dataset	σ (pb)	$\int \mathcal{L}\,dt$ (fb^{-1})
2016	ZJetsToNuNu_HT-100To200_13TeV-madgraph	344.83	70.39
	ZJetsToNuNu_HT-200To400_13TeV-madgraph	95.53	259.19
	ZJetsToNuNu_HT-400To600_13TeV-madgraph	13.20	747.31
	ZJetsToNuNu_HT-600To800_13TeV-madgraph	3.148	1831.99
	ZJetsToNuNu_HT-800To1200_13TeV-madgraph	1.451	1495.71
	ZJetsToNuNu_HT-1200To2500_13TeV-madgraph	0.355	1447.84
	ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph	0.009	47414.35
2017	ZJetsToNuNu_HT-100To200_13TeV-madgraph	344.83	65.74
	ZJetsToNuNu_HT-200To400_13TeV-madgraph	95.53	225.69
	ZJetsToNuNu_HT-400To600_13TeV-madgraph	13.20	686.16
	ZJetsToNuNu_HT-600To800_13TeV-madgraph	3.148	1789.28
	ZJetsToNuNu_HT-800To1200_13TeV-madgraph	1.451	1396.10
	ZJetsToNuNu_HT-1200To2500_13TeV-madgraph	0.355	929.88
	ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph	0.009	722.32
2018	ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph	0.009	39950.11
	ZJetsToNuNu_HT-100To200_13TeV-madgraph	344.83	68.60
	ZJetsToNuNu_HT-200To400_13TeV-madgraph	95.53	242.81
	ZJetsToNuNu_HT-400To600_13TeV-madgraph	13.20	203.97
	ZJetsToNuNu_HT-600To800_13TeV-madgraph	3.148	1813.18
	ZJetsToNuNu_HT-800To1200_13TeV-madgraph	1.451	1410.14
	ZJetsToNuNu_HT-1200To2500_13TeV-madgraph	0.355	951.46

FIGURE B.4: SM MC Z+jets samples used in analysis. The cross sections are calculated to NNLO.

Year	Dataset	σ (pb)	$\int \mathcal{L}\,dt$ (fb^{-1})
2016	ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8_TuneCUETP8M1	3.340	116.20
	ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1	136.02	493.35
	ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1	80.95	479.44
	ST_tW_antitop_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1	19.47	167.27
	ST_tW_top_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1	19.47	167.29
2017	ST_s-channel_4f_leptonDecays_TuneCP5_PSweights_13TeV-amcatnlo-pythia8	3.340	1154.17
	ST_t-channel_top_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8	136.02	43.13
	ST_t-channel_antitop_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8	80.95	48.67
	ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	19.47	272.59
	ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	19.47	237.87
2018	ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-madgraph-pythia8	3.340	2327.69
	ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8	19.47	55.30
	ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8	19.47	55.27
	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.60	212.18
	ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.60	267.14

FIGURE B.5: SM MC single top samples used in analysis. The cross sections are calculated to NLO.

Year	Dataset	σ (pb)	$\int \mathcal{L}\,dt$ (fb^{-1})
2016	DYJetsToLL_M-50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	181.30	60.77
	DYJetsToLL_M-50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	50.42	190.59
	DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	6.984	1392.58
	DYJetsToLL_M-50_HT-600to800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1.681	4932.14
	DYJetsToLL_M-50_HT-800to1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.775	3447.37
	DYJetsToLL_M-50_HT-1200to2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.186	3200.91
	DYJetsToLL_M-50_HT-2500toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.004	91105.26
2017	DYJetsToLL_M-50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8	181.30	78.00
	DYJetsToLL_M-50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8	50.42	226.83
	DYJetsToLL_M-50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8	6.984	1512.70
	DYJetsToLL_M-50_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8	1.681	4791.67
	DYJetsToLL_M-50_HT-800to1200_TuneCP5_13TeV-madgraphMLM-pythia8	0.775	3920.47
	DYJetsToLL_M-50_HT-1200to2500_TuneCP5_13TeV-madgraphMLM-pythia8	0.186	3267.32
	DYJetsToLL_M-50_HT-2500toInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.004	84618.62
2018	DYJetsToLL_M-50_HT-100to200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	181.30	63.46
	DYJetsToLL_M-50_HT-2500toInf_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	0.004	92287.49
	DYJetsToLL_M-50_HT-200to400_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	50.42	221.89
	DYJetsToLL_M-50_HT-400to600_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	6.984	1373.18
	DYJetsToLL_M-50_HT-600to800_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	1.681	5230.89
	DYJetsToLL_M-50_HT-800to1200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	0.775	4005.59
	DYJetsToLL_M-50_HT-1200to2500_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	0.186	2678.52

FIGURE B.6: SM MC DY+jets samples used in analysis. The cross sections are calculated to NNLO.

Year	Dataset	σ (pb)	$\int \mathcal{L}\,dt$ (fb^{-1})
2016	GJets_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	9226.00	1.10
	GJets_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2300.00	8.93
	GJets_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	277.40	18.24
	GJets_HT-600ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	93.38	54.41
	GJets_DR-0p4_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	5000.00	2.96
	GJets_DR-0p4_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1079.00	46.34
	GJets_DR-0p4_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	125.90	92.95
	GJets_DR-0p4_HT-600ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	43.36	269.54
2017	GJets_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	9226.00	1.08
	GJets_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	2300.00	8.05
	GJets_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	277.40	16.70
	GJets_HT-600ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	93.38	34.98
	GJets_DR-0p4_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8_v2	5000.00	3.19
	GJets_DR-0p4_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8_v2	1079.00	46.18
	GJets_DR-0p4_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8_v2	125.90	106.00
	GJets_DR-0p4_HT-600ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	43.36	192.11
2018	GJets_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	9226.00	1.06
	GJets_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	2300.00	8.28
	GJets_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	277.40	16.76
	GJets_HT-600ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	93.38	53.15
	GJets_DR-0p4_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	5000.00	3.08
	GJets_DR-0p4_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	1079.00	45.80
	GJets_DR-0p4_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	125.90	108.76
	GJets_DR-0p4_HT-600ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	43.36	286.14

FIGURE B.7: SM MC γ+jets samples used in analysis. The cross sections are calculated to LO.

Appendix C

Results using only 2016 data corresponding to luminosity 35.9 fb^{-1}

The figures C.1 and C.2 show statistical interpretations for simplified gluino and squark models based on SUSY search done using 35.9 fb^{-1} data from 2016.

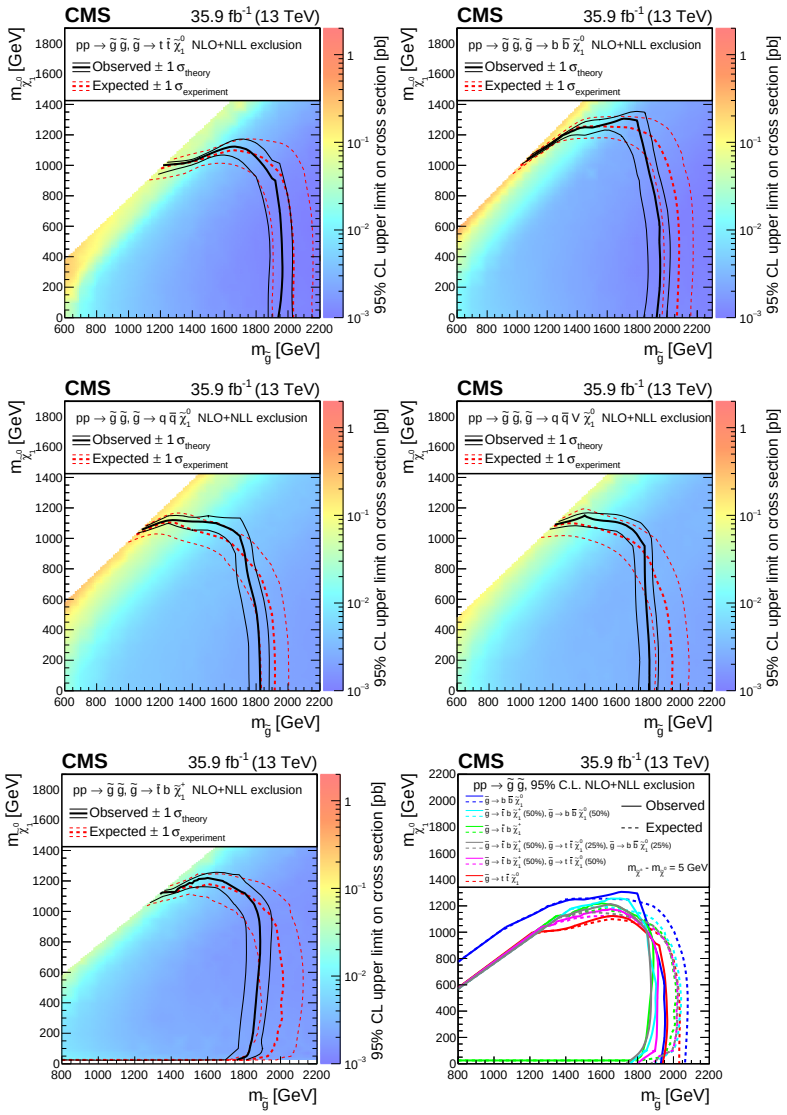

FIGURE C.1: 95% confidence level upper limit on crosssection for gluino pair production models T1tttt, T1bbbb, T1qqqq, T5qqqqVV and T1tbtb. Thick solid black line shows observed exclusion while dotted black line shows limits after varying theoretical crosssection by associated uncertainty. Thick red dashed line shows expected exclusion under background only hypothesis while thin red lines show region containing 68% and 95% distribution in limit under this hypothesis.

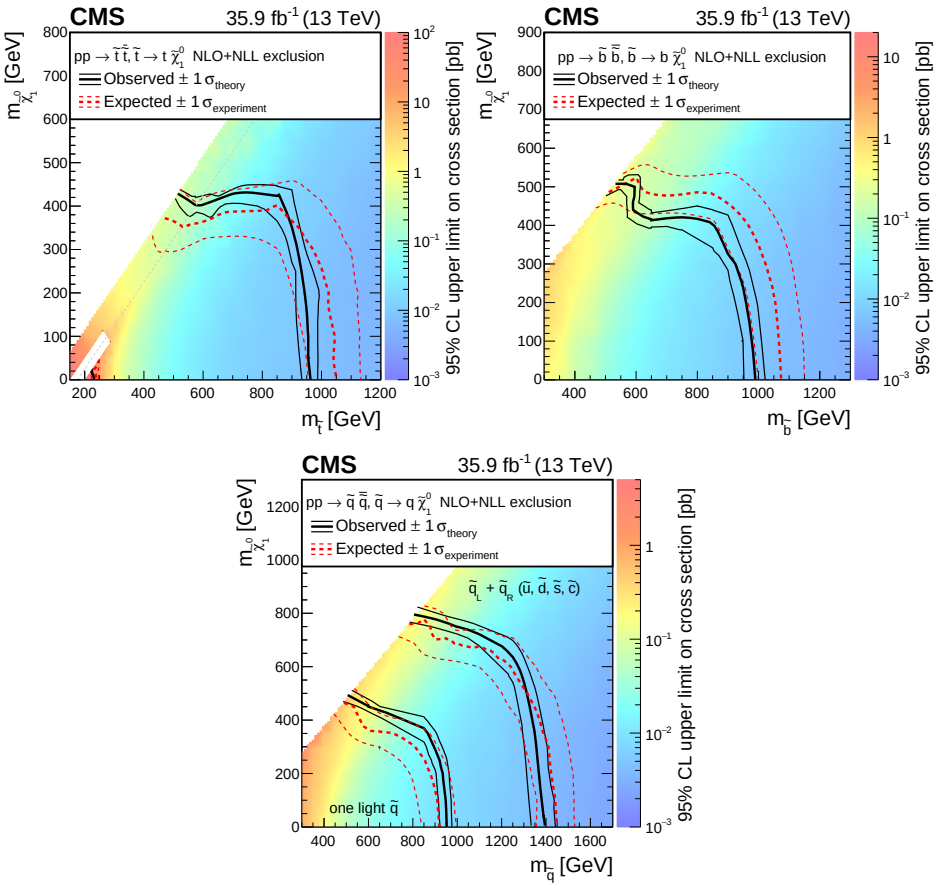

FIGURE C.2: 95% confidence level upper limit on crosssection for squark pair production models T2tt, T2bb, T2qq. Thick solid black line shows observed exclusion while dotted black line shows limits after varying theoretical crosssection by associated uncertainty. Thick red dashed line shows expected exclusion under background only hypothesis while thin red lines show region containing 68% and 95% distribution in limit under this hypothesis.

