

# Search for scalar top quark using 13 TeV proton proton collision data from CMS, LHC

*By*

**Koushik Mandal**

PHYS11201204010

National Institute of Science Education and Research

Jatni, Odisha - 752050, India

*A thesis submitted to the*

*Board of Studies in Physical Sciences*

*In partial fulfillment of requirements*

*for the Degree of*

**DOCTOR OF PHILOSOPHY**

*of*

**HOMI BHABHA NATIONAL INSTITUTE**

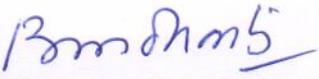
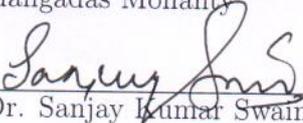
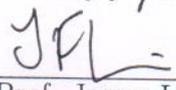
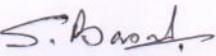


**July, 2018**

# Homi Bhabha National Institute

## Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by **Koushik Mandal** entitled "Search for scalar top quark using 13 TeV proton proton collision data from CMS, LHC" and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.

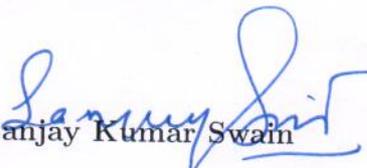
	<u>4/1/2019</u>
Chairman - Prof. Bedangadas Mohanty	Date
	<u>4/1/2019</u>
Guide / Convener - Dr. Sanjay Kumar Swain	Date
	<u>4/1/2019</u>
External Examiner - Prof. James Libby	Date
	<u>Jan 4, 2019</u>
Member 1 - Dr. Prolay Kumar Mal	Date
	<u>4/1/19</u>
Member 2 - Dr. Subhasish Basak	Date
	<u>4.1.19</u>
External Member - Dr. Seema Bahinipati	Date

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.

I hereby certify that I have read this thesis prepared under my direction and recommend that it may be accepted as fulfilling the thesis requirement.

Date: 4/1/2019

Place: NISER

  
Dr. Sanjay Kumar Swain

## STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Koushik Mandal

## DECLARATION

I, Koushik Mandal, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree/diploma at this or any other Institution/University.

Koushik Mandal

## List of publications arising from the thesis

### Journal

#### Published

1. Search for supersymmetry in the all-hadronic final state using top quark tagging in pp collisions at  $\sqrt{s} = 13\text{TeV}$ ,  
CMS Collaboration, (V. Khachatryan *et al.*), Phys. Rev. D **96**, 012004 (2017),
2. Search for supersymmetry in proton-proton collisions at 13 TeV using identified top quarks,  
CMS Collaboration (A.M. Sirunyan *et al.*), Phys. Rev. D **97**, 012007 (2018).

#### Symposium and conference

1. Search for scalar top quark production in all hadronic channel,  
Koushik Mandal (for the CMS Collaboration), ICHEP 2016, 38<sup>th</sup> International Conference on High Energy Physics, 3-10 Aug 2016, Chicago, IL (United States).
2. All hadronic scalar top quark search in CMS detector at 13 TeV pp collision at LHC,  
Koushik Mandal (for CMS Collaboration), DAE-HEP Symposium: XXII DAE-BRNS High Energy Physics Symposium, 12-16 Dec 2016, University of Delhi, Delhi (India).

Koushik Mandal

*Dedicated*  
*to*  
*All the people who inspired me to come this far.*

## Acknowledgements

I would like to begin by thanking my supervisor, Sanjay Kumar Swain for his invaluable guidance and decisive advice. I am grateful to him for giving me the opportunity to work in CMS collaboration, especially with the LPCSUSY group. I am thankful to Fermilab scientists, Daniel Elvira and Rick Cavanaugh for their support. My heartfelt gratitude to Kenichi Hatakeyama for his constant help. His passion and enthusiasm inspired me to cross a lot of hurdles during the course of my research work. Another person I can never thank enough is Hounxuan Liu. His encouragement and attention to every detail of my work enable me to complete my analysis work. It is an honor to work with LPCSUSY analysis team members: Joe, Zhenbin, Florent, Nadja, Scarlet, Hua, Joshi, and Seema. I learned a lot from you. I shall thank my thesis monitoring committee members for their constructive suggestions. I would like to thank DAE (India) and HBNI (India) for providing me the financial support. I would like to express my gratitude to the NISER and SPS where I learned to develop a research career.

I owe a lot of thanks to all of my friends and my family, without their constant support I could hardly survive the last few struggling years.

Koushik Mandal

# Contents

Synopsis	xi
List of Figures	xix
List of Tables	xxix
<b>1 Introduction</b>	<b>1</b>
<b>2 Standard Model and Supersymmetry</b>	<b>6</b>
2.1 Standard Model . . . . .	6
2.2 Supersymmetry . . . . .	12
2.3 Simplified Supersymmetric Model . . . . .	16
<b>3 LHC and CMS Experiment</b>	<b>19</b>
3.1 The Large Hadron Collider . . . . .	19
3.1.1 LHC Accelerator Chain . . . . .	20
3.1.2 LHC Layout and Operation . . . . .	21
3.1.3 LHC parameters . . . . .	23
3.2 Compact Muon Solenoid Experiment . . . . .	25
3.2.1 CMS Detector . . . . .	25
3.2.2 Event Reconstruction . . . . .	35

3.2.3	Event Simulation . . . . .	39
<b>4</b>	<b>Search for scalar top quark using top quark tagging</b>	<b>44</b>
<b>5</b>	<b>2015 Data Analysis</b>	<b>49</b>
5.1	Trigger . . . . .	49
5.2	Top Tagger . . . . .	52
5.3	Baseline Selection . . . . .	58
5.4	Search Regions . . . . .	60
5.5	Signal and Background MC Samples . . . . .	63
5.6	Background Estimation . . . . .	67
5.6.1	Hadronic Tau Background . . . . .	67
5.6.2	Lost Lepton Background . . . . .	82
5.6.3	$Z_{\nu\nu}$ Background . . . . .	86
5.6.4	QCD Background . . . . .	88
5.6.5	$t\bar{t}Z$ and other rare SM process . . . . .	90
5.7	Results . . . . .	91
<b>6</b>	<b>2016 Data Analysis</b>	<b>101</b>
6.1	Trigger . . . . .	102
6.2	Top Tagger . . . . .	104
6.2.1	Merged Top Tagger . . . . .	105
6.2.2	Resolved Top Tagger . . . . .	107
6.3	Baseline Selections . . . . .	109
6.4	Search Regions . . . . .	110
6.5	Signal and Background MC Samples . . . . .	110
6.6	Background Estimation . . . . .	112

6.6.1	Lost Lepton Background . . . . .	113
6.7	Results . . . . .	125
<b>7</b>	<b>Conclusions</b>	<b>130</b>
<b>A</b>	<b>Appendices</b>	<b>132</b>
A.1	MVA Based Resolved Top tagger . . . . .	132
A.1.1	MVA training . . . . .	133
A.1.2	MVA optimisation and tuning study . . . . .	135
A.1.3	MVA performance . . . . .	136

# Synopsis

The Standard Model (SM) of particle physics has been extremely successful in describing the physics of the particles and their interactions in atomic and subatomic realm. Despite its success, the SM suffers from shortcomings such as hierarchy problem, need for unnatural fine-tuning in order for the Higgs boson mass to be in electroweak scale [1]. Additionally, the existence of dark matter has been observed in astrophysical experiments [2] but fundamental dark matter particle is not included in SM theory. Supersymmetry (SUSY) [3] is one of the motivated and compelling theory beyond the Standard Model (SM) physics. It provides suitable solutions for different unexplained problems of SM by proposing a supersymmetric partner (sparticle) for each SM particle, with same quantum number except for spin which differs by half integer unit. The loop corrections to the Higgs boson mass due to these sparticles are opposite in sign to those of the SM particles predicting a finite value for the Higgs boson mass [3, 4]. This behavior can survive the breaking of SUSY, which is necessary to explain the non-observation of superpartners with exactly the same mass as their SM counterparts, provided that the superpartners are not themselves too heavy. In R-parity conserving SUSY model [5], the lightest supersymmetric particle (LSP) is stable and interacts weakly, thus, could be a good dark matter candidate.

The superpartners of quarks and gluons are squarks and gluinos respectively, while

neutralinos and charginos are mixtures of the superpartners of Higgs and electroweak gauge bosons. In order for SUSY to resolve fine-tuning problem naturally, masses of gluinos, top squarks, and bottom squarks are no heavier than a few TeV [4]. This opens up the opportunity to search for gluino and top squark within the energy range achieved by LHC. Provided that R-parity is conserved, the squarks and gluinos are produced in pairs. Searches for top squark pair production have been performed by the ATLAS Collaboration [6] and the CMS Collaboration [7] at the LHC using the Run 1 data. We, in particular, search for top squark which is a scalar particle sometime called scalar top quark or stop, using 13 TeV data collected by CMS detector. Inclusion of both stop production processes from pp collision, direct and mediated by gluino allows us to probe on the existence of gluino. In 2015 data analysis, target signals were T2tt ( $\tilde{t}\tilde{t}^* \rightarrow t\bar{t}\tilde{\chi}_1^0\tilde{\chi}_1^0$ ) and T2tb ( $\tilde{t}\tilde{t}^* \rightarrow tb\tilde{\chi}_1^0\tilde{\chi}_1^\pm$ ) based on direct stop production [8]. With 2016 data, along with T2tt, gluino mediated stop is also taken into consideration targeting T1tttt ( $\tilde{g}\tilde{g}^* \rightarrow t\bar{t}t\bar{t}\tilde{\chi}_1^0\tilde{\chi}_1^0$ ) signal [9]. Stop decays to SM top and  $\tilde{\chi}_1^0$  in case of T2tt and T1tttt or SM bottom quark and  $\tilde{\chi}_1^\pm$  in case of T2tb.  $\tilde{\chi}_1^\pm$  then decays to W boson and  $\tilde{\chi}_1^0$ .  $\tilde{\chi}_1^0$  is assumed to be stable, weakly interacting and LSP. It does not produce a signal in the detector, generating missing transverse momenta ( $p_T^{miss}$ ). Top quark subsequently decays to hadrons which are reconstructed as jets. Search for top squark is therefore performed in full hadronic final states with large imbalance in transverse momenta ( $p_T^{miss}$ ) and multiple jets. The knowledge of SM processes which can give the same final state as that of the stop signal is essential in order to comprehend the signature of stop in search region. The contribution from such SM processes in search region is estimated using 13 TeV data. A novel top quark algorithm is employed to identify hadronically decaying top quarks produced in the decay chain. The algorithm makes use of the facts that a top quark essentially always decays to a bottom quark and a W boson, and the W boson decays to a quark-

antiquark ( $q\bar{q}$ ) pair. The algorithm recognizes three different types of decay topology for the top quark. In order of increasing Lorentz boost for the top quark, these are: (i) three distinct jets with no more than one of them identified as a bottom quark jet (b jet), where two non-b jets represent the decay products of the  $q$  and  $\bar{q}$  coming from the W boson decay; (ii) two distinct jets, one of which corresponds to the b quark and the other to the merged  $q\bar{q}$  decay products from the W boson; and (iii) a single jet representing the merged decay products of the b quark and W boson. By accounting for these three different topologies, the algorithm achieves high detection efficiency over a wide range of top quark transverse momentum. The decay topology mentioned above in (i) is known as resolved scenario.

Events are selected with certain requirements with the aim of reducing background SM processes and conserving signal acceptance. We reject the event having well reconstructed and isolated muons and electrons. The requirement of minimum four jets ( $p_T > 30$  GeV), one b-tagged jet and one reconstructed top quark is implemented. A threshold on  $p_T^{miss}$  is applied to account for data trigger efficiency. An additional variable namely  $M_{T2}$  is derived from the kinematic of reconstructed top and events are selected with minimum optimized  $M_{T2}$  value. Likewise, a minimum threshold is imposed on  $H_T$  which is the scalar sum of the transverse momentum of all selected jets.

In order to increase the signal-to-background sensitivity, search region is divided into a number of bins based on the number of reconstructed top quarks and the number of b-tagged jets, as well as the value of missing transverse momentum ( $p_T^{miss}$ ) and  $M_{T2}$ . For gluino mediated signal,  $M_{T2}$  is replaced by  $H_T$  as one of search bin variables. After the application of baseline selection, the contribution from all the possible SM backgrounds is estimated in each search bin. Pair of top quarks, W+jets, and single top are such backgrounds contributing in search region when an

electron or a muon is non-isolated, non-reconstructed or out of acceptance region or W decays into hadronically decaying tau lepton ( $\tau$ ). The first scenario is called lost lepton background and the later one is called hadronic tau decay ( $\tau_h$ ) background. Hadronic tau is one of the largest components of expected backgrounds. We search the signal in full hadronic final state, so we apply electron and muon veto to reduce the background events containing leptons. If we could apply the tau veto it would reduce total leptonic background event. But vetoing reconstructed tau we can not get rid of all of the hadronic tau events because of tau reconstruction efficiency. So remnant events are to be estimated by the background estimation method. Instead of using tau identification we rather use a special technique called 'isolated charged track veto' to reduce mostly one prong hadronic tau event. When tau decays into one charged hadron the hadronic track has less activity around it compared to hadrons coming from other processes. So we isolate the charged tracks and veto the events with identified isotracks. This technique removes around 35% of total hadronic tau events and also removes the events with electron and muon.

After applying isotrack veto along with other baseline selections, the residual hadronic tau events are estimated in search region. Following basic method used in 8 TeV data analysis [10], new technique and parametrization have been implemented to achieve more accurate background estimation.  $\tau_h + jets$  events are estimated from the  $\mu + jets$  control sample taken from data. Because the  $\mu + jets$  and  $\tau_h + jets$  events arise from the same physics processes, the hadronic component of the two samples is the same except for the response of the detector to the muon or  $\tau_h$  jet. To account for this difference, the muon in data is replaced by a simulated  $\tau_h$  jet with the  $p_T$  sampled randomly from a response function for a hadronically decaying  $\tau$  lepton. Various correction factors are applied to get rectified hadronic tau event yield. A new correction factor is included to compensate the mis-tag rate of  $\tau_h$  jet faked as b tagged jet.

Closure test of the method has been done with 13 TeV MC samples. Hadronic tau prediction with systematic uncertainty in the search bins has been estimated with full available data collected in 2015.

Other backgrounds according to their contributions in descending order are lost lepton ( $t\bar{t}$ ,  $W + jets$ , single top),  $Z(\nu\bar{\nu}) + jets$ , QCD and rare SM processes like diboson, triboson etc. All the predicted backgrounds are added up and compared with the observed data yield. Good agreement between total SM prediction and data observation within uncertainty has been seen. Therefore, exclusion limit has been imposed on stop and neutralino mass following the models within Simplified Model Spectra (SMS) [11].

2015 top tagger algorithm is effective to select top in wide  $p_T$  range but it has quite high fake rate. It is found that resolved scenario when top is reconstructed with three jet combinations is responsible for high fake rate due to combinatorial backgrounds. In 2016 data analysis, new top quark reconstruction technique is employed in order to reduce fake rate while preserving overall efficiency of 2015 tagger. This goal is accomplished by replacing cut and count method used previously with a multivariate analysis technique. An extensive study has been done to optimize the technique for better performance. The adaptation of new tagger algorithm leads to the integration of new hadronic tau and lost lepton background estimation method in 2016. These two backgrounds are estimated with a new technique called transfer factor (TF) method. Other SM backgrounds are evaluated and total SM prediction and observed data are compared. No statistically significant deviation between the data and background prediction is observed. Mass exclusion limit for stop, gluino, and LSP is derived. The stronger exclusion limit is obtained from 2016 data analysis compared to 2015 data analysis. I, in particular, extensively worked on the hadronic tau background estimation in 2015 data analysis and the improvement and implementation

of resolved top quark tagging algorithm in 2016 data analysis. I also contributed to the W decay background estimation using TF method in 2016 analysis.

# Bibliography

- [1] R. K. Kaul and P. Majumdar, "Cancellation of quadratically divergent mass corrections in globally supersymmetric spontaneously broken gauge theories", Nucl. Phys. B 199 (1982) 36, doi:10.1016/0550-3213(82)90565-X.
- [2] V. Trimble, "Existence and Nature of Dark Matter in the Universe", Ann. Rev. Astron. Astrophys. 25 (1987) doi:10.1146/annurev.aa.25.090187.002233.
- [3] H. P. Nilles, "Supersymmetry, supergravity and particle physics", Phys. Rep. 110 (1984) 1, doi:10.1016/0370-1573(84)90008-5.
- [4] R. Barbieri and G. Giudice, "Upper Bounds on Supersymmetric Particle Masses", Nucl. Phys. B 306 (1988) 63, doi:10.1016/0550-3213(88)90171-X.
- [5] G. R. Farrar and P. Fayet, "Phenomenology of the Production, Decay, and Detection of New Hadronic States Associated with Supersymmetry", Phys. Lett. B 76 (1978) 575, doi:10.1016/0370-2693(78)90858-4.
- [6] ATLAS Collaboration, "Search for a supersymmetric partner to the top quark in final states with jets and missing transverse momentum at  $\sqrt{s} = 7\text{TeV}$  with the ATLAS detector", arXiv:1208.1447. doi: 10.1103/PhysRevLett.109.211802.

- [7] CMS Collaboration, "Searches for third generation squark production in fully hadronic final states in proton-proton collisions at  $\sqrt{s}=8$  TeV", arXiv:1503.08037. doi: 10.1007/JHEP06(2015)116.
- [8] CMS Collaboration,"Search for supersymmetry in the all-hadronic final state using top quark tagging in pp collisions at  $\sqrt{s} = 13$  TeV", arXiv:1701.01954, doi: 10.1103/PhysRevD.96.012004.
- [9] CMS Collaboration, "Search for supersymmetry in proton-proton collisions at 13 TeV using tagged top quarks", SUS-16-050.
- [10] CMS Collaboration,"Search for Direct Scalar Top-quark Pair Production in the All-hadronic Channel at 8 TeV using a Top-tagger", SUS-13-015.
- [11] CMS Collaboration, "Interpretation of searches for supersymmetry with simplified models", Phys. Rev. D 88 (2013) 052017, doi:10.1103/PhysRevD.88.052017, arXiv:1301.2175.

# List of Figures

2.1	Fundamental particles in Standard Model [4] . . . . .	7
2.2	Fermionic one loop correction to Higgs. [9] . . . . .	9
2.3	The running of three coupling constants with the energy of the interactions are shown [14]. Inverse of coupling constants which are the measure of interaction strength are plotted with the energy. The running of coupling constants are shown in the case of SM (left) and with the minimal supersymmetric extension of SM (right). . . . .	11
2.4	Superpartners of SM particles [20]. . . . .	13
2.5	Fermionic (a) and bosonic (b) one loop correction to Higgs [21]. . . . .	13
3.1	LHC full accelerator chain [2] . . . . .	21
3.2	LHC layout [3] . . . . .	22
3.3	Satellite view of LHC with the location of four experiments [7]. . . . .	26
3.4	CMS coordinate system [8]. Z axis is along the beam direction towards mount Jura, x axis points towards the center LHC ring and y axis is pointing upward. If P is vector representing the direction of a particle traverses the detector, picture shows how angle $\theta$ and $\phi$ are determined. . . . .	27
3.5	Different layers of CMS detector [9]. . . . .	28
3.6	CMS silicon tracker system. Left picture [10] shows a 3D view of full tracker system while the right one [11] shows only pixel part. . . . .	29

3.7	Cross-sectional view of HCAL (left) [14] and Muon chambers (right) [15] in r-z plane . . . . .	32
3.8	Partial cross-sectional view of CMS subdetectors and thier interaction status with the particles come from the collision [18]. . . . .	36
4.1	Signal models of direct stop production: top squark pair production with the top squark decaying into a top quark and neutralino (left) or into a bottom quark and chargino (right). The SUSY simplified model topology shown at the left is referred to as T2tt and the right one as T2tb. . . . .	45
4.2	Signal models of gluino mediated processes: gluino pair production with the gluino decaying into neutralinos and top quarks through off-shell top squarks (top left), or the gluino decaying into a top quark and on-shell top squark which decays to a top quark and a neutralino (top left), or gluino decaying into off-shell top or bottom squark (bottom left), or gluino decaying into on-shell stop which decays to charm quark and neutralino (bottom right). The SUSY simplified model topology shown at the top left is referred to as T1tttt, top right one as T5tttt, the bottom left one as T1ttbb, and the bottom right model as T5ttcc. . . . .	46
5.1	The trigger efficiency for $HLT\_PFHT350\_PFMET100*$ as a function of $p_T^{miss}$ (the axis label $E_T^{miss}$ actually means $p_T^{miss}$ )(left) and $H_T$ (right). In the first case, $H_T > 500$ GeV cut is applied to ensure full efficiency. In the second case, $p_T^{miss} > 200$ GeV cut is applied. The dashed (solid) blue lines show the distributions for the samples used to calculate the denominator (numerator). . . . .	51

5.2	Efficiency of the $\mu + H_T$ trigger as a function of $H_T$ and $\mu p_T$ . This is the trigger employed to select the muon + jets control sample for the the hadronically decaying $\tau$ lepton background estimation. . . . .	52
5.3	The 2D distributions of $\frac{m_{23}}{m_{123}}$ versus $\arctan(\frac{m_{13}}{m_{12}})$ for $t\bar{t}$ (left) and QCD (right). The three horizontal solid red lines which form a band demonstrate the first criteria in Eq. 5.1 with the central line taken as the nominal ratio of $\frac{m_W}{m_{top}}$ and the other two lines showing the boundaries defined by the $R_{min}$ and $R_{max}$ . Similarly the second and third criteria in Eq. 5.1 are in groups of solid lines. . . . .	54
5.4	The tagging efficiency of the top quark tagger in term of of the generator-level hadronically decaying top quark $p_T$ (black points). The efficiency was computed using the T2tt (850,100) signal model where $m_{\tilde{t}} = 850$ GeV and $m_{\tilde{\chi}_1^0} = 100$ GeV, and it is similar for $t\bar{t}$ events. The vertical bars depict the statistical uncertainty. The colored lines show the expected hadronically decaying top quark $p_T$ distribution from $t\bar{t}$ (red solid line), the T2tt(500, 100) (blue short-dashed line) and T2tt(750, 50)(green long-dashed line) signal models , the T1tttt (1200, 800) signal model with $m_{\tilde{g}} = 1200$ GeV and $m_{\tilde{\chi}_1^0} = 800$ GeV (purple long-dash-dotted line), and the T1tttt(1500, 100) signal model (orange short-dash-dotted line). The last bin contains the overflow entries and the top quark $p_T$ distributions are normalized to unit area. . . . .	56

5.5	Comparison of the distributions between simulated SM backgrounds and several signal models for $N_t$ , $N_b$ , $E_T^{miss}$ and $M_{T2}$ (clock-wise), after the baseline selection applied. Observed data events are shown by black points. Total SM backgrounds and signal models distributions have been normalized to the same area as the data distribution. Here $E_T^{miss}$ is synonymous with $p_T^{miss}$ . . . . .	62
5.6	Search bin definitions and bin numbers for 37 SBs after baseline selection. The number indicates bin indices. . . . .	63
5.7	Search bin definitions and bin numbers for 45 SBs after baseline selection. The number indicates bin indices. . . . .	64
5.8	The $m_T$ distribution in $t\bar{t}$ (black) and for the T2tt signal point $m_{\tilde{t}}=850$ GeV, $m_{LSP}=100$ GeV (red). Both curves are scaled to $2.3 fb^{-1}$ . . .	69
5.9	Closest $\Delta R$ 's between tau lepton "visible part" and jet as a function of $f_{VE}$ for different $p_T$ ranges of the generated $\tau$ lepton. . . . .	70
5.10	Tau jet visible energy fraction templates in tau lepton $p_T$ bins. . . . .	71
5.11	The rate for tau-jets to be misidentified as a b jet as a function of the tau jet $p_T$ . . . . .	71
5.12	Isolated track tagging efficiency of $\tau_h$ as a function of $N_{jets}$ and $N_b$ . Left plot has three $N_b$ bins and right plot has two $N_b$ bins where last two bins of left plot are merged. The tagging efficiency varies from 0.15 to 0.4 with the bin configuration. The average value is given in Tab. 5.4. . . . .	73
5.13	Muon acceptance as a function of search bin, in case of 45 bins (left) and 37 bins (right). . . . .	76

5.14	Left plot shows $m_T$ selection efficiency as a function of $N_{jets}$ and $p_T^{miss}$ . Right plot illustrates the $\tau \rightarrow \mu$ contamination, i.e., the fraction of $\mu$ coming from $\tau$ to the all $\mu$ coming from both $\tau$ and $W$ , as a function of $N_{jets}$ and $p_T^{miss}$ . . . . .	77
5.15	Closure tests with simulated data for $p_T^{miss}$ , $M_{T2}$ , $N_b$ and $N_t$ distributions based on comparisons between the predicted $\tau_h$ background using the $\tau_h$ template response method (red points) and the expected number of events using MC truth (blue line). Only statistical uncertainties are shown and properly propagated to the uncertainty of the ratio. . .	78
5.16	The predicted yield from simulated events (blue line) compared to the expected $\tau_h$ background yield (red points) in the case of 45 bins(left) and 37 bins(right). Only statistical uncertainties are shown. . . . .	78
5.17	Total Relative systematic uncertainty for hadronic tau prediction in the case 37 bins . . . . .	81
5.18	Predicted hadronic tau background yield for a $2.3 \text{ fb}^{-1}$ sample in the case of 45 bins (left) and 37 bins (right). Only statistical uncertainties are shown. . . . .	82
5.19	Sketch of the requirements electrons and muons from $W$ decays must meet in order to be rejected by the explicit lepton veto. . . . .	83
5.20	The lost-lepton background in all the search regions of the analysis as predicted directly from $t\bar{t}$ , $W$ +jets and single top simulation (in red) and as predicted by applying the lost-lepton background procedure to simulated muon control sample (in black). The lower panel shows the ratio between the true and predicted yields. The left (right) plot is for the case with 45 (37) search bins. . . . .	85

5.21	The distribution of $N_b$ (left) and $p_T^{miss}$ (right) from data and simulation in the loose dimuon control region after applying the $S_{DY}(N_j)$ scale factor to the simulation. The ratio is shown in the lower panels. Only statistical uncertainties are shown. The values in parentheses in the legend represent the integrated yield for each given process. . . . .	88
5.22	QCD multijet events directly from MC simulation (red) and predicted events (blue) using the background estimation procedure on simulation. Lower pannel shows the ratio. Closure in 37 SB (left) and 45 SB (right) are shown. . . . .	91
5.23	Data are shown as black points. The total predictions are shown in filled solid area. Expected yields from various direct stop production signal models are denoted by red and green lines. The bottom plot shows the ratio of data over total background prediction in each search bin. Statistical uncertainties associated with backgrounds are presented by grey and blue hatched bands in both panels. Within uncertainties, a good agreement between prediction and observation is found across all SBs. . . . .	93
5.24	Data are shown as black points. The total predictions are shown in filled solid area. Expected yields from various direct stop production signal models are denoted by red and green lines. The bottom plot shows the ratio of data over total background prediction in each search bin. Statistical uncertainties associated with backgrounds are presented by grey hatched bands in both panels. Within uncertainties, a good agreement between prediction and observation is found across all SBs. . . . .	95

5.25	95% CL exclusion limit for simplified signal models T2tt( top left) and T2tb (top right)and T1tttt (bottom). The solid black curves indicate the observed exclusion contour and corresponding uncertainties with $\pm 1$ standard deviation whereas dashes red curves represent similar expected exclusion contours. The observed limits are below the expected ones because of our background estimation, specially the QCD estimation associates huge uncertainty. . . . .	98
6.1	The trigger efficiency, denote by the black point, as a function of the offline $p_T^{miss}$ for (left) $300 < H_T < 1000$ and (right) $H_T > 1000$ . The error bar indicates the statistical uncertainty of the trigger efficiency. The dash blue line represents the denominator passing the selection, while the solid blue histogram represents the numerator where the denominator events also trigger the search triggers. . . . .	103
6.2	The efficiency of top tagging algorithm in term of generated top quark $p_T$ computed using simulated T2tt signal sample where $m_{\tilde{t}} = 850$ GeV and $m_{\tilde{\chi}_1^0} = 100$ . Blue circles represent the efficiency of full combined tagger where as other three plots shows the efficiency of three individual taggers. Red boxes indicate monojet boosted tagger, magenta triangles are for semiboosted dijet and green triangles present the resolved trijet tagger. . . . .	109
6.3	Search bin definitions and bin numbers for 84 SBs after baseline selection. The number indicates bin indices. . . . .	111
6.4	The pie chart of the contribution from SM background processes in search region. The pie chart in computed in the bin of $N_t$ and $N_b$ . . .	113

6.5	Shape comparison of $H_T$ (1st row), $N_{jets}$ (2nd row), $p_T^{miss}$ (3rd row) and $M_{T2}$ (4th row) for muon CS (left column) and electron CS (right column). . . . .	117
6.6	Shape comparison of $N_t$ (top row) and $N_b$ (bottom row) for muon CS (left column) and electron CS (right column). . . . .	118
6.7	Shape comparison of $H_T$ (1st row), $N_{jets}$ (2nd row), $p_T^{miss}$ (3rd row) and $M_{T2}$ (4th row) for muon CS (left column) and electron CS (right column) after the various corrections on MC samples. . . . .	119
6.8	Shape comparison of $N_t$ (top row) and $N_b$ (bottom row) for muon CS (left column) and electron CS (right column) after the various corrections on MC samples. . . . .	120
6.9	Translation factors for the $\tau_h$ (left) and lost lepton (right) background prediction with their uncentainties from both muon and electron CS. . . . .	121
6.10	Validation of TF method in the data sideband using the muon channel (left) and electron channel (right). The black points are observed data. The light blue is predicted $t\bar{t}$ , W+jets and single-top events using TF method. All other backgrounds directly come from MC yields. The errors include only statistical uncertainties. . . . .	123
6.11	Predicted hadronic tau background yield for $35.9 \text{ fb}^{-1}$ data for all the search bins. Right plot is a zoomed version of left plot. Both statistical and total systematic uncertainties are shown. . . . .	124
6.12	Predicted lost lepton background yield for $35.9 \text{ fb}^{-1}$ data for all the search regions. Right plot is a zoomed version of left plot. Both statistical and total systematic uncertainties are shown. . . . .	124

6.13	Data are shown as black points. The total predictions are shown in filled solid area. The bottom plot shows the ratio of data over total background prediction in each search bin. The hatched bands show uncertainties of total predictions with blue for systematic uncertainty and grey for statistical uncertainty. Within uncertainties, a good agreement between prediction and observation is found across all SBs . . . . .	125
6.14	95% CL exclusion limit for simplified signal model T2tt. The solid black curves indicate the observed exclusion contour and corresponding uncertainties with $\pm 1$ standard deviation whereas dashes red curves represent similar expected exclusion contours. . . . .	126
6.15	95% CL exclusion limit for simplified signal models T1tttt (top left) and T5tttt (top right), T1ttbb (bottom left) and T5ttcc (bottom right). The solid black curves indicate the observed exclusion contour and corresponding uncertainties with $\pm 1$ standard deviation whereas dashes red curves represent similar expected exclusion contours. . . . .	128
A.1	The ROC curves for random forest with depth 14 trained using lab frame and rest frame variables. . . . .	135
A.2	The discriminator for signal and background from the final trained random forest separated between signal- and background-like candidates.	137

A.3 The efficiency (left) and fake rate (right) of the resolved top tagging algorithm in term of generated top quark  $p_T$  and event  $p_T^{miss}$  respectively. Blue lines represnt the MVA algorithm and the red one indicates original cut-based algorithm. The efficiency and fake rate are measured according to the equation 5.3 in simulated semileptonic  $t\bar{t}$  and Z+jets events respectively. The legend 'ICHEP 2016 new code' means cut-based algorithm runs on 2016 ICHEP samples and 'MVA + AK8' indicates the new algorithm which is combined with AK8 tagger but the results presented here are from MVA tagger alone and with discriminator threshold of 0.7. . . . . . 137

# List of Tables

2.1	Force mediator bosons in SM. . . . .	8
3.1	Description of beam parameters which define the luminosity. . . . .	24
5.1	Sample selection cut flow and event yields from MC simulated samples. All numbers are normalized to $2.2fb^{-1}$ . Only statistical uncertainties are shown. Rare process refer to ttW, di-boson and tri-boson events.	61
5.2	Standard model Monte Carlo samples used in the analysis. . . . .	66
5.3	Cross sections for a couple of mass points for the T2tt and T1tttt Models.	67
5.4	Isolated track veto effect on one prong, three prong, and total $\tau_h$ event.	72
5.5	Contributions from different sources of systematic uncertainty to the $\tau_h$ background prediction. . . . .	81
5.6	Observed yields from the full luminosity ( $2.3 fb^{-1}$ ) of data and total background predictions with their statistical and systematic uncertain- ties in 37 search bins. . . . .	92
5.7	Observed yields from the full luminosity ( $2.3 fb^{-1}$ ) of data and total background predictions with their statistical and systematic uncertain- ties in 45 search bins. . . . .	94
6.1	Contributions from different sources of systematic uncertainty to the $\tau_h$ and lost lepton background prediction. . . . .	122

# Chapter 1

## Introduction

The Standard Model (SM) of particle physics has been extremely successful in describing the physics of the particles and their interactions in the atomic and subatomic realm. The glorious discoveries of elementary particles over past 50 years provided the excellent consistency with the prediction of SM theory. In 2012, the discovery of Higgs boson by CMS and ATLAS experiments [1, 2, 3] at LHC added one more feather to its cap by finding the last missing link of SM.

Despite its success SM, however, is not regarded as a complete theory since the SM suffers from shortcomings such as hierarchy problem, need for unnatural fine-tuning in order for the Higgs boson mass to be in electroweak scale [4]. Additionally, existence of dark matter has been observed in astrophysical experiments [5] but fundamental dark matter particle is not included in SM theory. The Neutrinos, being considered as massless within SM, can not be counted as the massive dark matter candidate.

Supersymmetry (SUSY) [6] is one of the motivated and compelling theories beyond the SM physics. It provides suitable solutions for different unexplained problems of SM by proposing a supersymmetric partner (sparticle) for each SM particle, with same quantum number except for spin which differs by half integer unit. The quan-

tum loop corrections to the Higgs boson mass due to these sparticles are opposite in sign to those of the SM particles predicting a finite value for the Higgs boson mass [7, 8]. This behaviour can survive the soft breaking of SUSY [9, 10], which is necessary to explain the non-observation of superpartners with exactly same mass as their SM counterparts, provided that the superpartners are not themselves too heavy. In R-parity conserving SUSY model [11], the lightest supersymmetric particle (LSP) is stable and interacts weakly, thereby could be a good dark matter candidate. Thus supersymmetry motivates us to explore the theory experimentally.

The Large Hadron Collider (LHC) at CERN offers to achieve energy frontier at TeV scale. It has been designed to operate at a center-of-mass energy of 14 TeV, the highest energy in the history of human particle physics experiments so far. This high energy collider machine provides the unique opportunity to investigate for new physics. The Compact Muon Solenoid (CMS) is a multi-purpose general detector which collects good quality collision data to be used for physics analysis. My analysis is based on LHC-CMS proton proton collision data collected in 2015 and 2016 at 13 TeV center-of-mass energy.

The superpartners of quarks and gluons are squarks and gluinos respectively, while neutralinos and charginos are mixtures of the superpartners of Higgs and electroweak gauge bosons. In order for SUSY to resolve the fine-tuning problem naturally, masses of gluinos, top squarks, and bottom squarks are no heavier than a few TeV [12, 13, 14]. This opens up the opportunity to search for gluino and top squark within the energy range achieved by LHC. Provided that R-parity is conserved, the squarks and gluinos are produced in pairs. Searches for top squark pair production have been performed by the ATLAS Collaboration [15, 16, 17] and the CMS Collaboration [18, 19, 20] at the LHC using the Run 1 data. Of particular interest are the searches for top squark which is a scalar particle sometimes called scalar top quark or stop. The inclusion

of both stop production processes from pp collision, direct and mediated by gluino allows us to probe on the existence of gluino. In this thesis, the search for direct and gluino mediated production of stop using 13 TeV data collected by CMS detector is presented.

Search for top squark is performed in full hadronic final states with large imbalance in transverse momenta ( $p_T^{miss}$ ) and multiple jets. The knowledge of SM processes which can give the same final state as that of the stop signal is essential in order to comprehend the signature of stop in the search region. The contribution from such SM processes in the search region is estimated using 13 TeV data. A novel top quark algorithm is employed to identify hadronically decaying top quarks produced in the decay chain.

The organization of the dissertation is described as follows. An overview of standard model and supersymmetry with a discussion of simplified models is given in chapter 2. Chapter 3 gives a brief description of LHC and CMS experiment. In chapter 4, an introduction of scalar top quark search is outlined followed by the disquisition of 2015 and 2016 data analysis in chapter 5 and 6 respectively. Finally, the summary is concluded in chapter 7.

# Bibliography

- [1] (ATLAS Collaboration) G. Aad *et al.*, Phys. Lett. B 716 1 (2012).
- [2] (CMS Collaboration) S. Chatrchyan *et al.*, Phys. Lett. B 716 30 (2012).
- [3] (CMS Collaboration) S. Chatrchyan *et al.*, JHEP 06 081 (2013).
- [4] R. K. Kaul and P. Majumdar, Nucl. Phys. B 199 36 (1982).
- [5] V. Trimble, Ann. Rev. Astron. Astrophys. 25 (1987).
- [6] H. P. Nilles, "Supersymmetry, supergravity and particle physics", Phys. Rep. 110 1 (1984).
- [7] B. de Carlos and J. Casas, Phys. Lett. B 309 320 (1993).
- [8] R. Barbieri and G. Giudice, Nucl. Phys. B 306 63 (1988).
- [9] S. Dimopoulos and G. Giudice, Phys. Lett. B 357 573 (1995).
- [10] L. J. Hall, J. D. Lykken *et al.*, Phys. Rev. D 27 2359 (1983).
- [11] G. R. Farrar and P. Fayet, Phys. Lett. B 76 575 (1978).
- [12] C. Brust, A. Katz *et al.*, JHEP 03 103 (2012).
- [13] A. Delgado *et al.*, Eur. Phys. J. C 73 2370 (2013).

- [14] J. L. Feng, *Ann. Rev. Nucl. Part. Sci.* 63 351 (2013).
- [15] (ATLAS Collaboration) G. Aad *et al.*, *Phys. Rev. Lett.* 109, 211802 (2012).
- [16] (ATLAS Collaboration) G. Aad *et al.*, *Phys. Rev. Lett.* 109, 211803 (2012).
- [17] (ATLAS Collaboration) G. Aad *et al.*, *JHEP* 10 189 (2013).
- [18] (CMS Collaboration) S. Chatrchyan *et al.*, *Eur. Phys. J. C* 73 2677 (2013).
- [19] (CMS Collaboration) V. Khachatryan *et al.*, *JHEP* 06 116 (2015).
- [20] (CMS Collaboration) V. Khachatryan *et al.* *Eur. Phys. J. C* 76 460 (2016).

# Chapter 2

## Standard Model and Supersymmetry

This chapter presents an overview of the theoretical background and motivation of the search documented in the thesis. A theory of elementary particle, known as Standard Model (SM) of particle physics, is introduced here. A brief review of SM theory and its limitation, a possible extension of SM, i.e., supersymmetry (SUSY) theory and few experimental SUSY model are outlined in succeeding sections.

### Standard Model

The SM of particle physics is a mathematical model which chronicles the properties of elementary particles and three of their interactions namely weak, electromagnetic and strong. SM is formulated in quantum field theory framework based on gauge symmetry group  $SU(3) \otimes SU(2) \otimes U(1)$ , where  $SU(3)$  is for color symmetry in strong interaction and  $SU(2) \otimes U(1)$  represents the combined symmetry in weak and electromagnetic interactions. SM states that the building block of matter is made of

elementary fermions ( with their antiparticles) which interact through the above mentioned three force fields. The mediators of these fields are elementary gauge bosons which acquire mass through spontaneous symmetry breaking by interacting with a scalar boson called Higgs boson [1, 2, 3]. All the fundamental particles predicted by SM are shown in figure 2.1. Twelve fermions are divided into two different classes:

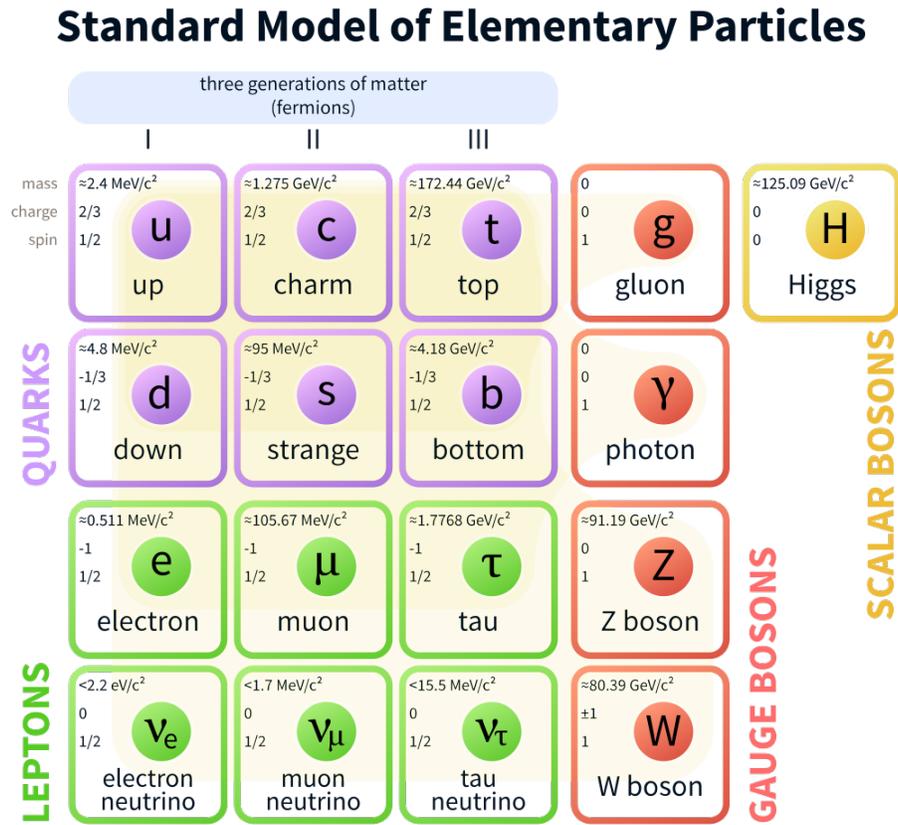


Figure 2.1: Fundamental particles in Standard Model [4]

six leptons (and antileptons) and six quarks (and antiquarks), each is organized in three generations. Leptons include chargeless neutrinos which interact through weak interaction and charged leptons which take part both in electromagnetic and in weak interactions. Quarks interact via weak, electromagnetic and strong forces. The list of

force carrier bosons is given in table 2.1. Unlike leptons, quarks are not observed as free particles and they form bound state known as hadrons. Hadrons are categorized into two types: three quarks bound state baryon and two quark bound state meson. In higher  $N_c$  limit of QCD theory, bound state with four quarks or higher can be formed. Recently a tetra quark state has been experimentally observed. But whether its a tetra quark bound state or two meson states is yet to be confirmed [5]. Among all SM fermions, the top quark is the heaviest with a very short lifetime of  $5 \times 10^{-25}$  s [6]. It decays to subsequent particles (b and W with almost 100% branching ratio) before any bound state formation.

Table 2.1: Force mediator bosons in SM.

Force	Mediator boson
Strong	gluon (g)
Electromagnetic	Photon ( $\gamma$ )
Weak	$W^\pm$ (charged), Z (neutral)

The development of the SM theory has been started since the 1960s and it has been asserted as the most successful theory to date. The predictions of SM have been proven to be consistent with the observations from various particle physics experiments. Though SM is extremely successful, it has the limitations. Neutrinos are considered massless particles in SM, though neutrino oscillation experiments [7, 8] prove the existence of neutrino with non-zero mass. Moreover, there are few unresolved issues in SM which motivate to formulate new theory beyond SM.

- Hierarchy problem: The huge difference in strength between weak force and gravitational force is known as the hierarchy problem in SM. This discrepancy can be quantified from the very large ratio of  $\frac{M_P}{M_W}$  where,  $M_P$  ( $10^{18}$ GeV) is Plank mass and  $M_W$  ( $10^2$ GeV) represents the Z or W mass which define weak

force. The Plank mass scale is defined by the following equation,

$$M_P = \sqrt{\frac{\hbar c}{G}} \quad (2.1)$$

where  $c$  is the speed of light in a vacuum,  $G$  is the gravitational constant, and  $\hbar$  is the reduced Planck constant. Since Higgs boson gives mass to gauge bosons, the more technical way to see the hierarchy problem is the discrepancy between Plank mass and measured Higgs boson mass. SM adds radiative quantum corrections to the square of Higgs mass which in turn makes the Higgs mass huge and comparable to Plank mass scale. The first order loop correction due to a fermion is shown by the diagram in figure 2.2 and expressed by equation 2.2.

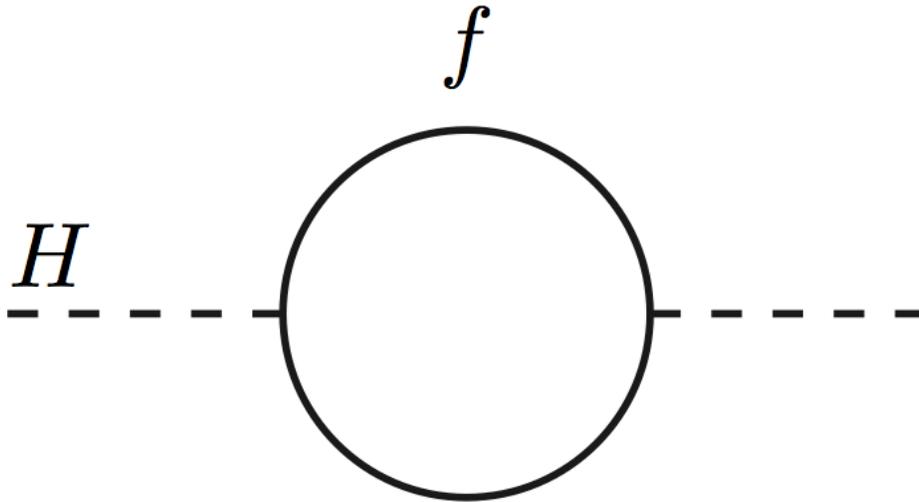


Figure 2.2: Fermionic one loop correction to Higgs. [9]

$$\Delta m_H^2 = -\frac{|\lambda_f|^2}{8\pi^2} \Lambda_{cutoff}^2 \quad (2.2)$$

where  $\lambda_f$  is the Yukawa coupling of Higgs with fermion and  $\Lambda_{cutoff}$  is the energy scale up to which SM is valid. Experimentally measured Higgs boson mass is

around 125 GeV [10, 11] which is much less than Plank cutoff scale ( $10^{19}$ ). An unnatural fine-tuning is required to make observed Higgs mass consistent with radiative corrected mass. There are several theories like supersymmetry (SUSY), little Higgs models [12], extra dimension [13] etc. to solve this issue.

- Dark matter: The presence of dark matter is proved by cosmological and astronomical observations. However, SM does not have any proper particle suitable for dark matter candidate. The neutrinos (even with mass) can only explain a tiny fraction of dark matter. SUSY, however, predicts a suitable cold dark matter candidate.
- Unification of three forces: Following the concept of electro-weak unification, a grand unification of electro-weak and strong interactions would be desirable to accomplish. SM can not unify the gauge coupling constants of three type of interactions which is shown in the left diagram of figure 2.3. The right side diagram of figure 2.3 shows the minimal supersymmetric extension of SM allows the unification at the energy scale of about  $10^{16}$  GeV by introducing superpartners of SM particles which change the strength of coupling.

There are other several issues in SM such as the theoretical prediction of CP violation rate lower than the observed experimental value. The supersymmetric extension of SM increase the CP violation rate in broken symmetric scenario [15]. The gravitational force is also not included in SM as constructing a gauge theory of gravity is very difficult due to renormalization problem.

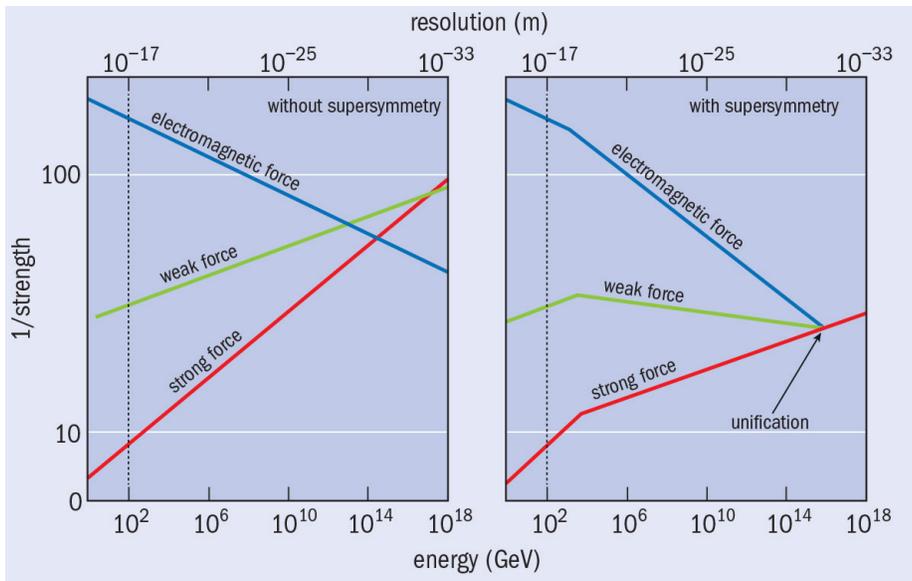


Figure 2.3: The running of three coupling constants with the energy of the interactions are shown [14]. Inverse of coupling constants which are the measure of interaction strength are plotted with the energy. The running of coupling constants are shown in the case of SM (left) and with the minimal supersymmetric extension of SM (right).

## Supersymmetry

During the past fifty years, several theories have been developed as the extension of SM. Among those theories, supersymmetry has become the one that gained major attention both in theoretical and in experimental point of view. This is mainly because of the fact that supersymmetry provides potential and elegant solutions to the SM unresolved problems mentioned in the previous section. The basic aspect of the supersymmetry theory is the new symmetry relating to bosons and fermions. The gist of supersymmetry is reviewed here based on the detail description in the framework of field theory given in Ref. [16, 17]. The mathematical formulation of the theory is based on superalgebra [18, 19] which introduces anti-commuting generators (usually denoted by  $Q$ ) that can change the spin of a SM particle by  $\frac{1}{2}$  unit. The particle represented by the new spin state is known as superpartner of the SM particle. SM particles and their superpartner are organized in supermultiplets which are the irreducible presentation of superalgebra. As the anti-commuting operator  $Q$  commute with gauge transformation generators and square mass operator, fermions and bosons in each supermultiplet share the same mass and quantum numbers. By convention, bosonic (spin 0) superpartners of SM leptons and quarks are called sleptons and squarks whereas the fermionic superpartners of SM bosons are indicated by adding 'ino' after their names and called higgsinos and gauginos. Figure 2.4 shows the list of superparticles. It is important to check how SUSY solve the fine-tuning problem naturally. The loop correction to the Higgs square mass due to a fermion is given by equation 2.2. Now the quantum loop correction due to a scalar field is given by (figure 2.5 ),

$$\Delta m_H^2 = \frac{\lambda_s}{16\pi^2} \Lambda_{cutoff}^2 \quad (2.3)$$

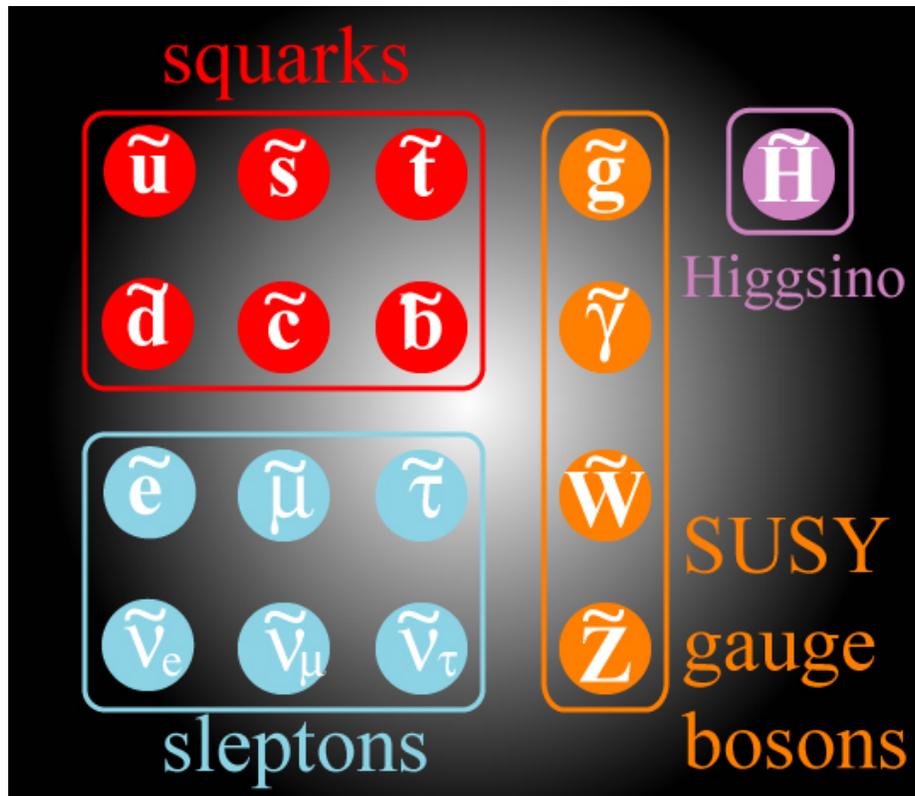


Figure 2.4: Superpartners of SM particles [20].

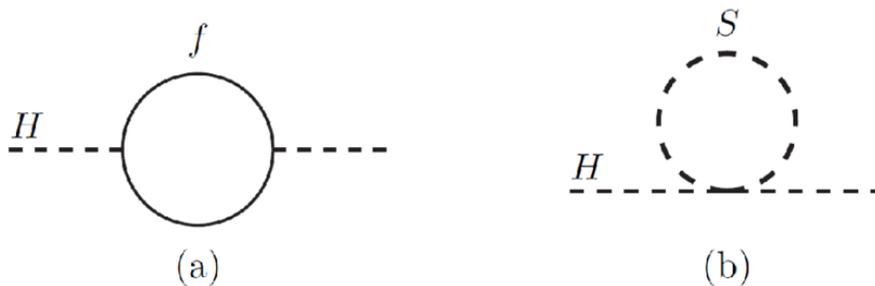


Figure 2.5: Fermionic (a) and bosonic (b) one loop correction to Higgs [21].

Supersymmetry introduces two scalar fields for each SM fermion and also predicts that  $\lambda_s = |\lambda_f|^2$  as the Yukawa couplings to the Higgs are exactly same except they differs by sign because of spin-statistics theorem. Therefore the quantum correction will cancel out due to the sign difference and Higgs mass retain the weak scale mass value after the radiative correction.

SM elementary particles have been observed experimentally but none of their superpartners has been found yet. Therefore SUSY must be a broken symmetry resulting in the mass difference between SM particles and their superpartners. This symmetry breaking should be soft so that it preserves the cancellation of quantum loop correction property. In other words, softly broken supersymmetry theory still must be able to solve the fine-tuning problem. The features and consequences of symmetry breaking will be briefly discussed here in the context of minimum supersymmetric standard model (MSSM) [18] which is the theoretical framework of my analysis.

In MSSM, the electroweak gauginos and Higgsinos mix with each other to create new mass eigenstates after the electroweak symmetry breaking. As the outcome, four neutralinos ( $\tilde{\chi}_i^0$ ) and two charginos ( $\tilde{\chi}_i^\pm$ ) are formed after mixing neutral Higgsinos ( $\tilde{H}_u^0, \tilde{H}_d^0$ ) with neutral gauginos ( $\tilde{Z}$  or  $\tilde{W}^3, \tilde{\gamma}$  or  $\tilde{B}^0$ ) and charged Higgsinos ( $\tilde{H}_u^\pm, \tilde{H}_d^\pm$ ) with winos ( $\tilde{W}^\pm$ ) respectively. Gluino, being on color octet, does not mix with other gauginos. The lightest neutralino ( $\tilde{\chi}_1^0$ ) will be a weakly interacting particle under the constraint of R-parity [22] conserving scenario and can be considered as dark matter candidate. R-parity is a multiplicative quantum number in MSSM theory and defined as,

$$P_R = (-1)^{3(B-L)+3s} \quad (2.4)$$

where B, L, and s are the baryon number, lepton number, and spin respectively. In MSSM Lagrangian, the gauge coupling terms originated from the interaction between

SM and SUSY particles would cause the B and L violation. The inclusion of new symmetry, R-parity, eliminates B and L violating term from the Lagrangian. All the SUSY particles have  $R_P = -1$  and SM particles have  $R_P$  equal to  $+1$ . Following the R-parity conservation, the lightest supersymmetric particle (LSP) provides the dark matter candidate which motivates to search for LSP ( $\tilde{\chi}_1^0$ ) in R-parity conserving SUSY models. This conservation has two other important phenomenological consequences.

- SUSY particles produced in pairs at LHC.
- Each SUSY particle will decay in an odd number of SUSY particles.

These features reflect in the decay topology of the SUSY signals considered in the analysis.

Naturalness principle [23] of MSSM theory sets the constraint on the masses of third generation squarks which makes them interesting to search in the experimental energy limit. Natural SUSY requires the masses of stop and gluino to be at TeV scale which is within the reach of current LHC energy range. This peculiarity motivates strongly to look for stop and gluino with LHC collision data. Moreover, search for stop and gluino in the hadronic channel has a relatively high probability as the cross-section of a hadronic process is high in LHC.

MSSM theory has too many free parameters to be probed experimentally. The number of free parameters is reduced by utilizing various phenomenological approaches. Based on these methods, there are few models like mSUGRA [24], GMSB [25], pMSSM [26] etc. In CMS analyses, we use further simplified models where the mass of the SUSY particle is considered as parameters to probe.

## Simplified Supersymmetric Model

SUSY simplified models [27] are comprised of a small number of parameters (usually 2 or 4) and are the useful tool to characterize the new physics using experimental data. A SUSY simplified model is defined by a particular set of SUSY particles, their production and decay topologies. The constraints are imposed on the mass of SUSY particles which are the free parameters of the models. Two important and flexible features of this approach are the comparison ability of analyses results from different experiments and quality of combining the simplified model interpretation in the framework of a complete model like MSSM. Experimental results are interpreted with the number of simplified models known as Simplified Model Spectra (SMS) [28]. In each model, production of only one type of particle is considered. The primary particle then decays to an LSP directly or via cascade mode through a new particle. The masses of primary particle and LSP are treated as free parameters. The simplified models we are interested in are models of squark-antisquark production (T2 models) and models of gluino pair production (T1, T5 models). More detail on the signal topologies and the constraints on their production cross section is again discussed in chapter 4, 5 and 6.

# Bibliography

- [1] Peter W. Higgs, Phys. Rev. Lett. 13, 508 (1964).
- [2] F. Englert and R. Brout, Phys. Rev. Lett. 13, 321 (1964).
- [3] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, Phys. Rev. Lett. 13, 585 (1964).
- [4] Wikipedia, [https://en.wikipedia.org/wiki/Standard\\_Model](https://en.wikipedia.org/wiki/Standard_Model).
- [5] Wolfgang Lucha *et al.*, Euro. Phys. J. C 77:866 (2017).
- [6] C. Patrignani *et al.* (Particle Data Group), Chin. Phys. C, 40 , 100001 (2016).
- [7] U Dore and D Orestano, Rep. Prog. Phys. 71 106201 (2008).
- [8] M.C.Gonzalez-Garcia, Michele Maltoni *et al.*, Nucl. Phys. B 908 199 (2016).
- [9] Bèranger Dumont, Higgs, supersymmetry and dark matter after Run I of the LHC (2014).
- [10] (CMS Collaboration) S. Chatrchyan *et al.*, Phys. Lett. B 716 30 (2012)
- [11] (ATLAS Collaboration) G. Aad *et al.*, Phys. Lett. B 716 1-29 (2012)
- [12] Martin Schmaltz and David Tucker-Smith, Ann. Rev. Nucl. Part. Sci., 55 229 (2005).

- [13] Lisa Randall and Raman Sundrum, Phys. Rev. Lett. 83, 3370 (1999).
- [14] S James Gates Jr. , <http://live.iop-pp01.agh.sleek.net/2014/09/25/sticking-with-susy/>.
- [15] Yosef Nir, “CP Violation In and Beyond the Standard Model”, arXiv:hep-ph/9911321 (1999).
- [16] S. P. Martin, Adv. Ser. Direct. High Energy Phys. 18 (1998) 1.
- [17] H. Mller-Kirsten and A. Wiedemann, ‘Introduction to Supersymmetry’, World Scientific, 2010.
- [18] Ian J R Aitchison, ‘Supersymmetry and the MSSM: An Elementary Introduction’, (2005)
- [19] Yu. A. Gol’fand, E. P. Likhtman, JETP Lett. 13 323 (1971).
- [20] <https://skullsinthestars.com/2013/08/22/supersymmetry-in-optics/>.
- [21] Tianyou Xie, Supersymmetry And Gauge Couplings Correction.
- [22] S. P. Martin, Phys. Rev. D 46, R2769(R) (1992).
- [23] M. Papucci, J. T. Ruderman *et al.*, JHEP 1209 035 (2012).
- [24] A. Chamseddine, R. Arnowitt *et al.*, Phys. Rev. Lett. 49, 970 (1982).
- [25] A Arbey, M Battaglia *et al.*, Phys. Lett. B. 3. 708 (2012).
- [26] A Djouadi, S Rosier-Lees *et al.*, France. Rep. PM/98-45.
- [27] Daniele Alves, Nima Arkani-Hamed *et al.*, J. Phys. G39, 105005 (2012).
- [28] (CMS Collaboration) S. Chatrchyan *et al.*, Phys. Rev. D 88, 052017 (2013).

# Chapter 3

## LHC and CMS Experiment

This chapter outlines the experimental setup and analysis infrastructure for the SUSY search. Starting from large hadron collider, compact muon solenoid detector, it's subdetectors, data collection, object reconstruction and software framework are recounted in the following sections.

### The Large Hadron Collider

The Large Hadron Collider (LHC) [1], located at the European Organization for Nuclear Research (CERN) in Geneva, Switzerland, is the largest and most energetic particle accelerator in the world. It was built in an existing 27 km circumference tunnel that was constructed for the LEP experiment. The LHC consists of two 27 km rings located at 100 meters underground belongs to the area of two countries: Switzerland and France. Two beams of very high energetic particles (proton or heavy ion like lead and xenon) travel in opposite direction along two beam pipe with a velocity very close to the velocity of light. The motion of charged particle beam is controlled by superconducting electromagnets which are kept at  $-271.3^{\circ}\text{C}$  temperature with the

aid of liquid helium. In order to avoid the collision of charged particles with the gas molecules, an ultra high vacuum is maintained inside the beam pipe. The pressure inside the pipe is of the order of  $10^{-10}$  to  $10^{-11}$  mbar. Besides the pipe, two other vacuum systems equivalent to  $10^{-6}$  mbar pressure are used to insulate cooled cryogenic magnets and helium distribution line. There is a large number (around 9000) of different kinds of magnets including dipoles, quadrupoles and higher order used in LHC to control the motion the beam around the accelerator as well as the beam collision at interaction points. The tunnel has 8 straight sections where the detector and accelerator systems are installed and 8 arcs where the bending magnets are placed to rotate the beam. Currently, LHC hosts four big detectors namely ATLAS, CMS, LHCb and ALICE at four beam interaction points. LHC was designed to accelerate the proton beam to the energy of 7 TeV resulting in the proton proton collision at the 14 TeV center-of-mass energy. However, the highest collision energy achieved so far is 13 TeV during 2015-2016 operation. The proton beam is accelerated to such high energy not in a single step but with the chain of accelerators.

## **LHC Accelerator Chain**

CERN accelerator compound consists of several accelerators. LHC is the last and largest accelerator in this chain. A simplified scheme of multistep proton acceleration is shown in figure 3.1.

The bare protons are obtained from the hydrogen gas by stripping the electron from hydrogen atom with strong electric field inside a metal cylinder. These protons are first accelerated with a DC power supply to 90 keV and then they are focused and energized to 750 keV by a Radio Frequency Quadrupole (RFQ). Thereafter proton beam is placed in a linear accelerator (LINAC2) to increase the energy to 50 MeV. The beam is boosted in a succession of circular accelerators starting from proton

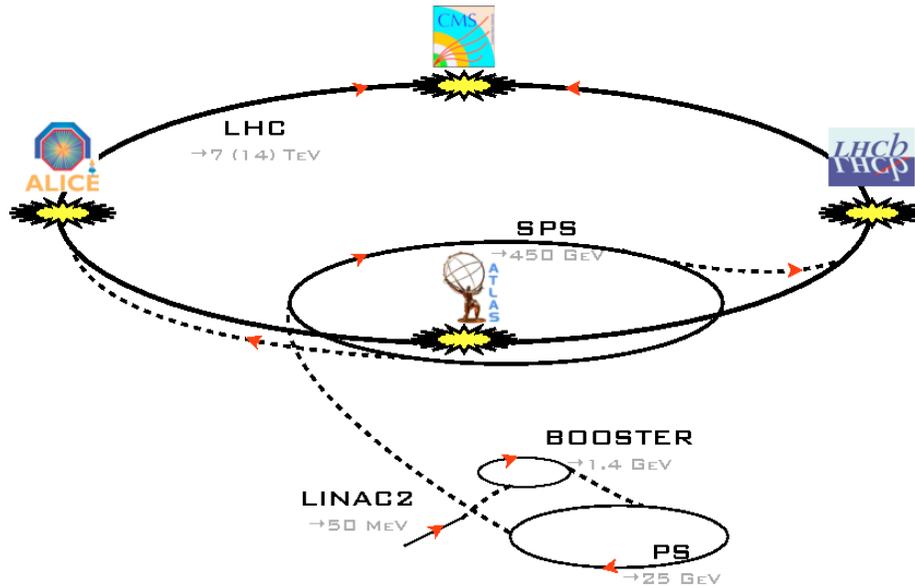


Figure 3.1: LHC full accelerator chain [2]

synchrotron booster (PSB) where the beam energy increases up to 1.4 GeV. After that beam is injected into the proton synchrotron (PS) where it is boosted to 25 GeV. PS delivers the beam into the super proton synchrotron (SPS) which accelerates the beam to 450 GeV. Finally, the beam is injected into the LHC and accelerated up to 20 minutes to achieve 6.5 TeV energy. The proton beam is split into bunches in smaller machines and filled in the LHC. One fill of beam has 2808 bunches in a ring with the number of proton of the order of  $10^{11}$  in each bunch.

## LHC Layout and Operation

LHC ring does not have exact circular shape, rather it made of eight straight and eight arc sections. The dipole magnets are placed at the arcs and the straight sections are used for beam injection, cleaning, dumping as well as colliding the beams where the detectors are installed. The whole tunnel is divided into eight octants as shown in figure 3.2. Each octant consists of a straight section and two 'half of arc' sections

at either end of the straight part.

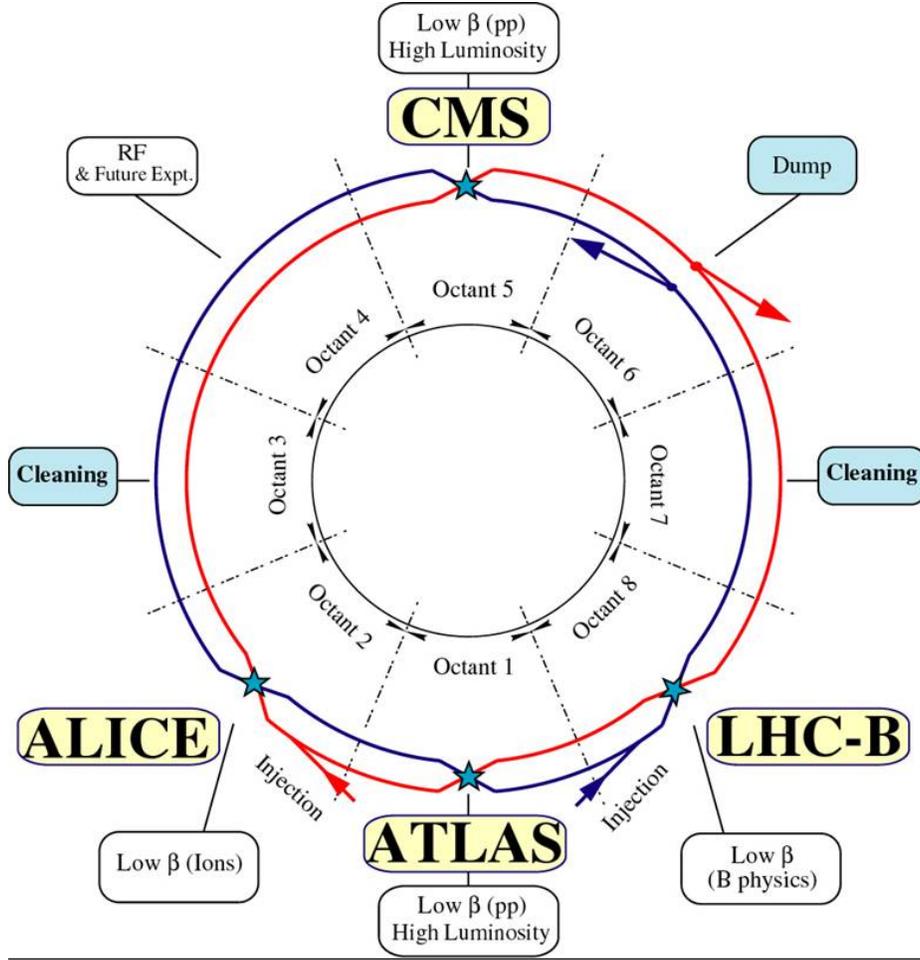


Figure 3.2: LHC layout [3]

The ATLAS detector is located at point 1 while the CMS detector is at point 5. The ALICE detector and clockwise beam injection point are at point 2 while anti-clockwise beam injection point and the LHCb detector are located at point 8. Point 3 and 7 host the place for Collimation systems. The collimation system conserves the beam by cleaning beam transverse momentum dispersion and beam betatron emission. Beam dumping system is situated at point 6 where the beam is extracted from the pipe and swept in a quasi-circular figure by two sets of orthogonally deflecting

dilution kickers (MKB) to debilitate the beam intensity, and finally being absorbed by carbon cylinder. The accelerator system is placed in only one octant, point 4. It contains two radio frequency (RF) systems to boost both beams.

LHC has the different modes of operation. The accelerator mode indicates the summary of LHC machine. This mode is operated with and without beam. Without the presence of beam, accessing and testing of both accelerator and detector part are done. Beam setup mode means the beam is under preparation. In the presence of beam, there are different modes like machine testing, calibrating etc. Finally, stable beam mode is on when detectors are fully functional to collect the data from the colliding beams. The time synchronization between the accelerator and detector systems is achieved and also tested during the different operation modes because it is very important for smooth data taking in stable beam physics run.

## **LHC parameters**

The main ingredients of LHC are discussed in the preceding sections. Here I'll review the principal parameters of LHC. One of the main goals of LHC is to search for new physics by analyzing the rare event occurred in the beam collision. The probability of rare event occurrence depends on the beam energy and on a quantity that measures how well and precisely beams can collide with each other. The number of any particular type of events can be expressed mathematically by the following equation:

$$N = \sigma \mathcal{L} \tag{3.1}$$

where  $\sigma$  is the cross section of that particular type of process obtained from the collision. The  $\sigma$  depends on the collision energy which is limited by the LHC geometry. The second term  $\mathcal{L}$  is known as luminosity, a quantity that measures the likeliness

of particle collision by assessing how perfectly particles can be squeezed into a given space in given time.  $\mathcal{L}$  depends on the beam properties which can be controlled by the accelerator setup. The LHC luminosity can be described by the following equation,

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F \quad (3.2)$$

The description of these parameters are given in table 3.1.

Table 3.1: Description of beam parameters which define the luminosity.

Item	Description
$N_b$	the number particles per bunch
$n_b$	the number of bunches per beam
$f_{rev}$	revolution frequency
$\gamma_r$	relativistic gamma factor for the beam
$\epsilon_n$	the normalized beam emittance, a measure of the spread of the beam in the plane transverse to its motion
$\beta^*$	the impact parameter at the point of collision, a measure of the focusing strength of the magnets at the collision point
$F$	the geometric luminosity reduction factor due to the crossing angle at the interaction point

$F$  is given by equation 3.3 under the assumption that the bunches are much shorter than  $\beta^*$ .  $\theta_c$  is the bunch crossing angle at interaction point,  $\sigma_z$  is RMS bunch length and  $\sigma^*$  is RMS beam size in the transverse plane.

$$F = \left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2\right)^{-1} \quad (3.3)$$

$\mathcal{L}$  is called instantaneous luminosity. LHC was designed for peak luminosity value of  $10^{34} cm^{-2} s^{-1}$  and it was already achieved during 2016 run. Besides the instantaneous luminosity, a common term used in data taking is integrated luminosity. It is defined as,

$$\mathcal{L}_{int} = \int \mathcal{L} dt \quad (3.4)$$

The time duration depends on the status of a stable beam in LHC. The data collected by any detector like CMS is expressed in term of integrated luminosity. It means

that collision data is collected during the time when all parts of CMS detector are functional and LHC provides stable beam. We can convert the integrated luminosity into the total number of collected events using equation 3.1.

Both proton beam and lead ion beam are used in LHC for different physics purposes. My physics analysis based on data coming from proton proton collision at 13 TeV center of mass energy.

## **Compact Muon Solenoid Experiment**

LHC hosts four major experiments as mentioned in the previous section (figure 3.3). CMS experiment [4, 5] is one such big particle physics experiment at LHC. It is based on the CMS detector. All the CMS physics analyses are performed with the data collected by CMS detector. Ideally, the data delivered by LHC should be same as the collected data. But in some cases the CMS detector is unable to take data because of problem either in data acquisition chain or in subdetectors. So the data actually logged by CMS detector is referred as recorded data [6] which are made ready for analysis through a number of steps with the help of dedicated groups of physicists, scientists, and engineers. The experiment consists of people accountable for all the aspects starting from the detector operation to the physics publication. In total CMS employes around 3500 people from all around the world. The detail description of each part of the experiment is beyond the scope of this review. Therefore, the topics relevant to my analysis are outlined in the following sections.

### **CMS Detector**

The Compact Muon Solenoid (CMS) is one of two general purpose detectors along with ATLAS detector. The position of CMS at LHC is shown in figure 3.3. The de-



Figure 3.3: Satellite view of LHC with the location of four experiments [7].

tector coordinate system is shown in figure 3.4. CMS follows a cylindrical coordinate system due to its cylindrical shape. Instead of azimuth angle  $\theta$ , pseudorapidity ( $\eta$ ) as defined in equation 3.5 is conveniently used for angular measurement. Pseudorapidity is used as a spatial coordinate describing the angle of a particle relative to the beam axis. In LHC, for high energetic relativistic particles, pseudorapidity difference ( $\Delta\eta$ ) becomes invariant under longitudinal Lorentz boost. Pseudorapidity can be measured only from the momentum (three-momentum and its longitudinal component) of the particle according to equation 3.6. In CMS, common and convenient choice of a plane for measuring any quantity is the transverse plane, i.e., x-y plane. For example,  $p_T$  is a widely used kinematic variable which is the transverse component of total momentum.

$$\eta = -\ln \left[ \tan \frac{\theta}{2} \right] \quad (3.5)$$

$$\eta = \frac{1}{2} \ln \left( \frac{|p| + p_L}{|p| - p_L} \right) \quad (3.6)$$

CMS is huge detector with 21.6 m length and 14.6 m diameter and with the total weight of 13800 tons. To fulfill its 'general purpose' criteria, CMS is assembled of

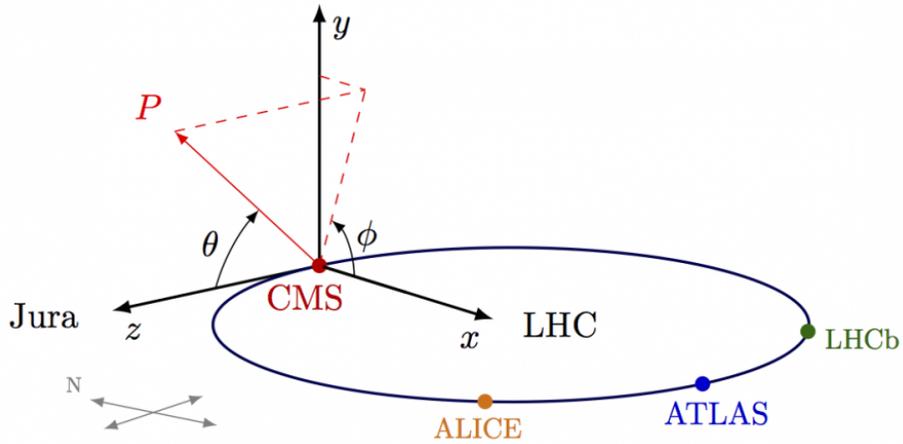


Figure 3.4: CMS coordinate system [8]. Z axis is along the beam direction towards mount Jura, x axis points towards the center LHC ring and y axis is pointing upward. If  $P$  is vector representing the direction of a particle traverses the detector, picture shows how angle  $\theta$  and  $\phi$  are determined.

multiple layers of subdetectors: silicon based inner tracker, crystal based calorimeter and gas detector based muon system. A superconducting solenoid is placed after the calorimeter to provide magnetic field throughout the detector. Compared to its weight, CMS size relatively small, hence the name compact comes with muon solenoid. The cylindrical part of the detector is called the barrel region which is extended by endcaps located at both sides of the barrel. The schematic layout of the CMS detector is shown in figure 3.5 and the key elements are briefly discussed below.

- **Tracking system:**

Tracking system is the most inner layer of the detector. The purpose is to track the path of the particles passing through the detector. As charged particles bend by the magnetic field created by the solenoid, tracker reconstructs the trajectory of the charged particles. From the trajectory, momenta of the charged particles are determined. The first characteristic of the tracker is to record the particle tracks accurately with minimal change of their kinematics. Moreover, it has to

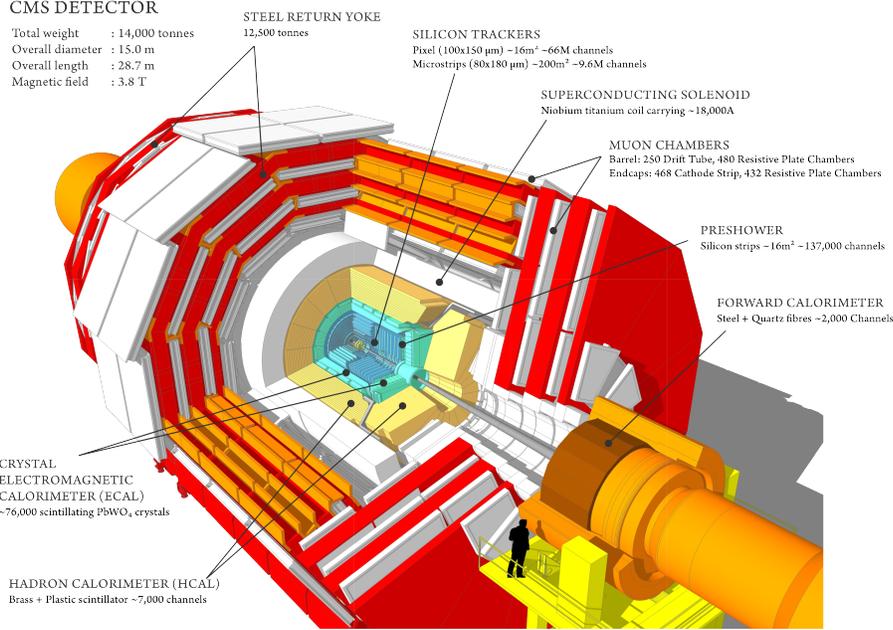


Figure 3.5: Different layers of CMS detector [9].

be highly granular and fast responsive since lots of particles come from each bunch crossing and the collision happens in 25 ns interval. The second feature is to be radiation hard as the tracker is the closest section of the detector to the LHC beam. These requirements make silicon to be a suitable choice for the tracking system. CMS tracker consists of silicon pixel detectors and silicon strip detectors.

– **Silicon pixel:**

Silicon pixel system is the closest tracker to the collision point. It has three barrel layers at distances of 4.4 cm, 7.3 cm, and 10.2 cm from the beam axis and four endcap disks located at  $\pm 34.5$  cm and  $\pm 46.5$  cm along the z axis. Each layer is segmented into tiny silicon tiles of  $100 \mu m \times 150 \mu m$  dimension. When charged particles pass through a silicon tile, it generates

an electric signal which is collected by the electronics attached to the tile. Altogether, the pixel system has 65 million pixels.

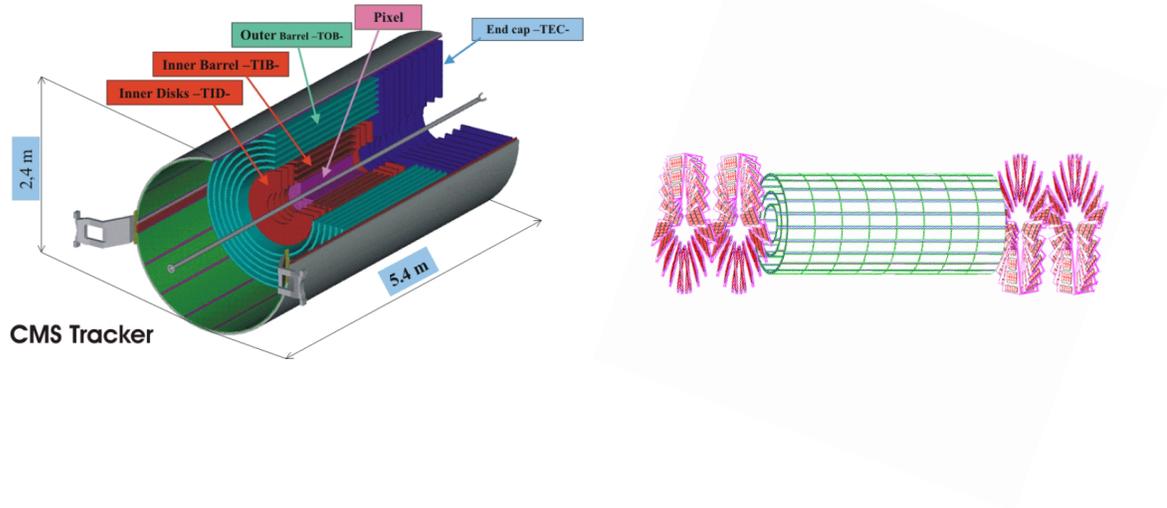


Figure 3.6: CMS silicon tracker system. Left picture [10] shows a 3D view of full tracker system while the right one [11] shows only pixel part.

– **Silicon strips:**

Silicon strips are located at the outer side of pixel system. Strips detector contains four subsections:

- \* four layers of tracker inner barrel (TIB),
- \* three tracker inner disks (TID),
- \* six layers of tracker outer barrel (TOB) and
- \* nine tracker endcap disks (TEC).

There are 10 million silicon sensors made of silicon strip of size  $320 \mu m$  or  $500 \mu m$ . Set of sensors are connected to one microelectronic chip which read the signal in silicon strip.

- **Calorimeter:**

A calorimeter is a part of the detector where the energy of traversing particles are measured. Calorimeter is constructed in such a way that it stops the particle in order to get full energy deposition of that particle. CMS calorimeter is split into two parts which are described below.

- **Electromagnetic calorimeter:**

CMS Electromagnetic calorimeter (ECAL) is a homogeneous calorimeter made of lead tungstate ( $PbWO_4$ ) crystals. Electron and photon lose their energy inside the crystal by making electromagnetic shower caused by bremsstrahlung radiation and pair production. The scintillating crystal is ionized or excited by the shower resulting in light emission as the signal. This signal is amplified and read by APD in the barrel and VPT in endcap region. The Barrel ECAL contains 61200 23 cm long crystals with  $22 \times 22 \text{ mm}^2$  cross-section while 22 cm long 7324 crystals with an area of  $28.6 \times 28.6 \text{ mm}^2$  form the endcap section. Preshower detectors are installed in front of ECAL endcaps to discriminate between pair of photons from neutral pion decay and single prompt photon. Preshower consists of two lead radiators, each followed by a silicon microstrip detector. The electromagnetic shower caused by a photon is detected and measured by silicon sensors. The energy resolution of CMS ECAL can be represented by following expression [12]:

$$\frac{\sigma(E)}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{120 \text{ MeV}}{E} \oplus 0.3\% \quad (3.7)$$

where the first term is called stochastic term arised from shower fluctuation, second tern is for noise and ther third one ia constant term. The

measured values are obtained from beam test with electron. The average ECAL energy resolution is of the order of 2-5% for electron and 1-5% for photon [13].