Appendix D

Validation of Average Transfer Factor method

D.1 Average TF method against event by event method (2016)

As a validation check, hadtau prediction in data obtained using Average TF method is compared against prediction obtained with event by event approach and used for 2016 published results. Figure D.1 shows a comparison where 174 bin hadtau prediction in data (solid black points) is compared against the prediction from event by event method (template method) used for 2016 published results. No systematic change is observed and two methods agree within statistical uncertainties.

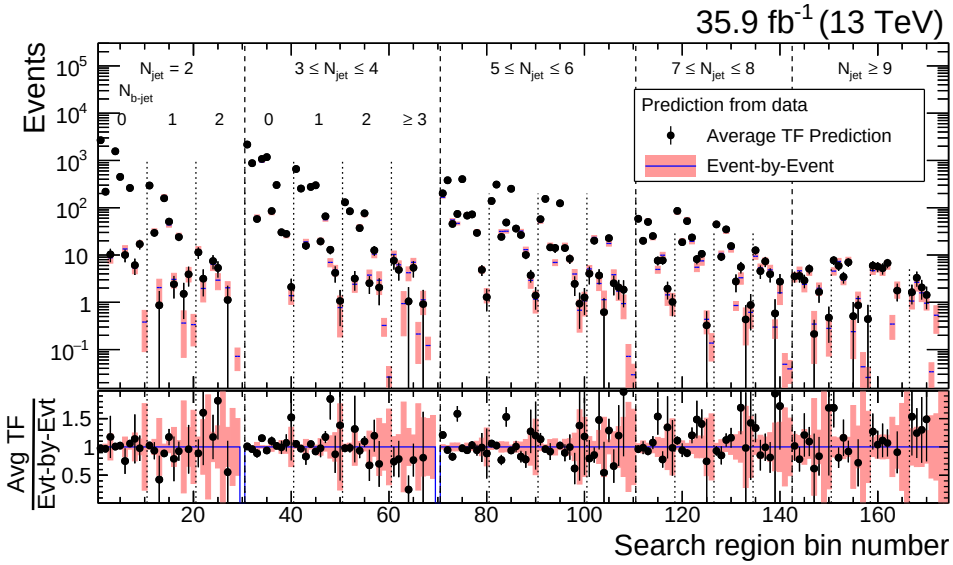

FIGURE D.1: Comparison of hadtau prediction in data obtained by TF method and event by event method

Further performance of Average TF implementation was checked by comparing how the expected and observed limits look after replacing Lost Lepton and Hadtau background components and associated systematics from 2016 analysis with combined Lost lepton estimate with Average TF and associated systematics scaled to 35.9 fb^{-1}. From figure D.2 it is observed that expected limit is almost consistent with previous limit implying sensitivity with new method is similar to that with event by event approach. Observed limit looked weaker over intermediate range of gluino, neutralino mass range. For further understanding of what is causing this change in observed limit, the particular search bins which are leading to significant change in

observed limit are found. As shown in figure D.3, dropping of low sensitivity search bins like 127, 140 showed hardly any impact on both observed and expected limits. But dropping of bin 172(173) lead to stronger(weaker) observed limit as compared to one using all 174 search bins. Expected limit became weaker equally after removing either 172/ 173.

For expected limits, observed data counts are not used. Bins with expected signal strength much higher than SM Bkg Prediction improve expected limit. Bins 172, 173 fall in this category. Hence, skipping of those bins made expected limit worse.

For observed limits, it is the interplay of data, signal strength and background predictions.

- Bins with Data $<$ SM Bkg, help to constrain observed limits. Hence skipping such bins makes observed limit worse (ex. bins 173 and 140).
- Bins with Data $>$ SM Bkg allow possibility to include signal, making observed limit weaker. Hence skipping of such bins improve observed limit.

  (a) If signal $>>$ SM Bkg, impact of such bins on observed limit is higher (ex. Bin 172).

  (b) If signal $<$ SM Bkg, impact of such bins on observed limit is not significant (ex. Bin 140).

Thus, high sensitivity bins like 172 and 173 which show significant opposite impact on observed limits are causing change in observed limit for Average TF method against event by event approach.

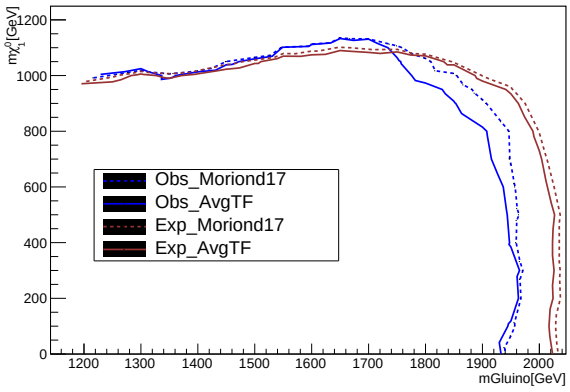

FIGURE D.2: Comparison of expected and observed limit for T1tttt model obtained using event by event background prediction of Lost Lepton and Hadtau as two separate background components (brown) against those obtained after using combined Lost Lepton+Hadtau prediction using average TF method (blue)

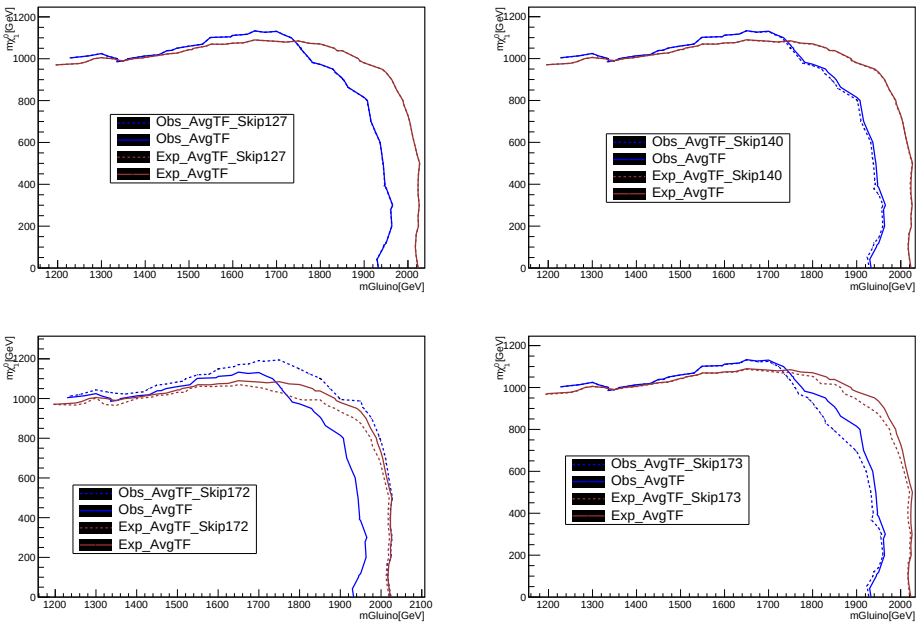

FIGURE D.3: Comparison of expected and observed limit for T1tttt model obtained using Lost Lepton+Hadtau background prediction using average TF method (solid line) against those after skipping one of the bins (127, 140, 172 or 173) from limit calculation (dotted line).

Appendix E

Numerical values for results

E.1 Numerical results for full set of search bins

The following tables give observed number of events in data and pre-fit background predictions in search bins corresponding to different N_{jet} categories. For the background predictions, the first uncertainty is statistical and the second is systematic.

Bin	$H_{\text{T}}^{\text{miss}}$ [GeV]	H_{T} [GeV]	N_{jet}	$N_{\text{b-jet}}$	Lost-lepton background	$\text{Z} \to \nu\overline{\nu}$ background	QCD background	Total background	Observed
1	300–350	300–600	2–3	0	$38\,872^{+320+580}_{-320-580}$	$89\,092^{+190+2600}_{-190-2500}$	$1828^{+990+1200}_{-990-840}$	$129\,792^{+1100+2900}_{-1100-2700}$	130 718
2	300–350	600–1200	2–3	0	2760^{+61+39}_{-60-39}	$4972^{+45+150}_{-45-150}$	$332^{+180+160}_{-180-150}$	$8064^{+200+220}_{-200-210}$	7820
3	300–350	≥1200	2–3	0	181^{+17+3}_{-16-3}	308^{+12+19}_{-12-17}	62^{+34+27}_{-34-27}	552^{+40+33}_{-39-32}	514
4	350–600	350–600	2–3	0	$26\,230^{+240+540}_{-240-540}$	$77\,996^{+180+2200}_{-180-2100}$	$659^{+360+300}_{-360-300}$	$104\,886^{+460+2300}_{-460-2200}$	100 828
5	350–600	600–1200	2–3	0	5319^{+81+78}_{-80-78}	$14\,569^{+77+440}_{-77-420}$	$205^{+110+99}_{-110-94}$	$20\,093^{+160+460}_{-160-440}$	19 319
6	350–600	≥1200	2–3	0	279^{+21+6}_{-20-6}	689^{+17+41}_{-17-36}	29^{+16+13}_{-16-13}	997^{+32+43}_{-30-38}	933
7	600–850	600–1200	2–3	0	1221^{+43+25}_{-42-25}	$6286^{+52+370}_{-52-360}$	$11.1^{+6.0+5.4}_{-6.0-5.1}$	$7519^{+68+370}_{-67-360}$	6786
8	600–850	≥1200	2–3	0	52^{+10+2}_{-8-2}	240^{+11+15}_{-10-15}	$0.73^{+0.65+0.31}_{-0.65-0.07}$	293^{+15+16}_{-13-15}	277
9	≥850	850–1700	2–3	0	116^{+15+3}_{-13-3}	$1088^{+23+100}_{-23-96}$	$0.35^{+0.21+0.15}_{-0.21-0.14}$	$1205^{+28+100}_{-27-96}$	933
10	≥850	≥1700	2–3	0	$1.8^{+4.1+0.1}_{-1.5-0.1}$	$48.9^{+5.3+5.2}_{-4.8-5.0}$	$0.02^{+0.02+0.01}_{-0.02-0.00}$	$50.7^{+6.7+5.2}_{-5.0-5.0}$	50
11	300–350	300–600	2–3	1	$5591^{+100+97}_{-99-97}$	$9809^{+21+1500}_{-21-1500}$	$363^{+200+330}_{-200-160}$	$15\,763^{+220+1500}_{-220-1500}$	15 272
12	300–350	600–1200	2–3	1	436^{+25+6}_{-24-6}	616^{+6+95}_{-6-95}	99^{+54+79}_{-54-45}	$1151^{+60+120}_{-59-110}$	1177
13	300–350	≥1200	2–3	1	$27.4^{+7.9+0.4}_{-6.3-0.4}$	$38.4^{+1.5+6.2}_{-1.4-6.1}$	18^{+10+14}_{-10-8}	84^{+13+15}_{-12-10}	71
14	350–600	350–600	2–3	1	3237^{+75+99}_{-74-99}	$8564^{+20+1300}_{-20-1300}$	124^{+67+96}_{-67-57}	$11\,925^{+100+1300}_{-100-1300}$	11 121
15	350–600	600–1200	2–3	1	757^{+33+14}_{-31-14}	$1782^{+10+270}_{-9-270}$	48^{+27+38}_{-27-21}	$2587^{+43+280}_{-42-270}$	2530
16	350–600	≥1200	2–3	1	$36.7^{+8.9+0.5}_{-7.3-0.5}$	86^{+2+14}_{-2-14}	$9.1^{+5.0+6.9}_{-5.0-4.1}$	132^{+10+16}_{-9-14}	127
17	600–850	600–1200	2–3	1	162^{+17+4}_{-16-4}	712^{+6+120}_{-6-110}	$2.3^{+1.3+1.8}_{-1.3-1.0}$	$876^{+18+120}_{-17-110}$	728
18	600–850	≥1200	2–3	1	$2.7^{+3.5+0.1}_{-1.7-0.1}$	$29.5^{+1.3+4.8}_{-1.3-4.8}$	$0.12^{+0.10+0.09}_{-0.10-0.02}$	$32.3^{+3.8+4.8}_{-2.1-4.8}$	31
19	≥850	850–1700	2–3	1	$8.7^{+5.2+0.2}_{-3.5-0.2}$	124^{+3+22}_{-3-22}	$0.10^{+0.07+0.07}_{-0.07-0.02}$	133^{+6+22}_{-4-22}	112
20	≥850	≥1700	2–3	1	$0.0^{+3.6+0.0}_{-0.0-0.0}$	$6.0^{+0.7+1.1}_{-0.6-1.1}$	$0.03^{+0.04+0.02}_{-0.03-0.00}$	$6.0^{+3.6+1.1}_{-0.6-1.1}$	5
21	300–350	300–600	2–3	≥2	706^{+37+13}_{-36-13}	935^{+2+290}_{-2-290}	66^{+68+72}_{-66-0}	$1708^{+77+300}_{-76-290}$	1787
22	300–350	600–1200	2–3	≥2	96^{+14+1}_{-12-1}	71^{+1+22}_{-1-22}	19^{+11+19}_{-11-8}	186^{+18+29}_{-17-23}	148
23	300–350	≥1200	2–3	≥2	$3.5^{+4.7+0.1}_{-2.3-0.1}$	$4.4^{+0.2+1.4}_{-0.2-1.4}$	$2.2^{+1.3+2.1}_{-1.3-0.9}$	$10.2^{+4.8+2.5}_{-2.6-1.7}$	11
24	350–600	350–600	2–3	≥2	362^{+27+14}_{-26-14}	811^{+2+250}_{-2-250}	13^{+8+13}_{-8-5}	$1186^{+28+250}_{-27-250}$	1159
25	350–600	600–1200	2–3	≥2	166^{+18+5}_{-17-5}	201^{+1+61}_{-1-61}	$5.1^{+3.3+5.1}_{-3.3-1.8}$	373^{+18+62}_{-17-62}	322
26	350–600	≥1200	2–3	≥2	$6.0^{+4.8+0.1}_{-2.9-0.1}$	$9.9^{+0.2+3.1}_{-0.2-3.1}$	$1.5^{+0.9+1.5}_{-0.9-0.6}$	$17.5^{+4.9+3.4}_{-3.1-3.1}$	13
27	600–850	600–1200	2–3	≥2	$17.5^{+7.6+0.3}_{-5.6-0.3}$	72^{+1+22}_{-1-22}	$0.09^{+0.09+0.09}_{-0.09-0.00}$	89^{+8+22}_{-6-22}	50
28	600–850	≥1200	2–3	≥2	$0.0^{+2.9+0.0}_{-0.0-0.0}$	$3.4^{+0.1+1.0}_{-0.1-1.0}$	$0.08^{+0.08+0.07}_{-0.08-0.00}$	$3.4^{+2.9+1.0}_{-0.2-1.0}$	4
29	≥850	850–1700	2–3	≥2	$0.0^{+4.4+0.0}_{-0.0-0.0}$	$12.5^{+0.3+4.0}_{-0.3-4.0}$	$0.09^{+0.07+0.09}_{-0.07-0.02}$	$12.6^{+4.5+4.0}_{-0.3-4.0}$	9
30	≥850	≥1700	2–3	≥2	$0.0^{+3.7+0.0}_{-0.0-0.0}$	$0.68^{+0.07+0.22}_{-0.07-0.22}$	$0.04^{+0.04+0.03}_{-0.04-0.00}$	$0.7^{+3.7+0.2}_{-0.1-0.2}$	0

FIGURE E.1: Numerical results for N_{jet} bins [2,3]

Bin	H_T^{miss} [GeV]	H_T [GeV]	N_{jet}	$N_{b\text{-jet}}$	Lost-lepton background	$Z \to \nu\bar{\nu}$ background	QCD background	Total background	Observed
31	300–350	300–600	4–5	0	$8719^{+110+120}_{-110-120}$	$13\,926^{+73+600}_{-73-580}$	$633^{+350+410}_{-350-290}$	$23\,278^{+370+740}_{-370-660}$	23 241
32	300–350	600–1200	4–5	0	2989^{+48+54}_{-48-54}	$3962^{+39+150}_{-39-150}$	$488^{+260+240}_{-260-230}$	$7439^{+270+290}_{-270-270}$	7277
33	300–350	≥1200	4–5	0	216^{+14+5}_{-13-5}	317^{+12+19}_{-11-17}	$225^{+120+97}_{-120-97}$	$759^{+120+99}_{-120-99}$	726
34	350–600	350–600	4–5	0	$5228^{+85+160}_{-84-160}$	$11\,407^{+67+460}_{-66-440}$	$184^{+100+84}_{-100-82}$	$16\,819^{+150+500}_{-150-480}$	16 720
35	350–600	600–1200	4–5	0	4654^{+59+68}_{-59-68}	$9002^{+59+360}_{-59-340}$	$211^{+110+100}_{-110-97}$	$13\,866^{+140+380}_{-140-360}$	13 837
36	350–600	≥1200	4–5	0	364^{+17+6}_{-16-6}	680^{+17+40}_{-17-35}	104^{+56+45}_{-56-45}	1148^{+61+61}_{-61-57}	1141
37	600–850	600–1200	4–5	0	428^{+19+9}_{-18-9}	1592^{+25+96}_{-25-92}	$5.1^{+2.8+2.4}_{-2.8-2.3}$	2025^{+32+96}_{-31-93}	2028
38	600–850	≥1200	4–5	0	$72.2^{+8.1+1.1}_{-7.3-1.1}$	225^{+10+14}_{-10-14}	$1.9^{+1.1+0.8}_{-1.1-0.8}$	299^{+13+14}_{-12-14}	291
39	≥850	850–1700	4–5	0	$42.4^{+6.9+0.8}_{-6.0-0.8}$	351^{+13+33}_{-12-31}	$0.13^{+0.09+0.06}_{-0.09-0.04}$	393^{+15+33}_{-14-31}	360
40	≥850	≥1700	4–5	0	$6.1^{+3.3+0.1}_{-2.3-0.1}$	$38.4^{+4.4+4.4}_{-4.0-4.3}$	$0.06^{+0.05+0.02}_{-0.05-0.01}$	$44.6^{+5.5+4.4}_{-4.6-4.3}$	51
41	300–350	300–600	4–5	1	4217^{+69+77}_{-68-77}	$2846^{+15+450}_{-15-450}$	$224^{+120+200}_{-120-100}$	$7287^{+140+500}_{-140-470}$	7157
42	300–350	600–1200	4–5	1	1389^{+35+23}_{-34-23}	847^{+8+130}_{-8-130}	$261^{+140+210}_{-140-120}$	$2496^{+150+250}_{-150-180}$	2387
43	300–350	≥1200	4–5	1	93^{+10+3}_{-9-3}	69^{+3+11}_{-2-11}	93^{+50+71}_{-50-43}	255^{+51+72}_{-51-44}	229
44	350–600	350–600	4–5	1	2068^{+50+41}_{-49-41}	$2326^{+14+370}_{-14-370}$	64^{+35+49}_{-35-29}	$4458^{+63+370}_{-62-370}$	4317
45	350–600	600–1200	4–5	1	1777^{+40+29}_{-39-29}	$1912^{+13+300}_{-12-300}$	92^{+50+73}_{-50-42}	$3782^{+65+310}_{-65-300}$	3822
46	350–600	≥1200	4–5	1	112^{+11+3}_{-10-3}	148^{+4+24}_{-4-24}	45^{+24+34}_{-24-21}	305^{+27+42}_{-27-32}	350
47	600–850	600–1200	4–5	1	107^{+11+3}_{-10-3}	332^{+5+54}_{-5-54}	$1.8^{+1.1+1.5}_{-1.1-0.8}$	441^{+12+55}_{-11-54}	388
48	600–850	≥1200	4–5	1	$23.1^{+5.5+0.4}_{-4.6-0.4}$	$48.6^{+2.2+8.0}_{-2.1-8.0}$	$0.78^{+0.51+0.59}_{-0.51-0.27}$	$72.5^{+6.0+8.1}_{-5.0-8.0}$	74
49	≥850	850–1700	4–5	1	$9.4^{+4.0+0.3}_{-3.0-0.3}$	73^{+3+13}_{-3-13}	$0.12^{+0.09+0.09}_{-0.09-0.03}$	82^{+5+13}_{-4-13}	73
50	≥850	≥1700	4–5	1	$1.0^{+2.3+0.0}_{-0.8-0.0}$	$8.3^{+1.0+1.6}_{-0.9-1.6}$	$0.03^{+0.04+0.02}_{-0.03-0.00}$	$9.4^{+2.5+1.6}_{-1.2-1.6}$	14
51	300–350	300–600	4–5	2	1806^{+49+30}_{-48-30}	468^{+2+79}_{-2-79}	68^{+45+74}_{-45-24}	$2342^{+66+110}_{-65-87}$	2505
52	300–350	600–1200	4–5	2	687^{+26+10}_{-25-10}	144^{+1+24}_{-1-24}	71^{+39+70}_{-39-32}	902^{+47+75}_{-47-41}	864
53	300–350	≥1200	4–5	2	$34.0^{+7.4+0.7}_{-6.2-0.7}$	$12.0^{+0.4+2.1}_{-0.4-2.1}$	24^{+13+23}_{-13-11}	70^{+15+23}_{-14-11}	72
54	350–600	350–600	4–5	2	820^{+35+20}_{-34-20}	381^{+2+64}_{-2-64}	17^{+10+17}_{-10-7}	1218^{+36+69}_{-35-67}	1208
55	350–600	600–1200	4–5	2	794^{+29+12}_{-28-12}	324^{+2+54}_{-2-54}	23^{+13+23}_{-13-10}	1141^{+32+60}_{-31-56}	1180
56	350–600	≥1200	4–5	2	$47.8^{+8.2+1.1}_{-7.2-1.1}$	$25.6^{+0.6+4.4}_{-0.6-4.4}$	12^{+7+12}_{-7-5}	85^{+11+12}_{-10-7}	78
57	600–850	600–1200	4–5	2	$37.1^{+8.0+0.7}_{-6.7-0.7}$	$55.5^{+0.9+9.6}_{-0.9-9.6}$	$0.45^{+0.30+0.45}_{-0.30-0.16}$	$93.1^{+8.0+9.7}_{-6.8-9.6}$	98
58	600–850	≥1200	4–5	2	$8.8^{+5.3+0.1}_{-3.5-0.1}$	$8.4^{+0.4+1.5}_{-0.3-1.5}$	$0.20^{+0.18+0.19}_{-0.18-0.02}$	$17.4^{+5.3+1.5}_{-3.6-1.5}$	15
59	≥850	850–1700	4–5	2	$1.2^{+2.8+0.0}_{-1.0-0.0}$	$12.0^{+0.4+2.3}_{-0.4-2.2}$	$0.09^{+0.07+0.09}_{-0.07-0.02}$	$13.3^{+2.8+2.3}_{-1.1-2.2}$	15
60	≥850	≥1700	4–5	2	$0.0^{+2.6+0.0}_{-0.0-0.0}$	$1.44^{+0.16+0.29}_{-0.15-0.28}$	$0.04^{+0.04+0.03}_{-0.04-0.00}$	$1.5^{+2.6+0.3}_{-0.1-0.3}$	1
61	300–350	300–600	4–5	≥3	147^{+15+2}_{-14-2}	40^{+0+14}_{-0-14}	$4.4^{+4.2+6.1}_{-4.2-0.2}$	192^{+15+16}_{-14-15}	222
62	300–350	600–1200	4–5	≥3	$76.7^{+9.5+1.3}_{-8.5-1.3}$	$13.5^{+0.1+4.8}_{-0.1-4.8}$	9^{+6+12}_{-6-3}	99^{+11+13}_{-10-6}	92
63	300–350	≥1200	4–5	≥3	$5.8^{+3.9+0.1}_{-2.5-0.1}$	$1.14^{+0.04+0.41}_{-0.04-0.41}$	$3.7^{+2.2+4.7}_{-2.2-1.5}$	$10.6^{+4.5+4.7}_{-3.3-1.5}$	5
64	350–600	350–600	4–5	≥3	73^{+11+1}_{-10-1}	33^{+0+12}_{-0-12}	$1.2^{+1.1+1.6}_{-1.1-0.1}$	107^{+11+12}_{-10-12}	111
65	350–600	600–1200	4–5	≥3	92^{+11+2}_{-10-2}	30^{+0+11}_{-0-11}	$3.2^{+2.0+4.2}_{-2.0-1.2}$	125^{+11+12}_{-10-11}	138
66	350–600	≥1200	4–5	≥3	$5.0^{+3.4+0.1}_{-2.2-0.1}$	$2.45^{+0.06+0.88}_{-0.06-0.87}$	$1.8^{+1.2+2.3}_{-1.2-0.6}$	$9.3^{+3.6+2.5}_{-2.5-1.1}$	5
67	600–850	600–1200	4–5	≥3	$1.3^{+2.9+0.0}_{-1.1-0.0}$	$4.9^{+0.1+1.8}_{-0.1-1.8}$	$0.10^{+0.12+0.13}_{-0.10-0.00}$	$6.3^{+2.9+1.8}_{-1.1-1.8}$	5
68	600–850	≥1200	4–5	≥3	$0.0^{+2.6+0.0}_{-0.0-0.0}$	$0.79^{+0.04+0.28}_{-0.03-0.28}$	$0.10^{+0.12+0.13}_{-0.10-0.00}$	$0.9^{+2.6+0.3}_{-0.1-0.3}$	0
69	≥850	850–1700	4–5	≥3	$0.0^{+3.2+0.0}_{-0.0-0.0}$	$1.05^{+0.04+0.38}_{-0.04-0.38}$	$0.10^{+0.09+0.13}_{-0.09-0.02}$	$1.2^{+3.2+0.4}_{-0.1-0.4}$	1
70	≥850	≥1700	4–5	≥3	$0.0^{+2.3+0.0}_{-0.0-0.0}$	$0.13^{+0.01+0.05}_{-0.01-0.05}$	$0.04^{+0.05+0.05}_{-0.04-0.00}$	$0.2^{+2.3+0.1}_{-0.0-0.1}$	0