– **Hadronic calorimeter:**

After the ECAL, CMS has hadronic calorimeter (HCAL) which measure the energy of hadrons created in the collision. HCAL is built with alternative layer of plastic scintillator and brass or steel absorber. Hadrons interact with dense absorber and develop the hadronic shower. Plastic scintillators emits photons being excited by the shower. These light signals are collected and amplified by photo detectors called HPDs. After that, analog signals are sent to FrontEnd electronic section where the charge pulse are integrated and digitized by QIE chips. HCAL has four sections (figure 3.7) as mentioned below:

- \* Hadron Barrel (HB): It is located just outside ECAL and inside the solenoid. HB covers  $|\eta| < 1.392$  range.
- \* Hadron Endcap (HE): HE located at the two ends of HB, extending the  $|\eta|$  range from 1.3 to 3.
- \* Hadron Outer (HO): This outer calorimeter is located outside of the magnet and plays a complementary role to HB in collecting energy when the hadronic shower extends beyond HB.
- \* Hadron Forward (HF): HF detectors sit at  $\pm 11.2\text{m}$  along z axis and cover up to  $|\eta| < 5$  region. It is placed to detect the Cherenkov radiation resulting from particle shower. HF is made of steel absorbers and quartz fiber radiators.

Each section of HCAL is segmented into tiny  $\Delta\eta \times \Delta\phi$  tiles called HCAL

cells. HCAL energy measurement has larger uncertainty compared to ECAL because of energy loss in absorber and sampling design of HCAL. Even after non-linearity correction, fluctuation remains and causes poor resolution. CMS HCAL + ECAL energy resolution for pion is [12]

$$\frac{\sigma(E)}{E} = \frac{84.7\%}{\sqrt{E}} \oplus 7.4\% \quad (3.8)$$

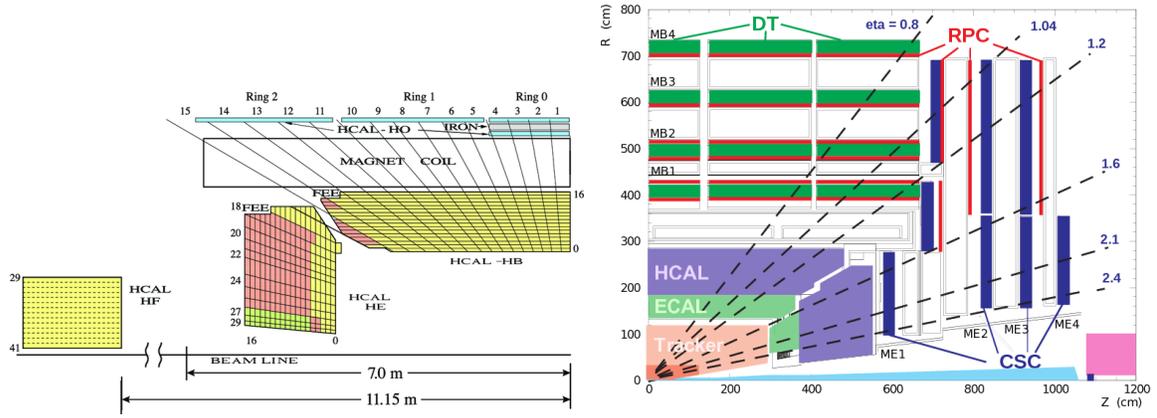


Figure 3.7: Cross-sectional view of HCAL (left) [14] and Muon chambers (right) [15] in r-z plane

- **Muon system:** Since muons can pass through all the inner sub detectors and solenoid, muon chambers are installed outside the solenoid magnet to detect the muon. A cross-sectional view of CMS muon system is shown in figure 3.7. To achieve optimal performances, the muon system makes use of three types of gaseous detectors as listed below:

- **Drift tube:** Drift tubes (DT) form the gas chambers in the barrel region. There are 250 DT chambers each of which is made of tiny DT cells. A DT cell itself a 42 mm wide gas filled tube with a positively charged wire

inside. When muon traverses the chambers, it knocks off the electrons of gas atom. These electrons are collected by the anode wire to be read as a signal. The passing position is determined by drift time. There are 4 stations each having 12 layers of DT cells. Middle 4 cells measure the position along z axis while outer 8 cells measure x-y coordinate. Thus muon tracks can be reconstructed from the position information provided by DT cells.

- **Cathode strip chambers:** In the endcap region, muon system consists of 540 radiation hard cathode strip chambers (CSC). CSC is made of anode wires placed perpendicular to cathode copper stripes inside gas volume. Muon knocks off the electron from gas atom which creates an avalanche while moving towards anode and ions also move towards cathode. Anode wires determine azimuthal angle whereas cathode strips measure the transverse distance from z axis.
- **Resistive plate chambers:** Resistive plate chambers (RPC) are installed both in barrel and in endcap section combined with DT and CSC to construct muon trigger system. RPC has two oppositely charged, high resistive parallel plastic plates with gas filled between them. Like CSC, an avalanche of electrons is developed when a muon pass. The electrons are collected by metallic strips. The pattern of strip hits quickly assesses the muon momentum for trigger decision. RPC has 1 ns time resolution which makes it a suitable choice for muon trigger system.

These gas chambers are interleaved in the iron yoke which return the magnetic flux created by solenoid magnet.

- **Solenoid:**

In order to measure the momentum from the curvature of charged particle trajectory, strong magnetic fields are generated using superconducting solenoid magnet. The solenoid is 13 meters in length and 6 meters in diameter. It is placed at outside the HCAL and before muon system. In a fully functional state, the magnet can produce a magnetic field of 3.8 Tesla inside the solenoid. The magnetic field outside the solenoid is maintained by the iron yoke interleaved with muon chambers. This iron yoke produces 2 Tesla magnetic field in the muon system by returning the magnetic flux coming from the coil. The directions of the fields at inside and outside of the solenoid are opposite to each other.

- **Trigger:**

With the 25 ns bunch crossing, intrinsic collision rate in LHC is 40 MHz which means roughly 40 million events are generated per second in CMS detector. CMS data acquisition system does not allow to record that huge amount of data. Moreover, most of those events are already understood and not interesting for physics analysis. Therefore we need a framework to reduce the event rate while selecting the interesting events. CMS has this framework called trigger system which reduces the event rate from 40 MHz to 1 kHz. Only the triggered data are stored from the collision. The rate reduction is done in two steps. In the first step, a hardware based FPGA programmable level one trigger (L1T) reduces the rate to 100 kHz. Then in the next step, a software based high level trigger (HLT), running on huge event filter computing cluster, decrease the event rate to 1 kHz.

L1T combines the local information from subdetectors like muon system and calorimeters into global information and then roughly reconstructs physics objects by FPGA program. The global triggers based on these roughly recon-

structed objects decide to block or pass the event.

HLT reads the event information from the whole detector after the event is triggered by L1T. Using the detector information, HLT software promptly reconstruct physics objects on which further selections are applied to filter out the events for various physics analysis. HLT filter or selection is a software based algorithm which stores or abandons an event. These filters are called HLT paths which are used to select the events to form the primary dataset (PD). Events in a PD can pass one trigger path or a group of trigger paths.

The data triggered by L1T from central data acquisition (DAQ) system are filtered in HLT farm and made ready for offline reconstruction. The output from HLT is called RAW data and already filtered into various PD based on L1T and physics analysis. For example, HTMHT or MET dataset is used for all hadronic SUSY search. RAW data in different PD stream are transferred to CERN Tier-0 or any other Tier-1 for offline event reconstruction.

## **Event Reconstruction**

The First step in offline reconstruction is unpacking the RAW dataset to DIGI dataset where electronic based counts are changed into detector based digital hits. In next step, the digi hits are converted into reconstructed hits incorporating all the information about calibration, hit-detector map etc. Then physics objects are reconstructed. The dataset containing RecHits and offline physics objects are called RECO. RECO dataset are further reduced in size by dropping the information not used for physics analysis in order to form an Analysis Object Data format called miniAOD. The reconstruction algorithm of physics objects in an event is briefly described here.

Event reconstruction includes vertex determination, particle identification, and higher level objects like jet, met (missing transverse momentum) etc. reconstruction. CMS

employs particle flow (PF) [16, 17] algorithm for event reconstruction. PF algorithm utilizes the information from different sub detectors to identify stable particles starting from track and energy cluster reconstruction. Different kind of particles interact differently with each subdetector and charged particles follow curved trajectories under the magnetic field (figure 3.8). Based on these two facts, PF algorithm combines each sub detector information to identify five stable type of particles: muon, electron, photon, charged hadron and neutral hadron.

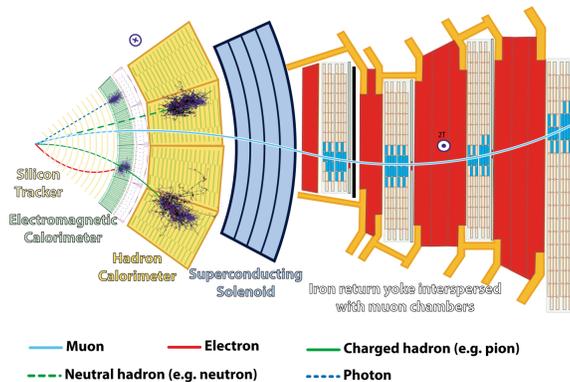


Figure 3.8: Partial cross-sectional view of CMS subdetectors and their interaction status with the particles come from the collision [18].

As mentioned above, track reconstruction is the starting point of this process. An iterative track finding algorithm [19] is used to reconstruct the tracks. This strategy is very reliable to identify tracks with wide  $p_T$  range from a few hundred MeV to a few hundred GeV with almost 100% efficiency and very low fake rate. The tracks, formed with pixels and strips hits at silicon tracker, are used for vertex reconstruction and momentum measurement. The primary vertex (PV) or proton proton collision point is reconstructed with the associated tracks. There can be multiple PVs due to pile up [20], but for the analysis, one PV is taken for which sum of tracks  $p_T^2$  is maximum. A secondary vertex which is the decay point of any long-lived particle is also recon-

structed using the information of tracks displaced from PV. The momenta measured from charged tracks provide the key information to form four vector (Lorentz vector) of PF particles. In the muon system, Tracks are also reconstructed which are used to identify PF muons.

Combining the information of energy deposition in ECAL and HCAL is known as calorimeter clustering and it is another cornerstone of PF process. The clustering is done locally in different part ECAL and HCAL. A cell in ECAL or HCAL with the energy deposition above a certain threshold is chosen as cluster seed. Next, a topological cluster is constructed by aggregating the cells around the seed cell. Each seed cell in a topological cluster gives rise to a particle flow cluster. The final energy and position of a particle flow cluster are determined by resolving the energy sharing with an iterative method.

The charged tracks, ECAL clusters, HCAL clusters, and muon system tracks are treated as the PF elements which are connected together to make PF blocks by a link algorithm. The final set of particles are reconstructed from these PF block of elements by particle identification algorithm. The link algorithm also ensures the null probability of double counting of a particle in an event. Five elementary particles such as muon, electron, photon, charged and neutral hadrons along with higher level objects like jet and met are reconstructed by PF algorithm. A brief description of our analysis oriented objects is given below.

- Muons: Muons pass through ECAL and HCAL with minimal energy deposit in the calorimeter. A block linking between a charged track in tracker and in muon system gives a global PF muon. The energy of the muon is determined by calculating the momentum of its track. The corresponding track is removed from the set of blocks. In our analysis, muons are selected with Physics Objects

Group (POG) recommended PF mini-isolation criteria [21].

- **Electrons:** Electrons deposit energy in ECAL by inducing shower in ECAL. Therefore an ECAL cluster linked with a charged track is used to identify a PF electron. For the analysis, electrons selected by POG recommended 'Cut Based VETO' [22].
- **Jets:** Jets are the experimental signature of collimated stable hadrons originated from a quark or a gluon by hadronization. PF jets are formed by clustering the four momenta of all stable PF candidates. There are various jet clustering algorithms available in HEP community. CMS uses an algorithm known as *anti - k<sub>T</sub>* [23] to combine PF candidates in order to make a PF jet. The size of the jet is determined by a parameter of the algorithm known as R which is related to the distance between two objects to be clustered through following equation.

$$d_{ij} = \min(p_{Ti}^{-2}, p_{Tj}^{-2}) \frac{\Delta R^2}{R^2}, \quad (3.9)$$

$$\Delta R = \sqrt{\delta\eta^2 + \delta\phi^2}$$

where  $d_{ij}$  is the chosen distance parameter in the algorithm,  $p_{Ti}$  and  $p_{Tj}$  are the transverse momenta of two objects  $i$  and  $j$  respectively and  $\Delta R$  is the angular distance between them in  $\eta, \phi$  space. Currently, CMS supports jets clustered with  $R = 0.4$  and  $0.8$  which are known as AK4 and AK8 jets accordingly. Jet energy calibration is performed so that it can represent the source quark or gluon energy properly and uniformly over the detector. Energy is also corrected to account for pileup contribution in jet energy measurement. In our analysis, JET-MET [24] recommended, energy-corrected [25, 26] jets are used.

- **b-Jets:** A jet originated from b quark or contains b quark is known as b-jet. The b hadrons decay at a point displaced from PV due to certain lifetime of

approximately 1.6 ps. Based on this displaced vertex along with other features such as jet mass close to b quark mass, wider jet size, higher multiplicity of jet constituents found in b-jet, several algorithms are developed to identify b-jets. In our analysis b-jets are selected using the combined secondary vertex algorithm [27].

- **Missing Transverse Momentum:** Missing transverse momentum ( $p_T^{miss}$ ) assesses the imbalance in transverse momenta in an event. It is the magnitude of negative vector sum of all reconstructed PF candidates'  $\vec{p}_T$ . Like jets,  $p_T^{miss}$  is also corrected by various filters and energy correction factors as mentioned in Ref. [28]. In our SUSY search analyses,  $p_T^{miss}$  is an important variable as the stable lightest supersymmetric particles being undetected by the detector cause an imbalance in transverse momenta.

## Event Simulation

Event simulation is an important part of any physics analysis. Before analyzing the collisional data collected by CMS detector, any analysis is designed based on the data produced by computer simulation. The characteristic of actual collisional data is assessed in prior by evaluating the simulated data. So each step of the actual data event reconstruction process starting from proton proton collision to PF objects reconstruction is replicated in computer simulation by the dedicated software programs. It is basically performed in two steps: physics process simulation and detector response simulation.

In the first step, physics process events are generated from proton proton collision according to their production cross sections and branching ratios using parton distribution function (PDF) information [29]. Then these primarily generated particles

decay to secondary particles which in turn decays to other particles. This process continues until the stable particles are obtained. The particle production, decay, and quark gluon hadronization are modeled by various theory module based software package. In our analysis, MadGraph [30] or POWHEG [31] is used as parton level generator and PYTHIA [32] is used for parton shower and hadronization to get stable particles in any event.

In next step, these particles are set to pass through the detector where their interactions with the detector and the detector responses as the electronic signals are simulated. This part is done by a software package called GEANT4. [33] It has the exact detector geometry, material, and electronic configuration as CMS has. The simulated detector responses are treated as detector signal as we get from the real detector and then used for offline event reconstruction.

One privilege in simulated events is generated level particle information is known, so any discrepancy originated due to offline reconstruction can be probed by comparing reco level particles with gen level ones. In HEP, simulated events are also called Monte Carlo (MC) events.

# Bibliography

- [1] <https://home.cern/topics/large-hadron-collider>.
- [2] Anna Henrichs, [http://inspirehep.net/record/1119569/files/figures\\_experiment\\_accelerator\\_chain.png](http://inspirehep.net/record/1119569/files/figures_experiment_accelerator_chain.png).
- [3] The Application of the Monte Carlo Code FLUKA in Radiation Protection Studies for the Large Hadron Collider - Scientific Figure on ResearchGate. Available from: [https://www.researchgate.net/Schematic-layout-of-the-LHC\\_fig1\\_284972340](https://www.researchgate.net/Schematic-layout-of-the-LHC_fig1_284972340).
- [4] <https://cms.cern/>.
- [5] (CMS Collaboration) S. Chatrchyan *et al.*, JINST 3 S08004 (2008).
- [6] <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [7] <https://sciencesprings.wordpress.com/2015/03/08/from-bbc-lhc-restart-we-want-to-break-physics/cern-lhc-map-30/>.
- [8] Izaak Neutelings, [https://wiki.physik.uzh.ch/cms/\\_detail/latex:cms\\_coordinate\\_system.png?id=latex%3Aexample\\_spherical\\_coordinates](https://wiki.physik.uzh.ch/cms/_detail/latex:cms_coordinate_system.png?id=latex%3Aexample_spherical_coordinates).
- [9] <https://cms.cern/detector>.

- [10] Search For New Physics In The Compact Muon Solenoid (CMS) Experiment And The Response Of The CMS Calorimeters To Particles And Jets - Scientific Figure on ResearchGate. Available from:[https://www.researchgate.net/The-CMS-tracker-is-shown\\_fig4\\_41217625](https://www.researchgate.net/The-CMS-tracker-is-shown_fig4_41217625).
- [11] Search For New Physics In The Compact Muon Solenoid (CMS) Experiment And The Response Of The CMS Calorimeters To Particles And Jets - Scientific Figure on ResearchGate. Available from:[https://www.researchgate.net/The-CMS-pixel-detector-is-depicted\\_fig5\\_41217625](https://www.researchgate.net/The-CMS-pixel-detector-is-depicted_fig5_41217625).
- [12] Francesca Cavallari, J. Phys.: Conf. Ser. 293 012001, (2011)
- [13] (CMS Collaboration) S. Chatrchyan *et al.*, JINST 8 P09009 (2013).
- [14] (CMS Collaboration) S. Chatrchyan *et al.*, JHEP 01 163 (2014).
- [15] (CMS Collaboration) Min Suk Kim *et al.*, JINST 8 T03001 (2013).
- [16] (CMS Collaboration), A.M. Sirunyan *et al.*, JINST 12 P10003 (2017).
- [17] (CMS Collaboration), CMS-PAS-PFT-09-001 (2009).
- [18] David Barney, <https://cds.cern.ch/record/2120661>.
- [19] (CMS Collaboration), CMS Note CMS-NOTE-2006-041 (2006).
- [20] A Perloff (CMS Collaboration), Journal of Physics: Conference Series 404 012045 (2012).
- [21] (CMS Collaboration) S. Chatrchyan *et al.*, JINST 8 P11002 (2013).
- [22] (CMS Collaboration) V. Khachatryan *et al.*, JINST 10 P06005 (2015).
- [23] Matteo Cacciari, Gavin P. Salam, Gregory Soyez, JHEP 0804:063 (2008).

- [24] CMS JME Results, <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsJME>.
- [25] (CMS Collaboration) V. Khachatryan *et al.*, JINST 12 P02014 (2017).
- [26] M. Cacciari and G. P. Salam, Phys. Lett. B 659 119 (2007).
- [27] (CMS Collaboration) S. Chatrchyan *et al.*, JINST 8 P04013 (2013).
- [28] (CMS Collaboration) S. Chatrchyan *et al.*, JINST 6 09001 (2011).
- [29] R. Placakyte, ‘Parton Distribution Functions’, XXXI PHYSICS IN COLLISION (2011).
- [30] J. Alwall *et al.*, JHEP 07 079 (2014).
- [31] S. Alioli, P. Nason *et al.*, JHEP 11 070 (2007)
- [32] T. Sjostrand *et al.*, ‘An introduction to PYTHIA 8.2’, Comput. Phys. Commun. 191 159 (2015).
- [33] GEANT4 Collaboration, Nucl. Instrum. and Methods 506 (2003).

# Chapter 4

## Search for scalar top quark using top quark tagging

The theoretical background of SUSY search, the experimental tools, instruments and the physics objects used in CMS analysis have been discussed in the previous chapters. This chapter presents the preface to the stop and gluino search in all hadronic channels. We consider SUSY signals where stops or gluinos are produced in pair and decay to four or more hadronic jets in the final state. The assumption of R-parity conservation gives stable and weakly interacting lightest SUSY particles (LSP). Therefore LSPs (neutralinos) leave our detector being undetected resulting in an imbalance in transverse momenta ( $p_T^{miss}$ ). So the SUSY signals we search are characterized by multiple jets and large  $p_T^{miss}$ . The target signal models considered in the analysis are shown in figure 4.1 and figure 4.2.

There are several processes predicted by SM that can end up with a similar final state as our signals. Those are called SM backgrounds to the signals. The first step is to define a search region (SR) with suitable selection criterion to mitigate the contribution from SM backgrounds in SR while retaining signal events as much as

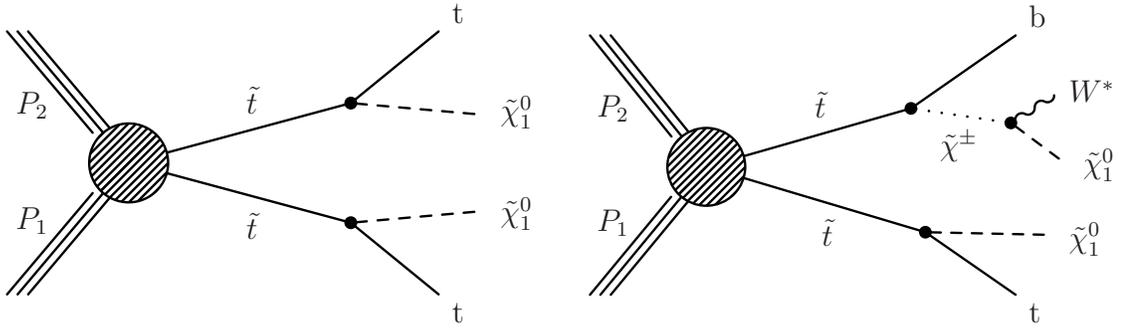


Figure 4.1: Signal models of direct stop production: top squark pair production with the top squark decaying into a top quark and neutralino (left) or into a bottom quark and chargino (right). The SUSY simplified model topology shown at the left is referred to as T2tt and the right one as T2tb.

possible. A novel top quark reconstruction algorithm is employed to identify the top quarks produced in signal decay chain. Signal processes also have b quarks which reconstructed as tagged b-jets. The scalar sum of jets transverse momenta is called  $H_T$  which is defined in following equation,

$$H_T = \sum_{jets} |p_T^{\vec{}}| \quad (4.1)$$

$H_T$  can be used as discriminating variables between background and signal events when signals have higher  $H_T$  value due to numbers of high  $p_T$  jets. To improve background suppression an extended transverse mass variable called stranverse mass ( $M_{T2}$ ) [1, 2] is added. The SR is defined in term of number of reconstructed top quarks ( $N_t$ ), number of tagged b-jets ( $N_b$ ),  $p_T^{miss}$ ,  $M_{T2}$  and  $H_T$  after the several selection requirements for background reduction. In general, T2tt signal has higher sensitivity than T2tb and T1tttt is observed to have more sensitivity compared to other gluino induced signal models in the search region. However, because of having various  $N_t$ ,  $N_b$  multiplicity and different range of  $p_T^{miss}$ , signal sensitivity changes

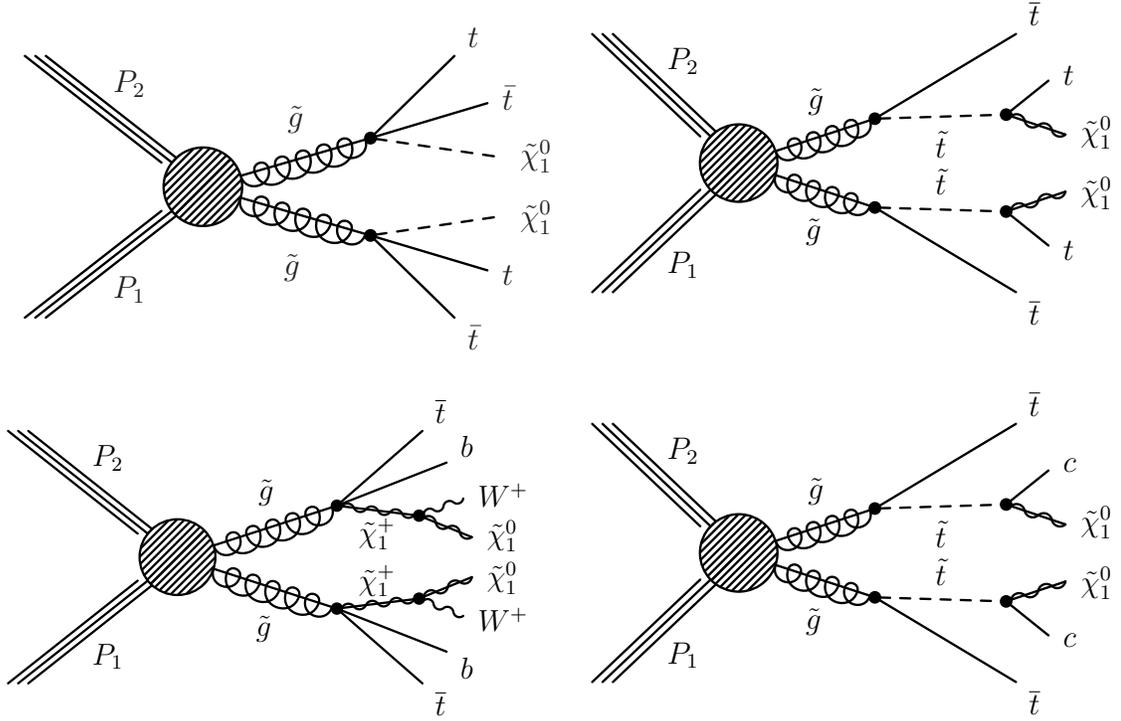


Figure 4.2: Signal models of gluino mediated processes: gluino pair production with the gluino decaying into neutralinos and top quarks through off-shell top squarks (top left), or the gluino decaying into a top quark and on-shell top squark which decays to a top quark and a neutralino (top right), or gluino decaying into off-shell top or bottom squark (bottom left), or gluino decaying into on-shell stop which decays to charm quark and neutralino (bottom right). The SUSY simplified model topology shown at the top left is referred to as T1tttt, top right one as T5tttt, the bottom left one as T1ttbb, and the bottom right model as T5ttcc.

with  $N_t$ ,  $N_b$ ,  $p_T^{miss}$  etc. The SR is therefore, divided into a number of search bins (SB) based on these search variables. The next step is to estimate the remaining SM background events in the SBs. The main background contribution comes from  $t\bar{t}$ ,  $W + jets$ , and single top processes. The  $Z + jets$ , QCD,  $t\bar{t}Z$ , and other rare processes also contribute significantly in some search bins. Then the observed data is compared with the total estimated SM processes in all SBs. Finally, data is interpreted in term of simplified models and the limits are set on model parameters.

2015 [3] and 2016 [4] data analyses are based on the data collected in 2015 and 2016

respectively. Both analyses were done blindly to avoid possible experimental bias caused by selection cuts tuning and background method optimization. Though both analyses follow the similar strategy as described above, there are few changes in the case of signal topology, search variables and quite significant differences in top quark tagging algorithm and background estimation. The detailed description is presented in the following two chapters.

# Bibliography

- [1] C. G. Lester and D. J. Summers, Phys. Lett. B 463 99 (1999).
- [2] A. Barr, C. Lester, and P. Stephens, J. Phys. G 29 2343 (2003).
- [3] (CMS Collaboration) V. Khachatryan *et al.*, Phys. Rev. D 96, 012004 (2017).
- [4] (CMS Collaboration) (A.M. Sirunyan *et al.*), Phys. Rev. D 97, 012007 (2018).

# Chapter 5

## 2015 Data Analysis

The analysis is designed for the maximum sensitivity to the SUSY simplified model T2tt topology (Figure 4.1) resulting in final states with multiple jets, b-tagged jets, no leptons, and large  $p_T^{miss}$ . The search optimization also considers sensitivities to T2tb and T1tttt models. The data used in this study corresponds to  $2.3 \text{ fb}^{-1}$  pp collision data collected in 2015 at a center-of-mass energy of 13 TeV. Targeting the hadronic final states, data are initially selected by requiring a number of jets and b-jets ( $N_{jets}$  and  $N_b$ ) and a large  $p_T^{miss}$ .

The data selection process starts with the triggers followed by baseline selections and proceeds to the definition of the search bins. The top reconstruction and identification procedure (top tagging) is described in this section 5.2 and the Monte Carlo (MC) samples that model signal and backgrounds are cataloged in section 5.5.

### Trigger

Data are collected with hadronic triggers for the events in the search region as well as for the single lepton and multi jets events used in lost lepton ( 5.6.2) and QCD background ( 5.6.4) estimation respectively. Following set of  $H_T$  ( 4.1) and  $p_T^{miss}$

based triggers is used as search triggers,

- *HLT\_PFH350\_PFMET100\_NoiseCleaned\_v\**,
- *HLT\_PFH350\_PFMET100\_JetIDCleaned\_v\**,
- *HLT\_PFH350\_PFMET100\_v\**.

In general, CMS trigger name contains the object the trigger is made of, the trigger threshold, an abbreviation of object reconstruction and also special filters if they are applied to form the trigger path. Here, our search triggers require two thresholds on  $H_T$  and  $p_T^{miss}$  with the values of 350 GeV and 100 GeV respectively. 'NoiseCleaned' and 'JetIDCleaned' indicate the filters associated with the triggers, used to remove the calorimeter noise and badly reconstructed jets.

Trigger efficiencies are estimated in term of offline  $H_T$  and  $p_T^{miss}$  with the base sample events collected by an independent trigger,

- *HLT\_Ele27\_eta2p1\_WPLoose\_Gsf\_v\**.

The additional requirements of having a single electron reconstructed with loose working point using Gaussian Sum Filter (GSF) with  $p_T > 27$  GeV,  $|\eta| < 2.1$  threshold, one jet with  $p_T > 50$  GeV, four jets with  $p_T > 30$  GeV are also applied. The efficiency is determined as the probability of accepting events by the primary triggers from the base sample events. Trigger efficiencies as a function of  $H_T$  and  $p_T^{miss}$  are shown in fig. 5.1.

Fig. 5.1 shows that this method yields an unbiased estimate of the trigger efficiency in all-hadronic events with  $p_T^{miss} > 200$  GeV and  $H_T > 500$  GeV, which are baseline selection requirements for this analysis. After these two cuts, the measured efficiency is above 97% in the search region.

Hadronically decaying  $\tau$  lepton background events are estimated from a single muon control sample collected with the following trigger,

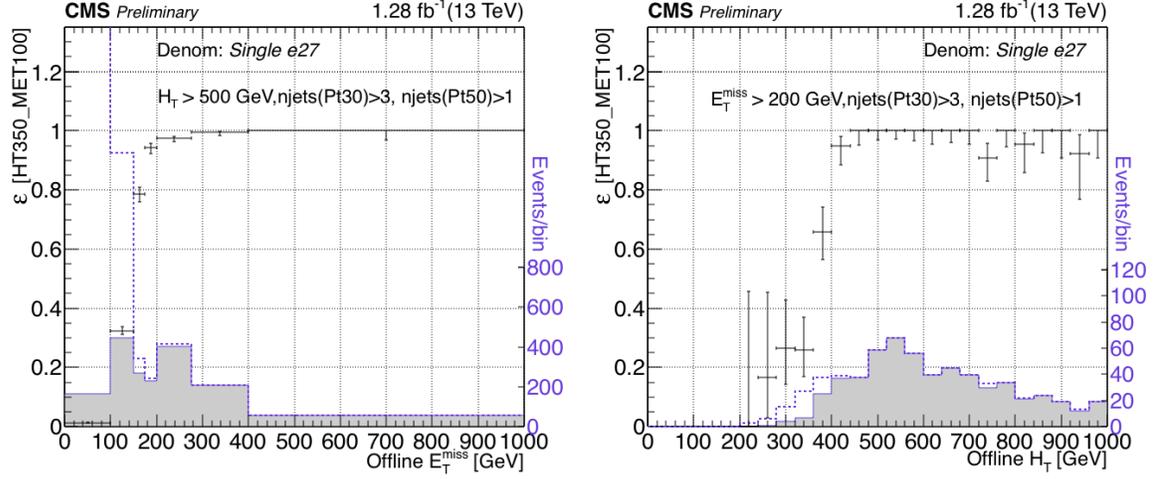


Figure 5.1: The trigger efficiency for  $HLT\_PFHT350\_PFMET100*$  as a function of  $p_T^{miss}$  (the axis label  $E_T^{miss}$  actually means  $p_T^{miss}$ ) (left) and  $H_T$  (right). In the first case,  $H_T > 500$  GeV cut is applied to ensure full efficiency. In the second case,  $p_T^{miss} > 200$  GeV cut is applied. The dashed (solid) blue lines show the distributions for the samples used to calculate the denominator (numerator).

- $HLT\_Mu15\_IsoVVVL\_PFHT350\_v*$ .

The efficiency of the trigger is measured as a function of muon  $p_T$  and  $H_T$  using base sample collected by  $p_T^{miss}$  trigger. As shown in the fig. 5.2, the trigger is more than 95% efficient in the search region.

Di-muon control sample events used for the validation in  $Z(\nu\nu)+jets$  background estimation are selected by a single muon trigger,

- $HLT\_Mu45\_eta2p1\_v*$ .

This trigger has higher efficiency than that of a di-muon trigger to select events with opposite charged leptons after baseline selection.

Trigger efficiencies are taken into account when SM background events are estimated

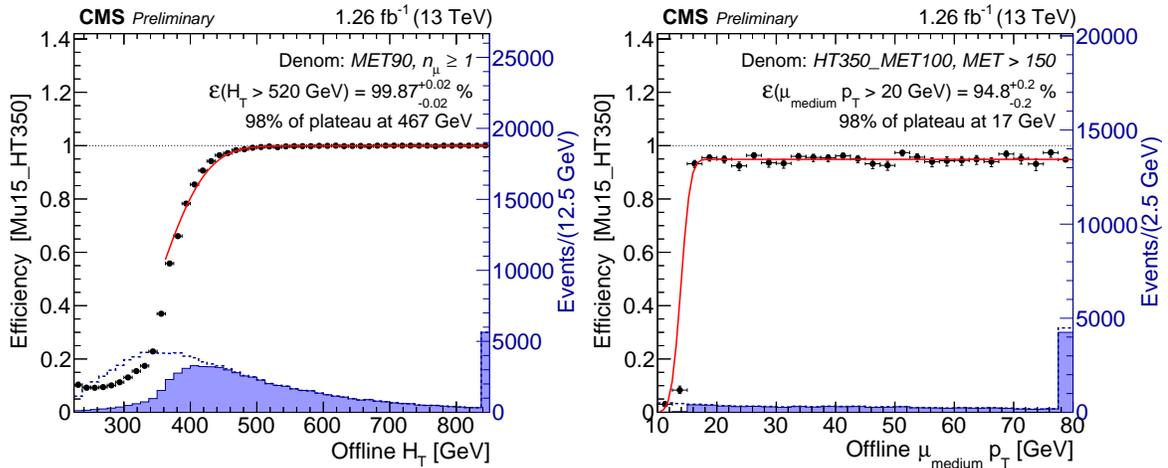


Figure 5.2: Efficiency of the  $\mu + H_T$  trigger as a function of  $H_T$  and  $\mu p_T$ . This is the trigger employed to select the muon + jets control sample for the the hadronically decaying  $\tau$  lepton background estimation.

either from MC simulated events or from a data control sample selected with triggers other than search triggers. The uncertainty on trigger efficiency is propagated to evaluate systematic error for both signal and background yields.

## Top Tagger

The presence of top quarks in every signal model decay chain motivates to reconstruct top quark from the final state jets come from top decay. The number of identified top quarks and their kinematic properties are used to enhance the signal over background sensitivity. The top reconstruction method follows the similar procedure used in 8 TeV analysis [1] with an additional algorithm to tag the top in the boosted scenario when decay products from W boson or top quark are merged into a single jet. The tagging algorithm recognizes three different types of decay topology for the top quark. In order of increasing Lorentz boost for the top quark, these are: (i) resolved scenario: three

distinct jets with no more than one of them identified as a bottom quark jet (b-jet), where two non-b jets represent the decay products of the  $q$  and  $\bar{q}$  coming from the W boson decay; (ii) partially merged: two distinct jets, one of which corresponds to the b quark and the other to the merged  $q\bar{q}$  decay products from the W boson; and (iii) fully merged: a single jet representing the merged decay products of the b quark and W boson. By accounting for these three different topologies, the algorithm achieves high detection efficiency over a wide range of top quark transverse momentum. AK4 jets with  $p_T > 30$  GeV and  $|\eta| < 5$  are clustered into three categories of top candidates: trijet, dijet and monojet to be used as input to tagging algorithm for three scenarios.

- Resolved case: the trijet system must satisfy the following requirements: (i) Each jet lies within a cone of radius 1.5 in  $(\eta, \phi)$  space, centered at the direction defined by the trijet combination. The radius requirement [2] implies a moderate Lorentz boost of the top quark as expected for events with large  $\Delta m$  between stop and neutralino targeted in this search. (ii) The trijet system mass ( $m_{3\text{-jet}}$ ) must be within the range 100-250 GeV. (iii) The trijet system must satisfy one of the three following mass relation conditions [3]:

$$\begin{aligned}
\text{a) } & 0.2 < \arctan\left(\frac{m_{13}}{m_{12}}\right) < 1.3 \quad \text{and} \quad R_{min} < \frac{m_{23}}{m_{3\text{-jet}}} < R_{max}, \\
\text{b) } & R_{min}^2 \left(1 + \left(\frac{m_{13}}{m_{12}}\right)^2\right) < 1 - \left(\frac{m_{23}}{m_{3\text{-jet}}}\right)^2 < R_{max}^2 \left(1 + \left(\frac{m_{13}}{m_{12}}\right)^2\right), \\
\text{c) } & R_{min}^2 \left(1 + \left(\frac{m_{12}}{m_{13}}\right)^2\right) < 1 - \left(\frac{m_{23}}{m_{3\text{-jet}}}\right)^2 < R_{max}^2 \left(1 + \left(\frac{m_{12}}{m_{13}}\right)^2\right).
\end{aligned} \tag{5.1}$$

Here,  $m_{12}$ ,  $m_{13}$ , and  $m_{23}$  are the dijet masses, where the jet indices 1, 2, and 3 reflect a decreasing order in  $p_T$ , while  $m_{3\text{-jet}}$  represents the trijet mass.

The numerical constants have values  $R_{min} = 0.85(m_W/m_t)$  and  $R_{max} = 1.25(m_W/m_t)$ , with  $m_W = 80.4$  GeV and  $m_t = 173.4$  GeV [4]. The dijet and trijet mass distri-

butions taken from 8 TeV stop analysis [1] are shown in Fig. 5.3 upholding the aforementioned conditions.

- Partially merged: the dijet system, one jet from merged W decay must be within the mass window of 70-110 GeV additionally need to pass the mass ratio,

$$R_{min} < \frac{m_{Wjet}}{m_{dijet}} < R_{max} \quad (5.2)$$

- Fully merged: one single jets identified as top jet if the jet mass lie within the range 110-220 GeV.

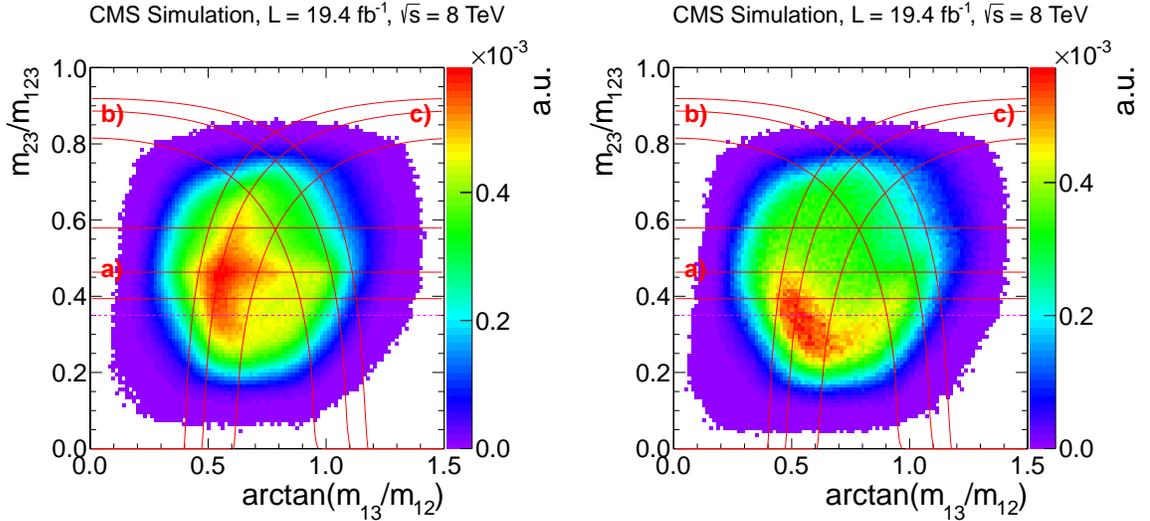


Figure 5.3: The 2D distributions of  $\frac{m_{23}}{m_{123}}$  versus  $\arctan(\frac{m_{13}}{m_{12}})$  for  $t\bar{t}$ (left) and QCD (right). The three horizontal solid red lines which form a band demonstrate the first criteria in Eq. 5.1 with the central line taken as the nominal ratio of  $\frac{m_W}{m_{top}}$  and the other two lines showing the boundaries defined by the  $R_{min}$  and  $R_{max}$ . Similarly the second and third criteria in Eq. 5.1 are in groups of solid lines.

After obtaining all types of top candidates, an extra condition requiring a single candidate must not have more than one b-jet is imposed. Now the final set of the

fully reconstructed tops is formed after removing the jet sharing between the top candidates by favouring the candidates with mass closer to the top quark mass.

With the inclusion of merged topology, the tagger shows good efficiency in wide range of top quark  $p_T$ , from 30% at 200 GeV to 80% at 1 TeV with mistag rate around 40% in all  $p_T^{miss}$  range above 200 GeV. The efficiency and mistag rate are defined by the following equations:

$$\begin{aligned} \text{Efficiency} &= \frac{\text{Number of generator-level top quarks matched to a tagged top}}{\text{Number of generator-level top quarks in an event}}, \\ \text{Fake rate} &= \frac{\text{Number of events having at least one tagged top}}{\text{Total number of events}}, \end{aligned} \tag{5.3}$$

A tagged top is matched with a generator-level top quark within a cone of  $\Delta R < 0.4$  in  $(\eta, \phi)$  space. The tagging efficiency is measured using the T2tt signal sample with  $m_{\tilde{t}} = 850$  GeV and  $m_{\tilde{\chi}_1^0} = 100$  GeV as it has a wide top quark  $p_T$  spectrum. The tagging efficiency was also determined using SM  $t\bar{t}$  background and other signal models, and was found to be consistent with the T2tt measurement within statistical uncertainties. Fig. 5.4 shows the tagging efficiency measured in T2tt(850, 100) as a function of generator-level hadronically decaying top quark  $p_T$ . Mistag rate is checked in  $Z(\nu\nu)+\text{jets}$  events.