FIGURE E.2: Numerical results for N_{jet} bins [4,5]

Bin	H_T^{miss} [GeV]	H_T [GeV]	N_{jet}	$N_{b\text{-jet}}$	Lost-lepton background	$Z \to \nu\bar{\nu}$ background	QCD background	Total background	Observed
71	300–350	300–600	6–7	0	686^{+29+11}_{-28-11}	761^{+17+63}_{-17-62}	144^{+83+92}_{-83-61}	$1590^{+89+110}_{-89-87}$	1480
72	300–350	600–1200	6–7	0	967^{+25+14}_{-25-14}	873^{+18+66}_{-18-63}	$275^{+140+130}_{-140-130}$	$2114^{+140+150}_{-140-150}$	1993
73	300–350	≥1200	6–7	0	$121.5^{+9.1+2.8}_{-8.5-2.8}$	$116.8^{+7.5+9.3}_{-7.1-9.1}$	172^{+86+74}_{-86-74}	410^{+87+75}_{-87-75}	362
74	350–600	350–600	6–7	0	353^{+21+8}_{-20-8}	514^{+14+41}_{-14-40}	33^{+20+15}_{-20-13}	901^{+32+45}_{-31-43}	847
75	350–600	600–1200	6–7	0	1219^{+28+28}_{-28-28}	$1542^{+24+110}_{-24-110}$	130^{+65+63}_{-65-63}	$2891^{+75+130}_{-74-130}$	2842
76	350–600	≥1200	6–7	0	208^{+11+4}_{-11-4}	258^{+11+19}_{-10-18}	81^{+40+35}_{-40-35}	547^{+43+40}_{-43-39}	553
77	600–850	600–1200	6–7	0	$76.1^{+7.2+1.0}_{-6.6-1.0}$	182^{+8+15}_{-8-14}	$1.70^{+0.88+0.82}_{-0.88-0.81}$	259^{+11+15}_{-10-14}	245
78	600–850	≥1200	6–7	0	$29.7^{+4.4+0.5}_{-3.9-0.5}$	$72.8^{+5.8+5.8}_{-5.4-5.6}$	$2.3^{+1.2+1.0}_{-1.2-1.0}$	$104.8^{+7.4+5.9}_{-6.7-5.8}$	122
79	≥850	850–1700	6–7	0	$18.5^{+3.8+0.3}_{-3.2-0.3}$	$35.2^{+3.8+3.9}_{-3.4-3.8}$	$0.10^{+0.07+0.04}_{-0.07-0.02}$	$53.8^{+5.4+3.9}_{-4.7-3.8}$	55
80	≥850	≥1700	6–7	0	$4.3^{+2.0+0.2}_{-1.4-0.2}$	$12.7^{+2.5+1.9}_{-2.1-1.9}$	$0.05^{+0.04+0.02}_{-0.04-0.01}$	$17.0^{+3.2+1.9}_{-2.6-1.9}$	20
81	300–350	300–600	6–7	1	675^{+25+12}_{-24-12}	248^{+6+45}_{-6-45}	42^{+22+27}_{-22-20}	965^{+34+54}_{-33-51}	946
82	300–350	600–1200	6–7	1	950^{+26+15}_{-25-15}	289^{+6+52}_{-6-52}	115^{+58+55}_{-58-55}	1355^{+63+77}_{-63-77}	1282
83	300–350	≥1200	6–7	1	$105.6^{+9.1+2.7}_{-8.4-2.7}$	$39.3^{+2.5+7.1}_{-2.4-7.1}$	57^{+28+24}_{-28-24}	201^{+30+26}_{-30-26}	197
84	350–600	350–600	6–7	1	252^{+16+5}_{-16-5}	168^{+5+30}_{-4-30}	$9.5^{+5.0+4.3}_{-5.0-4.3}$	429^{+18+31}_{-17-31}	425
85	350–600	600–1200	6–7	1	1050^{+28+19}_{-27-19}	510^{+8+91}_{-8-91}	53^{+27+26}_{-27-26}	1614^{+39+97}_{-39-96}	1521
86	350–600	≥1200	6–7	1	155^{+11+4}_{-10-4}	86^{+4+15}_{-3-15}	26^{+13+11}_{-13-11}	268^{+17+20}_{-17-20}	269
87	600–850	600–1200	6–7	1	$34.7^{+5.4+0.6}_{-4.8-0.6}$	60^{+3+11}_{-3-11}	$0.69^{+0.41+0.33}_{-0.41-0.28}$	95^{+6+11}_{-6-11}	90
88	600–850	≥1200	6–7	1	$25.9^{+4.7+0.4}_{-4.0-0.4}$	$24.4^{+1.9+4.4}_{-1.8-4.4}$	$0.59^{+0.34+0.26}_{-0.34-0.25}$	$50.9^{+5.1+4.4}_{-4.4-4.4}$	49
89	≥850	850–1700	6–7	1	$7.9^{+2.9+0.1}_{-2.2-0.1}$	$11.5^{+1.2+2.3}_{-1.1-2.2}$	$0.05^{+0.04+0.02}_{-0.04-0.00}$	$19.4^{+3.2+2.3}_{-2.5-2.2}$	17
90	≥850	≥1700	6–7	1	$1.5^{+2.0+0.0}_{-1.0-0.0}$	$4.29^{+0.85+0.96}_{-0.72-0.95}$	$0.04^{+0.05+0.02}_{-0.04-0.00}$	$5.9^{+2.2+1.0}_{-1.2-0.9}$	7
91	300–350	300–600	6–7	2	376^{+19+8}_{-18-8}	64^{+2+13}_{-1-13}	$9.8^{+5.5+6.3}_{-5.5-4.2}$	450^{+20+17}_{-19-16}	450
92	300–350	600–1200	6–7	2	693^{+23+10}_{-22-10}	76^{+2+15}_{-2-15}	34^{+17+16}_{-17-16}	803^{+28+25}_{-28-25}	797
93	300–350	≥1200	6–7	2	$46.7^{+6.4+0.7}_{-5.7-0.7}$	$10.5^{+0.7+2.1}_{-0.6-2.1}$	$18.7^{+9.4+8.1}_{-9.4-8.1}$	76^{+11+8}_{-11-8}	84
94	350–600	350–600	6–7	2	120^{+12+2}_{-11-2}	$43.6^{+1.2+8.9}_{-1.2-8.9}$	$2.1^{+1.2+0.9}_{-1.2-0.9}$	165^{+12+9}_{-11-9}	188
95	350–600	600–1200	6–7	2	661^{+23+11}_{-22-11}	134^{+2+27}_{-2-27}	$14.6^{+7.5+7.0}_{-7.5-7.0}$	809^{+24+30}_{-24-30}	762
96	350–600	≥1200	6–7	2	$66.6^{+7.7+2.2}_{-7.0-2.2}$	$22.8^{+0.9+4.6}_{-0.9-4.6}$	$7.5^{+3.8+3.2}_{-3.8-3.2}$	$96.9^{+8.7+6.0}_{-8.0-6.0}$	106
97	600–850	600–1200	6–7	2	$19.3^{+4.7+0.3}_{-3.9-0.3}$	$15.7^{+0.7+3.2}_{-0.7-3.2}$	$0.15^{+0.10+0.07}_{-0.10-0.05}$	$35.2^{+4.7+3.2}_{-4.0-3.2}$	32
98	600–850	≥1200	6–7	2	$8.0^{+3.2+0.2}_{-2.4-0.2}$	$6.5^{+0.5+1.3}_{-0.5-1.3}$	$0.09^{+0.07+0.04}_{-0.07-0.01}$	$14.5^{+3.3+1.3}_{-2.4-1.3}$	14
99	≥850	850–1700	6–7	2	$1.8^{+1.7+0.0}_{-1.0-0.0}$	$2.98^{+0.32+0.65}_{-0.29-0.65}$	$0.05^{+0.04+0.02}_{-0.04-0.01}$	$4.8^{+1.8+0.7}_{-1.0-0.7}$	9
100	≥850	≥1700	6–7	2	$0.5^{+1.2+0.0}_{-0.4-0.0}$	$1.15^{+0.23+0.28}_{-0.19-0.28}$	$0.02^{+0.02+0.01}_{-0.02-0.00}$	$1.7^{+1.2+0.3}_{-0.5-0.3}$	1
101	300–350	300–600	6–7	≥3	$67.8^{+8.8+1.6}_{-7.9-1.6}$	$8.8^{+0.2+3.7}_{-0.2-3.7}$	$1.4^{+1.0+0.9}_{-1.0-0.4}$	$78.0^{+8.9+4.1}_{-8.0-4.0}$	86
102	300–350	600–1200	6–7	≥3	136^{+11+2}_{-10-2}	$10.5^{+0.2+4.3}_{-0.2-4.3}$	$7.4^{+4.2+3.6}_{-4.2-3.2}$	154^{+11+6}_{-11-6}	167
103	300–350	≥1200	6–7	≥3	$15.7^{+4.1+0.2}_{-3.4-0.2}$	$1.44^{+0.09+0.59}_{-0.09-0.59}$	$3.9^{+2.2+1.7}_{-2.2-1.7}$	$21.1^{+4.7+1.8}_{-4.0-1.8}$	16
104	350–600	350–600	6–7	≥3	$20.6^{+5.3+0.5}_{-4.3-0.5}$	$6.0^{+0.2+2.5}_{-0.2-2.5}$	$0.68^{+0.62+0.31}_{-0.62-0.07}$	$27.2^{+5.4+2.5}_{-4.4-2.5}$	28
105	350–600	600–1200	6–7	≥3	137^{+11+4}_{-10-4}	$18.5^{+0.3+7.6}_{-0.3-7.6}$	$2.8^{+1.6+1.4}_{-1.6-1.2}$	158^{+11+9}_{-10-9}	115
106	350–600	≥1200	6–7	≥3	$15.4^{+4.4+0.6}_{-3.5-0.6}$	$3.1^{+0.1+1.3}_{-0.1-1.3}$	$1.7^{+1.0+0.8}_{-1.0-0.7}$	$20.2^{+4.5+1.6}_{-3.7-1.6}$	23
107	600–850	600–1200	6–7	≥3	$4.1^{+2.5+0.0}_{-1.7-0.0}$	$2.16^{+0.10+0.89}_{-0.09-0.89}$	$0.05^{+0.06+0.02}_{-0.05-0.00}$	$6.3^{+2.5+0.9}_{-1.7-0.9}$	6
108	600–850	≥1200	6–7	≥3	$2.1^{+2.0+0.0}_{-1.1-0.0}$	$0.89^{+0.07+0.37}_{-0.07-0.37}$	$0.07^{+0.06+0.03}_{-0.06-0.01}$	$3.0^{+2.0+0.4}_{-1.1-0.4}$	2
109	≥850	850–1700	6–7	≥3	$0.0^{+1.2+0.0}_{-0.0-0.0}$	$0.41^{+0.04+0.17}_{-0.04-0.17}$	$0.05^{+0.04+0.02}_{-0.04-0.01}$	$0.5^{+1.2+0.2}_{-0.1-0.2}$	1
110	≥850	≥1700	6–7	≥3	$0.0^{+1.9+0.0}_{-0.0-0.0}$	$0.16^{+0.03+0.07}_{-0.03-0.07}$	$0.02^{+0.02+0.01}_{-0.02-0.00}$	$0.2^{+1.9+0.1}_{-0.0-0.1}$	1

FIGURE E.3: Numerical results for N_{jet} bins [6,7]

Bin	H_T^{miss} [GeV]	H_T [GeV]	N_{jet}	$N_{b\text{-jet}}$	Lost-lepton background	$Z \to \nu\bar{\nu}$ background	QCD background	Total background	Observed
111	300–350	600–1200	8–9	0	$139.5^{+9.8+1.9}_{-9.2-1.9}$	$60.0^{+4.7+9.8}_{-4.4-9.7}$	58^{+29+28}_{-29-28}	258^{+31+30}_{-31-30}	245
112	300–350	≥1200	8–9	0	$31.0^{+4.5+1.1}_{-4.0-1.1}$	$25.1^{+3.7+2.8}_{-3.2-2.7}$	57^{+28+24}_{-28-24}	113^{+29+25}_{-29-25}	88
113	350–600	600–1200	8–9	0	$136.1^{+9.5+1.7}_{-9.0-1.7}$	123^{+7+14}_{-7-13}	30^{+15+14}_{-15-14}	289^{+19+20}_{-19-20}	280
114	350–600	≥1200	8–9	0	$49.9^{+5.5+0.9}_{-5.0-0.9}$	$52.2^{+5.0+5.6}_{-4.6-5.3}$	27^{+14+12}_{-14-12}	129^{+16+13}_{-15-13}	104
115	600–850	600–1200	8–9	0	$6.6^{+2.3+0.2}_{-1.8-0.2}$	$13.9^{+2.6+1.5}_{-2.2-1.5}$	$0.37^{+0.21+0.18}_{-0.21-0.16}$	$20.9^{+3.5+1.6}_{-2.9-1.5}$	28
116	600–850	≥1200	8–9	0	$6.1^{+2.1+0.1}_{-1.6-0.1}$	$12.9^{+2.6+1.6}_{-2.2-1.6}$	$0.79^{+0.44+0.34}_{-0.44-0.34}$	$19.7^{+3.3+1.7}_{-2.7-1.6}$	22
117	≥850	850–1700	8–9	0	$1.1^{+1.1+0.0}_{-0.6-0.0}$	$4.1^{+1.5+0.6}_{-1.2-0.6}$	$0.06^{+0.04+0.03}_{-0.04-0.02}$	$5.3^{+1.9+0.6}_{-1.3-0.6}$	2
118	≥850	≥1700	8–9	0	$1.5^{+1.2+0.1}_{-0.7-0.1}$	$2.2^{+1.3+0.3}_{-0.9-0.3}$	$0.02^{+0.02+0.01}_{-0.02-0.00}$	$3.7^{+1.8+0.3}_{-1.1-0.3}$	1
119	300–350	600–1200	8–9	1	183^{+11+3}_{-11-3}	37^{+3+11}_{-3-11}	27^{+13+13}_{-13-13}	247^{+18+17}_{-17-17}	229
120	300–350	≥1200	8–9	1	$43.8^{+5.5+0.7}_{-5.0-0.7}$	$13.8^{+2.0+3.8}_{-1.8-3.8}$	24^{+12+10}_{-12-10}	82^{+13+11}_{-13-11}	68
121	350–600	600–1200	8–9	1	176^{+11+3}_{-10-3}	75^{+4+21}_{-4-21}	$10.9^{+5.5+5.3}_{-5.5-5.3}$	262^{+13+22}_{-12-22}	224
122	350–600	≥1200	8–9	1	$68.4^{+6.7+1.2}_{-6.2-1.2}$	$29.5^{+2.8+8.2}_{-2.6-8.1}$	$9.8^{+5.0+4.2}_{-5.0-4.2}$	$107.8^{+8.8+9.3}_{-8.3-9.2}$	90
123	600–850	600–1200	8–9	1	$3.4^{+2.0+0.2}_{-1.4-0.2}$	$8.7^{+1.6+2.4}_{-1.4-2.4}$	$0.10^{+0.08+0.05}_{-0.08-0.02}$	$12.2^{+2.6+2.4}_{-2.0-2.4}$	7
124	600–850	≥1200	8–9	1	$8.3^{+2.8+0.1}_{-2.1-0.1}$	$8.1^{+1.6+2.3}_{-1.4-2.3}$	$0.31^{+0.18+0.13}_{-0.18-0.12}$	$16.7^{+3.2+2.3}_{-2.6-2.3}$	15
125	≥850	850–1700	8–9	1	$0.0^{+1.2+0.0}_{-0.0-0.0}$	$2.08^{+0.79+0.61}_{-0.59-0.61}$	$0.05^{+0.04+0.02}_{-0.04-0.01}$	$2.1^{+1.5+0.6}_{-0.6-0.6}$	2
126	≥850	≥1700	8–9	1	$1.0^{+1.3+0.0}_{-0.7-0.0}$	$1.35^{+0.81+0.41}_{-0.54-0.40}$	$0.02^{+0.02+0.01}_{-0.02-0.00}$	$2.4^{+1.5+0.4}_{-0.8-0.4}$	2
127	300–350	600–1200	8–9	2	169^{+11+4}_{-10-4}	$11.0^{+0.9+4.1}_{-0.8-4.1}$	$9.5^{+4.9+4.6}_{-4.9-4.6}$	190^{+12+7}_{-11-7}	193
128	300–350	≥1200	8–9	2	$28.9^{+4.7+0.5}_{-4.1-0.5}$	$5.5^{+0.8+1.9}_{-0.7-1.9}$	$10.1^{+5.1+4.4}_{-5.1-4.4}$	$44.6^{+7.0+4.8}_{-6.6-4.8}$	53
129	350–600	600–1200	8–9	2	146^{+10+2}_{-10-2}	$23.1^{+1.3+8.2}_{-1.2-8.1}$	$4.5^{+2.4+2.2}_{-2.4-2.1}$	174^{+11+9}_{-10-9}	158
130	350–600	≥1200	8–9	2	$42.9^{+5.6+0.9}_{-5.0-0.9}$	$11.0^{+1.1+3.9}_{-1.0-3.9}$	$4.1^{+2.1+1.8}_{-2.1-1.8}$	$58.0^{+6.1+4.4}_{-5.5-4.3}$	74
131	600–850	600–1200	8–9	2	$3.6^{+2.4+0.2}_{-1.6-0.2}$	$2.52^{+0.47+0.89}_{-0.40-0.89}$	$0.09^{+0.08+0.04}_{-0.08-0.01}$	$6.2^{+2.5+0.9}_{-1.6-0.9}$	7
132	600–850	≥1200	8–9	2	$8.0^{+2.9+0.3}_{-2.2-0.3}$	$2.30^{+0.46+0.82}_{-0.39-0.82}$	$0.08^{+0.09+0.04}_{-0.09-0.00}$	$10.4^{+3.0+0.9}_{-2.3-0.9}$	9
133	≥850	850–1700	8–9	2	$0.7^{+1.6+0.0}_{-0.6-0.0}$	$0.96^{+0.37+0.35}_{-0.27-0.35}$	$0.05^{+0.04+0.02}_{-0.04-0.01}$	$1.7^{+1.6+0.3}_{-0.7-0.3}$	0
134	≥850	≥1700	8–9	2	$2.5^{+3.3+0.1}_{-1.7-0.1}$	$0.40^{+0.24+0.15}_{-0.16-0.15}$	$0.02^{+0.02+0.01}_{-0.02-0.00}$	$2.9^{+3.4+0.2}_{-1.7-0.2}$	2
135	300–350	600–1200	8–9	≥3	$46.8^{+6.1+0.7}_{-5.5-0.7}$	$3.8^{+0.3+2.3}_{-0.3-2.3}$	$3.7^{+2.6+1.8}_{-2.6-1.2}$	$54.3^{+6.6+3.0}_{-6.1-2.7}$	57
136	300–350	≥1200	8–9	≥3	$17.3^{+4.0+0.5}_{-3.3-0.5}$	$1.26^{+0.18+0.76}_{-0.16-0.76}$	$3.6^{+2.0+1.5}_{-2.0-1.5}$	$22.2^{+4.4+1.8}_{-3.8-1.8}$	17
137	350–600	600–1200	8–9	≥3	$44.4^{+5.9+1.0}_{-5.3-1.0}$	$7.5^{+0.4+4.6}_{-0.4-4.6}$	$1.31^{+0.81+0.63}_{-0.81-0.51}$	$53.2^{+6.0+4.7}_{-5.4-4.7}$	36
138	350–600	≥1200	8–9	≥3	$15.2^{+3.6+0.3}_{-2.9-0.3}$	$2.8^{+0.3+1.7}_{-0.2-1.7}$	$1.17^{+0.68+0.51}_{-0.68-0.49}$	$19.2^{+3.6+1.8}_{-3.0-1.8}$	23
139	600–850	600–1200	8–9	≥3	$0.0^{+1.7+0.0}_{-0.0-0.0}$	$0.88^{+0.16+0.54}_{-0.14-0.53}$	$0.04^{+0.04+0.02}_{-0.04-0.00}$	$0.9^{+1.7+0.5}_{-0.1-0.5}$	2
140	600–850	≥1200	8–9	≥3	$2.7^{+2.2+0.1}_{-1.3-0.1}$	$0.83^{+0.17+0.51}_{-0.14-0.51}$	$0.05^{+0.05+0.02}_{-0.05-0.00}$	$3.6^{+2.2+0.5}_{-1.3-0.5}$	2
141	≥850	850–1700	8–9	≥3	$0.8^{+2.0+0.0}_{-0.7-0.0}$	$0.18^{+0.07+0.11}_{-0.05-0.11}$	$0.05^{+0.04+0.02}_{-0.04-0.01}$	$1.1^{+2.0+0.1}_{-0.7-0.1}$	0
142	≥850	≥1700	8–9	≥3	$0.0^{+1.8+0.0}_{-0.0-0.0}$	$0.14^{+0.08+0.08}_{-0.05-0.08}$	$0.02^{+0.02+0.01}_{-0.02-0.00}$	$0.2^{+1.8+0.1}_{-0.1-0.1}$	0