The fully reconstructed top quarks are used to define our search region. Using the kinematic properties of tagged tops,  $M_{T2}$  [5, 6] variable is derived.  $M_{T2}$  is an extension of the transverse mass variable in the case of events with pair production of heavy particles, each of which decays to an invisible particle. In order to illustrate  $M_{T2}$  calculation, let us take the process  $pp \rightarrow t\tilde{t}^* \rightarrow t\tilde{t}^* \tilde{\chi}_1^0 \tilde{\chi}_1^0$  as an example. Here we have two simultaneous decays of a particle of unknown mass ( $\tilde{t}$  or  $\tilde{t}^*$ ) into another

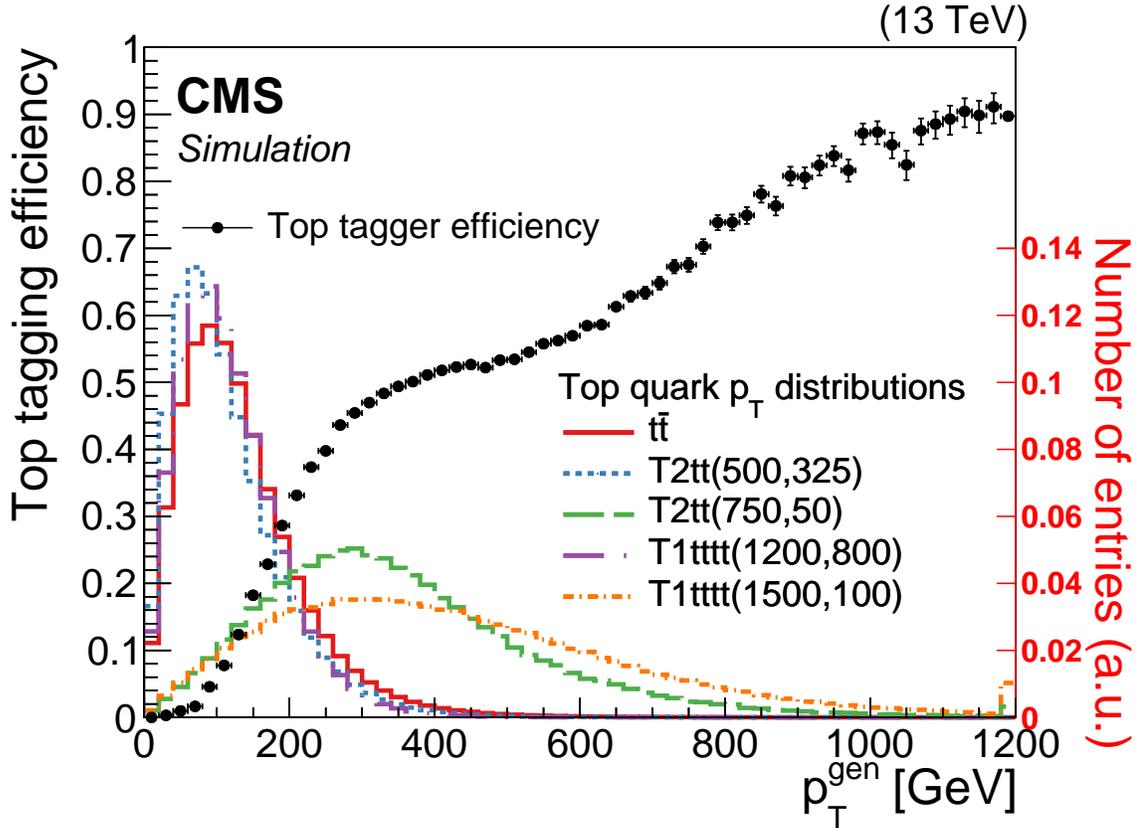


Figure 5.4: The tagging efficiency of the top quark tagger in term of of the generator-level hadronically decaying top quark  $p_T$  (black points). The efficiency was computed using the T2tt (850,100) signal model where  $m_{\tilde{t}} = 850$  GeV and  $m_{\tilde{\chi}_1^0} = 100$  GeV, and it is similar for  $t\bar{t}$  events. The vertical bars depict the statistical uncertainty. The colored lines show the expected hadronically decaying top quark  $p_T$  distribution from  $t\bar{t}$  (red solid line), the T2tt(500, 100) (blue short-dashed line) and T2tt(750, 50)(green long-dashed line) signal models , the T1tttt (1200, 800) signal model with  $m_{\tilde{g}} = 1200$  GeV and  $m_{\tilde{\chi}_1^0} = 800$  GeV (purple long-dash-dotted line), and the T1tttt(1500, 100) signal model (orange short-dash-dotted line). The last bin contains the overflow entries and the top quark  $p_T$  distributions are normalized to unit area.

visible particle ( $t$  or  $t^*$ ) and invisible particle ( $\tilde{\chi}_1^0$ ). Now  $p_T^{miss}$  in this events is the sum of  $\vec{p}_T$  of two  $\tilde{\chi}_1^0$ . The variable  $M_{T2}$  is defined as:

$$M_{T2} \equiv \min_{\vec{q}_T^{(1)} + \vec{q}_T^{(2)} = \vec{p}_T} [\max\{m_T^2(\vec{p}_T^{(1)}; m_{p^1}, \vec{q}_T^{(1)}; m_{\chi_1^0}), m_T^2(\vec{p}_T^{(2)}; m_{p^2}, \vec{q}_T^{(2)}; m_{\chi_1^0})\}] \quad (5.4)$$

where  $\vec{p}_T^{(i)}$  and  $m_{p^i}$  are the transverse momentum and mass of visible daughter while  $\vec{q}_T^{(i)}$  and  $m_{\chi_1^0}$  are the same of invisible one. The  $m_T^2$  is transverse mass squared which is defined as

$$m_T^2(\vec{p}_T^{t(1)}, \vec{q}_T^{(1)}; m_{\chi_1^0}) \equiv m_{t(1)}^2 + m_{\chi_1^0}^2 + 2(E_T^{t(1)} E_T^{(1)} - \vec{p}_T^{t(1)} \cdot \vec{q}_T^{(1)}) \quad (5.5)$$

$M_{T2}$  is a minimization of two transverse masses with a constraint that the sum of the transverse momenta of both  $\chi_1^0$ 's is equal to the missing transverse momentum of the event. The mass of an invisible particle is assumed to be zero to make it consistent with the use of neutrino as an invisible particle while calculating  $M_{T2}$  in SM backgrounds. For the direct pair production of  $\tilde{t}$ ,  $M_{T2}$  has the kinematic upper limit at  $\tilde{t}$  mass. For gluino pair production  $M_{T2}$  depends on decay topology. However for all signals,  $M_{T2}$  tends to have higher values than  $t\bar{t}$  and other SM backgrounds due to high  $p_T^{miss}$  and high  $p_T$  top quarks produced in signal decay.

In the case of two fully reconstructed top objects in an event,  $M_{T2}$  is calculated using the pair of tagged tops and  $p_T^{miss}$ . If there are more than two top objects present in an event,  $M_{T2}$  is calculated for all two tops combinations and the smallest one is chosen. If only one top is identified, then another top candidate is partially reconstructed with the remaining part of the event using b-jet as a seed. The fully reconstructed top and the partial one are used as the visible part in  $M_{T2}$  measurement. The partial top is formed by merging a b-jet with a nearby jet requiring the dijet invariant mass

satisfy the mass window of 50 GeV to  $m_t$ . If no dijet combination satisfies the mass condition, b-jet is treated as the remaining system to form the partial top object.

## Baseline Selection

The search is performed on a sample of multi jets events with b-jets coming from top quarks, large  $p_T^{miss}$  and no leptons. The following cuts along with the filters define the baseline selection:

- Pass all filters that remove detector and beam related noise:

- Muon veto:

Events having a muon with  $p_T > 10$  GeV and  $|\eta| < 2.4$  are rejected. The muon candidates are ‘Medium Muon’ selected with PF mini isolation criteria ( 3.2.2).

- Electron veto:

Events with an electron having  $p_T > 10$  GeV and  $|\eta| < 2.5$  are vetoed. The electron candidates are selected with ‘Cut Based VETO’ algorithm using PF mini isolation ( 3.2.2).

- Isolated track veto:

The lepton ( $\mu$  and e) veto reduces the background contribution from  $t\bar{t}$ , W+jets and single top events with  $W \rightarrow l\nu$  where l can be  $\mu$ , e or leptonically decaying  $\tau$ . To further suppress these SM backgrounds, a new veto technique is employed. Events with at least one isolated charged track are rejected. This method is very effective to throw the tauonic ( $W \rightarrow \tau$ ) events where  $\tau$  subsequently decays into charged hadrons (sec. 5.6.1). However events with muon and electron track are also rejected by isotrack veto. The track isolation is calculated from charged PF candidates consistent with the reconstructed primary vertex ( $|dz(PV)| < 0.1$

cm). The requirements are different for muon, electron and charged hadron tracks. For both electron and muon tracks, the isolated track requirements are:  $p_T > 5$  GeV,  $|\eta| < 2.5$  and relative isolation less than 0.2. For charged hadron tracks, the  $p_T$  requirement is raised to be at least 10 GeV and the relative isolation value to be less than 0.1. Because of better reconstruction efficiencies of muon and electron in low  $p_T$ , smaller  $p_T$  threshold is chosen for them compared to the charged hadrons. To avoid signal event reduction, events with one isolated track, as defined above, are only rejected if they satisfy

$$m_T(tk, p_T^{miss}) = \sqrt{2p_T^{tk}p_T^{miss}(1 - \cos \Delta\phi)} < 100 \text{ GeV} \quad (5.6)$$

where  $p_T^{tk}$  is the transverse momentum of the track and  $\Delta\phi$  is the azimuthal separation between the track and  $p_T^{miss}$  vector. The above criteria is based on the fact that the upper limit of  $m_T$  for the charged track comes from W decay will be W mass but in case of signal events,  $m_T$  tends to have higher values due to larger  $p_T^{miss}$  carried by LSP. The relevant distribution is given in Fig. 5.8.

- $N_{jets} \geq 4$ :

Stop signal contains six jets in its final state from the hadronic decay of each stop in case of T2tt and T2tb topology. T1tttt has more jet multiplicity. Not all the jets pass the jet selection and sometimes two or more jets are merged to count as single jet. Therefore at least four jets requirement is chosen. Each AK4 jet is required to have  $p_T > 30$  GeV and  $|\eta| < 2.4$  and additionally passes loose jet id. filter as mentioned in above list. The leading two jets are required to have  $p_T > 50$  GeV since SUSY predicts centrally produced jets with high  $p_T$ .

- $p_T^{miss} \geq 200$  GeV:

The cut threshold is decided by trigger efficiency ( 5.1).

- $H_T \geq 500$  GeV:

Jets in the  $H_T$  calculation must meet the same jet selection criteria defined above. The cut is constrained by trigger efficiency ( 5.1).

- $N_b \geq 1$ :

B-jets are identified using CSV algorithm, medium working point ( 3.2.2).

- Angular cut: QCD multijet process should have zero  $p_T^{miss}$  or negligible  $p_T^{miss}$  when quark decays leptonically. But the multijet events can pass high  $p_T^{miss}$  cut if transverse momenta of jets are mis-reconstructed resulting in high  $p_T^{miss}$ . However the direction of  $p_T^{miss}$  vector is found to be close to one of the leading jets. Therefore A cut on the angle between  $p_T^{miss}$  and the first three leading jets,  $\Delta\phi(p_T^{miss}, j_{1,2,3}) > 0.5, 0.5, 0.3$ , is applied to remove events coming from QCD processes.

After these pre-selections, events are required to pass following two additional cuts based on top reconstruction algorithm.

- $N_t \geq 1$ ,

- $M_{T2} \geq 200$  GeV.

The threshold on  $M_{T2}$  is chosen to suppress  $t\bar{t}$  events.

The baseline selection preserves 2-20% of the signal events as shown in table 5.1.

## Search Regions

The events passing the baseline selection are classified into search regions defined in terms of  $N_t$ ,  $N_b$ ,  $p_T^{miss}$  and  $M_{T2}$ . Fig. 5.5 shows the comparison between total SM backgrounds from simulation and several signal points for the four search bin variables

Table 5.1: Sample selection cut flow and event yields from MC simulated samples. All numbers are normalized to  $2.2fb^{-1}$ . Only statistical uncertainties are shown. Rare process refer to ttW, di-boson and tri-boson events.

Samples	$t\bar{t}$	$W \rightarrow l\nu$	single top	$Z \rightarrow \nu\bar{\nu}$	QCD	ttZ	rate	sum SM	T2tt	T2tt	T2tt	T1tttt	T1tttt
Cuts	$9.77 \times 10^5$ (100%)	$4.62 \times 10^6$ (100%)	$1.53 \times 10^5$ (100%)	$9.88 \times 10^5$ (100%)	$6.39 \times 10^{10}$ (100%)	$1.69 \times 10^4$ (100%)	$3.87 \times 10^7$ (100%)	$6.39 \times 10^{10}$ (100%)	$1.12 \times 10^3$ (100%)	$1.12 \times 10^3$ (100%)	$41$ (100%)	$(1200, 800)$	$(1500, 100)$
No cut	$9.64 \times 10^5$ (99%)	$4.54 \times 10^6$ (98%)	$1.52 \times 10^5$ (99%)	$9.81 \times 10^5$ (99%)	$6.34 \times 10^{10}$ (99%)	$1.66 \times 10^4$ (99%)	$3.81 \times 10^7$ (99%)	$6.34 \times 10^{10}$ (99%)	$1.11 \times 10^3$ (99%)	$1.11 \times 10^3$ (99%)	$41$ (100%)	$184$ (100%)	$31$ (100%)
event filter	$6.18 \times 10^5$ (64%)	$3.39 \times 10^6$ (75%)	$1.21 \times 10^5$ (80%)	$9.80 \times 10^5$ (100%)	$6.34 \times 10^{10}$ (100%)	$1.26 \times 10^4$ (76%)	$3.27 \times 10^7$ (86%)	$6.34 \times 10^{10}$ (100%)	$8.75 \times 10^2$ (79%)	$8.75 \times 10^2$ (79%)	$32$ (79%)	$183$ (99%)	$30$ (99%)
$\mu$ veto	$3.46 \times 10^5$ (56%)	$2.35 \times 10^6$ (69%)	$9.69 \times 10^4$ (80%)	$9.77 \times 10^5$ (100%)	$6.33 \times 10^{10}$ (100%)	$9.61 \times 10^3$ (76%)	$2.82 \times 10^7$ (86%)	$6.33 \times 10^{10}$ (100%)	$6.92 \times 10^2$ (79%)	$6.92 \times 10^2$ (79%)	$25$ (78%)	$76$ (65%)	$12$ (64%)
e veto	$2.20 \times 10^5$ (63%)	$1.81 \times 10^6$ (77%)	$7.65 \times 10^4$ (79%)	$9.34 \times 10^5$ (96%)	$6.06 \times 10^{10}$ (96%)	$7.30 \times 10^3$ (76%)	$2.41 \times 10^7$ (85%)	$6.06 \times 10^{10}$ (96%)	$5.52 \times 10^2$ (80%)	$5.52 \times 10^2$ (80%)	$23$ (93%)	$57$ (75%)	$11$ (88%)
iso1Tk veto	$1.04 \times 10^5$ (48%)	$8.21 \times 10^4$ (5%)	$3.61 \times 10^4$ (47%)	$3.38 \times 10^4$ (4%)	$5.65 \times 10^8$ (1%)	$6.43 \times 10^2$ (88%)	$1.96 \times 10^4$ (8%)	$5.65 \times 10^8$ (1%)	$4.50 \times 10^2$ (81%)	$4.50 \times 10^2$ (81%)	$21$ (90%)	$57$ (100%)	$11$ (100%)
$N_{\text{res}} \geq 4$	$8.49 \times 10^4$ (81%)	$1.23 \times 10^4$ (15%)	$2.52 \times 10^4$ (70%)	$5.72 \times 10^4$ (17%)	$8.49 \times 10^7$ (15%)	$5.21 \times 10^2$ (81%)	$4.52 \times 10^7$ (23%)	$8.50 \times 10^7$ (15%)	$3.65 \times 10^2$ (81%)	$3.65 \times 10^2$ (81%)	$17$ (83%)	$54$ (95%)	$10$ (96%)
$N_b \geq 1$	$1.63 \times 10^4$ (19%)	$2.99 \times 10^3$ (24%)	$5.53 \times 10^3$ (22%)	$1.50 \times 10^3$ (26%)	$6.46 \times 10^6$ (8%)	$2.74 \times 10^2$ (63%)	$9.59 \times 10^2$ (21%)	$6.49 \times 10^6$ (8%)	$1.27 \times 10^2$ (35%)	$1.27 \times 10^2$ (35%)	$16$ (94%)	$47$ (88%)	$10$ (100%)
$H_T \geq 500$ GeV	$1727 \pm 3$ (11%)	$333 \pm 3$ (11%)	$111 \pm 3$ (2%)	$395 \pm 3$ (26%)	$6773 \pm 1053$ (0%)	$21.7 \pm 0.4$ (8%)	$27 \pm 2$ (3%)	$9386 \pm 1053$ (0%)	$52.9 \pm 0.4$ (42%)	$52.9 \pm 0.4$ (42%)	$14.4 \pm 0.1$ (88%)	$21.2 \pm 0.2$ (45%)	$9.3 \pm 0.1$ (91%)
$p_T^{\text{mass}} \geq 200$ GeV	$988 \pm 3$ (57%)	$195 \pm 3$ (59%)	$53 \pm 2$ (48%)	$283 \pm 2$ (72%)	$1051 \pm 729$ (16%)	$16.6 \pm 0.4$ (76%)	$16 \pm 1$ (60%)	$2604 \pm 729$ (28%)	$32.9 \pm 0.3$ (62%)	$32.9 \pm 0.3$ (62%)	$13.0 \pm 0.1$ (91%)	$16.9 \pm 0.1$ (80%)	$7.57 \pm 0.05$ (82%)
tagger + $N_T \geq 1$	$451 \pm 2$ (46%)	$33 \pm 1$ (17%)	$17 \pm 1$ (32%)	$48 \pm 1$ (17%)	$52 \pm 6$ (5%)	$9.0 \pm 0.3$ (54%)	$4.5 \pm 0.6$ (28%)	$615 \pm 7$ (24%)	$18.6 \pm 0.2$ (57%)	$18.6 \pm 0.2$ (57%)	$8.56 \pm 0.04$ (66%)	$15.4 \pm 0.1$ (91%)	$6.89 \pm 0.05$ (91%)
$M_{T2} \geq 200$ GeV	$289 \pm 1$ (64%)	$28 \pm 1$ (83%)	$13 \pm 1$ (77%)	$42 \pm 1$ (87%)	$39 \pm 5$ (75%)	$7.9 \pm 0.3$ (89%)	$3.5 \pm 0.5$ (76%)	$422 \pm 6$ (69%)	$14.9 \pm 0.2$ (80%)	$14.9 \pm 0.2$ (80%)	$8.39 \pm 0.04$ (98%)	$14.5 \pm 0.1$ (94%)	$6.77 \pm 0.04$ (98%)

after baseline cuts. All the variables show good discrimination power. Observed data are also shown and total SM backgrounds as well signals as are scaled to the data yield to ensure a shape comparison. Events in the search region are first distributed

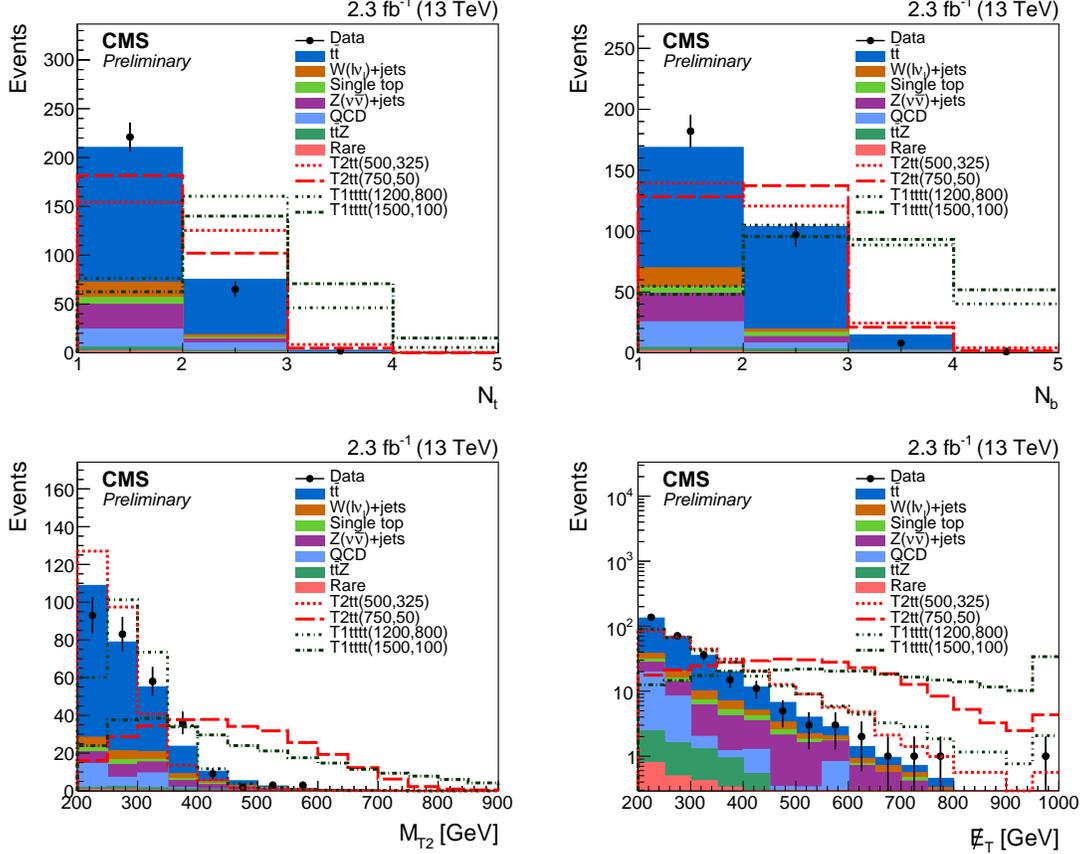


Figure 5.5: Comparison of the distributions between simulated SM backgrounds and several signal models for  $N_t$ ,  $N_b$ ,  $E_T^{miss}$  and  $M_{T2}$  (clock-wise), after the baseline selection applied. Observed data events are shown by black points. Total SM backgrounds and signal models distributions have been normalized to the same area as the data distribution. Here  $E_T^{miss}$  is synonymous with  $p_T^{miss}$ .

into the bins of  $N_t$  and  $N_b$ . Now each  $N_t, N_b$  bin is further divided into the bins of  $p_T^{miss}$  and  $M_{T2}$ . Two sets of search bin optimizations are used in order to get better sensitivity for direct stop production models (T2tt, T2tb) and gluino mediated production model (T1tttt).  $p_T^{miss}$ ,  $M_{T2}$  bin edges are regularized and adjusted to have

reasonable number of control sample events for major SM background predictions. Fig. 5.6 illustrate the bin numbering for 37 bins used to interpret T2tt and T2tb models. Top tagged and b tagged jet objects are binned as  $N_t = 1, N_t \geq 2$  and  $N_b = 1, N_b \geq 2$ . Due to the higher multiplicity of b-jets and top objects produced in gluino mediated stop decay, bins are designed with  $N_{t \text{ or } b} = 1, 2$  and  $\geq 3$  conditions. Fig. 5.7 shows the bin indices for 45 bins used to interpret T1tttt model. All the SM backgrounds are estimated in both sets of search bins.

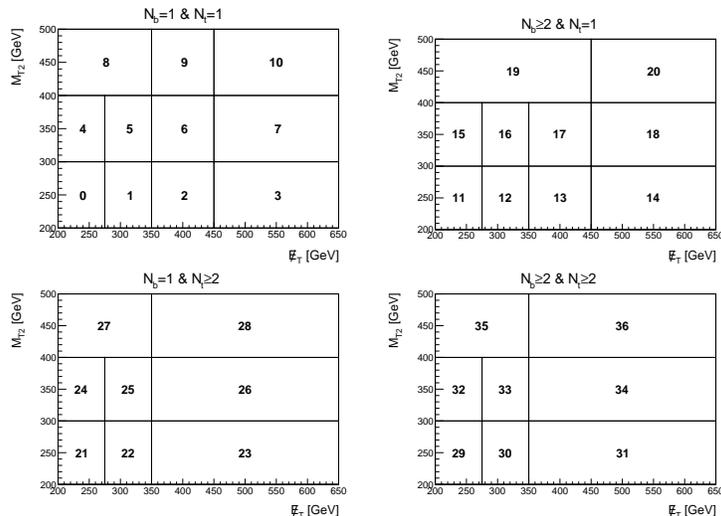


Figure 5.6: Search bin definitions and bin numbers for 37 SBs after baseline selection. The number indicates bin indices.

## Signal and Background MC Samples

Monte Carlo samples for SM processes are used for background estimation method development and prediction as well. SUSY signal samples are used to check the distribution of analysis variables and most importantly to interpret the result in term of simplified models. All SM background processes except single top are generated using the *MADGRAPH 5* [30] program. Single top quark events produced in the tW chan-

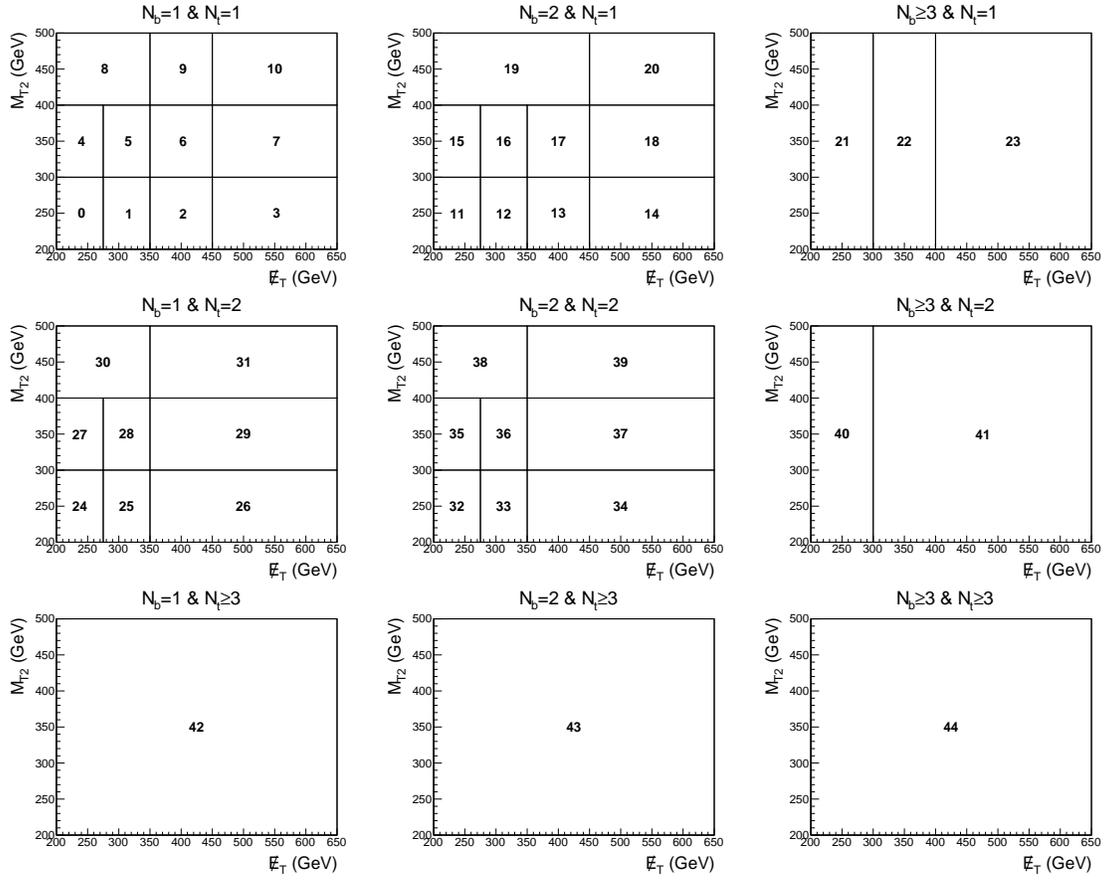


Figure 5.7: Search bin definitions and bin numbers for 45 SBs after baseline selection. The number indicates bin indices.

nel with *POWHEG* [31, 7] generator. The generation of these processes is based on parton distribution function (PDFs) modeled with *NNPDF3.0* [8]. The parton showering and hadronization are simulated with *PYTHIA v8* [32] using underlying-event tune *CUETP8M1* [9]. The detector response is simulated using Geant4-based [33] CMS simulation application and the events are reconstructed with CMSSW release 7.4 (Spring15) using PU20bx25 pileup scenario, which simulates a pileup distribution with an average of 20 interactions per bunch crossing and a 25 ns interval between bunches. In CMS, the simulated pile up describes the observed one quite well. However the remaining differences are compensated by correcting the simulated events with pile up weight. The complete list of these samples are given in table 5.2 where the mentioned cross sections are calculated at the next-to-next-to-leading-order (NNLO) unless otherwise noted.

All the signal production cross section calculated using NLO plus next-to-leading-logarithm (NLL) calculations [10]. The detector responses for the signal samples are simulated using Fast Simulation application [11]. The signal simplified models used in our analysis are shown in figure 4.1 and 4.2. Direct top squark-antisquark pair production is modeled in  $T2tt(x,y)$  with  $x$  and  $y$  be the top squark and neutralino ( $\tilde{\chi}_1^0$ ) masses respectively. If the gluino is taken to be within the LHC Run 2 reach, gluino-induced processes also become relevant to the analysis. We, in particular, consider  $T1tttt(x, y)$  with  $x$  and  $y$  be the gluino and  $\tilde{\chi}_1^0$  masses. In this model, each gluino first decays to an off-shell top squark and an on-shell top quark, then off-shell top squark decays to a top and  $\tilde{\chi}_1^0$  resulting in three-body decay of gluino into  $t\bar{t}$  and  $\tilde{\chi}_1^0$ . Cross-sections for a couple of mass points are shown in table 5.3 for the  $T2tt$  and  $T1tttt$  signal models.

Table 5.2: Standard model Monte Carlo samples used in the analysis.

Dataset	$\sigma$ (pb)	Luminosity( $\text{fb}^{-1}$ )
QCD MC samples (LO)		
QCD_HT200to300	1735000	0.01
QCD_HT300to500	366800	0.05
QCD_HT500to700	29370	0.67
QCD_HT700to1000	6524	2.30
QCD_HT1000to1500	1064	4.67
QCD_HT1500to2000	121.5	31.67
QCD_HT2000toInf	25.42	77.17
SM $t\bar{t}$ MC samples		
TTJets	816.0	13.90
TTJets_SingleLeptFromT	179.3	324.6
TTJets_SingleLeptFromTbar	179.3	335.7
TTJets_DiLept	86.66	351.2
TTJets_HT-600to800	2.615	1898
TTJets_HT-800to1200	1.077	3198
TTJets_HT-1200to2500	0.195	5063
TTJets_HT-2500toInf	0.002	218575
SM $W \rightarrow \nu l$ MC samples		
WJetsToLNu_HT-100To200	1635	6.20
WJetsToLNu_HT-200To400	437.0	11.97
WJetsToLNu_HT-400To600	59.50	31.96
WJetsToLNu_HT-600ToInf	22.80	45.44
WJetsToLNu_HT-600To800	15.50	257.1
WJetsToLNu_HT-800To1200	6.366	247.4
WJetsToLNu_HT-1200To2500	1.614	158.4
WJetsToLNu_HT-2500ToInf	0.037	6770
SM $Z \rightarrow \nu\bar{\nu}$ MC samples		
ZJetsToNuNu_HT-100To200	345.0	14.92
ZJetsToNuNu_HT-200To400	96.38	52.22
ZJetsToNuNu_HT-400To600	13.46	75.34
ZJetsToNuNu_HT-600ToInf	5.170	196.5
SM $Z \rightarrow l^+l^-$ MC samples		
DYJetsToLL_M-50	6025	1.50
DYJetsToLL_M-50_HT-100to200	171.5	15.31
DYJetsToLL_M-50_HT-200to400	52.58	18.18
DYJetsToLL_M-50_HT-400to600	6.761	155.0
DYJetsToLL_M-50_HT-600toInf	2.718	363.5
SM single-top MC samples		
ST_tW_antitop	35.80 (NLO)	27.93
ST_tW_top	35.80 (NLO)	27.81
SM diboson and other rare process MC samples		
ttHJetTobb	0.293	18269
TTZToLLNuNu	0.228	811.4
TTZToQQ	0.530	663.4
TTWJetsToLNu	0.204	635.6
TTWJetsToQQ	0.423	1018
ZH_HToBB_ZToNuNu	0.100	12116
WH_HToBB_WToLNu	0.260	4782
WWTo1L1Nu2Q	50.00	64.26
WWTo2L2Nu	12.18	158.5
WZTo1L1Nu2Q	10.71	1339
WZTo1L3Nu	3.058	305.7
ZZTo2Q2Nu	4.040	5556
ZZTo2L2Q	3.220	3706
WWZ	0.165	1341
WZZ	0.056	3938
ZZZ	0.014	15297

Table 5.3: Cross sections for a couple of mass points for the T2tt and T1tttt Models.

Dataset	$\sigma$ (pb)	Luminosity ( $\text{fb}^{-1}$ )
SMS-T2tt_mStop-500_mLSP-325	0.51848	748
SMS-T2tt_mStop-850_mLSP-100	0.01896	12694
SMS-T1tttt_mGluino-1500_mLSP-100	0.014	7268
SMS-T1tttt_mGluino-1200_mLSP-800	0.086	1719

## Background Estimation

The contribution from SM processes in the search region is reduced to a large extent by applying baseline selection. However, the presence of SM background events in the search bins cannot be fully abated. Search variables distributions for simulated backgrounds are shown in figure 5.5. The major background contribution comes from  $t\bar{t}$ ,  $W$ +jets, and single top events because of two effects. The first one happens when  $W$  boson decays to a tau lepton and tau then decays into hadrons which are reconstructed as jets and in the second case, muon and electron come from  $W$  decay are lost due to inefficiency of isolation and reconstruction or out of detector acceptance. Events pass the lepton veto in both cases. The first scenario is called hadronic tau ( $\tau_h$ ) and the second one is known as lost lepton background.  $Z$ +jets events enter our search region when  $Z$  decays into pair neutrinos. There is also background contribution from QCD multi jets events and small fraction comes from  $t\bar{t}Z$  and other rare processes as listed in table 5.2. All the background estimation methods are presented in following sections with a detail description of hadronic tau background.

### Hadronic Tau Background

The hadronic decay of  $\tau$  leptons ( $\tau_h$ ) is one of the largest components of the background from  $t\bar{t}$ ,  $W$ +jets, and single-top events contributing to the search regions. As

described in section 5.3 this analysis incorporates a veto on isolated tracks to reduce the hadronic  $\tau$  background while sustaining a minimal impact on signal efficiency. After applying the veto, the remnant hadronic  $\tau$  events are estimated using the method described in the succeeding sections.

### **Tau response function method:**

The procedure adapts the basic method described in Ref. [1]. When a W boson decays to a neutrino and a hadronically decaying  $\tau$  lepton ( $\tau_h$ ), the presence of neutrinos in the final state results in  $p_T^{miss}$ , and the event passes the lepton veto because the hadronically decaying  $\tau$  is reconstructed as a jet. This background is estimated from a control sample (CS) of  $\mu$  + jets events selected from data using a  $\mu$  +  $H_T$ -based trigger, *HLT\_Mu15\_IsoVVVL\_PFHT350\_v*, and requiring exactly one  $\mu$  with  $p_T^\mu > 20\text{GeV}$  and  $|\eta| < 2.4$ . A cut on the transverse mass of the W,  $m_T = \sqrt{2p_T^\mu p_T^{miss}(1 - \cos \Delta\phi)} < 100\text{ GeV}$ , is required in order to select events containing a  $W \rightarrow \mu\nu$  decay and to suppress possible new physics signal contamination, i.e., signal events present in the  $\mu$  + jets sample. Here,  $\Delta\phi$  is the azimuthal angle between the  $\vec{p}_T^\mu$  and the  $\vec{p}_T^{miss}$  directions. For the electroweak control sample  $p_T^{miss}$  originates from the neutrino,  $m_T$  represents the transverse W-mass, and therefore the distribution falls sharply above 80 GeV. The  $m_T$  distribution in  $t\bar{t}$  events and for a T2tt signal is shown in Fig. 5.8. Because the  $\mu$  + jets and  $\tau_h$  + jets events arise from the same physics processes, the hadronic component of the two samples is the same except for the response of the detector to the muon or the  $\tau_h$  jet. The trick consists of replacing the muon  $p_T$  by a random sample of a simulated  $\tau_h$  jet response "template" function for a hadronically-decaying  $\tau$  lepton. The global variables of the event are recalculated with this  $\tau_h$  jet, and the search selections are applied to predict the  $\tau_h$  background.

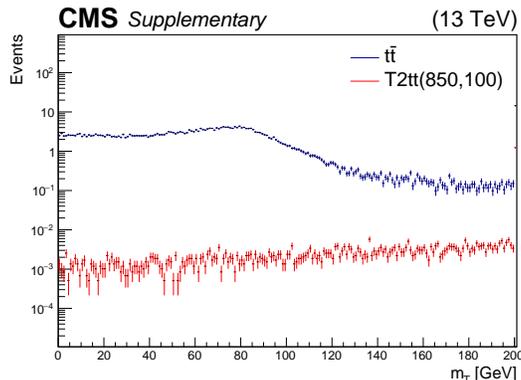


Figure 5.8: The  $m_T$  distribution in  $t\bar{t}$  (black) and for the T2tt signal point  $m_{\bar{t}}=850$  GeV,  $m_{LSP}=100$  GeV (red). Both curves are scaled to  $2.3 \text{ fb}^{-1}$

### The $\tau$ jet response template:

Tau jets are characterized by a low particle multiplicity as compared to jets from gluons or quarks. Typically, they consist of one or three mesons ( $\pi^\pm$ ), up to two neutral mesons ( $\pi^0 \rightarrow \gamma\gamma$ ), and a  $\nu_\tau$ . The tau response i.e., the fraction of visible energy ( $f_{VE}$ ), can be described by a MC template. A reconstructed jet is matched in  $\eta-\phi$  space ( $\Delta R(\text{jet}, \tau) < 0.1, 0.15, 0.2, 0.25$  for tau  $p_T > 100$  GeV,  $50 < p_T < 100$  GeV,  $30 < p_T < 50$  GeV and  $20 < p_T < 30$  GeV) to the "visible part" of a generated tau-lepton ( $p_T > 20 \text{ GeV}$  and  $|\eta| < 2.4$ ). This "visible part" contains all the daughter particles of the tau-lepton excluding the tau neutrino. We always choose the jet closest to the "visible part" of a generated tau-lepton for the template measurement. The larger matching  $\Delta R$  for lower  $p_T$  ranges is used to make sure more than 95% of the events are picked up. This is shown in Fig. 5.9. The jets are not used for the template measurement if they are within  $\Delta R < 0.4$  of one of the generator level b-quarks coming from the top decay. For matched tau-jet pairs, the fraction of visible energy, defined as the ratio of the reconstructed JES corrected tau-jet  $p_T$  to the generated tau-lepton  $p_T$ , is computed. The tau template is obtained from the  $t\bar{t}$ , W+jets, and single-top Spring15 MC samples mixed with the proper cross-section

scalings.

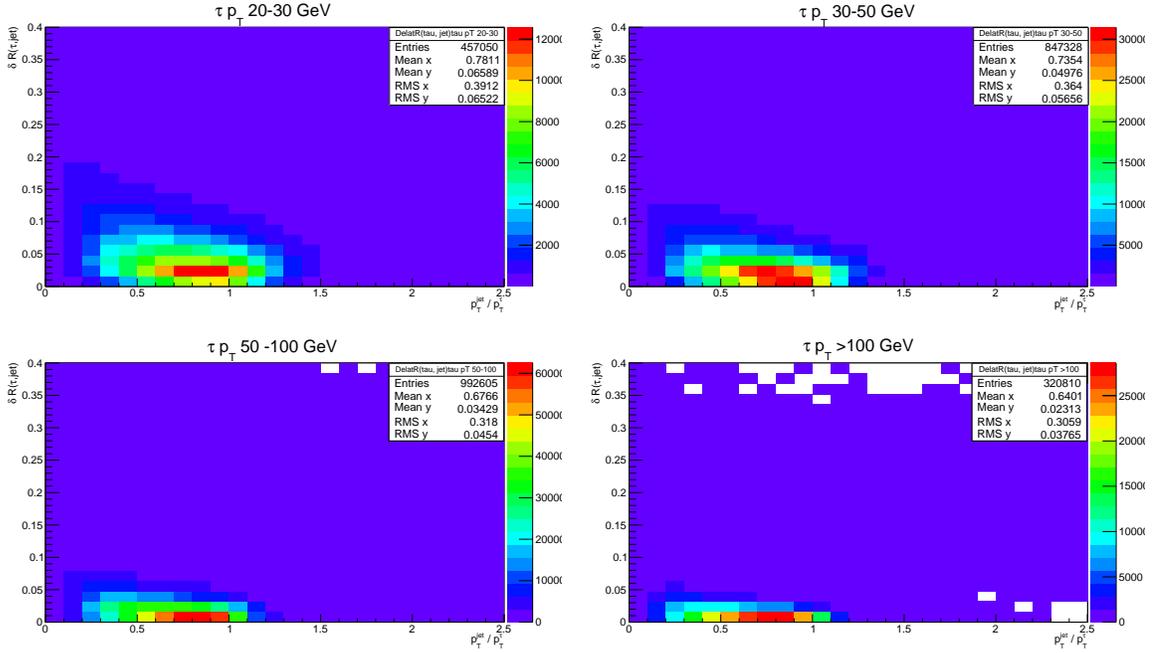


Figure 5.9: Closest  $\Delta R$ 's between tau lepton "visible part" and jet as a function of  $f_{VE}$  for different  $p_T$  ranges of the generated  $\tau$  lepton.

As shown in Fig. 5.10, the template is measured in four  $p_T$  bins to account for the  $p_T$  dependence of  $\tau$  jet response.