FIGURE E.4: Numerical results for N_{jet} bins [8,9]

Bin	H_T^{miss} [GeV]	H_T [GeV]	N_{jet}	$N_{b\text{-jet}}$	Lost-lepton background	$Z \to \nu\bar{\nu}$ background	QCD background	Total background	Observed
143	300–350	600–1200	≥10	0	$5.7^{+2.2+0.3}_{-1.7-0.3}$	$2.9^{+1.3+0.6}_{-1.0-0.5}$	$7.8^{+4.5+3.7}_{-4.5-3.3}$	$16.4^{+5.2+3.8}_{-4.9-3.3}$	17
144	300–350	≥1200	≥10	0	$5.7^{+2.5+0.2}_{-1.8-0.2}$	$2.5^{+1.5+0.4}_{-1.0-0.3}$	$12.6^{+6.3+5.4}_{-6.3-5.4}$	$20.8^{+7.0+5.4}_{-6.7-5.4}$	20
145	350–600	600–1200	≥10	0	$6.0^{+2.4+0.1}_{-1.8-0.1}$	$4.2^{+1.6+0.6}_{-1.2-0.6}$	$3.3^{+1.8+1.6}_{-1.8-1.5}$	$13.6^{+3.4+1.7}_{-2.8-1.6}$	12
146	350–600	≥1200	≥10	0	$10.7^{+2.9+0.2}_{-2.3-0.2}$	$6.5^{+2.1+0.9}_{-1.6-0.9}$	$6.0^{+3.1+2.6}_{-3.1-2.6}$	$23.2^{+4.7+2.8}_{-4.2-2.8}$	21
147	600–850	600–1200	≥10	0	$0.19^{+0.44+0.00}_{-0.17-0.00}$	$0.36^{+0.84+0.05}_{-0.30-0.05}$	$0.07^{+0.07+0.03}_{-0.07-0.00}$	$0.63^{+0.95+0.06}_{-0.35-0.05}$	2
148	600–850	≥1200	≥10	0	$2.0^{+1.6+0.0}_{-1.0-0.0}$	$1.5^{+1.2+0.2}_{-0.7-0.2}$	$0.15^{+0.13+0.06}_{-0.13-0.02}$	$3.6^{+2.0+0.2}_{-1.2-0.2}$	6
149	≥850	850–1700	≥10	0	$0.0^{+2.3+0.0}_{-0.0-0.0}$	$0.00^{+0.64+0.00}_{-0.00-0.00}$	$0.05^{+0.04+0.02}_{-0.04-0.01}$	$0.0^{+2.4+0.0}_{-0.0-0.0}$	0
150	≥850	≥1700	≥10	0	$0.00^{+0.91+0.00}_{-0.00-0.00}$	$0.42^{+0.96+0.07}_{-0.35-0.07}$	$0.02^{+0.02+0.01}_{-0.02-0.00}$	$0.4^{+1.3+0.1}_{-0.3-0.1}$	2
151	300–350	600–1200	≥10	1	$15.2^{+3.3+0.2}_{-2.8-0.2}$	$1.24^{+0.56+0.90}_{-0.40-0.90}$	$4.0^{+2.1+1.9}_{-2.1-1.9}$	$20.4^{+4.0+2.1}_{-3.5-2.1}$	22
152	300–350	≥1200	≥10	1	$11.2^{+3.2+0.4}_{-2.6-0.4}$	$1.05^{+0.63+0.76}_{-0.42-0.76}$	$6.9^{+3.5+3.0}_{-3.5-3.0}$	$19.2^{+4.8+3.1}_{-4.4-3.1}$	18
153	350–600	600–1200	≥10	1	$13.8^{+3.3+0.3}_{-2.7-0.3}$	$1.8^{+0.7+1.3}_{-0.5-1.3}$	$1.53^{+0.85+0.74}_{-0.85-0.68}$	$17.1^{+3.5+1.5}_{-2.9-1.5}$	9
154	350–600	≥1200	≥10	1	$16.2^{+3.4+0.4}_{-2.9-0.4}$	$2.7^{+0.9+2.0}_{-0.7-2.0}$	$2.6^{+1.3+1.1}_{-1.3-1.1}$	$21.5^{+3.8+2.3}_{-3.2-2.3}$	32
155	600–850	600–1200	≥10	1	$0.0^{+3.6+0.0}_{-0.0-0.0}$	$0.15^{+0.35+0.11}_{-0.13-0.09}$	$0.04^{+0.04+0.02}_{-0.04-0.00}$	$0.2^{+3.6+0.1}_{-0.1-0.1}$	0
156	600–850	≥1200	≥10	1	$1.3^{+1.3+0.0}_{-0.7-0.0}$	$0.61^{+0.49+0.44}_{-0.29-0.44}$	$0.06^{+0.05+0.03}_{-0.05-0.01}$	$2.0^{+1.4+0.5}_{-0.8-0.4}$	3
157	≥850	850–1700	≥10	1	$0.0^{+3.2+0.0}_{-0.0-0.0}$	$0.00^{+0.27+0.00}_{-0.00-0.00}$	$0.05^{+0.04+0.02}_{-0.04-0.01}$	$0.0^{+3.2+0.0}_{-0.0-0.0}$	0
158	≥850	≥1700	≥10	1	$0.7^{+1.5+0.0}_{-0.6-0.0}$	$0.18^{+0.41+0.13}_{-0.15-0.10}$	$0.03^{+0.04+0.01}_{-0.03-0.00}$	$0.9^{+1.6+0.1}_{-0.6-0.1}$	1
159	300–350	600–1200	≥10	2	$13.1^{+3.2+0.3}_{-2.6-0.3}$	$0.38^{+0.18+0.42}_{-0.13-0.36}$	$2.1^{+1.5+1.0}_{-1.5-0.6}$	$15.5^{+3.5+1.1}_{-3.0-0.8}$	15
160	300–350	≥1200	≥10	2	$10.8^{+3.0+0.4}_{-2.4-0.4}$	$0.33^{+0.19+0.36}_{-0.13-0.30}$	$3.3^{+1.7+1.4}_{-1.7-1.4}$	$14.4^{+3.5+1.5}_{-3.0-1.5}$	11
161	350–600	600–1200	≥10	2	$18.2^{+3.8+0.3}_{-3.2-0.3}$	$0.55^{+0.21+0.60}_{-0.16-0.53}$	$0.77^{+0.52+0.37}_{-0.52-0.26}$	$19.5^{+3.8+0.8}_{-3.3-0.7}$	11
162	350–600	≥1200	≥10	2	$13.7^{+3.2+0.3}_{-2.6-0.3}$	$0.85^{+0.27+0.92}_{-0.21-0.82}$	$1.15^{+0.66+0.50}_{-0.66-0.50}$	$15.7^{+3.3+1.1}_{-2.7-1.0}$	12
163	600–850	600–1200	≥10	2	$1.6^{+2.2+0.0}_{-1.2-0.0}$	$0.05^{+0.11+0.05}_{-0.04-0.03}$	$0.04^{+0.04+0.02}_{-0.04-0.00}$	$1.7^{+2.2+0.1}_{-1.2-0.0}$	0
164	600–850	≥1200	≥10	2	$0.9^{+1.2+0.0}_{-0.6-0.0}$	$0.19^{+0.15+0.21}_{-0.09-0.17}$	$0.06^{+0.05+0.03}_{-0.05-0.01}$	$1.2^{+1.2+0.2}_{-0.6-0.2}$	0
165	≥850	850–1700	≥10	2	$0.0^{+2.4+0.0}_{-0.0-0.0}$	$0.00^{+0.08+0.00}_{-0.00-0.00}$	$0.05^{+0.04+0.02}_{-0.04-0.01}$	$0.0^{+2.4+0.0}_{-0.0-0.0}$	0
166	≥850	≥1700	≥10	2	$0.0^{+1.5+0.0}_{-0.0-0.0}$	$0.05^{+0.13+0.06}_{-0.04-0.03}$	$0.02^{+0.02+0.01}_{-0.02-0.00}$	$0.1^{+1.5+0.1}_{-0.0-0.0}$	0
167	300–350	600–1200	≥10	≥3	$6.4^{+2.4+0.1}_{-1.8-0.1}$	$0.36^{+0.17+0.41}_{-0.12-0.34}$	$0.46^{+0.32+0.22}_{-0.32-0.14}$	$7.2^{+2.4+0.5}_{-1.8-0.4}$	13
168	300–350	≥1200	≥10	≥3	$3.8^{+2.1+0.1}_{-1.4-0.1}$	$0.31^{+0.19+0.35}_{-0.12-0.28}$	$1.50^{+0.87+0.65}_{-0.87-0.63}$	$5.6^{+2.3+0.7}_{-1.7-0.7}$	5
169	350–600	600–1200	≥10	≥3	$1.6^{+1.5+0.0}_{-0.9-0.0}$	$0.52^{+0.20+0.59}_{-0.15-0.50}$	$0.11^{+0.12+0.05}_{-0.11-0.00}$	$2.2^{+1.6+0.6}_{-0.9-0.5}$	3
170	350–600	≥1200	≥10	≥3	$4.2^{+2.1+0.1}_{-1.4-0.1}$	$0.81^{+0.26+0.90}_{-0.20-0.78}$	$0.71^{+0.44+0.31}_{-0.44-0.27}$	$5.7^{+2.1+0.9}_{-1.5-0.8}$	9
171	600–850	600–1200	≥10	≥3	$0.0^{+3.0+0.0}_{-0.0-0.0}$	$0.05^{+0.10+0.05}_{-0.04-0.03}$	$0.04^{+0.04+0.02}_{-0.04-0.00}$	$0.1^{+3.0+0.1}_{-0.1-0.0}$	0
172	600–850	≥1200	≥10	≥3	$0.0^{+1.4+0.0}_{-0.0-0.0}$	$0.18^{+0.14+0.20}_{-0.09-0.16}$	$0.04^{+0.04+0.02}_{-0.04-0.00}$	$0.2^{+1.4+0.2}_{-0.1-0.2}$	1
173	≥850	850–1700	≥10	≥3	$0.0^{+2.0+0.0}_{-0.0-0.0}$	$0.00^{+0.08+0.00}_{-0.00-0.00}$	$0.05^{+0.04+0.02}_{-0.04-0.01}$	$0.0^{+2.0+0.0}_{-0.0-0.0}$	0
174	≥850	≥1700	≥10	≥3	$0.0^{+1.3+0.0}_{-0.0-0.0}$	$0.05^{+0.12+0.06}_{-0.04-0.03}$	$0.02^{+0.02+0.01}_{-0.02-0.00}$	$0.1^{+1.3+0.1}_{-0.0-0.0}$	0

FIGURE E.5: Numerical results for N_{jet} bin ≥10

Appendix F

Background composition against search variables

The following tables give aggregate predicted pre-fit background predictions of Lost-lepton, $Z \to \nu\overline{\nu}$ and QCD in data obtained as a function of one or two dimensional search variables. These aggregate background predictions are obtained by combining the predictions in appropriate search bins as given by tables E.1, E.2, E.3, E.4, E.5.