### Misidentification of $\tau_h$ jets as b-jets:

The probability to mistag a  $\tau_h$  jet as a b-jet is significantly high. This mistag rate must be taken into account in order to accurately predict the  $N_b$  distribution of  $\tau_h$  background events, and correctly assign  $\tau_h$  background events to search bins depending on the b-jet multiplicity. The  $\tau_h$  to b-jet misidentification rate (b-mistag rate) as a function of  $\tau_h$  jet  $p_T$  is shown for a sample of simulated  $t\bar{t}$  and W+jets events in Fig. 5.11. The dependence of b-mistag rate with  $\tau_h$  jet  $p_T$  is larger for  $t\bar{t}$  events than for W+jets events. This is because of the overlap of the  $\tau_h$  jet with the nearby b-quark from the same top quark decay in case of the  $t\bar{t}$  sample. This overlap fraction

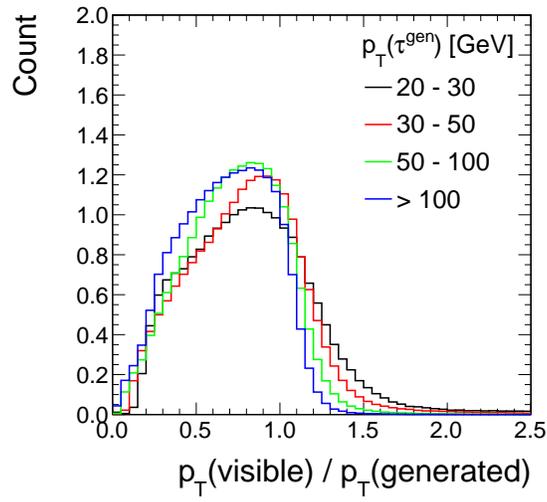


Figure 5.10: Tau jet visible energy fraction templates in tau lepton  $p_T$  bins.

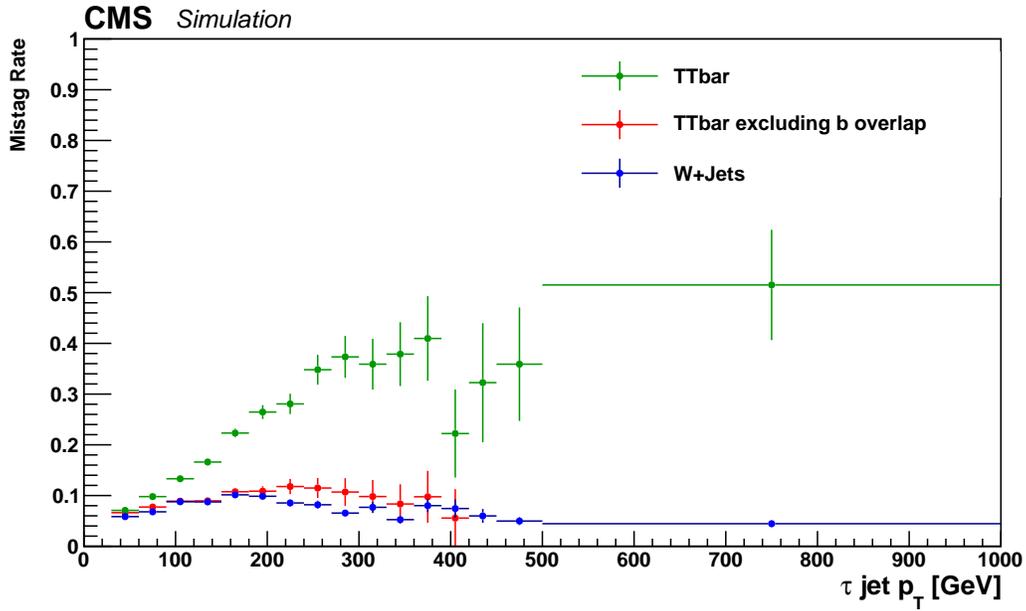


Figure 5.11: The rate for tau-jets to be misidentified as a b jet as a function of the tau jet  $p_T$ .

increases with the increase in  $\tau_h$  jet  $p_T$ . If the events with overlap are excluded from the calculation, the b-mistag rate from  $t\bar{t}$  agrees quite well with that from W+jets. In order to take into account this mistag rate in the  $\mu$ +jets control sample, we randomly select a simulated tau-jet and count it as a b-jet with the probability obtained from the simulated mistag rate W+jets for the corresponding  $\tau_h$  jet  $p_T$  bin.

### Isolated track veto:

The veto on isolated tracks helps to reject hadronically decaying  $\tau$  leptons, mostly one prong taus. However, it also vetoes the events containing isolated muons or electrons. Thus the veto cannot be directly applied to the  $\mu$ +jets control sample as part of the search sample selection. Then the isolated track tagging efficiency ( $f_{\text{isotrack}}$ ) and consequently the isolated track veto efficiency for  $\tau_h$  ( $\epsilon_{\text{isotrack}} = 1 - f_{\text{isotrack}}$ , i.e. the fraction of events that survive the veto) is measured and the  $\mu$ +jets control sample yield is multiplied by  $\epsilon_{\text{isotrack}}$  to get a prediction with the isolated track veto applied. This efficiency is determined from simulated  $t\bar{t}$ , W+jets, and single-top events by matching isolated tracks to a  $\tau_h$  jets and computing the ratio of the number of tracks passing the isolation criteria over the total. The average isolated track tagging efficiency for the  $\tau_h$  control sample after the baseline cuts is shown in Tab. 5.4. Isotrack tagging efficiency for  $\tau_h$  is measured as a function of  $N_{\text{jets}}$  and  $N_b$ . This two

Table 5.4: Isolated track veto effect on one prong, three prong, and total  $\tau_h$  event.

Sample category	Isolated track tagging efficiency ( $f_{\text{isotrack}}$ )
One prong $\tau_h$	40%
Three prong $\tau_h$	2.5%
All $\tau_h$	33%

dimensional efficiency is shown in Fig. 5.12

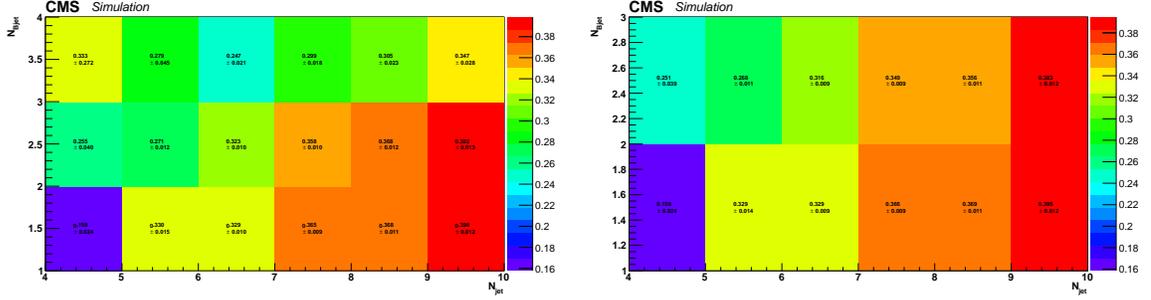


Figure 5.12: Isolated track tagging efficiency of  $\tau_h$  as a function of  $N_{jets}$  and  $N_b$ . Left plot has three  $N_b$  bins and right plot has two  $N_b$  bins where last two bins of left plot are merged. The tagging efficiency varies from 0.15 to 0.4 with the bin configuration. The average value is given in Tab. 5.4.

### Muon control sample selection:

The muon control sample is selected from collider data applying the following criteria.

- A  $\mu + H_T$  cross trigger, HLT\_Mu15\_IsoVVVL\_PFHT350\_v is used.
- Events are required to pass the same cleaning requirements described in section 5.3.
- Events are required to have one and only one isolated muon with  $p_t > 20 GeV$  and  $|\eta| < 2.4$ . The  $p_t$  cut ensures that the selected sample populates the trigger plateau while the  $abs_\eta$  requirement is the same as the one for the muon leg of the trigger. The offline muon identification follows the "medium muon" recommendation as mentioned in section 5.3.
- Events with electrons and/or additional muons are rejected according to the prescription in section 5.3.
- The transverse mass  $W m_T = \sqrt{2p_t^\mu p_T^{miss}(1 - \cos \Delta\phi)} < 100 GeV$  is enforced to remove most of the potential signal contamination.

The muon  $p_t(\mu)$  in each event passing the abovementioned selection is smeared with the  $\tau$  response  $f_{VE}$  taken from a template of the  $\tau$   $p_t$  visible energy. PF jets used in this analysis include any PF muon and their momenta are corrected by the standard jet energy corrections; therefore we subtract JES-corrected muon momenta from the PF muon-jet momenta and add the  $\tau_h$  response to emulate a  $\tau_h$  jet. A tau jet is counted as such if it satisfies the  $p_t > 30$  GeV and  $|abs\eta| < 2.4$  conditions. After the muon replacement by a  $\tau$ , the event  $p_T^{miss}$  is recalculated, as well as all the other kinematic variables. For events with a single muon in the  $\mu + jets$  sample,  $f_{VE}$  is taken from all bins of a corresponding  $\tau_h$  response template to emulate  $\tau_h$ +jet event. For a given  $\tau_h$  response, the event variables are recalculated and the probability from the template bin is taken as the event weight. The probability of a  $\tau_h$  jet to be misidentified as b-jet is also accounted for.

### Correction factors to $\tau_h$ simulated sample:

The measurement of  $\tau_h$  background events can be summarized by the following equation:

$$N_{\tau_h} = \sum_i^{N_{CS}^\mu} \left( \sum_j^{\text{Template bins}} (P_{\tau_h}^{\text{resp}}) \frac{\epsilon_{\tau \rightarrow \mu}}{\epsilon_{\text{trigger}}^\mu \epsilon_{\text{reco}}^\mu \epsilon_{\text{iso}}^\mu \epsilon_{\text{acc}}^\mu \epsilon_{m_T}^\mu} \frac{\mathcal{B}(W \rightarrow \tau_h)}{\mathcal{B}(W \rightarrow \mu)} \epsilon_{\text{isotrack}} \epsilon_{\text{dilepton}} \right) \quad (5.7)$$

where the first summation is over the events in the  $\mu + jets$  control sample, the second is over the bins of the  $\tau_h$  response template and  $P_{\tau_h}^{\text{resp}}$  is the probability of  $\tau_h$  response from each bin. The various correction factors applied to convert  $\mu + jets$  events into  $\tau_h + jets$  events to construct the final  $\tau_h$  sample are listed below.

- The branching ratio  $\frac{\mathcal{B}(W \rightarrow \tau_h)}{\mathcal{B}(W \rightarrow \mu)} = 0.65$ ;
- The muon reconstruction and identification efficiency  $\epsilon_{\text{reco}}^\mu$  and the muon isola-

tion efficiency  $\epsilon_{\text{iso}}^\mu [p_t(\mu)]$ , activity within the muon-jet. Section 5.6.2];

- The muon kinematic and geometric acceptance  $\epsilon_{\text{acc}}^\mu$  [search bins:  $N_t$ ,  $N_b$ ,  $M_{T2}$ ,  $p_T^{\text{miss}}$ ]. The acceptance reflects the effect of low  $p_t$  and high  $\eta$  muon on the  $\tau_h$  prediction. This value is calculated from generator level  $\mu$  control sample after smearing the muon  $p_t$  using the  $\tau_h$  template function. The acceptance is the ratio of  $\tau_h$  events converted from generator level muon events within muon kinematic acceptance to the  $\tau_h$  events converted from all gen level muon events. The muon acceptance in each search bin is shown in Fig. 5.13;
- The  $m_T$  selection efficiency  $\epsilon_{m_T} [N_{jets}, p_T^{\text{miss}}]$  as shown in Fig. 5.14;
- The contamination in the control sample by the  $\mu$ 's coming from  $\tau$  decays. The contamination factor  $f_{\tau \rightarrow \mu} \left( = \frac{\tau \rightarrow \mu\nu}{(\tau \rightarrow \mu\nu) + (W \rightarrow \mu\nu)} \right)$  is measured after the selection cuts are applied and the "extra" muon events corrected for with the multiplicative factor  $\epsilon_{\tau \rightarrow \mu} = 1 - f_{\tau \rightarrow \mu} [N_{jets}, p_T^{\text{miss}}]$ .  $f_{\tau \rightarrow \mu}$  is shown in Fig. 5.14;
- Isolated track veto efficiency for  $\tau_h$ ,  $\epsilon_{\text{isotrack}} [N_{jets}, N_b]$ . The isolated track veto is modeled by applying the  $\epsilon_{\text{isotrack}}$  multiplicative term. This is discussed in section. 5.6.1.4;
- $\tau_h$  contribution that overlaps with the lost-lepton prediction (double-counting) due to dileptonic event contamination in the control sample ( $F_{\text{dilepton}}$ ) where  $F_{\text{dilepton}} = 0.024$ ;
- A correction for  $\mu$  trigger efficiency,  $\epsilon_{\text{trigger}}^\mu$ . As trigger efficiency is 94.8% (sec. 5.1) and constant over  $\mu p_T$ ,  $\epsilon_{\text{trigger}}^\mu$  is taken to be 100/94.8.

The muon reconstruction and identification efficiency and the muon isolation efficiency are the same used for the lost-lepton background determination and are discussed in Sec. 5.6.2. Block parentheses indicate the variables the corrections are

parametrized in terms of. The data-MC difference in the  $N_{jets}$  distribution is taken care of in the uncertainty of the  $N_{jets}$  parametrization.

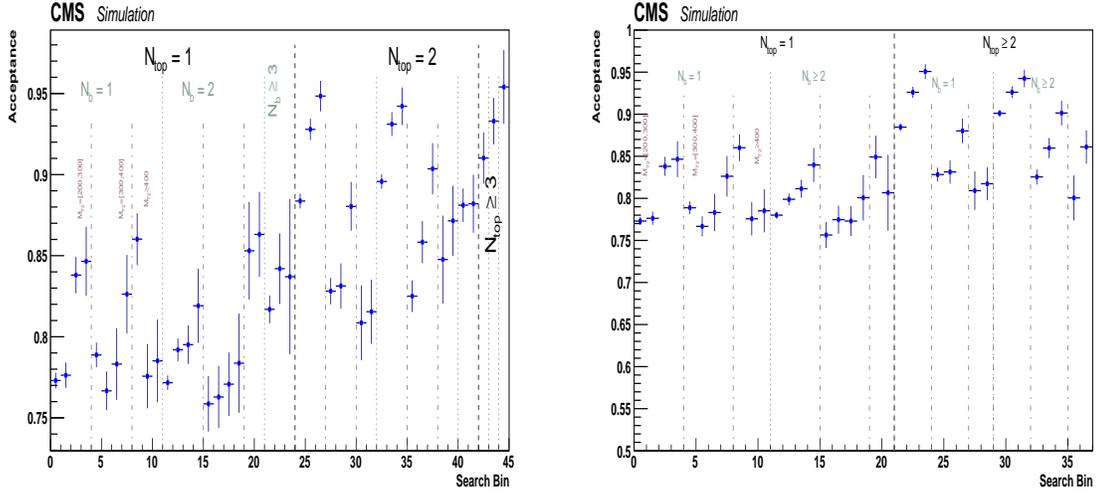


Figure 5.13: Muon acceptance as a function of search bin, in case of 45 bins (left) and 37 bins (right).

### Closure test of the method:

A closure test of  $\tau_h$  background estimation method is performed using the  $t\bar{t}$ ,  $W$ +jets, and single-top MC samples for all variables. The  $\tau_h$  background prediction (labeled by "treat simulation like data") is evaluated from  $\mu$  + jets MC events following exactly same procedure discussed in the previous section. The MC truth numbers (labeled by "direct from simulation") are obtained by applying the selection cuts and counting the number of  $\tau_h$  events that survive. The closure test which compares the  $\tau_h$  prediction and  $\tau_h$  expectation examines numerous corrections in the prediction method (see Eq. 5.7) and the muon  $p_T$  smearing based on the  $\tau_h$  response template. Many of the correction factors are not parameterized by search bins. For example,

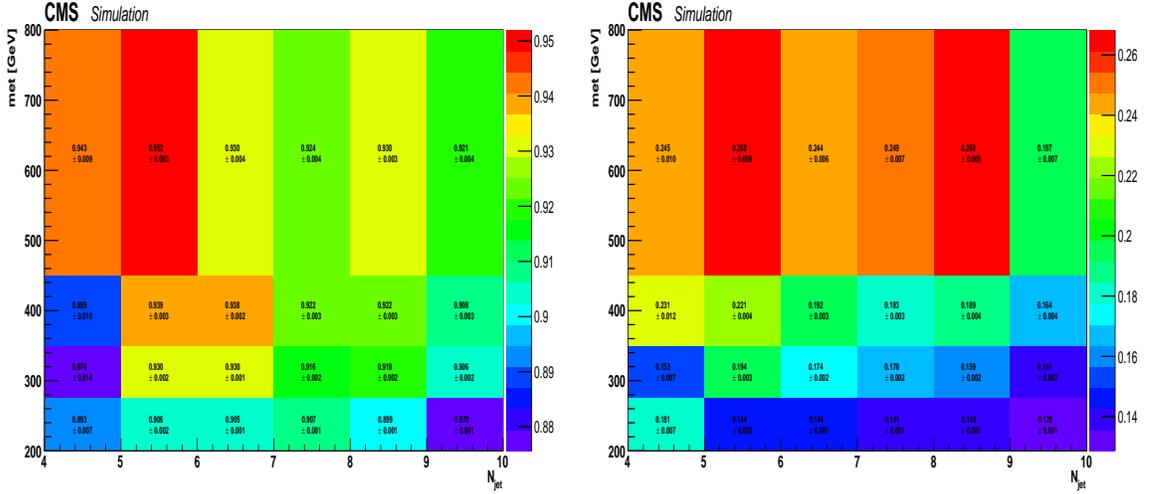


Figure 5.14: Left plot shows  $m_T$  selection efficiency as a function of  $N_{jets}$  and  $p_T^{miss}$ . Right plot illustrates the  $\tau \rightarrow \mu$  contamination, i.e., the fraction of  $\mu$  coming from  $\tau$  to the all  $\mu$  coming from both  $\tau$  and  $W$ , as a function of  $N_{jets}$  and  $p_T^{miss}$ .

the muon efficiencies are parameterized in muon  $p_T$  and activity variables so that it is easier to do the tag-and-probe efficiency measurement with reasonable statistics. If we do the tag-and-probe measurement in search bins, we run out of statistics at tight bins. This closure shows that this  $(p_T, \text{activity})$  parameterization of muon reconstruction and isolation efficiencies are good enough to model the search bin dependence. Also, the muon  $p_T$  smearing based on  $\tau_h$  response is binned in muon  $p_T$ , and all search variables are recomputed after the smearing. The closure test verifies that this smearing procedure models the kinematic differences between  $W(\rightarrow \mu\nu)+jets$  and  $W(\rightarrow \tau\nu)+jets$  events. Figure 5.15 shows closure tests of the  $p_T^{miss}$ ,  $M_{T2}$ ,  $N_b$  and  $N_t$  distributions for the baseline sample. Figure 5.16 shows the closure for each of the search bins.

As illustrated in Fig. 5.16, the method closes within statistical uncertainties in most of the search bins. For each search bin the larger one of either statistical uncer-

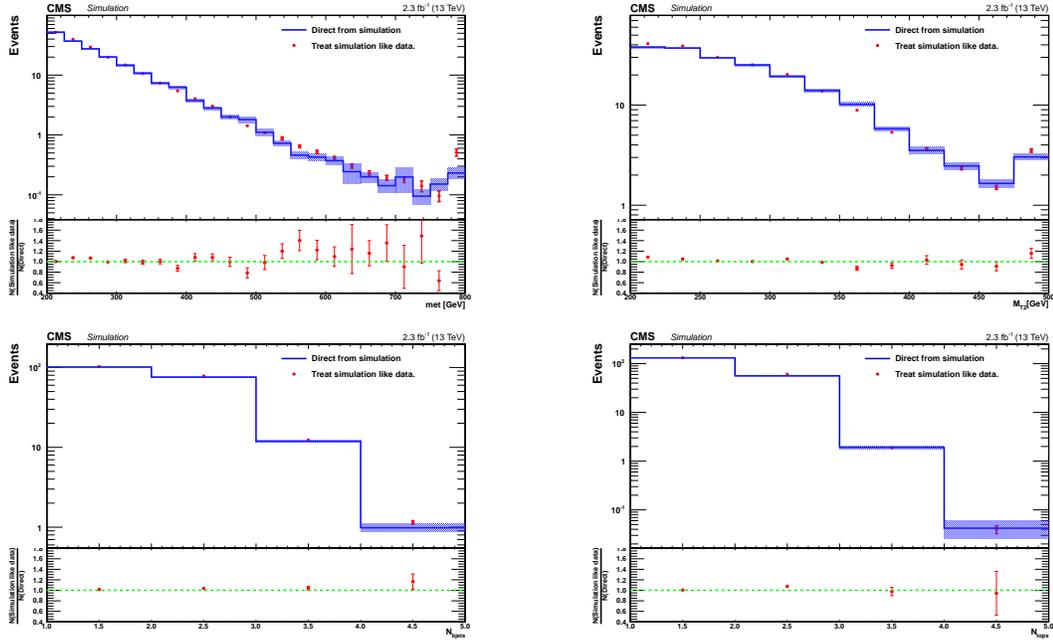


Figure 5.15: Closure tests with simulated data for  $p_T^{miss}$ ,  $M_{T2}$ ,  $N_b$  and  $N_t$  distributions based on comparisons between the predicted  $\tau_h$  background using the  $\tau_h$  template response method (red points) and the expected number of events using MC truth (blue line). Only statistical uncertainties are shown and properly propagated to the uncertainty of the ratio.

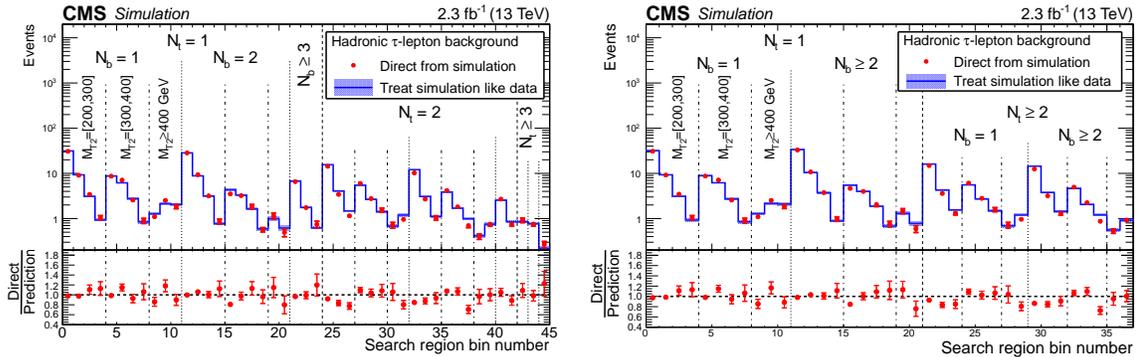


Figure 5.16: The predicted yield from simulated events (blue line) compared to the expected  $\tau_h$  background yield (red points) in the case of 45 bins (left) and 37 bins (right). Only statistical uncertainties are shown.

tainty of the closure test or the amount of the non-closure are assigned as systematic uncertainties.

### **Systematic uncertainties:**

Though  $\tau_h$  events are estimated from data control sample, the correction factors are evaluated from simulated events. They are the primary sources of the systematic uncertainties of the  $\tau_h$  prediction. The different sources of uncertainty are described below, and their effect on the  $\tau_h$  prediction summarized in table. 5.5. Statistical errors on efficiencies such as  $(\epsilon_{acc}^\mu, \epsilon_{m_T}, f_{\tau \rightarrow \mu}, \epsilon_{isotrack})$  are counted as contributions to systematics. However, since they are measured from the same MC samples, the error is added only once and not for each source. The uncertainties from each sources except closure are propagated to get the uncertainty on the prediction. Total relative systematic uncertainty in each search bin is shown in Fig. 5.17

- **Hadronic tau response template:** The uncertainty in response template comes from the potential data-MC energy scale difference for  $\tau_h$  jet. The change in template due to the variation of tau jet energy scale suggested by JetMET POG [12, 13] is evaluated.
- **Muon reconstruction and isolation efficiency:** Muon efficiencies are calculated from MC simulation. So the data-MC corrections from tag and probe method by the SUSY lepton scale factor (SF) group [14] are considered for the uncertainty.
- **Acceptance:** The uncertainty of the acceptance includes the uncertainty in the parton distribution functions (PDF) and in the renormalization and factorization scales used for the MC generation.

- **B mistagging rate of hadronic tau jet:** A constant conservative value is considered for the uncertainty of mistag rate after consulting with the Btag POG [15].
- **$m_T$  cut efficiency:**  $m_T$  is calculated with  $p_T^{miss}$  and muon  $p_t$ . The uncertainties on  $p_T^{miss}$  scale are the sources of changes in  $m_T$  cut efficiency.
- **Isolated track veto:** The major uncertainty on this method comes from data-MC correction for the efficiency of hadronic track veto on  $\tau_h$  jet. The isolation efficiency for hadronic tracks can not be validated directly, instead the tag and probe method for muon tracks is extrapolated to hadronic tracks. A conservative uncertainty due to the difference between muon and hadronic tau track isolation efficiencies is added. Other smaller sources like uncertainty on track reconstruction efficiency performed by tracking POG [16] and the uncertainty on  $m_T$  cut efficiency due to met scale variation are also taken into account..
- **Lost-lepton contamination:** 100% conservative uncertainty on this correction factor is considered.
- **Trigger efficiency:** The uncertainty on the efficiency measured for the muon  $p_t$  leg of  $\mu + H_T$  trigger is considered.
- **Closure:** In general non-closure coming from MC closure contributes in uncertainty calculation. But for low statistics bins, statistical uncertainties on closure need to be considered. The larger one between relative non-closure and statistical precision of the closure is taken to evaluate the uncertainty.

All these uncertainties except those come from closure and lost-lepton contamination factor are correlated across all search bins and are modeled with a single nuisance parameter.

Table 5.5: Contributions from different sources of systematic uncertainty to the  $\tau_h$  background prediction.

Process	Source	Effect on $\tau_h$ Prediction in %
$\tau$ response template	JES variation on $\tau$ jet	0 to 2
$\mu$ efficiency	Data-MC correction from tag and probe method. A constant value; 1.5%	1.5
Acceptance efficiency	PDF and MC scale variation	0 to 5
B-mistag rate of Had-tau	A conservative value; 50%	0 to 15
$m_T$ cut efficiency	Variation of $p_T^{miss}$ energy scale	0 to 0.5
Isolated track veto	Data-MC correction on hadronic track veto efficiency of $\tau_h$	4 to 6.5
$\tau \rightarrow \mu$ contamination factor	Statistical precision of the factor	0 to 1
Lost-lepton contamination factor	100% conservative uncertainty	2.4
Closure	Non-closure and statistical precision of the closure	2 to 28
Trigger	Uncertainty on trigger efficiency	1

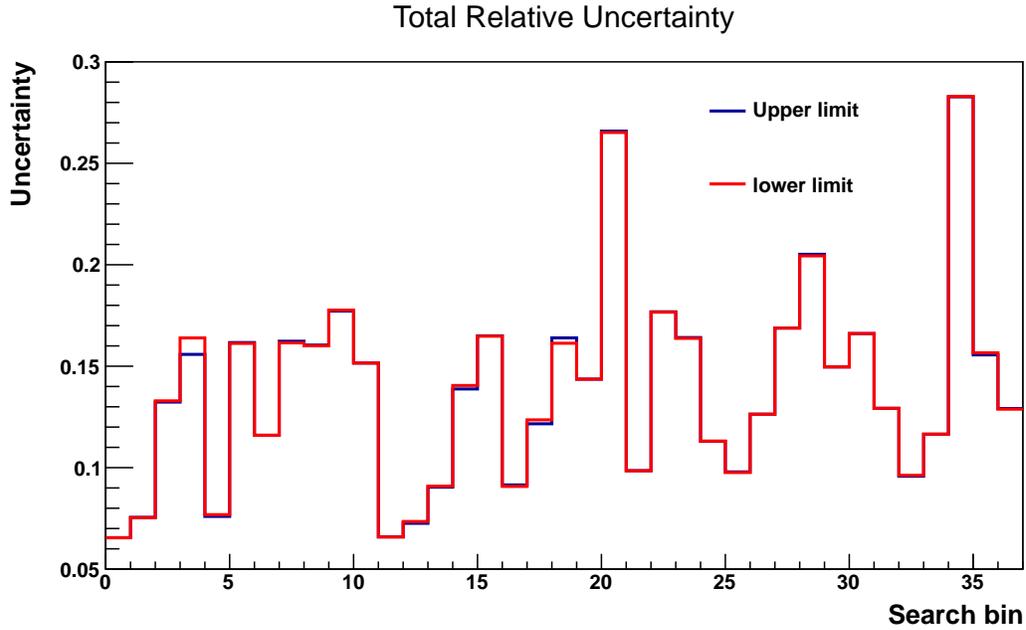


Figure 5.17: Total Relative systematic uncertainty for hadronic tau prediction in the case 37 bins

### Hadronic tau background prediction:

The hadronic tau background predictions from the single muon dataset scaled to a luminosity of  $2.3\text{fb}^{-1}$  are shown in figure 5.18 for all search bins. The error bars in Fig. 5.18 are statistical only. The predicted yield with systematic and statistical uncertainties listed in table 5.6 and table 5.7 for 37 and 45 search bins respectively. The statistical uncertainties include an additional term to account for zero observed events with high weights based on the Poisson statistical error on the zero observed as given by the Garwood interval [17], which is 1.84, multiplied by a factor of 0.30. This factor is determined by the statistical model tests and gives a good coverage of the background.

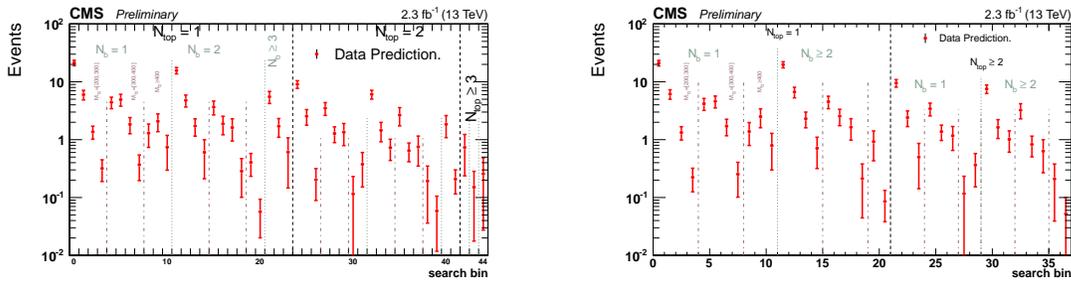


Figure 5.18: Predicted hadronic tau background yield for a  $2.3\text{fb}^{-1}$  sample in the case of 45 bins (left) and 37 bins (right). Only statistical uncertainties are shown.

### Lost Lepton Background

The lepton (electron and muon) veto does not succeed in rejecting events when the leptons (electrons or muons) are not isolated, not identified/reconstructed, or are out of the acceptance region (figure 5.19). These residual events due to "lost" leptons are estimated from a data control sample (CS) which mainly contains  $t\bar{t}$  events with a small fraction of  $W$ +jets and single top events. This control sample is collected using

search trigger (section 5.1) and defined by the same selection as the baseline criteria, but the muon veto is replaced by requiring exactly one well identified and isolated muon with  $p_T < 10$  GeV and  $|\eta| < 2.1$  and no isolated track veto is applied. In order to suppress the presence of signal while selecting the background events from  $W \rightarrow \mu\nu$  decay, data events with  $m_T < 100$  GeV are only considered (section 5.6.1.1). The "lost" leptons are modeled by weighting the CS events appropriately. The predicted number of  $t\bar{t}$ ,  $W$ +jets, and single top events with lost leptons,  $N_{LostLepton}$  contributing to each search bin is determined as

$$N_{LostLepton} = \sum_{CS} \left( \sum_{i=e,\mu} (F_{ISO}^i + F_{ID}^i + F_{Acc}^i) \times F_{dilepton}^i \right) \times \frac{\epsilon_{isotrack}}{\epsilon_{m_T}} \quad (5.8)$$

where  $\sum_{CS}$  is the sum over the  $\mu$  CS events in corresponding bin.  $F_{ISO}^i$ ,  $F_{ID}^i$  and  $F_{Acc}^i$  are the factor converting the number of events in the control sample to the number of lost electron and muon events due to isolation, reconstruction or acceptance criteria respectively.  $F_{dilepton}^i$  the correction factor for dilepton contribution,  $\epsilon_{m_T}$  the correction factor due to the  $m_T$  cut and  $\epsilon_{isotrack}$  the correction factor to compensate the isolated track veto. All these factors are described below.

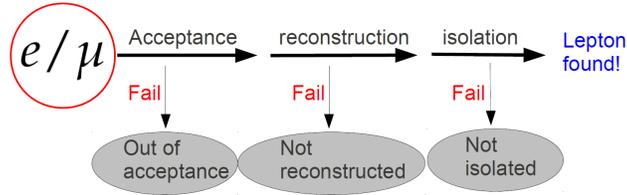


Figure 5.19: Sketch of the requirements electrons and muons from  $W$  decays must meet in order to be rejected by the explicit lepton veto.

- CS is corrected with  $\epsilon_{m_T}$  to compensate the inefficiency of  $m_T$  cut in selecting  $W \rightarrow \mu\nu$  events.  $\epsilon_{m_T}$  is obtained from the simulation as a function of muon  $p_T$

and its value is around 0.9.

- The control sample is weighted according to the lepton isolation efficiency in order to model the non-isolated leptons in the signal region ( $F_{ISO}^i$ ). The calculation is for muons and electrons depending on the superscript:

$$F_{ISO}^{e/\mu} = \frac{1 - \epsilon_{ISO}^{e/\mu}}{\epsilon_{ISO}^\mu} \cdot \frac{\epsilon_{ID}^{e/\mu}}{\epsilon_{ID}^\mu} \cdot \frac{\epsilon_{Acc}^{e/\mu}}{\epsilon_{Acc}^\mu}. \quad (5.9)$$

- To model the sample containing no identified electron or muons in the signal region, the control sample is weighted as follows:

$$F_{ID}^{e/\mu} = \frac{1}{\epsilon_{ISO}^\mu} \cdot \frac{1 - \epsilon_{ID}^{e/\mu}}{\epsilon_{ID}^\mu} \cdot \frac{\epsilon_{Acc}^{e/\mu}}{\epsilon_{Acc}^\mu}. \quad (5.10)$$

- To count for the leptons which are out of acceptance that means leptons have transverse momentum below the  $p_T$  threshold of lepton veto and/or they are emitted in the forward region with an out-of-acceptance  $|\eta|$ . This  $F_{Acc}^{e/\mu}$  factor is modeled according to the following equation:

$$F_{Acc}^{e/\mu} = \frac{1}{\epsilon_{ISO}^\mu} \cdot \frac{1}{\epsilon_{ID}^\mu} \cdot \frac{1 - \epsilon_{Acc}^{e/\mu}}{\epsilon_{Acc}^\mu} \quad (5.11)$$

- The contribution from di-leptonic events when both leptons are lost is taken into account by multiplying the factor  $F_{dilepton}$ . This factor is measured in simulated events and found to be  $(99.2 \pm 0.02)\%$  for muons and  $(96.9 \pm 0.02)\%$  for electrons respectively.
- Finally, the isolated track veto efficiency ( $\epsilon_{isotrack}$ ) is applied to get the lost lepton background prediction in search bins. The veto efficiency is measured in each search bin using MC samples.

$F_{ISO}^{e/\mu}$ ,  $F_{ID}^{e/\mu}$  and  $F_{Acc}^{e/\mu}$  are calculated from simulated  $t\bar{t}$ , W+jets, and single top events. The isolation and reconstruction efficiencies are measured using reconstructed muons and electrons after the baseline selection and parametrized as a function of the lepton  $p_T$  and the activity around the lepton, defined as the sum of the  $p_T$  of all PF particle candidates in an annulus outside a standard isolation cone divided by the  $p_T$  of the lepton:

$$A_{\mu/e} = \left( \sum_{\text{PFcands}}^{R_{\text{miniIso}} < r < 0.4} p_T \right) / p_T(\text{lep}) . \quad (5.12)$$

The acceptance efficiencies are derived for each search bin after the baseline selection. To assess whether the method can correctly predict the background, a closure test is performed by comparing the predicted number of events obtained using the lost lepton method to the simulated muon CS with the expected lost lepton events directly obtained from  $t\bar{t}$ , W+jets, and single top MC samples. Figure 5.20 illustrates the closure test result in the search region.

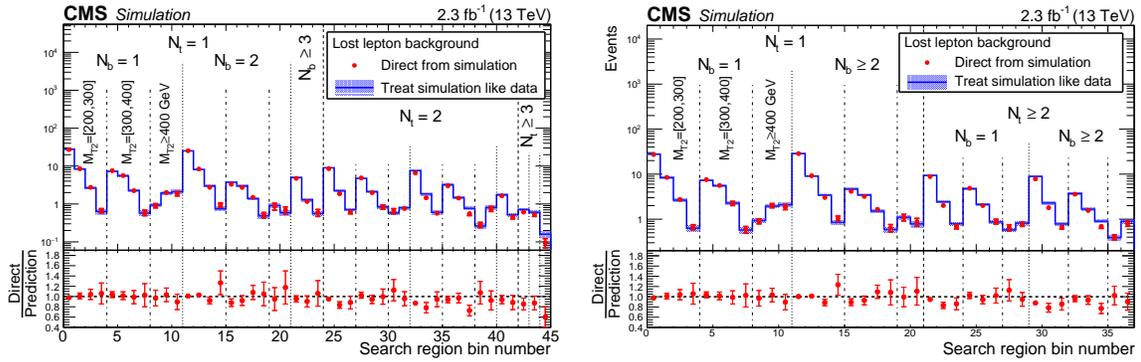


Figure 5.20: The lost-lepton background in all the search regions of the analysis as predicted directly from  $t\bar{t}$ , W+jets and single top simulation (in red) and as predicted by applying the lost-lepton background procedure to simulated muon control sample (in black). The lower panel shows the ratio between the true and predicted yields. The left (right) plot is for the case with 45 (37) search bins.

The major source of systematic uncertainty comes from closure test. Other sys-

tematic uncertainties are derived for all the factors used in lost lepton prediction in similar way as described in  $\tau_h$  section ( 5.6.1.8). The final lost lepton prediction with statistical and systematic uncertainty in all search bins using  $2.3 \text{ fb}^{-1}$  data is given in table 5.6 and table 5.7.

## $Z_{\nu\nu}$ Background

The production of  $Z$  bosons in association with jets, where the  $Z$  boson decays to a pair of neutrinos is an important background to the SUSY signals decaying to jets and  $p_T^{miss}$  final state. Ideally  $Z \rightarrow \nu\bar{\nu}$  events would be estimated from  $Z \rightarrow \mu\mu$  control sample (CS) events collected from data. The kinematics of  $Z_{\mu\mu}+\text{jets}$  events are indistinguishable from the kinematics of  $Z_{\nu\nu}+\text{jets}$  events. The behaviour under the search region selection and the characteristics of the distributions of physics observables would be preserved. However, because of limited statistics of  $Z_{\mu\mu}$  CS after the baseline selections, a method incorporating data-validated MC is used instead. The final prediction is taken from simulated  $Z_{\nu\bar{\nu}}$  events which are corrected for the differences between data and simulation. A  $Z_{\mu\mu}$  CS is used to derive the residual shape difference and overall normalization between data and simulation. First  $Z_{\mu\mu}$  events are corrected to account for any difference between  $Z_{\mu\mu}$  and  $Z_{\nu\nu}$ , then reweighted  $Z_{\mu\mu}$  are used to measure the factors correspond to data/MC difference. The  $Z \rightarrow \nu\bar{\nu}$  background prediction in each search bin  $B$  can be written as

$$\hat{N}_B = R_{\text{norm}} \cdot \sum_{\text{events} \in B} S_{DY}(N_{\text{jet}})w_{\text{MC}}, \quad (5.13)$$

with  $\hat{N}_B$  the number of predicted  $Z \rightarrow \nu\bar{\nu}$  background events in the search bin  $B$ . Here the summation is over all the simulated  $Z_{\nu\bar{\nu}}$  events fall in search bin  $B$ , and  $w_{\text{MC}}$  is MC event weight including the data/MC luminosity scale, the b-tag scale factors

and the measured trigger efficiency. Each MC event is corrected using two additional scale factors,  $R_{\text{norm}}$  which is an overall normalization factor for the  $Z \rightarrow \nu\bar{\nu}$  simulation and  $S_{DY}$  which corrects the shape of the simulated  $N_{jets}$  distribution. These two scale factors are measured in dimuon (DrellYan or DY) control sample contains two opposite charged muons with invariant mass in the  $81 < m_{\mu\mu} < 101$  GeV window around the Z boson mass. Two muons are treated as two neutrinos.

$S_{DY}$  is used to correct the shape of events variables to the simulation. A good amount of dimuon CS events are needed to assess the data/MC shape comparison. Hence a loose region is chosen removing selection on  $p_T^{miss}$ ,  $N_t$  and  $M_{T2}$  and relaxing  $H_T$  to 200 GeV. The main difference comes from mis-modeling of the jet multiplicity distribution in simulation. So  $S_{DY}$  is measured in the bin of  $N_{jets}$  as the ratio between data CS after subtracting non-DY contribution from it, and simulated DY CS. Non-DY contribution in dimuon data sample is dominated by  $t\bar{t}$ . Therefore  $t\bar{t}$  MC reweighted with  $N_{jets}$  dependent data/MC correction factor calculated in  $e\mu$  CS, are subtracted from DY data sample. After correcting the dimuon simulated events with  $S_{DY}(N_{jets})$ , the reasonable agreement in shape of search bin variables is observed. Figure 5.21 shows  $N_b$  and  $p_T^{miss}$  distributions in data and simulation after reweighting simulation with  $S_{DY}(N_{jets})$ . The deviation in  $p_T^{miss}$  around 400 GeV is caused by possible reconstruction issue in data. However, this disagreement is taken into account with a shape uncertainty equal to the magnitude of the disagreement and has a negligible effect on the final results.