TABLE F.1: Variation against $H_{\mathrm{T}}^{\mathrm{miss}}$, N_{jet}, $N_{\mathrm{b-jet}}$

variable	range	Lost-lepton	$Z \to \nu\overline{\nu}$	QCD
$H_{\mathrm{T}}^{\mathrm{miss}}$	300-350	74623	131156	6005
	350-600	21055	29919	1225
	600-850	2377	10027	32.82
	≥ 850	223	1838	1.98
N_{jet}	2-3	86647	219176	3897
	4-5	37151	51669	2873
	6-7	9332	6237	1148
	≥ 8	1547	534	339
$N_{\mathrm{b-jet}}$	0	101835	241468	6021
	1	23854	32000	1836
	≥ 2	8988	4187	500

TABLE F.2: Variation against N_{jet}, $N_{\mathrm{b-jet}}$while $H_{\mathrm{T}}>$ 300 GeV and $H_{\mathrm{T}}^{\mathrm{miss}}>$ 300 GeV

[N_{jet},$N_{\mathrm{b-jet}}$]	Lost-lepton	$Z \to \nu\bar{\nu}$	QCD
[2-3,0]	75031	195288	3127
[2-3,1]	10258	21766	663
[2-3,≥2]	1357	2120	107
[4-5,0]	22718	41500	1852
[4-5,1]	9796	8609	781
[4-5,≥2]	4636	1558	239
[6-7,0]	3683	4367	839
[6-7,1]	3257	1440	303
[6-7,≥2]	2391	429	105
[≥8,0]	402	311	203
[≥8,1]	542	183	87
[≥8,≥2]	603	78	48.8

TABLE F.3: Variation against N_{jet}, $N_{\mathrm{b-jet}}$while $H_{\mathrm{T}}>$ 600 GeV and $H_{\mathrm{T}}^{\mathrm{miss}}>$ 600 GeV

[N_{jet},$N_{\mathrm{b-jet}}$]	Lost-lepton	$Z \to \nu\bar{\nu}$	QCD
[2-3,0]	1390	7662	12.19
[2-3,1]	173	871	2.6
[2-3,≥2]	17.5	88	0.3
[4-5,0]	548	2206	7.18
[4-5,1]	140	461	2.7
[4-5,≥2]	48.4	84	1.12
[6-7,0]	128	302	4.15
[6-7,1]	70	100	1.37
[6-7,≥2]	35.8	29.9	0.5
[≥8,0]	17.5	35.4	1.53
[≥8,1]	14.7	21.17	0.66
[≥8,≥2]	20.8	8.78	0.72

Appendix G

Lost-lepton background composition against search variables

The following table shows the variation in composition of Lost-lepton background obtained from MC as a function of one search variable at a time, and scaled to 137 fb^{-1}luminosity.

TABLE G.1: Composition of various contributing processes to Lost-lepton background as a function of search variables H_{T}, $H_{\mathrm{T}}^{\mathrm{miss}}$, N_{jet}, $N_{\mathrm{b-jet}}$

variable	range	$t\bar{t}$	W+jets	single top
H_{T}	300-500	44052	149264	5131
	500-900	33485	56370	2684
	900-1500	5911	8979	512
	$\geq$ 1500	779	1367	93
$H_{\mathrm{T}}^{\mathrm{miss}}$	300-350	20276	57427	2030
	350-500	12631	44085	1411
	500-750	1317	7849	200
	$\geq$ 750	68	893	14.2
N_{jet}	2-3	18849	156860	3870
	4-5	43293	51747	3474
	6-7	18545	6825	956
	$\geq$ 8	3544	523	120
$N_{\mathrm{b-jet}}$	0	15695	196322	2300
	1	40420	17911	4279
	2	25071	1657	1675
	$\geq$ 3	2990	80	164

Bibliography

[1] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264. URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[2] Abdus Salam. "Gauge unification of fundamental forces". In: *Rev. Mod. Phys.* 52 (3 July 1980), pp. 525–538. DOI: 10.1103/RevModPhys.52.525. URL: https://link.aps.org/doi/10.1103/RevModPhys.52.525.

[3] Sheldon L. Glashow. "Partial-symmetries of weak interactions". In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-5582(61)90469-2. URL: http://www.sciencedirect.com/science/article/pii/0029558261904692.

[4] F. Zwicky. "Die Rotverschiebung von extragalaktischen Nebeln". In: *Helv. Phys. Acta* 6 (1933), p. 110. URL: http://adsabs.harvard.edu/abs/1933AcHPh..6..110Z.

[5] Douglas Clowe et al. "A direct empirical proof of the existence of dark matter". In: *Astrophys. J.* 648 (2006), pp. L109–L113. DOI: 10.1086/508162. arXiv: astro-ph/0608407 [astro-ph].

[6] Gary Steigman. "Observational Tests of Antimatter Cosmologies". In: *Annual Review of Astronomy and Astrophysics* 14.1 (1976), pp. 339–372. DOI: 10.1146/annurev.aa.14.090176.002011. eprint: https://doi.org/10.1146/annurev.aa.14.090176.002011. URL: https://doi.org/10.1146/annurev.aa.14.090176.002011.

[7] Laurent Canetti, Marco Drewes, and Mikhail Shaposhnikov. "Matter and Antimatter in the Universe". In: *New J. Phys.* 14 (2012), p. 095012. DOI: 10.1088/1367-2630/14/9/095012. arXiv: 1204.4186 [hep-ph].

[8] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett. B* 716 (2012), p. 1. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[9] Serguei Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Phys. Lett. B* 716 (2012), p. 30. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[10] N. Seiberg J. H. Schwarz. "String theory, supersymmetry, unification, and all that". In: *Rev. Mod. Phys.* 71 (1999), S112–S120.

[11] Vardan Khachatryan et al. "Inclusive search for supersymmetry using razor variables in pp collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. D* 95 (2017), p. 012003. DOI: 10.1103/PhysRevD.95.012003. arXiv: 1609.07658 [hep-ex].

[12] Vardan Khachatryan et al. "Search for new physics with the M_{T2} variable in all-jets final states produced in pp collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 10 (2016), p. 006. DOI: 10.1007/JHEP10(2016)006. arXiv: 1603.04053 [hep-ex].

[13] Vardan Khachatryan et al. "Search for supersymmetry in the multijet and missing transverse momentum final state in pp collisions at 13 TeV". In: *Phys. Lett. B* 758 (2016), p. 152. DOI: 10.1016/j.physletb.2016.05.002. arXiv: 1602.06581 [hep-ex].

[14] C Borschensky et al. "Squark and gluino production cross sections in pp collisions at $\sqrt{s} = 13, 14, 33 and 100 TeV$". In: *Eur. Phys. J. C* (2014), p. 22. DOI: 10.1140/epjc/s10052-014-3174-y. arXiv: 1407.5066 [hep-ex].

[15] CMS Collaboration. "Search for direct top squark pair production in events with one lepton, jets, and missing transverse momentum at 13 TeV with the CMS experiment ". In: *JHEP* (2019). arXiv: 1912.08887 [hep-ex].

[16] CMS Collaboration. "Searches for pair production of charginos and top squarks in final states with two oppositely charged leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* (2018). DOI: 10.1007/JHEP11(2018)079. arXiv: 1807.07799 [hep-ex].

[17] CMS Collaboration. "Search for supersymmetry with a compressed mass spectrum in events with a soft tau lepton, a highly energetic jet, and large missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. Lett.* (2019). arXiv: 1910.01185.

[18] CMS Collaboration. "Search for Dark Matter and Supersymmetry with a Compressed Mass Spectrum in the Vector Boson Fusion Topology in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV". In: *Phys. Rev. Lett.* (2017). DOI: 10.1103/PhysRevLett.118.021802. arXiv: 1605.09305.

[19] Albert M Sirunyan et al. "Search for supersymmetry in multijet events with missing transverse momentum in proton-proton collisions at 13 TeV". In: *Phys. Rev. D* 96 (2017), p. 032003. DOI: 10.1103/PhysRevD.96.032003. arXiv: 1704.07781 [hep-ex].

[20] Albert M Sirunyan et al. "Search for supersymmetry in proton-proton collisions at 13 TeV in final states with jets and missing transverse momentum". In: *JHEP* (2019), p. 59. arXiv: 1908.04722 [hep-ex].

[21] K. Nakamura and et al. (Particle Data Group). "Review of particle physics". In: *J.Phys.G 37 (2010, and 2011 partial update for the 2012 edition)* (). DOI: 10.1088/0954-3899/37/7A/075021.

[22] "LEP design report. CERN, Geneva". In: *Copies shelved as reports in LEP, PS and SPS libraries.* (1984).

[23] R. R. Wilson. "The Tevatron". In: *Phys.Today* 30N10 (1977), pp. 23–30.

[24] E. . Bruning et al. "LHC design report. Vol. I: The LHC main ring". In: *CERN-2004-003-V-1* ().

[25] Particle Data Group, J. Beringer, et al. "Review of Particle Physics". In: *Phys. Rev. D* 86 (2012), p. 010001. DOI: 10.1103/PhysRevD.86.010001.

[26] E. Noether. "Invariante Variationsprobleme". ger. In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1918 (1918), pp. 235–257. URL: http://eudml.org/doc/59024.

[27] David J. Gross and Frank Wilczek. "Ultraviolet Behavior of Non-Abelian Gauge Theories". In: *Phys. Rev. Lett.* 30 (26 June 1973), pp. 1343–1346. DOI: 10.1103/PhysRevLett.30.1343. URL: https://link.aps.org/doi/10.1103/PhysRevLett.30.1343.

[28] H. David Politzer. "Reliable Perturbative Results for Strong Interactions?" In: *Phys. Rev. Lett.* 30 (26 June 1973), pp. 1346–1349. DOI: 10.1103/PhysRevLett.30.1346. URL: https://link.aps.org/doi/10.1103/PhysRevLett.30.1346.

[29] G. Arnison et al. "Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV". In: *Physics Letters B* 122.1 (1983), pp. 103–116. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(83)91177-2. URL: http://www.sciencedirect.com/science/article/pii/0370269383911772.

[30] G. Arnison et al. "Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider". In: *Physics Letters B* 126.5 (1983), pp. 398–410. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(83)90188-0. URL: http://www.sciencedirect.com/science/article/pii/0370269383901880.

[31] M. Banner et al. "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider". In: *Physics Letters B* 122.5 (1983), pp. 476–485. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(83)91605-2. URL: http://www.sciencedirect.com/science/article/pii/0370269383916052.

[32] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (1964). [,160(1964)], pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[33] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (1964). [,157(1964)], pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[34] P. Corbelli E.; Salucci. "The extended rotation curve and the dark matter halo of M33". In: *Monthly Notices of the Royal Astronomical Society* 311 (2000), pp. 441–447. DOI: 10.1046/j.1365-8711.2000.03075.x.

[35] Riccardo Barbieri and G. F. Giudice. "Upper Bounds on Supersymmetric Particle Masses". In: *Nucl. Phys. B* 306 (1988), p. 63. DOI: 10.1016/0550-3213(88)90171-X.

[36] S. Dimopoulos and G. F. Giudice. "Naturalness constraints in supersymmetric theories with nonuniversal soft terms". In: *Phys. Lett. B* 357 (1995), p. 573. DOI: 10.1016/0370-2693(95)00961-J. arXiv: hep-ph/9507282 [hep-ph].

[37] Riccardo Barbieri and Duccio Pappadopulo. "S-particles at their naturalness limits". In: *JHEP* 10 (2009), p. 061. DOI: 10.1088/1126-6708/2009/10/061. arXiv: 0906.4546 [hep-ph].

[38] Michele Papucci, Joshua T. Ruderman, and Andreas Weiler. "Natural SUSY endures". In: *JHEP* 09 (2012), p. 035. DOI: 10.1007/JHEP09(2012)035. arXiv: 1110.6926 [hep-ph].

[39] Hans Peter Nilles. "Supersymmetry, supergravity and particle physics". In: *Phys. Rep.* 110 (1984), p. 1. DOI: 10.1016/0370-1573(84)90008-5.

[40] Stephen P. Martin. "A supersymmetry primer". In: *Adv. Ser. Direct. High Energy Phys.* 21 (2010), p. 1. DOI: 10.1142/9789814307505_0001. arXiv: hep-ph/9709356 [hep-ph].

[41] Bertolini M. "Lectures on Supersymmetry". In: (2019), p. 319. URL: https://people.sissa.it/~bertmat/susycourse.pdf.