The normalization scale factor  $R_{\text{norm}}$  is measured in tight region which is a very good proxy for our search region. The dimuon CS is selected with baseline cuts apart from requiring the presence of two muons and relaxing b-jet criteria. The normalization factor is determined as the ratio of non-DY subtracted dimuon data yield to the yield of  $S_{DY}$  weighted dimuon simulated CS. The data CS statistics are

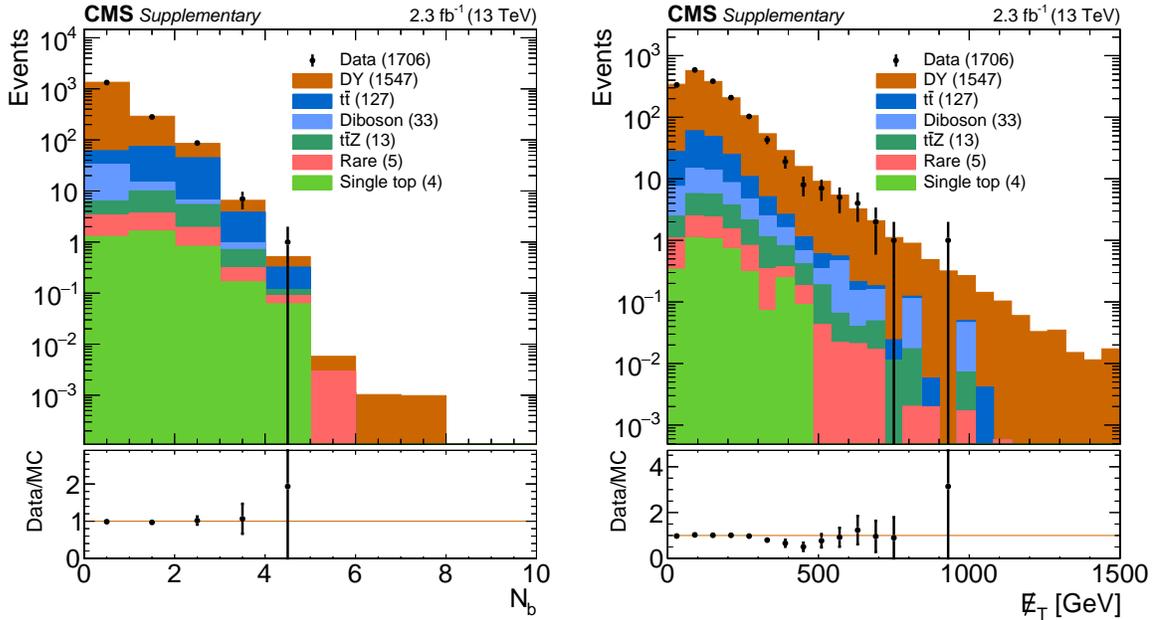


Figure 5.21: The distribution of  $N_b$  (left) and  $p_T^{miss}$  (right) from data and simulation in the loose dimuon control region after applying the  $S_{DY}(N_j)$  scale factor to the simulation. The ratio is shown in the lower panels. Only statistical uncertainties are shown. The values in parentheses in the legend represent the integrated yield for each given process.

not adequate to measure  $R_{\text{norm}}$  in each search bin, rather a single factor is calculated for entire search region.

The uncertainties associated to the use of MC simulation and the uncertainties due to the prediction method are two main source of systematic uncertainties for the  $Z \rightarrow \nu\bar{\nu}$  background prediction. The final  $Z \rightarrow \nu\bar{\nu}$  predicted backgrounds in each SB with statistical and systematic uncertainties, are shown in table 5.6 and table 5.7.

## QCD Background

The QCD multi jets events can mimic SUSY signal in  $p_T^{miss}$  and multiple jets final state due to mismeasurement of the energy of one or more jets in multi jets event. The probability of satisfying the search selection on b-jet and tagged top is very low

for such an event, but high QCD cross-section results in the presence of a number of QCD events in our signal region. The QCD background estimation procedure starts with selecting a signal depleted data control sample rich in QCD events. The non QCD contributions come from  $t\bar{t}$ , W+jets, and Z+jets are subtracted from CS to prevent the over-prediction in search region. A transfer factor which is calculated with simulated events and corrected with measurement in data sideband is used to extrapolate the number of QCD events from the control sample to the search region. The CS is selected from data with search trigger as described in section 5.1 and after applying full baseline criteria (5.3), except for the  $\Delta\phi(p_T^{miss}, j_{1,2,3})$  requirement which is inverted. The inverse  $\Delta\phi$  selection,  $\Delta\phi(p_T^{miss}, j_{1,2,3}) < 0.5, 0.5, 0.3$ , maximize the multijet events for which fake  $p_T^{miss}$  tends to align with one of the leading jets. The QCD events in inverted  $\Delta\phi$  region is estimated after subtracting the predicted  $t\bar{t}$ , W+jets and Z+jets events in this region. Events from  $t\bar{t}$ , W+jets are calculated following the hadronic tau and lostlepton method described in 5.6.1 and 5.6.2 respectively.  $Z_{inv}$  contribution is computed from simulation as it is expected to be very small. The remaining QCD events,  $N_{QCD}^{\Delta\bar{\phi}}$ , can be expressed as,

$$N_{QCD}^{\Delta\bar{\phi}} = N_{Data}^{\Delta\bar{\phi}} - N_{LL}^{\Delta\bar{\phi}} - N_{\tau_h}^{\Delta\bar{\phi}} - N_{Z\rightarrow\nu\nu}^{\Delta\bar{\phi}}, \quad (5.14)$$

where  $N_X^{\Delta\bar{\phi}}$  is the number of type X events in the inverted  $\Delta\phi$  region.

The translation factor,  $T_{QCD}$ , is defined as the ratio of multi jets events in  $\Delta\phi$  to those events in  $\Delta\bar{\phi}$  samples,

$$T_{QCD} = \frac{N_{QCD}^{\Delta\phi}}{N_{QCD}^{\Delta\bar{\phi}}}, \quad (5.15)$$

$T_{QCD}^{MC}$  is measured with simulated events after applying search selection.  $N_{MC-QCD}^{\Delta\phi}$  is obtained with baseline  $\Delta\phi$  cut where  $N_{QCD}^{\Delta\bar{\phi}}$  is computed with inverted  $\Delta\phi$  cut. In order to suppress the statistical uncertainty in  $T_{QCD}^{MC}$ , b-jet requirement is relaxed

while calculating  $N_{MC-QCD}^{\Delta\phi}$  and  $N_{QCD}^{\Delta\bar{\phi}}$ . The value of  $T_{QCD}^{MC}$  is then corrected by  $T_{QCD}^{Data}$  measured in data sideband with  $p_T^{miss}$  range from 175 GeV to 200 GeV.  $T_{QCD}^{Data}$  is given by following equation,

$$T_{QCD}^{Data} = \frac{N_{Data}^{\Delta\phi} - N_{LL}^{\Delta\phi} - N_{\tau_h}^{\Delta\phi} - N_{Z\rightarrow\nu\nu}^{\Delta\phi}}{N_{Data}^{\Delta\bar{\phi}} - N_{LL}^{\Delta\bar{\phi}} - N_{\tau_h}^{\Delta\bar{\phi}} - N_{Z\rightarrow\nu\nu}^{\Delta\bar{\phi}}}, \quad (5.16)$$

where the non-QCD contributions are computed as outlined above.  $T_{QCD}^{Data}$  is used to normalize the  $T_{QCD}^{MC}$  measured as a function of  $p_T^{miss}$  in two  $M_{T2}$  bins. The data corrected  $T_{QCD}^{Scale}$  is finally used to predict the QCD events in search region in following manner,

$$N_{QCD}^{SR} = N_{QCD}^{\Delta\bar{\phi}} \times T_{QCD}^{Scale}, \quad (5.17)$$

The main source of systematic uncertainties in QCD prediction is closure test which compares the expected number of QCD multijet events directly comes from simulation, with the predicted events obtained by applying the above mentioned method to the simulated CS events. Figure 5.22 shows the closure test performed in each search bins. Any observed nonclosure arised from relaxing b-jet requirement and using  $p_T^{miss}$  sideband in determining  $T_{QCD}^{Scale}$  is taken as systematic uncertainty. The closure uncertainty is taken from next lowest  $p_T^{miss}$  bin if any bin suffers from insufficient statistics to perform closure test. Another source of systematic uncertainty is the statistical uncertainty in measured  $T_{QCD}$ .

The QCD multi jets background prediction with statistical and systematic uncertainties are shown in table 5.6 and table 5.7.

## $t\bar{t}Z$ and other rare SM process

$t\bar{t}Z$  becomes an irreducible background to our signal when  $Z$  decays to a pair of neutrinos and both top quarks decay hadronically. The cross-section of  $t\bar{t}Z$  process

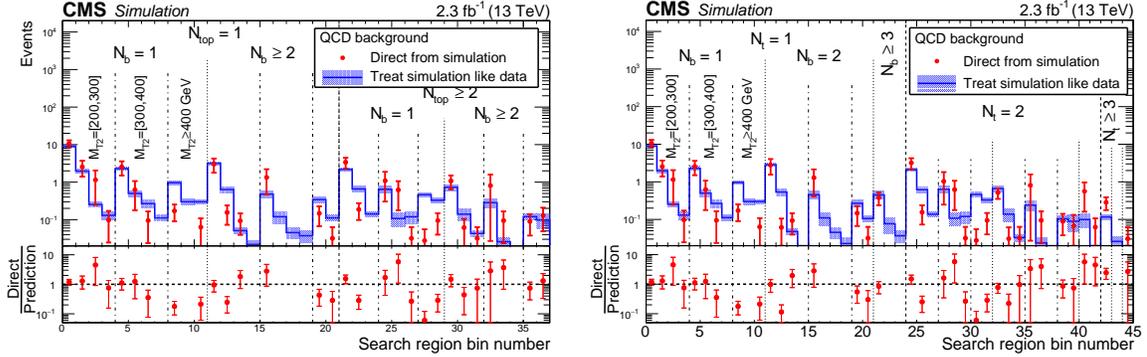


Figure 5.22: QCD multijet events directly from MC simulation (red) and predicted events (blue) using the background estimation procedure on simulation. Lower panel shows the ratio. Closure in 37 SB (left) and 45 SB (right) are shown.

is 783 fb at 13 TeV (computed at NLO level). The small cross section makes the expected  $t\bar{t}Z$  yield less than 10% of total SM backgrounds in the search region. So the final predictions in the search bins are obtained from simulation with a conservative uncertainty of 30% originated from the validation performed in tri-lepton data sample. The yields from diboson (WZ, ZZ), multibosons (WWW, WWZ, ZZZ) processes are directly calculated from MC simulation and added together into single background yield. The  $t\bar{t}Z$  and rare background predictions are shown in table 5.6 and table 5.7.

## Results

The predicted SM background events and the observed data events in 37 search bins, which are optimized for the production of direct top squark, are shown in table 5.6 and figure 5.23. Similarly, for the binning optimized for gluino mediated models, results are shown in table 5.7 and figure 5.24.

The major background contribution in the search regions comes from  $t\bar{t}$  and  $W$ +jets processes. Next largest one is  $Z_{\nu\bar{\nu}}$  + jets. In very high  $p_T^{miss}$  search bins,  $Z_{\nu\bar{\nu}}$  + jets can be dominant background. The QCD multi jets,  $t\bar{t}Z$ , and rare SM

Table 5.6: Observed yields from the full luminosity ( $2.3 fb^{-1}$ ) of data and total background predictions with their statistical and systematic uncertainties in 37 search bins.

SB	$N_t$	$N_b$	$M_{T2}$ [GeV]	$p_T^{miss}$ [GeV]	Obs.	Sum. Pred.	Lost. Lep.	Had. Tau	$Z(\nu\nu)+Jets$	QCD	ttZ	Rare
0	1	1	200-300	$3.33^{+6.34}_{-3.05}$	68	$53.54^{+3.78}_{-3.78}$	$14.97^{+2.80}_{-2.63}$	$21.10^{+2.39}_{-2.32}$	$5.74^{+0.09}_{-0.09}$	$11.08^{+1.39}_{-1.39}$	$0.51^{+0.04}_{-0.04}$	$0.14^{+0.02}_{-0.02}$
1	1	1	200-300	$2.18^{+2.81}_{-2.81}$	15	$14.98^{+2.18}_{-2.18}$	$2.84^{+1.57}_{-1.57}$	$6.15^{+1.29}_{-1.29}$	$2.11^{+0.05}_{-0.05}$	$3.56^{+0.80}_{-0.80}$	$0.20^{+0.02}_{-0.02}$	$0.12^{+0.02}_{-0.02}$
2	1	1	200-300	$1.16^{+0.90}_{-0.90}$	2	$4.87^{+1.16}_{-1.16}$	$1.75^{+1.07}_{-1.07}$	$1.33^{+0.67}_{-0.67}$	$1.01^{+0.03}_{-0.03}$	$0.64^{+0.29}_{-0.29}$	$0.09^{+0.02}_{-0.02}$	$0.05^{+0.01}_{-0.01}$
3	1	1	200-300	$0.43^{+0.43}_{-0.43}$	3	$1.19^{+0.43}_{-0.43}$	$0.00^{+0.88}_{-0.88}$	$0.22^{+0.06}_{-0.06}$	$0.54^{+0.02}_{-0.02}$	$0.34^{+0.21}_{-0.21}$	$0.01^{+0.00}_{-0.00}$	$0.08^{+0.02}_{-0.02}$
4	1	1	300-400	$1.53^{+0.92}_{-0.92}$	13	$9.80^{+1.53}_{-1.53}$	$3.02^{+1.17}_{-1.17}$	$4.18^{+1.11}_{-1.11}$	$1.96^{+0.05}_{-0.05}$	$0.45^{+0.09}_{-0.09}$	$0.14^{+0.02}_{-0.02}$	$0.05^{+0.01}_{-0.01}$
5	1	1	300-400	$1.78^{+0.89}_{-0.89}$	16	$13.18^{+2.39}_{-2.09}$	$5.75^{+2.07}_{-1.80}$	$4.60^{+1.19}_{-1.06}$	$2.24^{+0.06}_{-0.06}$	$0.15^{+0.05}_{-0.05}$	$0.24^{+0.03}_{-0.03}$	$0.19^{+0.04}_{-0.04}$
6	1	1	300-400	$1.08^{+0.88}_{-0.88}$	8	$5.04^{+1.07}_{-1.07}$	$1.65^{+1.51}_{-0.97}$	$1.71^{+0.77}_{-0.53}$	$1.51^{+0.04}_{-0.04}$	$0.02^{+0.05}_{-0.05}$	$0.15^{+0.02}_{-0.02}$	$0.01^{+0.00}_{-0.00}$
7	1	1	300-400	$0.47^{+0.47}_{-0.47}$	4	$1.29^{+1.11}_{-1.11}$	$0.00^{+0.99}_{-0.99}$	$0.25^{+0.12}_{-0.12}$	$0.84^{+0.03}_{-0.03}$	$0.10^{+0.07}_{-0.07}$	$0.10^{+0.02}_{-0.02}$	$0.01^{+0.00}_{-0.00}$
8	1	1	400+	$0.43^{+0.43}_{-0.43}$	2	$2.92^{+1.80}_{-1.80}$	$0.91^{+0.39}_{-0.52}$	$1.38^{+0.81}_{-0.59}$	$0.36^{+0.02}_{-0.02}$	$0.15^{+0.05}_{-0.05}$	$0.01^{+0.01}_{-0.01}$	$0.11^{+0.05}_{-0.05}$
9	1	1	400+	$1.24^{+0.89}_{-0.89}$	3	$6.14^{+2.11}_{-2.11}$	$2.15^{+1.85}_{-1.37}$	$2.51^{+1.02}_{-0.79}$	$1.26^{+0.04}_{-0.04}$	$0.02^{+0.05}_{-0.05}$	$0.18^{+0.02}_{-0.02}$	$0.01^{+0.00}_{-0.00}$
10	1	1	400+	$1.45^{+2.59}_{-2.59}$	3	$6.56^{+2.15}_{-2.15}$	$2.33^{+1.68}_{-1.37}$	$1.79^{+0.73}_{-0.44}$	$2.93^{+0.06}_{-0.06}$	$0.18^{+0.10}_{-0.10}$	$0.28^{+0.03}_{-0.03}$	$0.05^{+0.01}_{-0.01}$
11	1	2+	200-300	$4.15^{+4.84}_{-4.84}$	43	$44.30^{+4.15}_{-4.15}$	$16.25^{+3.17}_{-3.02}$	$19.81^{+2.44}_{-2.38}$	$1.38^{+0.04}_{-0.04}$	$6.33^{+1.10}_{-1.10}$	$0.52^{+0.04}_{-0.04}$	$0.00^{+0.00}_{-0.00}$
12	1	2+	200-300	$1.79^{+1.79}_{-1.79}$	10	$14.80^{+2.35}_{-2.35}$	$5.72^{+1.80}_{-1.80}$	$6.67^{+1.40}_{-1.40}$	$0.53^{+0.03}_{-0.03}$	$1.63^{+0.57}_{-0.57}$	$0.21^{+0.03}_{-0.03}$	$0.04^{+0.01}_{-0.01}$
13	1	2+	200-300	$0.58^{+0.58}_{-0.58}$	5	$3.61^{+1.89}_{-1.89}$	$0.39^{+0.39}_{-0.39}$	$2.30^{+0.76}_{-0.76}$	$0.26^{+0.02}_{-0.02}$	$0.49^{+0.26}_{-0.26}$	$0.14^{+0.02}_{-0.02}$	$0.04^{+0.02}_{-0.02}$
14	1	2+	200-300	$0.23^{+0.23}_{-0.23}$	1	$1.38^{+0.66}_{-0.66}$	$0.48^{+0.30}_{-0.30}$	$0.71^{+0.71}_{-0.71}$	$1.14^{+0.01}_{-0.01}$	$0.00^{+0.13}_{-0.13}$	$0.05^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$
15	1	2+	300-400	$0.95^{+0.95}_{-0.95}$	7	$7.58^{+1.38}_{-1.38}$	$2.06^{+1.35}_{-1.35}$	$4.53^{+1.22}_{-1.09}$	$0.49^{+0.03}_{-0.03}$	$0.26^{+0.06}_{-0.06}$	$0.24^{+0.03}_{-0.03}$	$0.00^{+0.00}_{-0.00}$
16	1	2+	300-400	$0.58^{+0.58}_{-0.58}$	10	$4.77^{+1.70}_{-1.70}$	$1.36^{+0.39}_{-0.39}$	$2.55^{+0.96}_{-0.79}$	$0.55^{+0.03}_{-0.03}$	$0.04^{+0.03}_{-0.03}$	$0.26^{+0.03}_{-0.03}$	$0.00^{+0.00}_{-0.00}$
17	1	2+	300-400	$0.38^{+0.38}_{-0.38}$	3	$2.78^{+0.85}_{-0.85}$	$0.54^{+0.34}_{-0.34}$	$1.64^{+0.66}_{-0.66}$	$0.35^{+0.02}_{-0.02}$	$0.03^{+0.05}_{-0.05}$	$0.21^{+0.03}_{-0.03}$	$0.00^{+0.00}_{-0.00}$
18	1	2+	300-400	$0.23^{+0.23}_{-0.23}$	2	$0.53^{+1.27}_{-1.27}$	$0.00^{+0.13}_{-0.13}$	$0.21^{+0.57}_{-0.13}$	$0.22^{+0.01}_{-0.01}$	$0.04^{+0.05}_{-0.05}$	$0.06^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$
19	1	2+	400+	$0.57^{+0.57}_{-0.57}$	2	$2.01^{+1.02}_{-1.02}$	$0.49^{+0.22}_{-0.22}$	$0.92^{+0.72}_{-0.72}$	$0.36^{+0.02}_{-0.02}$	$0.06^{+0.03}_{-0.03}$	$0.14^{+0.02}_{-0.02}$	$0.04^{+0.02}_{-0.02}$
20	1	2+	400+	$0.65^{+0.65}_{-0.65}$	1	$0.99^{+0.99}_{-0.99}$	$0.00^{+0.88}_{-0.88}$	$0.09^{+0.05}_{-0.05}$	$0.66^{+0.02}_{-0.02}$	$0.00^{+0.07}_{-0.07}$	$0.14^{+0.02}_{-0.02}$	$0.01^{+0.00}_{-0.00}$
21	2+	1	200-300	$3.31^{+3.31}_{-3.31}$	18	$20.13^{+2.43}_{-2.43}$	$5.42^{+1.68}_{-1.59}$	$9.50^{+1.83}_{-1.63}$	$0.67^{+0.03}_{-0.03}$	$4.27^{+0.87}_{-0.87}$	$0.17^{+0.02}_{-0.02}$	$0.11^{+0.04}_{-0.04}$
22	2+	1	200-300	$1.08^{+1.08}_{-1.08}$	3	$5.45^{+1.13}_{-1.13}$	$1.69^{+0.74}_{-0.74}$	$2.41^{+0.72}_{-0.72}$	$0.20^{+0.01}_{-0.01}$	$1.07^{+0.44}_{-0.44}$	$0.08^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$
23	2+	1	200-300	$0.20^{+0.20}_{-0.20}$	1	$1.14^{+0.89}_{-0.89}$	$0.32^{+0.39}_{-0.32}$	$0.50^{+0.65}_{-0.34}$	$0.09^{+0.01}_{-0.01}$	$0.20^{+0.18}_{-0.18}$	$0.03^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$
24	2+	1	300-400	$0.72^{+0.72}_{-0.72}$	10	$7.07^{+1.82}_{-1.82}$	$2.88^{+1.51}_{-1.29}$	$3.43^{+1.02}_{-0.86}$	$0.50^{+0.02}_{-0.02}$	$0.11^{+0.05}_{-0.05}$	$0.14^{+0.02}_{-0.02}$	$0.00^{+0.00}_{-0.00}$
25	2+	1	300-400	$0.49^{+0.49}_{-0.49}$	6	$4.03^{+1.52}_{-1.52}$	$2.10^{+0.99}_{-0.99}$	$1.38^{+0.41}_{-0.41}$	$0.37^{+0.02}_{-0.02}$	$0.03^{+0.02}_{-0.02}$	$0.11^{+0.02}_{-0.02}$	$0.04^{+0.02}_{-0.02}$
26	2+	1	300-400	$0.44^{+0.44}_{-0.44}$	2	$2.70^{+1.18}_{-1.18}$	$1.08^{+0.92}_{-0.65}$	$1.17^{+0.73}_{-0.48}$	$0.28^{+0.02}_{-0.02}$	$0.08^{+0.07}_{-0.07}$	$0.09^{+0.02}_{-0.02}$	$0.00^{+0.00}_{-0.00}$
27	2+	1	400+	$0.38^{+0.38}_{-0.38}$	2	$0.46^{+1.09}_{-1.09}$	$0.00^{+0.65}_{-0.65}$	$0.12^{+0.39}_{-0.39}$	$0.13^{+0.01}_{-0.01}$	$0.14^{+0.04}_{-0.04}$	$0.07^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$
28	2+	1	400+	$0.83^{+0.83}_{-0.83}$	3	$1.91^{+1.11}_{-1.11}$	$0.45^{+0.93}_{-0.45}$	$0.36^{+0.59}_{-0.22}$	$0.81^{+0.03}_{-0.03}$	$0.10^{+0.09}_{-0.09}$	$0.18^{+0.02}_{-0.02}$	$0.00^{+0.00}_{-0.00}$
29	2+	2+	200-300	$2.05^{+2.05}_{-2.05}$	6	$16.32^{+2.90}_{-2.79}$	$7.62^{+2.50}_{-2.13}$	$7.54^{+1.40}_{-1.36}$	$0.14^{+0.01}_{-0.01}$	$0.69^{+0.45}_{-0.45}$	$0.33^{+0.03}_{-0.03}$	$0.00^{+0.00}_{-0.00}$
30	2+	2+	200-300	$0.53^{+0.53}_{-0.53}$	1	$3.31^{+1.27}_{-1.27}$	$1.58^{+0.85}_{-0.85}$	$1.63^{+0.59}_{-0.59}$	$0.05^{+0.01}_{-0.01}$	$0.00^{+0.23}_{-0.23}$	$0.06^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$
31	2+	2+	200-300	$0.15^{+0.15}_{-0.15}$	0	$1.25^{+0.89}_{-0.89}$	$0.18^{+0.18}_{-0.18}$	$1.01^{+0.67}_{-0.38}$	$0.02^{+0.00}_{-0.00}$	$0.00^{+0.09}_{-0.09}$	$0.04^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$
32	2+	2+	300-400	$0.84^{+0.84}_{-0.84}$	10	$7.12^{+1.80}_{-1.80}$	$3.51^{+1.41}_{-1.41}$	$3.24^{+1.02}_{-1.02}$	$0.12^{+0.01}_{-0.01}$	$0.06^{+0.04}_{-0.04}$	$0.18^{+0.02}_{-0.02}$	$0.00^{+0.00}_{-0.00}$
33	2+	2+	300-400	$0.17^{+0.17}_{-0.17}$	2	$1.66^{+2.27}_{-2.27}$	$0.60^{+1.11}_{-0.60}$	$0.83^{+0.63}_{-0.31}$	$0.09^{+0.01}_{-0.01}$	$0.00^{+0.02}_{-0.02}$	$0.13^{+0.02}_{-0.02}$	$0.00^{+0.00}_{-0.00}$
34	2+	2+	300-400	$0.19^{+0.19}_{-0.19}$	1	$0.76^{+1.00}_{-1.00}$	$0.00^{+0.77}_{-0.77}$	$0.63^{+0.64}_{-0.33}$	$0.05^{+0.01}_{-0.01}$	$0.01^{+0.04}_{-0.04}$	$0.07^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$
35	2+	2+	400+	$0.05^{+0.05}_{-0.05}$	1	$0.27^{+1.16}_{-1.16}$	$0.00^{+0.90}_{-0.90}$	$0.21^{+0.16}_{-0.16}$	$0.03^{+0.00}_{-0.00}$	$0.00^{+0.04}_{-0.04}$	$0.03^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$
36	2+	2+	400+	$0.17^{+0.17}_{-0.17}$	1	$0.41^{+1.27}_{-1.27}$	$0.00^{+1.14}_{-1.14}$	$0.05^{+0.55}_{-0.55}$	$0.16^{+0.01}_{-0.01}$	$0.00^{+0.04}_{-0.04}$	$0.19^{+0.02}_{-0.02}$	$0.00^{+0.00}_{-0.00}$

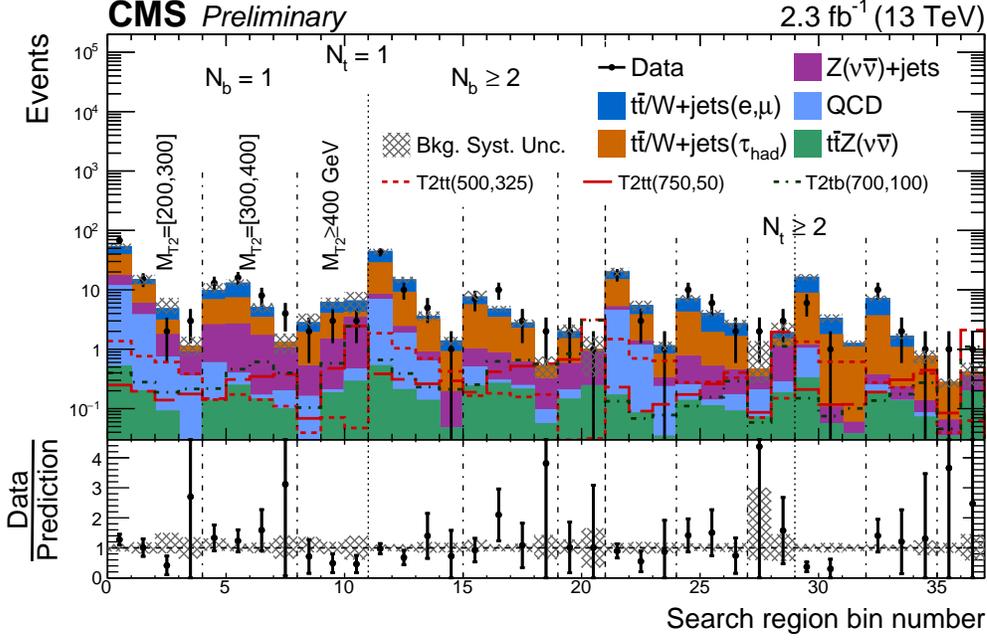


Figure 5.23: Data are shown as black points. The total predictions are shown in filled solid area. Expected yields from various direct stop production signal models are denoted by red and green lines. The bottom plot shows the ratio of data over total background prediction in each search bin. Statistical uncertainties associated with backgrounds are presented by grey and blue hatched bands in both panels. Within uncertainties, a good agreement between prediction and observation is found across all SBs.

backgrounds are subdominants across all search bins. No significant deviation from SM prediction is found in the observed data.

The result is interpreted by setting the upper limit on the cross-sections of the signal models sensitive to the analysis. The limit is calculated using  $CL_s$  method [18, 19] for signal models T2tt, T2tb and T1ttttt (chapter4) within SMS as mentioned in section 2.3 of chapter 2. Under the SMS assumption, all the properties of a SUSY model are reduced to the mass spectrum of SUSY particles, SUSY particle production cross section ( $\sigma$ ) and the decay branching ratio ( $\mathcal{B}$ ). For each model, SUSY signal is accessed by presenting its cross section times branching ratio ( $\sigma \times \mathcal{B}$ )

Table 5.7: Observed yields from the full luminosity ( $2.3 \text{ fb}^{-1}$ ) of data and total background predictions with their statistical and systematic uncertainties in 45 search bins.

Search Bin	$N_i$	$N_b$	$M_{T2}$ [GeV]	$p_T^{miss}$ [GeV]	Obs.	Sum. Pred.	Lost. Lep.	Had. Tau	$Z(\nu\nu)+\text{Jets}$	QCID	ttZ	rate
0	1	1	200-300	$3.93^{+6.36}_{-2.59}$	68	$14.97^{+2.80}_{-2.59}$	$21.02^{+2.38}_{-2.39}$	$5.74^{+0.09}_{-0.06}$	$11.07^{+1.38}_{-1.38}$	$0.51^{+0.04}_{-0.04}$	$0.14^{+0.02}_{-0.02}$	
1	1	1	200-300	$2.16^{+2.28}_{-1.84}$	15	$2.84^{+2.16}_{-1.21}$	$5.98^{+1.43}_{-1.13}$	$2.11^{+0.05}_{-0.05}$	$3.59^{+0.87}_{-0.80}$	$0.20^{+0.02}_{-0.02}$	$0.12^{+0.02}_{-0.02}$	
2	1	1	200-300	$1.65^{+2.44}_{-0.91}$	2	$1.75^{+1.65}_{-1.16}$	$1.23^{+0.65}_{-0.34}$	$1.01^{+0.03}_{-0.03}$	$0.65^{+0.32}_{-0.32}$	$0.09^{+0.02}_{-0.02}$	$0.05^{+0.01}_{-0.01}$	
3	1	1	200-300	$0.94^{+3.11}_{-0.32}$	3	$0.00^{+0.60}_{-0.30}$	$0.23^{+0.12}_{-0.12}$	$0.54^{+0.05}_{-0.05}$	$0.34^{+0.21}_{-0.21}$	$0.01^{+0.00}_{-0.00}$	$0.08^{+0.02}_{-0.02}$	
4	1	1	300-400	$1.85^{+2.42}_{-1.54}$	13	$3.02^{+1.37}_{-1.18}$	$4.27^{+0.98}_{-0.88}$	$2.96^{+0.05}_{-0.05}$	$0.45^{+0.09}_{-0.09}$	$0.14^{+0.02}_{-0.02}$	$0.15^{+0.01}_{-0.01}$	
5	1	1	300-400	$2.42^{+1.85}_{-1.12}$	16	$9.38^{+2.42}_{-2.12}$	$4.75^{+2.07}_{-1.80}$	$1.24^{+0.06}_{-0.06}$	$0.15^{+0.03}_{-0.03}$	$0.24^{+0.03}_{-0.03}$	$0.09^{+0.01}_{-0.01}$	
6	1	1	300-400	$1.11^{+0.89}_{-0.89}$	8	$5.09^{+1.11}_{-1.11}$	$1.65^{+0.97}_{-0.97}$	$1.51^{+0.04}_{-0.04}$	$0.02^{+0.05}_{-0.05}$	$0.15^{+0.02}_{-0.02}$	$0.01^{+0.00}_{-0.00}$	
7	1	1	300-400	$0.53^{+0.47}_{-0.47}$	4	$0.00^{+0.91}_{-0.17}$	$0.00^{+0.91}_{-0.17}$	$0.84^{+0.03}_{-0.03}$	$0.10^{+0.07}_{-0.07}$	$0.10^{+0.02}_{-0.02}$	$0.01^{+0.00}_{-0.00}$	
8	1	1	400+	$1.05^{+0.99}_{-0.99}$	2	$2.90^{+1.26}_{-1.13}$	$0.91^{+1.26}_{-1.13}$	$0.36^{+0.02}_{-0.02}$	$0.15^{+0.05}_{-0.05}$	$0.01^{+0.01}_{-0.01}$	$0.11^{+0.05}_{-0.05}$	
9	1	1	400+	$1.28^{+0.48}_{-0.48}$	3	$6.11^{+2.13}_{-1.54}$	$2.15^{+1.85}_{-1.26}$	$1.26^{+0.04}_{-0.04}$	$0.02^{+0.05}_{-0.05}$	$0.18^{+0.02}_{-0.02}$	$0.01^{+0.00}_{-0.00}$	
10	1	1	400+	$2.64^{+2.64}_{-1.46}$	3	$6.54^{+2.11}_{-1.46}$	$2.33^{+1.98}_{-1.37}$	$0.75^{+0.73}_{-0.45}$	$0.19^{+0.10}_{-0.10}$	$0.28^{+0.08}_{-0.08}$	$0.05^{+0.01}_{-0.01}$	
11	2	2	200-300	$3.94^{+3.94}_{-2.57}$	38	$13.34^{+3.94}_{-2.57}$	$15.98^{+2.05}_{-2.05}$	$1.27^{+0.04}_{-0.04}$	$5.08^{+1.01}_{-1.01}$	$0.43^{+0.04}_{-0.04}$	$0.00^{+0.00}_{-0.00}$	
12	1	2	200-300	$1.56^{+1.56}_{-1.56}$	7	$10.89^{+1.56}_{-1.56}$	$4.11^{+1.58}_{-1.52}$	$0.48^{+0.03}_{-0.03}$	$1.32^{+0.53}_{-0.53}$	$0.17^{+0.02}_{-0.02}$	$0.04^{+0.02}_{-0.02}$	
13	1	2	200-300	$0.80^{+0.80}_{-0.80}$	4	$3.54^{+1.32}_{-1.32}$	$0.40^{+1.21}_{-0.49}$	$0.23^{+0.02}_{-0.02}$	$0.57^{+0.26}_{-0.26}$	$0.11^{+0.02}_{-0.02}$	$0.04^{+0.02}_{-0.02}$	
14	1	2	300-400	$0.22^{+0.22}_{-0.22}$	1	$1.33^{+1.32}_{-0.64}$	$0.49^{+1.35}_{-0.64}$	$0.67^{+0.68}_{-0.12}$	$0.01^{+0.13}_{-0.13}$	$0.05^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$	
15	1	2	300-400	$0.77^{+0.77}_{-0.77}$	7	$6.37^{+1.64}_{-1.27}$	$1.97^{+1.19}_{-0.81}$	$3.52^{+1.12}_{-0.98}$	$0.45^{+0.02}_{-0.02}$	$0.23^{+0.06}_{-0.06}$	$0.00^{+0.00}_{-0.00}$	
16	1	2	300-400	$0.46^{+0.46}_{-0.46}$	10	$3.63^{+0.93}_{-0.93}$	$0.92^{+0.65}_{-0.65}$	$1.90^{+0.66}_{-0.66}$	$0.52^{+0.03}_{-0.03}$	$0.03^{+0.03}_{-0.03}$	$0.00^{+0.00}_{-0.00}$	
17	1	2	300-400	$0.41^{+0.41}_{-0.41}$	3	$2.62^{+0.87}_{-0.87}$	$0.55^{+1.41}_{-0.55}$	$0.33^{+0.02}_{-0.02}$	$0.00^{+0.04}_{-0.04}$	$0.16^{+0.02}_{-0.02}$	$0.00^{+0.00}_{-0.00}$	
18	1	2	300-400	$0.16^{+0.16}_{-0.16}$	2	$0.47^{+1.24}_{-0.47}$	$0.00^{+1.10}_{-0.47}$	$0.22^{+0.38}_{-0.38}$	$0.01^{+0.01}_{-0.01}$	$0.05^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$	
19	1	2	400+	$0.57^{+0.57}_{-0.57}$	2	$0.99^{+1.39}_{-0.21}$	$0.00^{+1.14}_{-0.21}$	$0.42^{+0.59}_{-0.59}$	$0.07^{+0.03}_{-0.03}$	$0.07^{+0.03}_{-0.03}$	$0.04^{+0.02}_{-0.02}$	
20	1	2	400+	$0.62^{+0.62}_{-0.62}$	1	$0.91^{+1.37}_{-0.05}$	$0.00^{+1.47}_{-0.05}$	$0.34^{+0.35}_{-0.35}$	$0.63^{+0.02}_{-0.02}$	$0.12^{+0.02}_{-0.02}$	$0.01^{+0.00}_{-0.00}$	
21	1	3+	200+	$0.33^{+0.33}_{-0.33}$	5	$12.02^{+2.30}_{-2.30}$	$4.55^{+1.86}_{-1.86}$	$0.19^{+0.02}_{-0.02}$	$1.88^{+0.53}_{-0.53}$	$0.13^{+0.02}_{-0.02}$	$0.00^{+0.00}_{-0.00}$	
22	1	3+	200+	$0.28^{+0.28}_{-0.28}$	3	$2.17^{+1.35}_{-0.72}$	$0.36^{+1.06}_{-0.36}$	$1.46^{+0.58}_{-0.23}$	$0.09^{+0.01}_{-0.01}$	$0.09^{+0.02}_{-0.02}$	$0.00^{+0.00}_{-0.00}$	
23	1	3+	200+	$0.25^{+0.25}_{-0.25}$	1	$1.35^{+1.59}_{-0.25}$	$0.60^{+1.33}_{-0.60}$	$0.54^{+0.70}_{-0.17}$	$0.02^{+0.04}_{-0.04}$	$0.09^{+0.02}_{-0.02}$	$0.00^{+0.00}_{-0.00}$	
24	2	1	200-300	$3.14^{+3.14}_{-2.26}$	16	$19.36^{+3.14}_{-2.26}$	$5.18^{+1.66}_{-1.56}$	$8.97^{+1.48}_{-1.32}$	$0.65^{+0.03}_{-0.03}$	$4.28^{+0.87}_{-0.87}$	$0.11^{+0.04}_{-0.04}$	
25	2	1	200-300	$0.98^{+0.98}_{-0.98}$	3	$5.20^{+1.36}_{-1.11}$	$1.65^{+0.92}_{-0.73}$	$2.33^{+0.32}_{-0.32}$	$0.19^{+0.01}_{-0.01}$	$0.84^{+0.22}_{-0.22}$	$0.00^{+0.00}_{-0.00}$	
26	2	1	200-300	$0.17^{+0.17}_{-0.17}$	1	$0.46^{+0.20}_{-0.20}$	$0.00^{+0.31}_{-0.31}$	$0.15^{+0.09}_{-0.09}$	$0.08^{+0.01}_{-0.01}$	$0.20^{+0.18}_{-0.18}$	$0.00^{+0.00}_{-0.00}$	
27	2	1	300-400	$1.08^{+1.08}_{-1.08}$	10	$7.04^{+1.82}_{-1.52}$	$2.89^{+1.31}_{-1.26}$	$3.40^{+0.84}_{-0.84}$	$0.50^{+0.02}_{-0.02}$	$0.11^{+0.05}_{-0.05}$	$0.00^{+0.00}_{-0.00}$	
28	2	1	300-400	$0.53^{+0.53}_{-0.53}$	6	$4.01^{+1.48}_{-1.07}$	$2.10^{+1.31}_{-0.93}$	$1.37^{+0.68}_{-0.68}$	$0.37^{+0.02}_{-0.02}$	$0.03^{+0.02}_{-0.02}$	$0.04^{+0.02}_{-0.02}$	
29	2	1	300-400	$0.37^{+0.37}_{-0.37}$	2	$2.68^{+0.84}_{-0.84}$	$1.08^{+0.65}_{-0.65}$	$1.16^{+0.56}_{-0.56}$	$0.08^{+0.02}_{-0.02}$	$0.08^{+0.02}_{-0.02}$	$0.00^{+0.00}_{-0.00}$	
30	2	1	400+	$0.19^{+0.19}_{-0.19}$	2	$0.45^{+1.09}_{-0.12}$	$0.00^{+0.83}_{-0.12}$	$0.12^{+0.36}_{-0.36}$	$0.13^{+0.01}_{-0.01}$	$0.07^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$	
31	2	1	400+	$0.82^{+0.82}_{-0.82}$	3	$1.90^{+0.17}_{-0.17}$	$0.45^{+0.92}_{-0.92}$	$0.36^{+0.39}_{-0.39}$	$0.80^{+0.03}_{-0.03}$	$0.18^{+0.02}_{-0.02}$	$0.00^{+0.00}_{-0.00}$	
32	2	2	200-300	$1.63^{+1.63}_{-1.63}$	6	$13.56^{+2.67}_{-2.54}$	$6.19^{+2.33}_{-2.25}$	$6.46^{+1.23}_{-1.10}$	$0.12^{+0.01}_{-0.01}$	$0.50^{+0.39}_{-0.39}$	$0.00^{+0.00}_{-0.00}$	
33	2	2	200-300	$0.46^{+0.46}_{-0.46}$	1	$3.09^{+1.29}_{-1.02}$	$1.54^{+1.00}_{-0.83}$	$1.47^{+0.77}_{-0.77}$	$0.04^{+0.01}_{-0.01}$	$0.00^{+0.23}_{-0.23}$	$0.00^{+0.00}_{-0.00}$	
34	2	2	200-300	$0.14^{+0.14}_{-0.14}$	0	$1.17^{+0.44}_{-0.44}$	$0.18^{+0.18}_{-0.18}$	$0.94^{+0.39}_{-0.39}$	$0.00^{+0.00}_{-0.00}$	$0.03^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$	
35	2	2	300-400	$0.87^{+0.87}_{-0.87}$	10	$5.28^{+1.60}_{-1.27}$	$2.47^{+1.26}_{-0.96}$	$2.42^{+0.39}_{-0.39}$	$0.11^{+0.01}_{-0.01}$	$0.09^{+0.04}_{-0.04}$	$0.00^{+0.00}_{-0.00}$	
36	2	2	300-400	$0.15^{+0.15}_{-0.15}$	2	$1.33^{+1.27}_{-0.99}$	$0.62^{+1.13}_{-0.99}$	$0.52^{+0.39}_{-0.39}$	$0.08^{+0.01}_{-0.01}$	$0.01^{+0.02}_{-0.02}$	$0.00^{+0.00}_{-0.00}$	
37	2	2	300-400	$0.15^{+0.15}_{-0.15}$	1	$0.67^{+0.99}_{-0.36}$	$0.00^{+0.74}_{-0.36}$	$0.56^{+0.66}_{-0.66}$	$0.05^{+0.01}_{-0.01}$	$0.05^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$	
38	2	2	400+	$0.04^{+0.04}_{-0.04}$	1	$0.20^{+0.87}_{-0.11}$	$0.00^{+0.87}_{-0.11}$	$0.15^{+0.36}_{-0.36}$	$0.03^{+0.00}_{-0.00}$	$0.00^{+0.02}_{-0.02}$	$0.00^{+0.00}_{-0.00}$	
39	2	2	400+	$0.16^{+0.16}_{-0.16}$	1	$0.38^{+0.97}_{-0.97}$	$0.00^{+1.30}_{-0.97}$	$0.05^{+0.05}_{-0.05}$	$0.01^{+0.01}_{-0.01}$	$0.18^{+0.02}_{-0.02}$	$0.00^{+0.00}_{-0.00}$	
40	3+	3+	200+	$0.53^{+0.53}_{-0.53}$	0	$4.34^{+1.59}_{-1.30}$	$2.41^{+1.30}_{-1.08}$	$1.84^{+0.30}_{-0.30}$	$0.03^{+0.01}_{-0.01}$	$0.06^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$	
41	2	3+	200+	$0.06^{+0.06}_{-0.06}$	0	$0.29^{+0.91}_{-0.74}$	$0.00^{+0.71}_{-0.74}$	$0.21^{+0.36}_{-0.36}$	$0.00^{+0.07}_{-0.07}$	$0.05^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$	
42	3+	1	200+	$0.23^{+0.23}_{-0.23}$	2	$1.67^{+1.23}_{-0.74}$	$0.75^{+0.95}_{-0.55}$	$0.72^{+0.49}_{-0.49}$	$0.06^{+0.01}_{-0.01}$	$0.12^{+0.13}_{-0.12}$	$0.00^{+0.00}_{-0.00}$	
43	3+	2	200+	$0.11^{+0.11}_{-0.11}$	0	$0.30^{+0.92}_{-0.19}$	$0.00^{+0.71}_{-0.19}$	$0.01^{+0.55}_{-0.55}$	$0.02^{+0.00}_{-0.00}$	$0.25^{+0.19}_{-0.19}$	$0.00^{+0.00}_{-0.00}$	
44	3+	3+	200+	$0.04^{+0.04}_{-0.04}$	0	$0.23^{+0.21}_{-0.21}$	$0.00^{+0.21}_{-0.21}$	$0.22^{+0.21}_{-0.21}$	$0.00^{+0.00}_{-0.00}$	$0.01^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$	

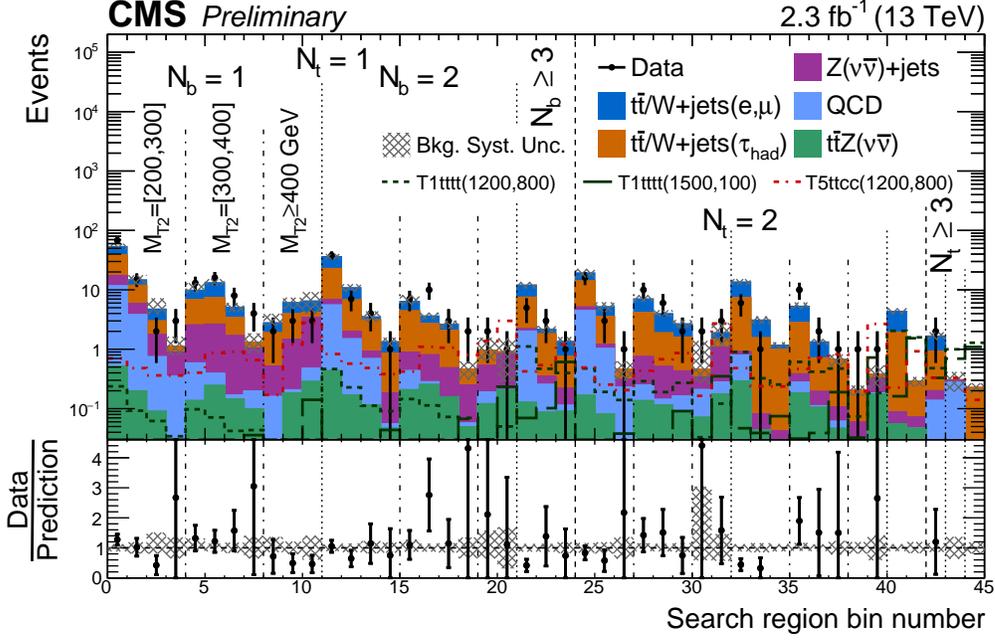


Figure 5.24: Data are shown as black points. The total predictions are shown in filled solid area. Expected yields from various direct stop production signal models are denoted by red and green lines. The bottom plot shows the ratio of data over total background prediction in each search bin. Statistical uncertainties associated with backgrounds are presented by grey hatched bands in both panels. Within uncertainties, a good agreement between prediction and observation is found across all SBs.

as a function of free mass parameters. In the experimental point of view, SUSY signal acceptance is measured after various experimental selection criteria. These accepted signal events are then translated into the signal cross section. By comparing experimentally measured  $\sigma \times \mathcal{B}$  from the collision data with the theoretical value of  $\sigma \times \mathcal{B}$ , an upper limit on  $\sigma \times \mathcal{B}$  is derived in phase space of mass parameters using Modified Frequentist ( $CL_s$ ) statistical method.

In the modified frequentist method confidence level of signal ( $CL_s$ ) is determined as

the ratio of two confidence levels,

$$CL_s = \frac{CL_{s+b}}{CL_b} \quad (5.18)$$

where  $CL_{s+b}$  and  $CL_b$  represent the confidence level in signal plus background hypothesis and in background only hypothesis respectively. The confidence level in a specific hypothesis  $x$  is given by the probability that the test-statistic function  $Q$  is less than or equal to the observed value in the data  $Q_{obs}$ :

$$CL_x = P_x(Q < Q_{obs}) \quad (5.19)$$

Based on the maximum discriminating power for a given significance level, the optimal choice of test-static is the likelihood function which is given by the Poisson probability in case of counting experiment. For example, test static for s+b hypothesis can be written as,

$$Q_{s+b} = L_{s+b} = \prod_{i=1}^N \frac{e^{-(s_i+b_i)} (s_i + b_i)^{n_i}}{n_i!} \quad (5.20)$$

where  $N$  is the total number of search channels or search bins and  $n_i$  is the observed data events in  $i$ th search bin. Final test static will be the likelihood ratio,  $\frac{L_{s+b}}{L_b}$ . The signal acceptance is estimated by maximizing the likelihood ratio. The expected number of background events ( $b$ ) is experimentally measured known quantity and is treated as a nuisance parameter while estimating signal events. In reality, the uncertainties related to signal and background predictions are taken into account in the formation of the likelihood function. These uncertainties are also treated as nuisance parameters. The nuisance parameters are taken care of by using profile likelihood method. The final likelihood ratio is maximized to obtain the signal yield which will be translated into the signal cross-section and corresponding confidence

level is obtained by utilizing the asymptotic limit of  $CL_s$  (Wilks's theorem [20]).

The 95% confidence level upper limit on the signal cross-section is determined for various signal mass points. The signal benchmark mass point represents the masses of top squark and LSP for T2tt and T2tb and the masses of gluino and LSP for T1tttt. The signal mass points for which the ratio of the measured cross-section to the theoretical cross section falls below 1 are excluded at 95% CL. Figure 5.25 shows the exclusion plots for T2tt, T2tb, and T1tttt signal models. For T2tb scenario, each of  $\tilde{t} \rightarrow t\tilde{\chi}_1^0$  and  $\tilde{t} \rightarrow b\tilde{\chi}_1^\pm$  decay modes has 50% branching fraction and 5 GeV mass difference between  $\tilde{\chi}_1^\pm$  and  $\tilde{\chi}_1^0$  is assumed as suggested by natural SUSY theory and phenomenology groups. In case of both T2tt and T2tb, no interpretation is provided for  $|m_{\tilde{t}} - m_{\tilde{\chi}_1^0} - m_t| < 25$  GeV and  $m_t < 275$  GeV due to a significant difference between FastSim and FullSim in these low  $p_T^{miss}$  regions and also because of the fact that similarity between signal and SM  $t\bar{t}$  events make the signal acceptance difficult to model. The analysis excludes the top squark mass up to 740 (610) GeV and LSP mass up to 240 (190) GeV in T2tt (T2tb) scenario. For T1tttt, gluino (LSP) mass is excluded up to 1550 (900) GeV.

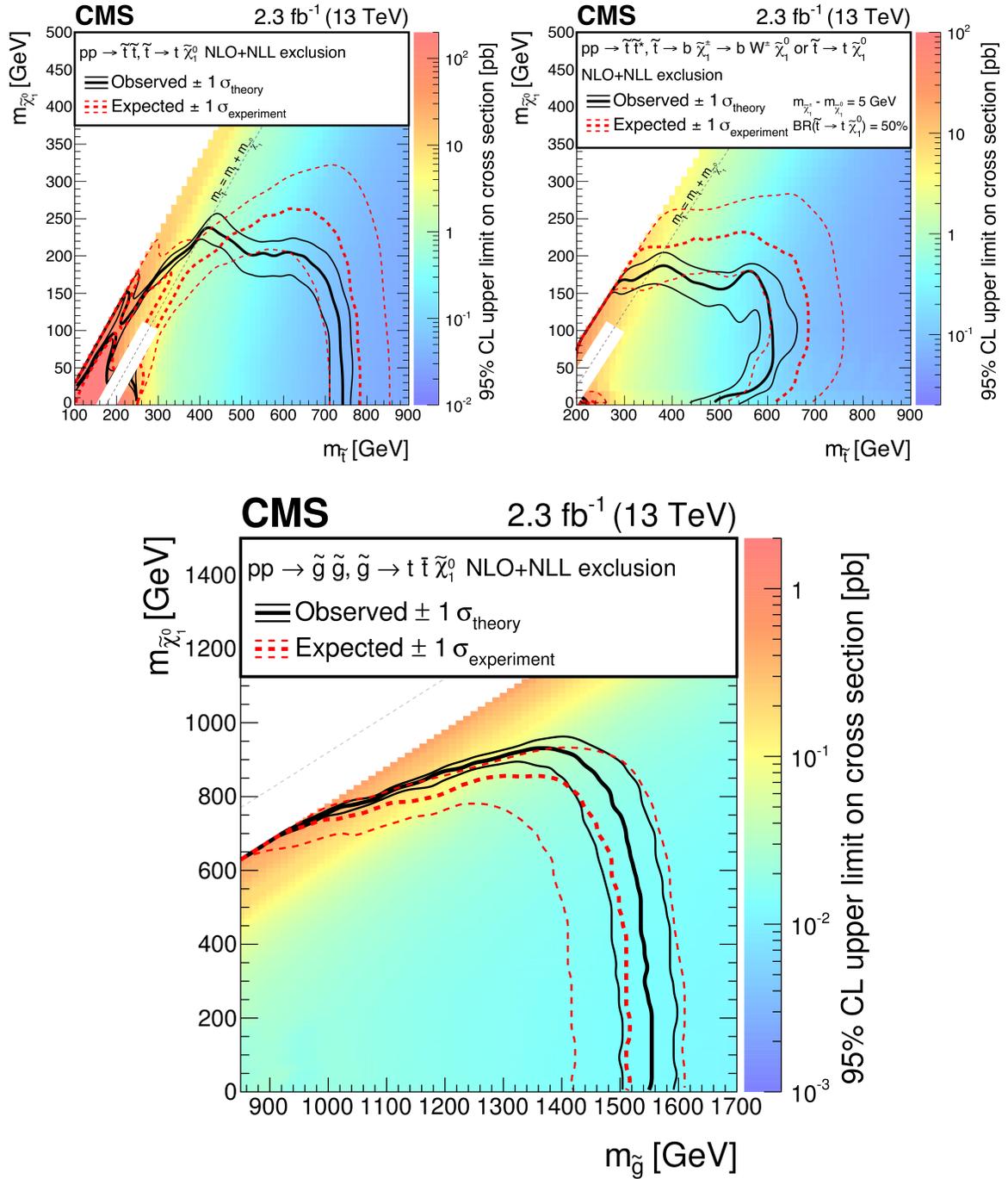


Figure 5.25: 95% CL exclusion limit for simplified signal models T2tt( top left) and T2tb (top right)and T1tttt (bottom). The solid black curves indicate the observed exclusion contour and corresponding uncertainties with  $\pm 1$  standard deviation whereas dashes red curves represent similar expected exclusion contours. The observed limits are below the expected ones because of our background estimation, specially the QCD estimation associates huge uncertainty. 98

# Bibliography

- [1] (CMS Collaboration) V. Khachatryan *et al.*, JHEP 06 116 (2015).
- [2] D. E. Kaplan, K. Rehermann *et al.*, Phys. Rev. Lett. 101 142001 (2008).
- [3] T. Plehn, M. Spannowsky *et al.*, JHEP 10 078 (2010).
- [4] C. Patrignani *et al.* (Particle Data Group), Chin. Phys. C, 40 , 100001 (2016).
- [5] C. G. Lester and D. J. Summers, Phys. Lett. B 463 99 (1999).
- [6] A. Barr, C. Lester, and P. Stephens, J. Phys. G 29 2343 (2003).
- [7] E. Re, “Single-top Wt-channel production matched with parton showers using the POWHEG method”, Eur. Phys. J. C 71 1547 (2011).
- [8] NNPDF Collaboration, JHEP 04 040 (2015).
- [9] CMS Collaboration, Eur. Phys. J. C 76 155 (2016).
- [10] C. Borschensky *et al.*, Eur. Phys. J. C 74 3174 (2014).
- [11] S. Abdullin *et al.*, J. Phys. Conf. Ser. 331 032049 (2011).
- [12] CMS JETMET POG, <https://twiki.cern.ch/twiki/bin/view/CMS/JetMET>.
- [13] CMS Webpage for Jet Energy Corrections, <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookJetEnergyCorrections>.

- [14] CMS SUSY Object Efficiency <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SUSICHEP20160bjectsEfficiency>.
- [15] CMS BTag POG <https://twiki.cern.ch/twiki/bin/view/CMS/BtagPOG>.
- [16] CMS Tracking Results <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsTRK>.
- [17] CMS Statistics Committee, “Asymmetric Error Bars for Poisson Event Counts”, (2015).
- [18] A. L. Read, J. Phys. G 28 2693 (2002).
- [19] T. Junk, Nucl. Instrum. Methods Phys. Res., Sect. A 434 435 (1999).
- [20] S. S. Wilks, Annals Math. Statist., 9(1):6062 (1938).

# Chapter 6

## 2016 Data Analysis

This analysis reports on a search for top squark (stop) production in multi jets events with a large imbalance in transverse momentum, based on 13 TeV data collected in 2016 with a corresponding integrated luminosity of  $35.9fb^{-1}$ . The main target signal process is the pair production of top squarks (T2tt topology) which is pictorially represented in figure 4.1. The search is also optimized for gluino mediated stop production process such as T1tttt (figure 4.2). In addition to these signals which were targetted in 2015 analysis, few other gluino mediated processes are also considered. These new signals are shown pictorially in figure 4.2.

Since the search is performed in the fully hadronic final state, analysis strategy very closely follow the one implemented in 2015 stop search analysis but with few changes and important improvements. The following sections report on the 2016 data analysis in a similar fashion to 2015 analysis description but focusing the detail

of new developments.

## Trigger

Data selection starts with collecting events with hadronic search triggers. Six triggers based on MET and MHT, and seeded by level 1 missing  $E_T$  trigger, are used for search region events as well as the events for lost lepton and QCD prediction. MET and MHT are defined as follows:

$$MET = \left| \sum_{\text{Reconstructed Particles}} -\vec{p}_T \right|; \quad MHT = \left| \sum_{\text{Jets}} -\vec{p}_T \right|. \quad (6.1)$$

Six MET and MHT based triggers with three thresholds (100, 110 and 120 GeV) requirement are:

- *HLT\_PFMET100\_PFMHT100\_IDTight\_v\**
- *HLT\_PFMET110\_PFMHT110\_IDTight\_v\**
- *HLT\_PFMET120\_PFMHT120\_IDTight\_v\**
- *HLT\_PFMETNoMu100\_PFMHTNoMu100\_IDTight\_v\**
- *HLT\_PFMETNoMu110\_PFMHTNoMu110\_IDTight\_v\**
- *HLT\_PFMETNoMu120\_PFMHTNoMu120\_IDTight\_v\**

The trigger efficiency of the search triggers is measured in a sample of events collected by the single-electron trigger,

- *HLTEle27WPTightv\**

with the requirement that events have at least one offline reconstructed electron with  $p_T > 30$  GeV and additionally following cuts to mimic baseline selections ( 6.3):

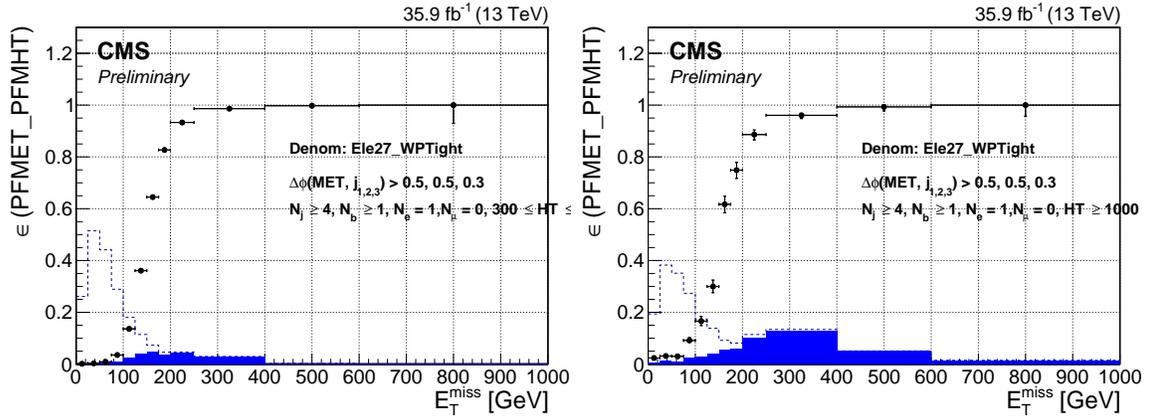


Figure 6.1: The trigger efficiency, denote by the black point, as a function of the offline  $p_T^{miss}$  for (left)  $300 < H_T < 1000$  and (right)  $H_T > 1000$ . The error bar indicates the statistical uncertainty of the trigger efficiency. The dash blue line represents the denominator passing the selection, while the solid blue histogram represents the numerator where the denominator events also trigger the search triggers.

- Pass all filters
- Veto reconstructed muon
- $N_{jets} \geq 4$
- $N_b \geq 1$
- $H_T \geq 300$  GeV
- $\Delta\phi(p_T^{miss}, j_{1,2,3}) > 0.5, 0.5, 0.3$  .

The trigger efficiency is measured in term of  $p_T^{miss}$  in low and high  $H_T$  region as shown in figure

Di-muon control sample events which are used to estimate  $Z_{\nu\bar{\nu}} + jets$  background are collected with following single muon triggers,

- $HLL\_IsoMu24\_eta2p1\_v*$ ,
- $HLL\_IsoTKMu24\_eta2p1\_v*$ ,

- *HLL\_Mu50\_eta2p1\_v\**.

## Top Tagger

Top quark reconstruction is the salient feature of the analysis. Tagging of top quark with low  $p_T$  to high  $p_T$  enable the top tagger algorithm to maintain good identification efficiency over the wide range of top quark  $p_T$ . In 2015 analysis, a custom tagger ( 5.2) was used to reconstruct top in boosted, semi-boosted and resolved decay topologies with 80% to 30% tagging efficiency. The existing algorithm is effective at giving an enriched signal region with more tops, however, it has a very high fake rate which is undesirable. A new algorithm is developed aiming to reduce this fake rate while still preserving the overall efficiency of the previous tagger. The new algorithm effectively combines three different methods to reconstruct the top in three types of top decay scenarios. A popular method for tagging highly boosted hadronically decaying tops is to use fat-jet algorithms [1]. Since the decay products of a boosted top quark end up in a very localized area in the detector, this method use an AK8 jet which is clustered with large jet cone ( $\Delta R = 0.8$ ) and then test whether the jet has three subjects consistent with  $t \rightarrow bq\bar{q}'$  ( $q\bar{q}'$  come via W) decay. This technique is very efficient to identify high  $P_T$  tops with small fake rate. However, because of using large cone jets, this strategy becomes incompetent for low  $P_T$  tops where decay products are no longer merged into a single AK8 jet. The cone size 0.8 requires that the top quark starts with at least 400 GeV of  $P_T$  to have the decay products fully captured in the jet cone. The strategy to tag the top quarks with  $P_T < 400$  GeV is combining the individual 3 jets with smaller cone size. The tagging algorithm for the resolved top decay uses 3 AK4 jets and utilizes the similar approach taken in 2015 top tagger in resolved case. The fake rate in 2015 tagger predominantly comes from the resolved

tagger due to the possibility of huge number of tri-jet combinations to be treated as the top candidates. In the new algorithm, a multivariate technique is adopted to achieve a sufficiently low fake rate and maintaining good efficiency. In order to cover the semi-boosted topology when the decay products of W that comes from top decay are merged together to form a single jet (W jet). In this case, one AK8 jet which contains the decay product of W is combined with an AK4 jet (supposed to originate from b) to reconstruct a top. More detail on boosted and semi-boosted top identification method known as merged tagging algorithm and on the resolved tagger are given in the following sections.

## Merged Top Tagger

The highly boosted top quarks are reconstructed with AK8 jets following the recommendation of JetMET group. The soft drop algorithm [2, 3] is used to remove soft constituents in the jet, and improve the jet mass resolution. Jets consistent with three-prong decay are selected by using the N-subjettiness variable  $\tau_N$  [4]. An AK8 jet is chosen as a top quark following a working point which corresponds to the following criteria:

- AK8 jet  $p_T > 400$  GeV,
- Soft drop mass between 105 and 210 GeV,
- $\frac{\tau_3}{\tau_2} < 0.65$ .

More detail on the algorithm and the optimization studies is discussed in Ref. [1] and Ref. [5].

A semi-boosted top is reconstructed by combining an AK4 jet with a W jet that is tagged from AK8 jets following the JetMET recommendation [5]. A similar technique

like boosted top tagging is employed to tag a boosted W jet. A working point with following requirements is chosen.

- AK8 jet  $p_T > 200$  GeV,
- Soft drop mass between 65 and 100 GeV,
- $\frac{\tau_2}{\tau_1} < 0.60$ .

The W-tagged jets are then combined with all possible AK4 jets to make dijet combinations and the top candidates are selected from dijet combinations if they satisfy the following conditions:

- the dijet mass is between 100 to 250 GeV,
- both jets must lie within a cone of  $\Delta R < 1.0$  around their combined  $p_T$  direction,
- the ratio of the AK8 W-tagged jet softdrop mass to the mass of the dijet system must be between  $R_{min} = 0.85(\frac{m_W}{m_t})$  and  $R_{min} = 1.25(\frac{m_W}{m_t})$ .

If multiple dijet top candidates meet these requirements and share a jet, the combination with the dijet mass closest to  $m_t$  is kept in favor of the other combinations.

We first start with AK8 jet collection to identify highly boosted tops with fully merged decay scenario. The used AK8 jets are removed from the AK8 jet collection that then takes part in semi boosted or partially merged top reconstruction. The AK4 jets used in W+jets semi boosted tagger are removed from AK4 jet collection. We also remove the AK4 jets matched to the subjets of an AK8 jet (used in two merged taggers) within a cone of radius  $\Delta R = 0.4$ . Now the cleaned AK4 jet collection is considered as the input of resolved top tagger described in next section.

## Resolved Top Tagger

A resolved top refers to a top which is reconstructed from three individual AK4 jets. In the new resolved tagging algorithm, the cut-and-count technique used in 2015 is replaced with multivariate analysis (MVA) technique. The MVA version of the top tagging algorithm uses the same general structure as the cut-and-count tagger. The top candidates are formed by clustering all combinations of three AK4 jets that have  $p_T > 30$  GeV and  $|\eta| < 5$ . Again the jets are required to fall into a cone of  $\Delta R < 1.5$  around the direction of their combined  $\vec{p}_T$  and have a combined mass between 100 and 250 GeV. The differences now arise in how the combinatoric background is rejected. The cut-based approach using the dijet and trijet masses is replaced with a random forest boosted decision tree algorithm [6] trained to discriminate between top-like triplets and background trijet combinations. The final top quark candidates are obtained by choosing the trijet combination having the random forest discriminator values above a specific threshold. Random forest training is performed with  $t\bar{t}$  and Z+jets simulated events. The signal sample contains the trijet combination from  $t\bar{t}$  events where the combination matches with a generator level hadronically decaying top quark and three jets match with three decay products of that top, geometrically in the  $\eta$ - $\phi$  space. The non-matched trijet combinations from  $t\bar{t}$  and all trijet combination from Z+jets events form the background sample. The following set of variables are considered for the training,

- trijet mass and mass of each dijet combination,
- the momentum of each jet and angular separation between jets measured in trijet rest frame,
- CSV discriminator value of jet,

- quark-gluon discriminator [7] value of each jet.

The trijet combinations that have the random forest discriminator value above 0.85 are selected as top candidates. The first step in choosing the discriminator threshold employs the knowledge of cut-and-count tagger performance. Few working points are considered where MVA tagger gives similar efficiency as that of the previous tagger. The final threshold value is chosen based on the optimization survey done in the context of SUSY signal significance study. The top candidate is allowed to have no more than one b-tagged jet among its three jets. The top candidates with one or more AK4 jets shared by another top candidate are removed in favor of the candidate with the higher MVA discriminator value. The remaining top candidates are considered to be reconstructed resolved tops. This approach gives a better performance than the cut-based approach in term of fake rate reduction by around 50% at the cost of efficiency by few percents. More detail description of resolved tagger is given in appendix A.1.

The final set of reconstructed tops contains the nonoverlapping candidates selected from three categories. The efficiency and fake rate of the top tagger are measured according to the equation 5.3. The fake rate or misidentification rate is calculated in term of  $p_T^{miss}$  using simulated  $Z_{\nu\bar{\nu}}+\text{jets}$  events and is found to be an average value of 20%. Figure 6.2 shows the efficiency measured in T2tt signal events. A similar result is observed for SM  $t\bar{t}$  events. Like 2015 analysis, the number of reconstructed top quarks are used as search variables and the kinematic properties of tops are utilized to derive  $M_{T2}$  variable.

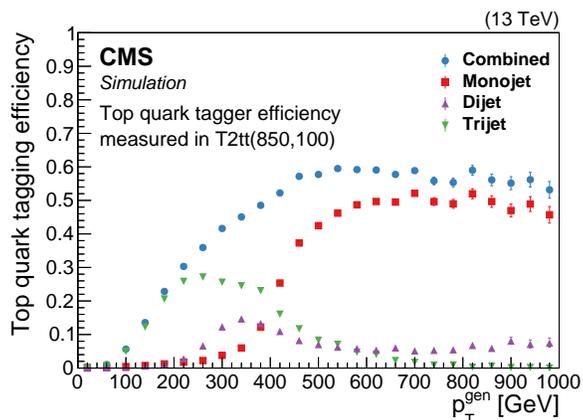


Figure 6.2: The efficiency of top tagging algorithm in term of generated top quark  $p_T$  computed using simulated T2tt signal sample where  $m_{\tilde{t}} = 850$  GeV and  $m_{\tilde{\chi}_1^0} = 100$ . Blue circles represent the efficiency of full combined tagger where as other three plots shows the efficiency of three individual taggers. Red boxes indicate monojet boosted tagger, magenta triangles are for semiboosted dijet and green triangles present the resolved trijet tagger.

## Baseline Selections

The set of following cuts define the baseline selection:

- Pass all filters that remove detector and beam related noise:
- Muon veto
- Electron veto
- Isolated track veto
- $N_{jets} \geq 4$
- $p_T^{miss} \geq 250$  GeV
- $H_T \geq 300$  GeV
- $N_b \geq 1$

- Angular cut:

$$\Delta\phi(p_T^{miss}, j_{1,2,3}) > 0.5, 0.5, 0.3$$

- $N_t \geq 1$
- $M_{T2} \geq 200$  GeV

The detail about these cuts are discusses in section 5.3.

## Search Regions

In 2016 analysis, search region is defined in term of five search variables,  $N_t$ ,  $N_b$ ,  $p_T^{miss}$ ,  $M_{T2}$  and  $H_T$ . In  $N_t \leq 2$  and  $N_b \leq 2$  regions, events are distributed in search bins (SB) based on four variables,  $N_t$ ,  $N_b$ ,  $p_T^{miss}$ ,  $M_{T2}$ . The regions with  $N_t \geq 3$  and  $N_b \geq 3$  are mostly sensitive to gluino mediated signal models where three or more top quarks produced.  $M_{T2}$  reconstruction may not always associate with correct reconstructed objects and results in a broad and flat  $M_{T2}$  distribution. In this region,  $H_T$  is found to be better discriminating variable compared to  $M_{T2}$  in selecting signal over background. Therefore, search regions with  $N_t \geq 3$  and  $N_b \geq 3$  are divided into  $p_T^{miss}$  and  $H_T$ . Altogether 84 search bins are constructed which are shown in figure 6.3

## Signal and Background MC Samples

Monte Carlo samples of SM processes reconstructed with CMSSW release 8.0 (Summer16) are used throughout this analysis. All samples use the PU25bx25 pileup scenario, which simulates a pileup distribution with an average of 25 interactions per bunch crossing and a 25 ns interval between bunches. The SM background processes

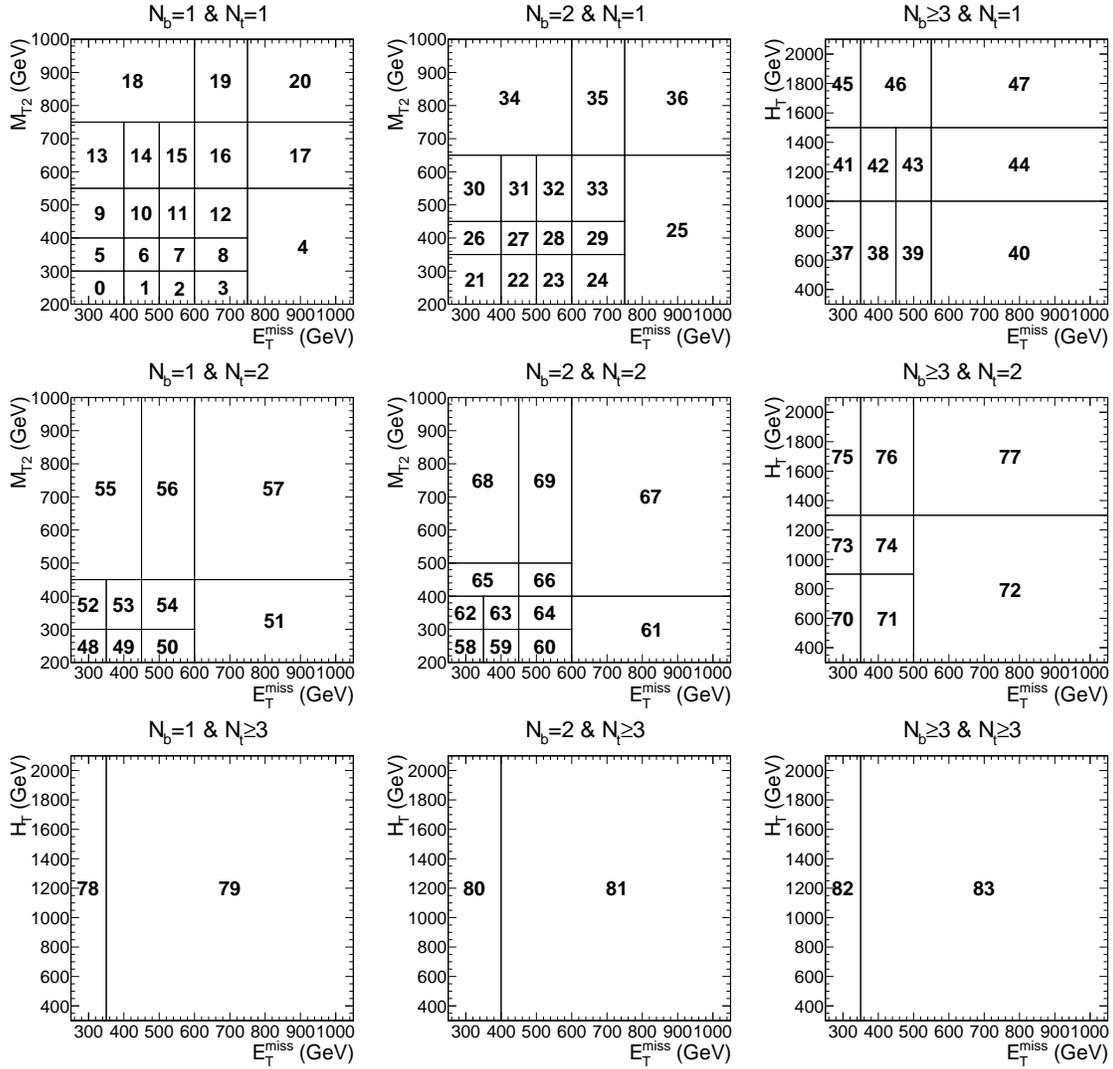


Figure 6.3: Search bin definitions and bin numbers for 84 SBs after baseline selection. The number indicates bin indices.

as listed in table 5.2 are also used in 2016 analysis.

The detail about signal samples for direct top squark production (T2tt) as well as one gluino mediated production (T1tttt) is already discussed in section 5.5. Decay topology of T5tttt model is similar to that of T1tttt model, except top squark produced from gluino is on-shell and decays to top quark and LSP ( $\tilde{\chi}_1^0$ ) with  $\Delta M(\tilde{t}, \tilde{\chi}_1^0)$  is 175 GeV. For  $\Delta M(\tilde{t}, \tilde{\chi}_1^0) < m_t$ , top squark cannot decay to an on-shell top quark and LSP. In this scenario,  $\tilde{t}$  decays to a charm quark and LSP which is modeled in T5ttcc signal topology with  $\Delta M(\tilde{t}, \tilde{\chi}_1^0) = 20$  GeV assumption. In T1ttbb signal model, the decay of gluino into off-shell top and bottom squarks is considered. Here both direct and cascade three-body decay of gluino are modeled. In direct decay, gluino decay to either  $\tilde{g} \rightarrow t\tilde{t}\tilde{\chi}_1^0$  or  $\tilde{g} \rightarrow b\tilde{b}\tilde{\chi}_1^0$  mode. In one step cascade mode gluino first decays to  $\tilde{t}b\tilde{\chi}_1^+$  or its charge conjugate and then  $\tilde{\chi}_1^+$  converts to a  $\tilde{\chi}_1^0$  through a  $W$  boson.  $\tilde{\chi}_1^+$  is the lightest chargino and considered nearly mass degenerate with LSP by constraining  $\Delta M(\tilde{\chi}_1^+, \tilde{\chi}_1^0) = 5$  GeV, hence resulting in off-shell  $W$  production. All signal samples are generated using Fast Simulation. Cross section for some mass points of T2tt and T1tttt signal models are listed in table 5.3.

## Background Estimation

The SM background processes are estimated in search region after applying baseline cuts (6.3). The contributions from different SM processes after the baseline selection are shown in figure 6.4. The events from  $t\bar{t}$ ,  $W$ +jets, and single top processes enter our search region due to lost lepton and hadronic tau effect. Due to the novel top tagger algorithm which makes use of two types of jets, tau template method become very difficult to implement in estimating hadronically decaying tau events. Therefore a new procedure is followed to estimate both lost lepton and hadronic tau events.

The background which comes from  $t\bar{t}$ , W+jets, and single top, predicted by this new method, is combinedly known as lost lepton background.  $Z_{\nu\bar{\nu}}$ + jets, QCD,  $t\bar{t}Z$  and other rare backgrounds are estimated following the exact same methods as discussed in 2015 analysis. So only lost lepton background estimation method is described in the succeeding section. Total predicted backgrounds from all SM processes using 2016 data and simulated events are shown in the result section.

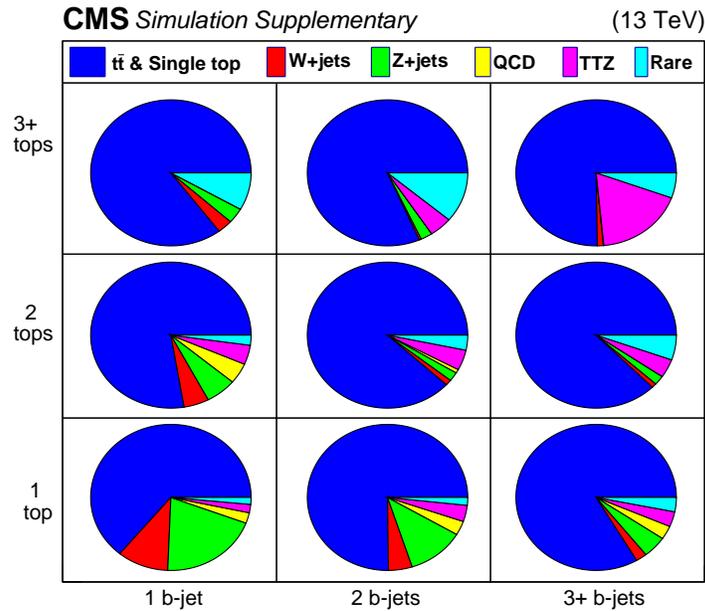


Figure 6.4: The pie chart of the contribution from SM background processes in search region. The pie chart is computed in the bin of  $N_t$  and  $N_b$ .

## Lost Lepton Background

When a W boson decays to a neutrino and a lepton which goes missing either in a form of hadronically decaying tau ( $\tau_h$ ) which is reconstructed as jet (section 5.6.1.1) or in the form of light electron or muon which gets lost (due to the reasons discussed in section 5.6.2), the event passes our lepton vetoes. Although an isolated track veto is applied, there are residual events passing because of the veto in-efficiency. These

background events coming from  $t\bar{t}$ , W+jets, and single top processes are estimated using "translation factor (TF) method".

**Translation factor method:**

Translation factor is the ratio between background yield and single lepton control sample (CS) yield in the search region (SR). The translation factor is measured in the simulated events come from  $t\bar{t}$ , W+jets and single top processes. Using the TF, hadronic tau and lost lepton events are predicted from the control region of single lepton events selected from data using search trigger. The procedure can be summarized in the following set of equations:

$$\begin{aligned}
 \text{a) } N_{Data}^{SR} &= TF_{simulation} * N_{Data}^{CS} \\
 \text{b) } TF_{simulation} &= \frac{N_{simulation}^{SR}}{N_{simulation}^{CS}}
 \end{aligned}
 \tag{6.2}$$

where  $N^{SR}$  is the number of either hadronic tau or lost lepton events in search region and  $N^{CS}$  is the events from either electron or muon control sample in the search region. Generally, simulation of the  $t\bar{t}$ , W + jets, and single-top have good agreement with data. However some differences are observed by comparing lepton CS events from data and MC. In order to make simulation agree better with data, we apply various officially provided data/MC scale factors. This is discussed in more detail in 6.6.1.3. Since the translation factor takes a ratio of events between signal region and CS in the same simulated samples, some data/MC scale factor effects cancel and have reduced effect on the measured factors. Two different translation factors are obtained from  $\mu$  and  $e$  CS separately. By applying these factors to the corresponding data CS, hadronic tau or lost lepton events are estimated in search region (6.6.1.6). Systematic uncertainties on this method are addressed in 6.6.1.4.

### Control Sample:

Lepton ( $\mu/e$ ) control sample is selected from either simulation or collider data applying following criteria.

- Search trigger ( 6.1) is used (for collider data)
- Events are required to pass the same cleaning requirements described in Section 6.3.
- Events are required to have one and only one identified and isolated lepton. For muon CS, offline muon identification follows the POG “medium muon” recommendation, including the mini-isolation requirement. The electron in electron CS is selected following POG recommended “Cut Based VETO” criteria and also including mini-isolation requirement.
- Events with additional leptons (electrons/muons) are rejected according to the prescription in section 6.3.
- The transverse mass  $m_T = \sqrt{2p_T^l p_T^{miss}(1 - \cos \Delta\phi)} < 100$  GeV is enforced to remove most of the potential signal contamination.

To investigate possible differences between data and simulated events, data CS is compared with the CS selected from  $t\bar{t}$ ,  $W$ +jet, and single-top simulated events following above requirements. In order to compare data and MC in a region as close as the search region, signal selection cuts except lepton veto and isolated track veto are applied. The shapes of  $N_{jets}$ ,  $H_T$ ,  $p_T^{miss}$ ,  $N_b$ ,  $N_t$  and  $M_{T2}$  distributions are compared separately for  $\mu$  and  $e$  CS. Figure 6.5 and 6.6 show the comparison between data and MC for various kinematic variables. For a shape comparison, the overall MC is scaled down by 73% for muon CS and 71% for electron CS. The agreement between

data and MC for  $N_b$  is quite good but the prominent shape differences are found in  $N_{jets}$ ,  $N_t$ .

To account for data and MC differences, MC sample is corrected by following three scale factors.

- **ISR jet re-weighting:** ISR jet re-weighting is applied to correct for the difference in hadronic part of the event between data and MC. It corrects for LO and NLO differences which results in the correction for  $N_{jet}$  distribution.
- **b-jet tagging:** The data-MC scale factor of b-jet tagging recommended by B-POG is applied.
- **Lepton efficiency:** Although in general, the lepton data/MC scale factors are close to 1.0, these SF are important because of the fact that leptons are treated differently between CS (leptons are selected) and signal region (leptons are vetoed). In the CS events, a reconstructed and isolated lepton is selected. For this lepton selection efficiency, there are differences between data and MC. Therefore the data/MC corrections of lepton reconstruction and isolation efficiencies from tag and probe method by the SUSY lepton scale factor (SF) group are considered. Implementation of lepton SF in the measurement of TF is discussed in section 6.6.1.3.