[42] Takumi Kuwahara Borut Bajc Junji Hisano and Yuji Omura. "Threshold Corrections to Dimension-six Proton Decay Operators in Non-minimal SUSY SU(5) GUTs". In: *Nuclear Physics B* (2016). DOI: 10.1016/j.nuclphysb.2016.06.017.

[43] Pierre Fayet. "Supergauge invariant extension of the Higgs mechanism and a model for the electron and its neutrino". In: *Nucl. Phys. B* 90 (1975), p. 104. DOI: 10.1016/0550-3213(75)90636-7.

[44] G. R. Farrar and P. Fayet. "Phenomenology of the production, decay, and detection of new hadronic states associated with supersymmetry". In: *Phys. Lett. B* 76 (1978), p. 575. DOI: 10.1016/0370-2693(78)90858-4.

[45] Nima Arkani-Hamed et al. "MARMOSET: The path from LHC data to the new standard model via on-shell effective theories". 2007.

[46] Johan Alwall et al. "Model-independent jets plus missing energy searches". In: *Phys. Rev. D* 79 (2009), p. 015005. DOI: 10.1103/PhysRevD.79.015005. arXiv: 0809.3264 [hep-ph].

[47] J. Alwall, P. Schuster, and N. Toro. "Simplified models for a first characterization of new physics at the LHC". In: *Phys. Rev. D* 79 (2009), p. 075020. DOI: 10.1103/PhysRevD.79.075020. arXiv: 0810.3921 [hep-ph].

[48] Daniele Alves et al. "Simplified models for LHC new physics searches". In: *J. Phys. G* 39 (2012), p. 105005. DOI: 10.1088/0954-3899/39/10/105005. arXiv: 1105.2838 [hep-ph].

[49] Serguei Chatrchyan et al. "Interpretation of searches for supersymmetry with simplified models". In: *Phys. Rev. D* 88 (2013), p. 052017. DOI: 10.1103/PhysRevD.88.052017. arXiv: 1301.2175 [hep-ex].

[50] S. Chatrchyan et al. "The CMS experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[51] PaulLujan. "CMS Luminosity-Public Results". https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. 2019.

[52] A. M. Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector". In: *JINST* 12 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[53] Richard Clinton Fernov. "Introduction to Experimental Particle Physics,Cambridge University Press, 1986". In: *Cambridge University Press* (1986), p. 421.

[54] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[55] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet user manual". In: *Eur. Phys. J. C* 72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[56] A. M. Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *JINST* 13 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158.

[57] Albert M Sirunyan et al. "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector". Accepted by *JINST*. 2019.

[58] Vardan Khachatryan et al. "The CMS trigger system". In: *JINST* 12 (2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366 [physics.ins-det].

[59] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[60] Johan Alwall et al. "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions". In: *Eur. Phys. J. C* 53 (2008), p. 473. DOI: 10.1140/epjc/s10052-007-0490-5. arXiv: 0706.2569 [hep-ph].

[61] Torbjörn Sjöstrand et al. "An Introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), p. 159. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[62] Vardan Khachatryan et al. "Event generator tunes obtained from underlying event and multiparton scattering measurements". In: *Eur. Phys. J. C* 76 (2016), p. 155. DOI: 10.1140/epjc/s10052-016-3988-x. arXiv: 1512.00815 [hep-ex].

[63] Albert M Sirunyan et al. "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements". Submitted to *Eur. Phys. J. C*. 2019.

[64] Simone Alioli et al. "NLO single-top production matched with shower in POWHEG: s- and t-channel contributions". In: *JHEP* 09 (2009). [Erratum: 10.1007/JHEP02(2010)011], p. 111. DOI: 10.1088/1126-6708/2009/09/111. arXiv: 0907.4076 [hep-ph].

[65] Emanuele Re. "Single-top Wt-channel production matched with parton showers using the POWHEG method". In: *Eur. Phys. J. C* 71 (2011), p. 1547. DOI: 10.1140/epjc/s10052-011-1547-z. arXiv: 1009.2450 [hep-ph].

[66] M. Beneke et al. "Hadronic top-quark pair production with NNLL threshold resummation". In: *Nucl. Phys. B* 855 (2012), p. 695. DOI: 10.1016/j.nuclphysb.2011.10.021. arXiv: 1109.1536 [hep-ph].

[67] Matteo Cacciari et al. "Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation". In: *Phys. Lett. B* 710 (2012), p. 612. DOI: 10.1016/j.physletb.2012.03.013. arXiv: 1111.5869 [hep-ph].

[68] Peter Bärnreuther, Michal Czakon, and Alexander Mitov. "Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t}+X$". In: *Phys. Rev. Lett.* 109 (2012), p. 132001. DOI: 10.1103/PhysRevLett.109.132001. arXiv: 1204.5201 [hep-ph].

[69] Michal Czakon and Alexander Mitov. "NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels". In: *JHEP* 12 (2012), p. 054. DOI: 10.1007/JHEP12(2012)054. arXiv: 1207.0236 [hep-ph].

[70] Michal Czakon and Alexander Mitov. "NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction". In: *JHEP* 01 (2013), p. 080. DOI: 10.1007/JHEP01(2013)080. arXiv: 1210.6832 [hep-ph].

[71] Michał Czakon, Paul Fiedler, and Alexander Mitov. "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(^4)$". In: *Phys. Rev. Lett.* 110 (2013), p. 252004. DOI: 10.1103/PhysRevLett.110.252004. arXiv: 1303.6254 [hep-ph].

[72] Ryan Gavin et al. "W physics at the LHC with FEWZ 2.1". In: *Comput. Phys. Commun.* 184 (2013), p. 208. DOI: 10.1016/j.cpc.2012.09.005. arXiv: 1201.5896 [hep-ph].

[73] Ryan Gavin et al. "FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order". In: *Comput. Phys. Commun.* 182 (2011), p. 2388. DOI: 10.1016/j.cpc.2011.06.008. arXiv: 1011.3540 [hep-ph].

[74] S. Agostinelli et al. "GEANT4—a simulation toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), p. 250. DOI: 10.1016/S0168-9002(03)01368-8.

[75] Wim Beenakker et al. "Squark and gluino production at hadron colliders". In: *Nucl. Phys. B* 492 (1997), p. 51. DOI: 10.1016/S0550-3213(97)00084-9. arXiv: hep-ph/9610490 [hep-ph].

[76] Anna Kulesza and L. Motyka. "Threshold resummation for squark-antisquark and gluino-pair production at the LHC". In: *Phys. Rev. Lett.* 102 (2009), p. 111802. DOI: 10.1103/PhysRevLett.102.111802. arXiv: 0807.2405 [hep-ph].

[77] Anna Kulesza and L. Motyka. "Soft gluon resummation for the production of gluino-gluino and squark-antisquark pairs at the LHC". In: *Phys. Rev. D* 80 (2009), p. 095004. DOI: 10.1103/PhysRevD.80.095004. arXiv: 0905.4749 [hep-ph].

[78] Wim Beenakker et al. "Soft-gluon resummation for squark and gluino hadroproduction". In: *JHEP* 12 (2009), p. 041. DOI: 10.1088/1126-6708/2009/12/041. arXiv: 0909.4418 [hep-ph].

[79] Wim Beenakker et al. "Squark and gluino hadroproduction". In: *Int. J. Mod. Phys. A* 26 (2011), p. 2637. DOI: 10.1142/S0217751X11053560. arXiv: 1105.1110 [hep-ph].

[80] S. Abdullin et al. "The fast simulation of the CMS detector at LHC". In: *J. Phys. Conf. Ser.* 331 (2011), p. 032049. DOI: 10.1088/1742-6596/331/3/032049.

[81] Andrea Giammanco. "The fast simulation of the CMS experiment". In: *J. Phys. Conf. Ser.* 513 (2014), p. 022012. DOI: 10.1088/1742-6596/513/2/022012.

[82] CMS Collaboration. *Jet performance in pp collisions at $\sqrt{s} = 7$ TeV*. CMS Physics Analysis Summary CMS-PAS-JME-10-003. 2010. URL: http://cdsweb.cern.ch/record/1279362.

[83] CMS Collaboration. *Jet algorithms performance in 13 TeV data*. CMS Physics Analysis Summary CMS-PAS-JME-16-003. 2017. URL: http://cds.cern.ch/record/2256875.

[84] Tejinder S. VIRDEE. "EXPERIMENTAL TECHNIQUES". In: (). URL: https://cds.cern.ch/record/454176/files/p347.pdf.

[85] M. RUDOWICZ G. GRINDHAMMER and S. PETERS. "The fast simulation of electromagnetic and hadronic showers". In: *Nuclear Instruments and Methods in Physics Research Section A* 290 (1990), pp. 469–488. DOI: 10.1016/0168-9002(90)90566-O.

[86] Rane A. "Searching for SUSY with Multijets and Missing Transverse Momentum". In: *Naimuddin M. (eds) XXII DAE High Energy Physics Symposium. Springer Proceedings in Physics, vol 203. Springer, Cham* 203 (2018), pp. 189–192. DOI: 10.1007/978-3-319-73171-1_42. URL: https://link.springer.com/chapter/10.1007/978-3-319-73171-1_42.

[87] Georges Aad et al. "Search for pair production of gluinos decaying via stop and sbottom in events with b-jets and large missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Rev. D* 94 (2016), p. 032003. DOI: 10.1103/PhysRevD.94.032003. arXiv: 1605.09318 [hep-ex].

[88] Morad Aaboud et al. "Search for squarks and gluinos in final states with jets and missing transverse momentum at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Eur. Phys. J. C* 76 (2016), p. 392. DOI: 10.1140/epjc/s10052-016-4184-8. arXiv: 1605.03814 [hep-ex].

[89] Vardan Khachatryan et al. "A search for new phenomena in pp collisions at $\sqrt{s} = 13$ TeV in final states with missing transverse momentum and at least one jet using the α_{T} variable". In: *Eur. Phys. J. C* 77 (2017), p. 294. DOI: 10.1140/epjc/s10052-017-4787-8. arXiv: 1611.00338 [hep-ex].

[90] PieterEveraerts. "LHC SUSY Cross Section Working Group". https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SUSYCrossSections. 2019.

[91] A. M. Sirunyan et al. "The Performance of the CMS Muon Detector in Proton-Proton Collisions at $\sqrt{}$ 13 TeV". In: *JINST 13.06* 72.2018 (2001), p. 06015. URL: http://link.aip.org/link/?RSI/72/4477/1.

[92] Vardan Khachatryan et al. "Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$". In: *JINST* 10 (2015), P08010. DOI: 10.1088/1748-0221/10/08/P08010. arXiv: 1502.02702 [physics.ins-det].

[93] Serguei Chatrchyan et al. "Search for new physics with jets and missing transverse momentum in pp collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 08 (2011), p. 155. DOI: 10.1007/JHEP08(2011)155. arXiv: 1106.4503 [hep-ex].

[94] Serguei Chatrchyan et al. "Search for new physics in the multijet and missing transverse momentum final state in proton-proton collisions at $\sqrt{s} = 8$ TeV". In: *JHEP* 06 (2014), p. 055. DOI: 10.1007/JHEP06(2014)055. arXiv: 1402.4770 [hep-ex].

[95] Alexis Kalogeropoulos and Johan Alwall. "The SysCalc code: A tool to derive theoretical systematic uncertainties". 2018.

[96] MauroVerzetti. "Methods to apply b-tagging efficiency scale factors". https://twiki.cern.ch/twiki/bin/viewauth/CMS/BTagSFMethods. 2019.

[97] Thomas Junk. "Confidence level computation for combining searches with small statistics". In: *Nucl. Instrum. Meth. A* 434 (1999), p. 435. DOI: 10.1016/S0168-9002(99)00498-2. arXiv: hep-ex/9902006 [hep-ex].

[98] A. L. Read. "Presentation of search results: the $CL_{(s)}$ technique". In: *J. Phys. G* 28 (2002), p. 2693. DOI: 10.1088/0954-3899/28/10/313.