After applying above scale factors, events variables shape comparison are shown in Figure 6.7 and 6.8. For the shape comparison, the muon MC CS is scaled by 84% and electron MC CS is scaled by 83%. We can clearly see the improvement of the  $N_{jets}$  and  $N_t$  distributions (mainly from the ISR re-weighting). For both the muon and electron CS, there seems to be residual shape difference between Data and simulation. However, since the TF method measures a ratio in MC between the SR and the

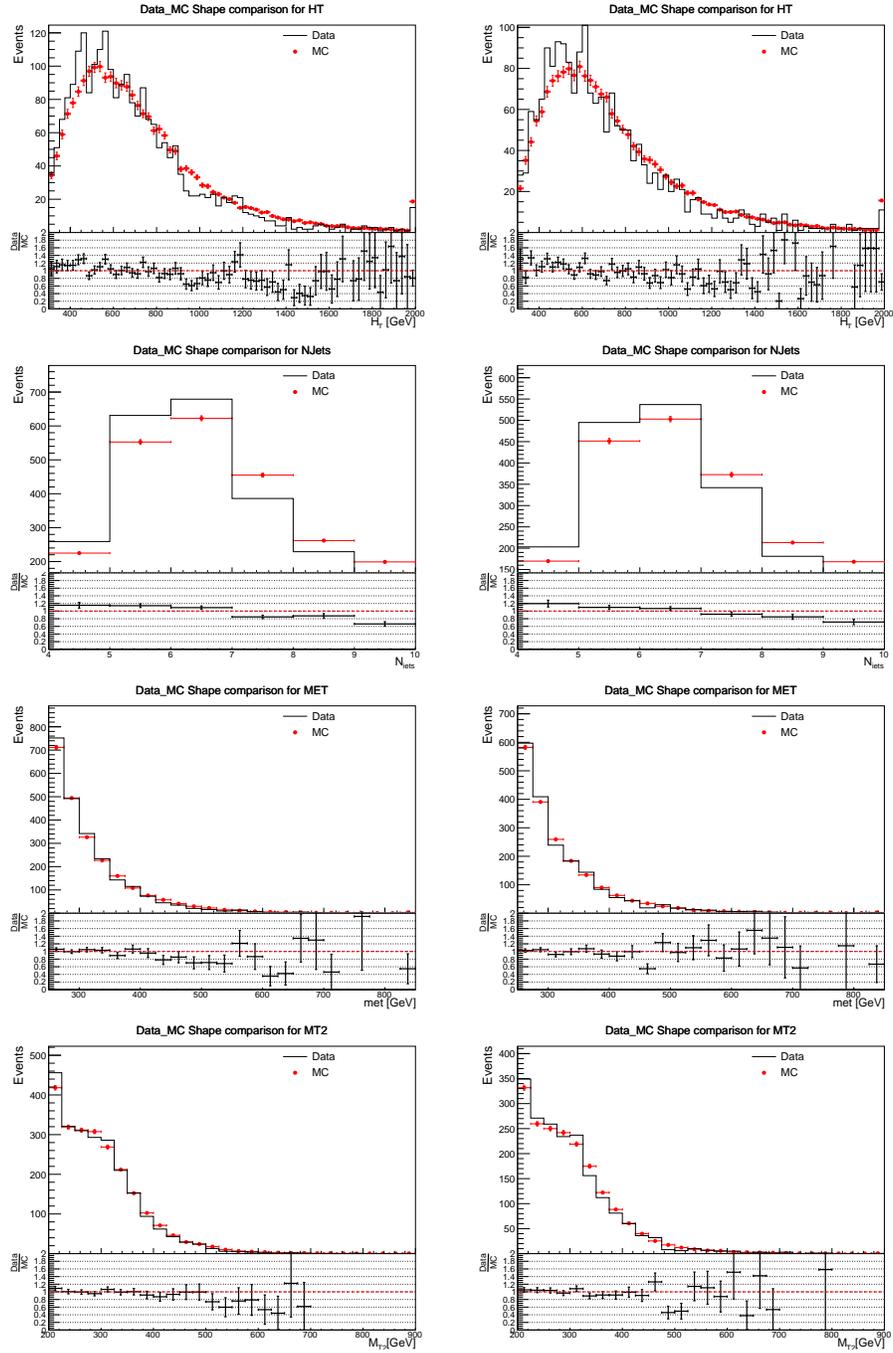


Figure 6.5: Shape comparison of  $H_T$  (1st row),  $N_{jets}$  (2nd row),  $p_T^{miss}$  (3rd row) and  $M_{T2}$  (4th row) for muon CS (left column) and electron CS (right column).

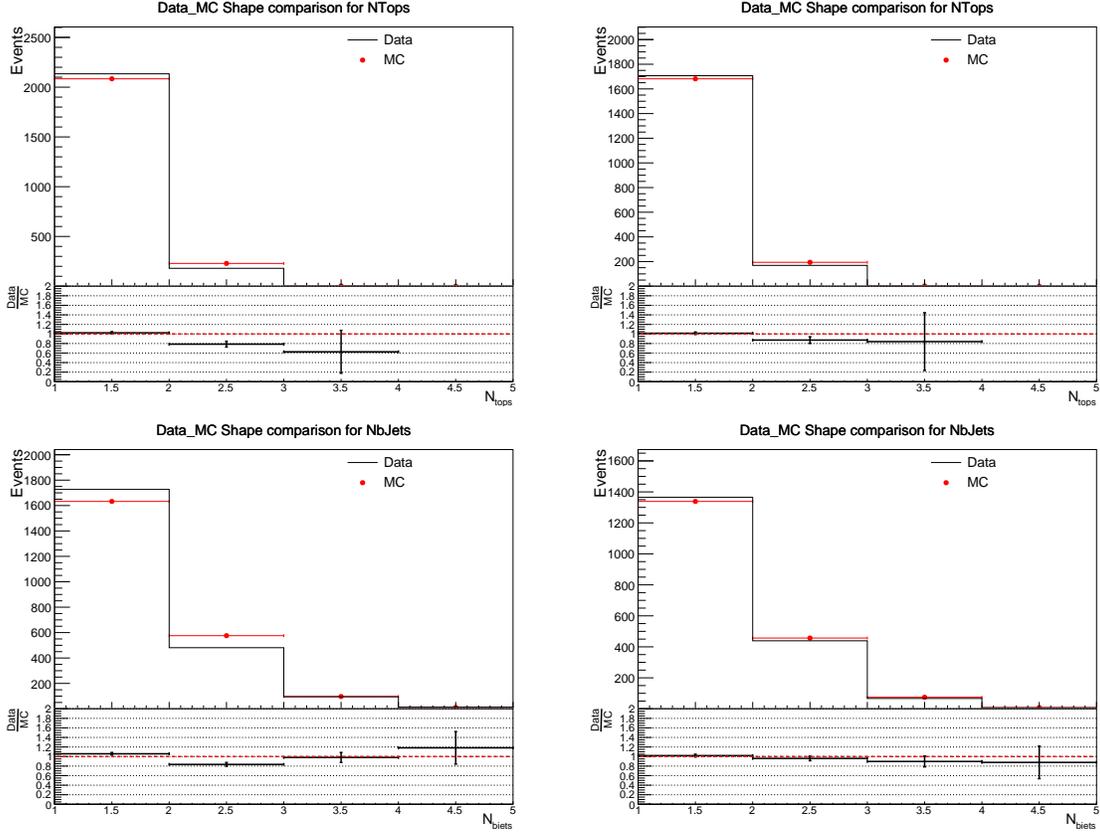


Figure 6.6: Shape comparison of  $N_t$  (top row) and  $N_b$  (bottom row) for muon CS (left column) and electron CS (right column).

corresponding CR in the bins of the kinematic variables the residual shape difference should have minimum impact on our measured TF.

### Translation factor measurement:

Translation factor is evaluated in each search bin by making the ratio between  $\tau_h$  or lost lepton events after full search selection cuts ( 6.3) and the CS events selected with criteria discussed in section 6.6.1.2. Two sets of TF are measured for  $\tau_h$  and e CS separately. To account for the difference in kinematic modeling between data and MC, three scale factors as mentioned in section 6.6.1.2 are applied both on  $\tau_h$  or lost lepton and CS events. We cannot directly apply the data/MC corrections on the

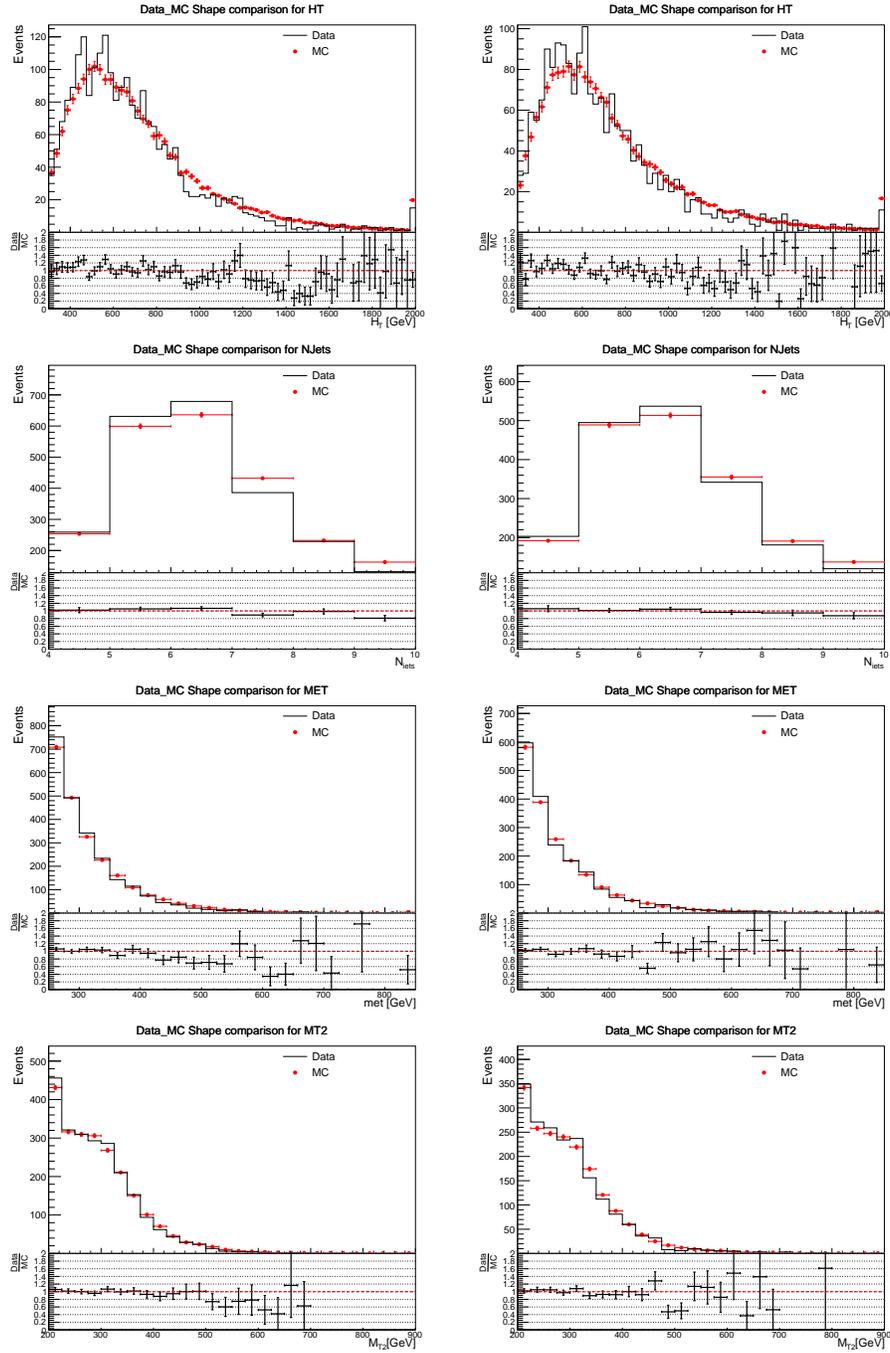


Figure 6.7: Shape comparison of  $H_T$  (1st row),  $N_{jets}$  (2nd row),  $p_T^{miss}$  (3rd row) and  $M_{T2}$  (4th row) for muon CS (left column) and electron CS (right column) after the various corrections on MC samples.

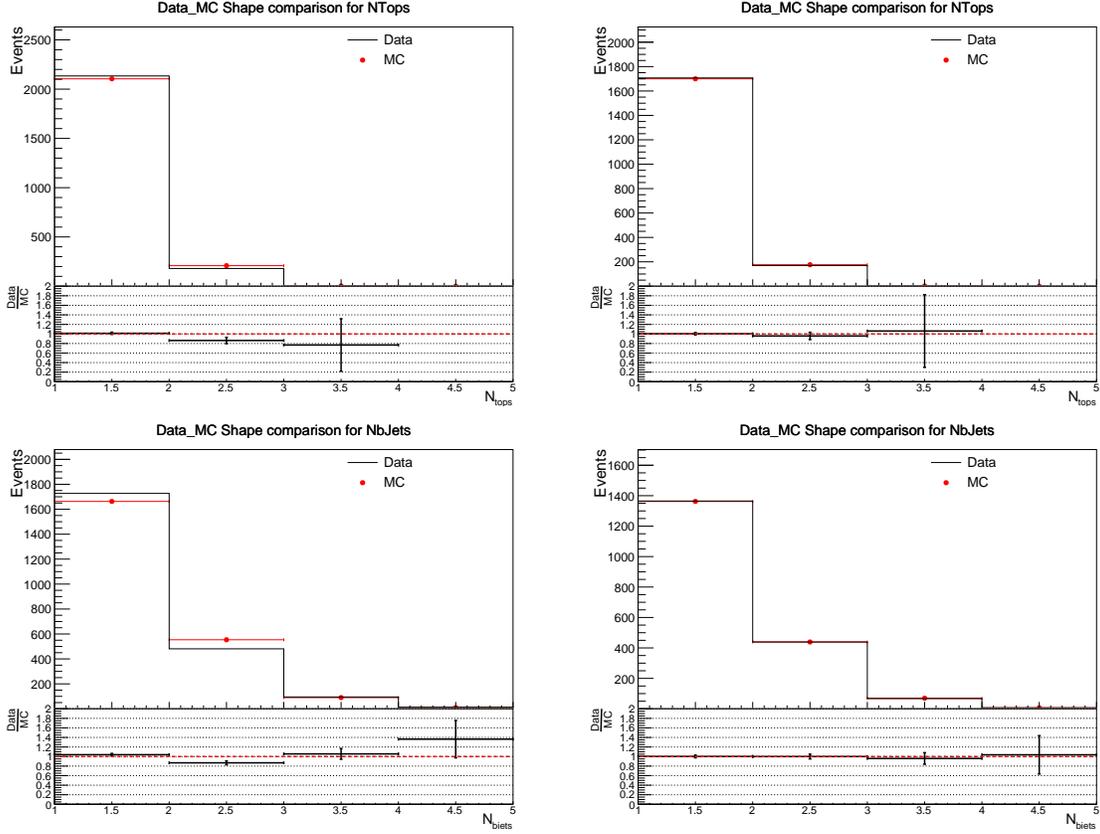


Figure 6.8: Shape comparison of  $N_t$  (top row) and  $N_b$  (bottom row) for muon CS (left column) and electron CS (right column) after the various corrections on MC samples.

events in the signal regions, because in the signal regions leptons are vetoed therefore we do not have a well-identified lepton to begin with. Therefore we use the following relation to propagate the correction factors:

$$N_{prod}^i = N_{lost}^i + N_{sel}^i \quad (6.3)$$

The  $N_{prod}^i$  is the total produced number of events in the  $i^{th}$  search bin with a  $W$  boson decaying into muon, electron or  $\tau_h$ .  $N_{lost}^i$  is the number of events that end up in our search bin after all the search cuts including the lepton and isolated track veto.  $N_{sel}^i$  is the number of events selected with identified muon, electron or isolated tracks for

vetoing. Note that  $N_{prod}^i$  remains unchanged regardless of corrections of the lepton data/MC scale factors. Therefore the change in simulated number of events due to lepton SF on the  $N_{sel}^i$  can be easily propagated to the quantity we are interested in, i.e., the  $N_{lost}^i$ . Therefore for electron and muon we apply the data/MC corrections from the SUSY lepton SF group. For isolated tracks we assume the scale factor is 1.0 and we will propagate the uncertainty on the isolated track identification efficiency as a systematic uncertainty on the final prediction. The overall effect of the lepton efficiency scale factor on the  $\tau_h$  prediction is small. However the lepton efficiency SF has a big impact on the prediction of the lost lepton events, i.e., lost electron or muon events, using the TF method.

The data-corrected translation factors for the lepton control sample are shown in figure 6.9 for the  $\tau_h$  and lost lepton for both muon and electron CS. As we expect, within uncertainties the TF from electron and muon CS follow similar trend across all the search bins.

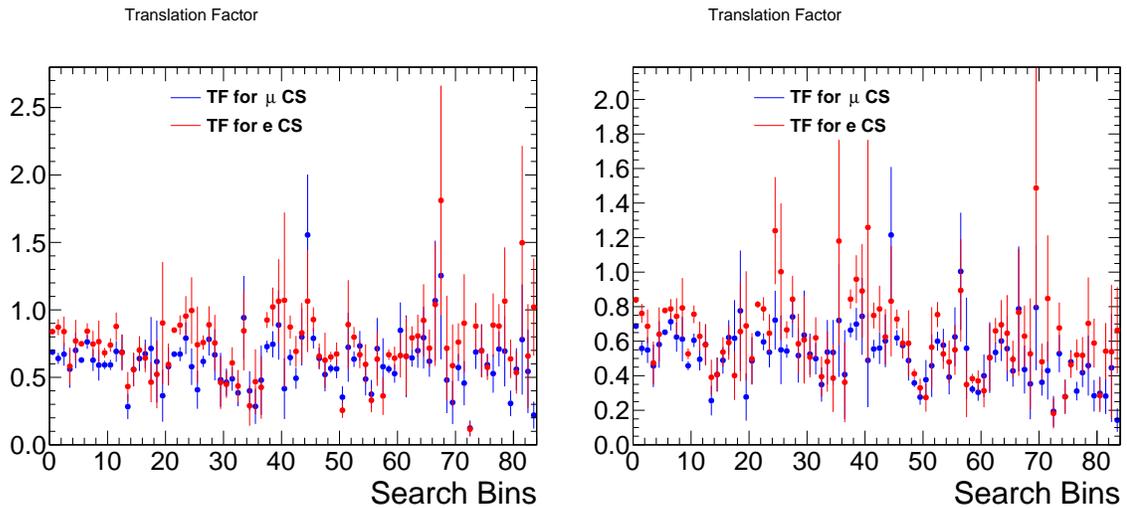


Figure 6.9: Translation factors for the  $\tau_h$  (left) and lost lepton (right) background prediction with their uncertainties from both muon and electron CS.

### Systematic Uncertainty:

The major source of systematic uncertainty of this method is the statistical error on the translation factor. A translation factor comprehends all the cuts, correction factors, and selection efficiencies. So the uncertainties on them result in the systematic uncertainties. The uncertainty on lepton efficiency affects both lepton selection and lepton veto. Change in isotrack veto efficiency varies the number of hadronic tau events. Jet energy scale uncertainty and btag SF uncertainty affect the jet and b-jet selection respectively. As the prediction is obtained by multiplying translation factor with the data CS, the change in prediction can be obtained by estimating the change in translation factor. Each factor folded into the translation factor is varied by its up and down uncertainties to determine the change in ratio. The different sources of uncertainties and their effect on the final prediction are summarized in Table 6.1

Table 6.1: Contributions from different sources of systematic uncertainty to the  $\tau_h$  and lost lepton background prediction.

Process	Source	Effect on $\tau_h$ Prediction in %	Effect on lost lepton Prediction in %
Translation factor statistical error	Statistics of MC SR and CS events	1 to 50	2 to 51
Lepton efficiency SF (including isolated tracks)	Data-MC correction from tag and probe method and studies	5 to 52	7 to 46
B-tag SF	Uncertainty on b-tag SF	0 to 1	0 to 2
$p_T^{miss}$ magnitude and $\phi$	Uncertainty related to $p_T^{miss}$ magnitude and $\phi$	0 to 54	0 to 40
JEC	Jet energy correction uncertainty	0 to 52	0 to 56
ISR	Variation of ISR weight	0 to 11	0 to 13
PDF	PDF uncertainty	0 to 31	0 to 32
Trigger	Uncertainty on trigger efficiency	0 to 1	0 to 1

### Validation of traslation factor method:

The method is validated in an orthogonal region selected by inverting  $N_t$  requirement, i.e.,  $N_t = 0$ . This data sideband is enriched with  $t\bar{t}$ , W+jets, and single-top events. In order to suppress the contributions from QCD and  $Z \rightarrow \nu\bar{\nu}$ , tighter requirements on  $N_b$  and  $\Delta\phi$ ,  $N_b \geq 2$  and  $\Delta\phi p_T^{miss}, j_{1,2,3,4} > 0.5$  are applied. The same TF method is employed to predict the  $t\bar{t}$ , W+jets, and single-top backgrounds. The estimated background is compared to the observed data in the sideband region which is divided

into  $p_T^{miss}$  intervals. The other backgrounds are taken from simulation. Figure 6.10 the comparison between data and total background in which  $t\bar{t}$ , W+jets, and single-top events are determined using TF method with single muon and electron CS separately.

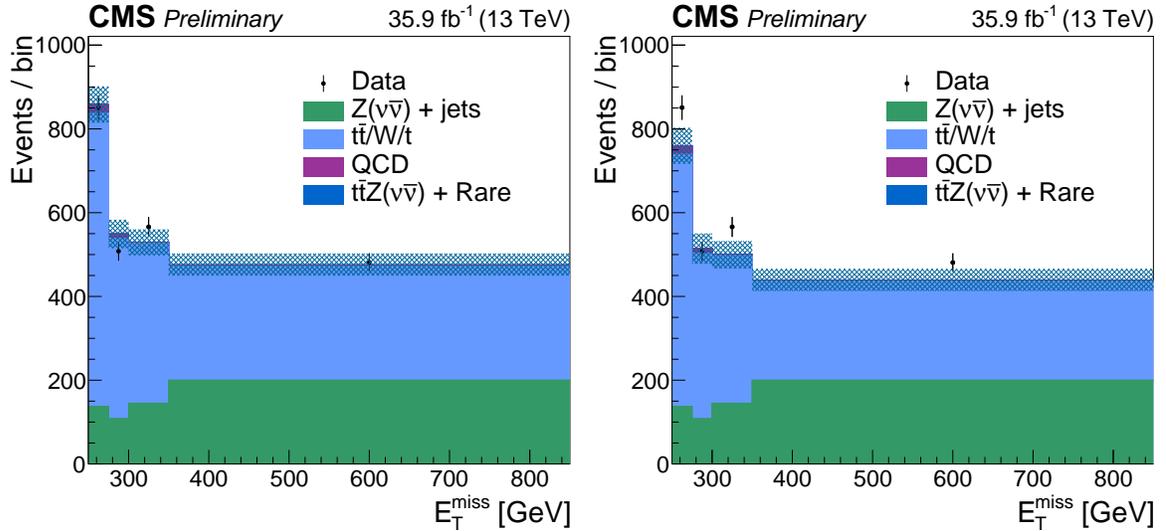


Figure 6.10: Validation of TF method in the data sideband using the muon channel (left) and electron channel (right). The black points are observed data. The light blue is predicted  $t\bar{t}$ , W+jets and single-top events using TF method. All other backgrounds directly come from MC yields. The errors include only statistical uncertainties.

### Hadronic tau and lost lepton background prediction:

The  $\tau_h$  and lost lepton background predictions in search region are obtained by applying the measured TF (as discussed in 6.6.1.3) on single lepton data control sample. Since there are two data CS, the final estimation is made by averaging the predictions from muon and electron CS. Figure 6.11 and figure 6.12 show the  $\tau_h$  and lost lepton predictions respectively in all search bins. The error bars in the figures include both statistic and total systematic uncertainties.

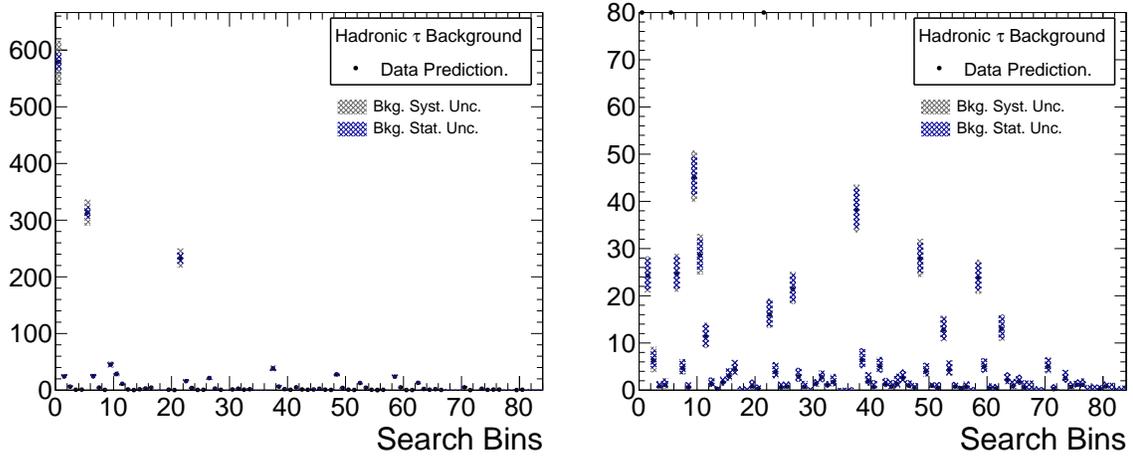


Figure 6.11: Predicted hadronic tau background yield for  $35.9 \text{ fb}^{-1}$  data for all the search bins. Right plot is a zoomed version of left plot. Both statistical and total systematic uncertainties are shown.

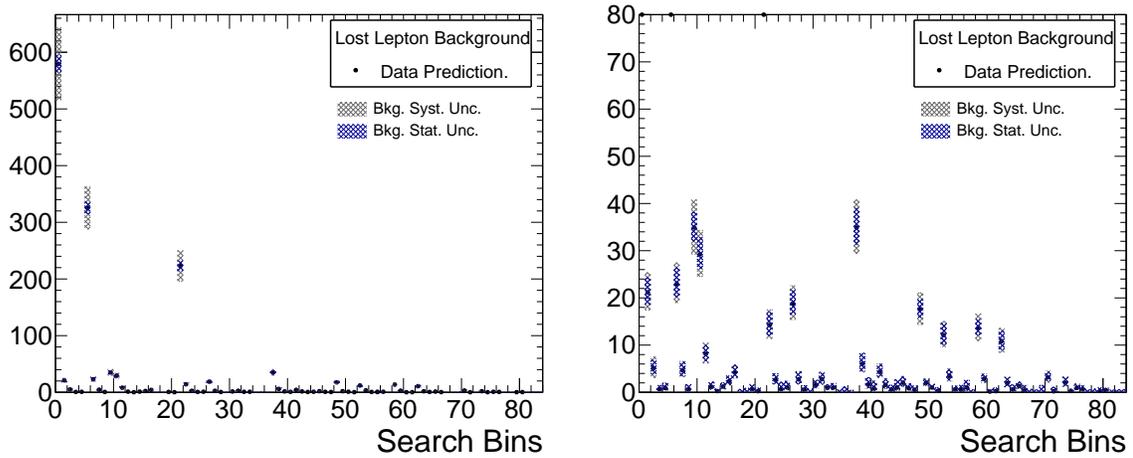


Figure 6.12: Predicted lost lepton background yield for  $35.9 \text{ fb}^{-1}$  data for all the search regions. Right plot is a zoomed version of left plot. Both statistical and total systematic uncertainties are shown.

# Results

The background yields from  $Z_{\nu\bar{\nu}} + \text{jets}$ , QCD,  $t\bar{t}Z$ , and other rare processes are estimated with corresponding  $35.9 \text{ fb}^{-1}$  luminosity following the same procedure as used in 2015 analysis. The observed data events and the predicted events from SM backgrounds in all 84 search bins are shown in figure 6.13.

The major background contribution in the search regions typically comes from  $t\bar{t}$  and

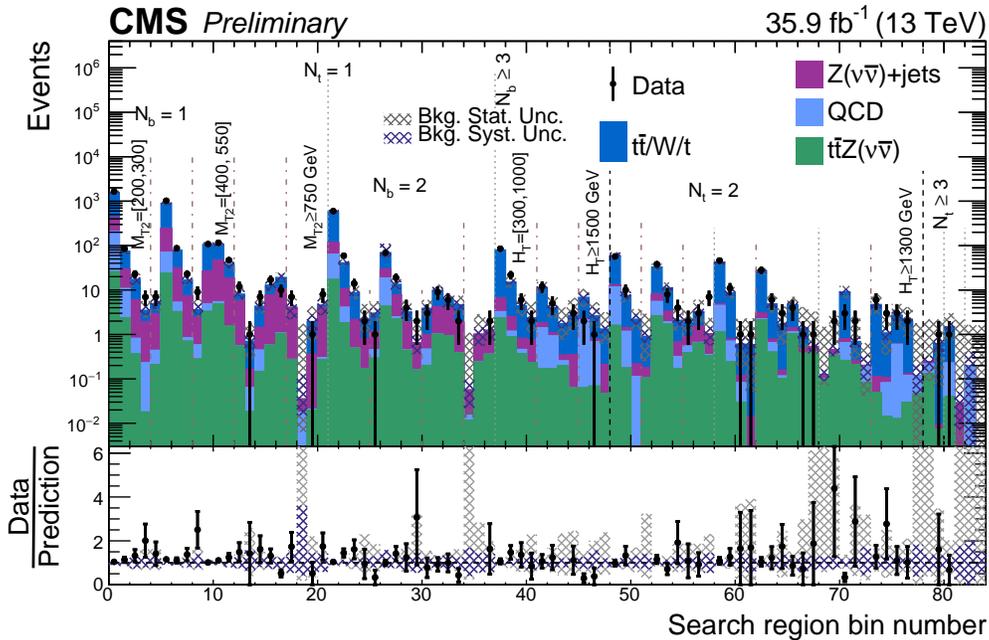


Figure 6.13: Data are shown as black points. The total predictions are shown in filled solid area. The bottom plot shows the ratio of data over total background prediction in each search bin. The hatched bands show uncertainties of total predictions with blue for systematic uncertainty and grey for statistical uncertainty. Within uncertainties, a good agreement between prediction and observation is found across all SBs

$W + \text{jets}$  processes followed by  $Z_{\nu\bar{\nu}} + \text{jets}$ . In very high  $p_T^{\text{miss}}$  search bins,  $Z_{\nu\bar{\nu}} + \text{jets}$  can be dominant background. The QCD multi jets,  $t\bar{t}Z$ , and rare SM backgrounds are subdominants across all search bins. No significant deviation between the SM

prediction and observed data is found.

The result is interpreted with the exclusion limit at 95% CL for the signal models sensitive to the analysis. Figure 6.14 shows the exclusion plots for T2tt and figure 6.15 present the same for T1tttt, T5tttt, T1ttbb and T5ttcc models.

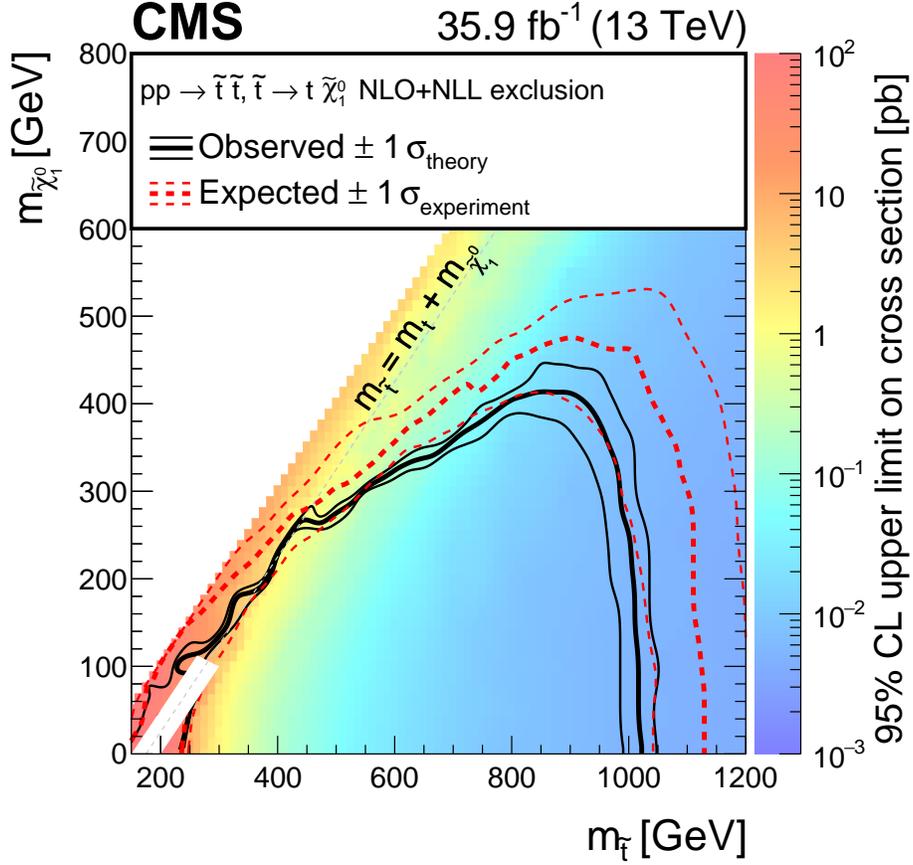


Figure 6.14: 95% CL exclusion limit for simplified signal model T2tt. The solid black curves indicate the observed exclusion contour and corresponding uncertainties with  $\pm 1$  standard deviation whereas dashes red curves represent similar expected exclusion contours.

Like 2015 analysis (section 5.7), no interpretation is given for the specific region of T2tt spectrum. For the very low mass of  $\tilde{\chi}_1^0$  in the case of T5tttt model which undergoes  $\tilde{t} \rightarrow t\tilde{\chi}_1^0$  decay, events tend to have smaller  $p_T^{miss}$  value and contaminate in  $t\tilde{t}$

background prediction. The larger signal contamination makes statistical treatment unreliable. That's why interpretation is not given if  $m_{\tilde{\chi}_1^0} < 50$  GeV for T5tttt. The analysis excludes top squark mass up to 1020 GeV and LSP mass up to 430 GeV for T2tt model. In the case of T1tttt, gluino mass is excluded up to 2040 GeV and LSP mass is up to 1150 GeV, corresponding exclusion limits for T5tttt are 2020 and 1150 GeV, 2020 and 1150 GeV for T1ttbb and 1810 and 1100 GeV for the T5ttcc signal model.

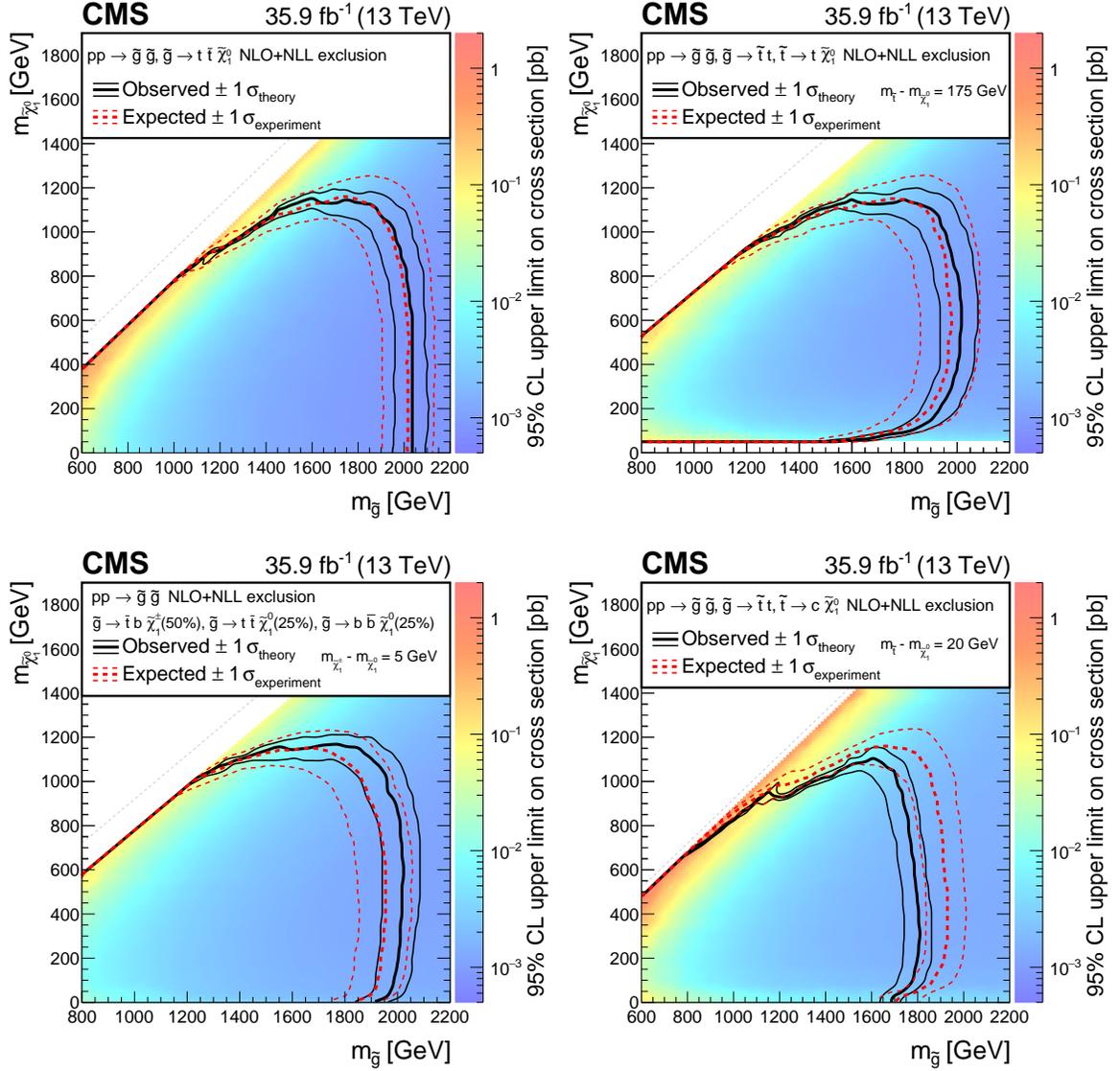


Figure 6.15: 95% CL exclusion limit for simplified signal models T1tttt (top left) and T5tttt (top right), T1ttbb (bottom left) and T5ttcc (bottom right). The solid black curves indicate the observed exclusion contour and corresponding uncertainties with  $\pm 1$  standard deviation whereas dashed red curves represent similar expected exclusion contours.

# Bibliography

- [1] CMS Collaboration, CMS Physics Analysis Summary CMS-PAS-JME-15-002 (2016).
- [2] M. Dasgupta, A. Fregoso *et al.*, JHEP 09 029 (2013).
- [3] A. J. Larkoski, S. Marzani *et al.*, JHEP 05 146 (2014).
- [4] J. Thaler and K. Van Tilburg, HEP 03 015 (2011).
- [5] CMS Collaboration, CMS Physics Analysis Summary CMS-PAS-JME-16-003 (2017).
- [6] T. K. Ho, Proceedings of 3rd International Conference on Document Analysis and Recognition 1 278 (1995).
- [7] CMS Collaboration, CMS Physics Analysis Summary CMS-PAS-JME-13-002 (2013).

# Chapter 7

## Conclusions

This dissertation presents the search for scalar top quark in all hadronic final state using the 13 TeV proton proton (pp) collision data collected by the CMS detector at LHC, CERN during the year of 2015 and 2016. The search analyses not only explore the possibility of scalar top or stop production in pp collision, but also sheds light on the creation of gluino from the collision. Following the theoretical motivation and experimental tools, stop and gluino searches are described in two separate analyses corresponding to 2015 data of  $2.3 \text{ fb}^{-1}$  integrated luminosity and 2016 data of  $35.9 \text{ fb}^{-1}$  integrated luminosity respectively. The data are analyzed to look for stop and gluino in the fully hadronic events distributed over search bins defined by search variables and optimized to have the maximal sensitivity to various stop and gluino signal models considering the contribution from Standard Model (SM) background processes. A novel top quark tagging algorithm has been developed in order to identify the top quarks in the event. The use of reconstructed top quark and  $M_{T2}$  which is derived from top kinematic makes the analysis distinct among other CMS hadronic SUSY analyses. The top tagger algorithm has been revised and improved in 2016 compared to that used in 2015. SM background events are estimated utilizing either

data driven method or data validated MC simulation.

No excess of events beyond the predicted SM backgrounds is observed in the data. The results are interpreted in light of the models within the Simplified Model Spectra as setting the upper limit on stop and gluino production cross section as the function of their masses and of the mass of LSP. The mass of scalar top quark is excluded up to 740 GeV in 2015 and up to 1020 GeV in 2016 analysis. The corresponding exclusion on the gluino mass is up to 1550 GeV as obtained in 2015 and up to 1810-2040 GeV, according to the models, in 2016 search. The analysis result of 2015 significantly extends the limits of previous searches and 2016 analysis further amplifies the exclusion limit.

The latest results already set a quite high mass limit for stop and gluino production. However, the mass limits are still within the natural SUSY bound. Moreover, the compressed region of mass phase-space where the masses of stop and neutralino are nearly degenerate is not probed within our analyses, leaves the possibility of finding them in the low mass region by different searches. The searches in coming years with the data currently being collected at LHC will possibly exclude or verify the existence of stop or gluino in TeV scale predicted by the natural SUSY theory. In a final note, I like to mention that the LHC will deliver more high luminosity data with the scope of drawing a crucial conclusion about the naturalness of MSSM theory and the opportunity of exploring various aspects of BSM physics.

# Appendix A

## Appendices

### MVA Based Resolved Top tagger

Among the myriad of MVA algorithms available, the final algorithm selected was a tree based approach given their simplicity of use, fast training and evaluation time, and robustness in classification problems. Out of a number of decision tree algorithms, the most robust against overtraining without losing discrimination power are ensemble algorithms which use a large number of decision trees to create a single discriminator. To this end, different algorithms were tested including gradient boost, adaboost, random forest, extreme gradient boost, and basic decision trees. Among these, gradient boost and random forest proved to be the strongest options giving an equivalent performance. The comparison between different algorithms is discussed in section A.1.2. The procedure starts with the training to discriminate between the trijet combinations come from top decay and the random background combinations. The training makes use of the various features or variables of the trijet combinations in order to develop the discriminator. A cut is placed on the discriminator to select top-like combinations. A specific threshold is chosen such that the top quark mis-

reconstruction rate gets significantly reduced keeping the tagging efficiency similar as compared to the previous cut-and-count algorithm. The training, parameter optimization, and the performance of the MVA algorithm are discussed in succeeding sections.

## MVA training

The training is performed with the random forest decision tree algorithm. A random forest decision tree is an ensemble of decision trees which are each independently trained on a different subset of the available training data. Additionally, at each branch point in the individual decision trees, the algorithm only considers a subset of the total number of sample features (input variables) when deciding which cut is best. The output of the random forest is the mode of the individual trees for a classifier or the mean of the individual trees for a regression. In order to study the effects of training on the MVA performance and to tune the parameters, the python based Scikit-learn package [1] was used. For the final integration into the analysis framework, another C++ based package, OpenCV 3.1.0 [2], was selected. The training data is prepared using a mixture of signal and background from 100000 single-lepton  $t\bar{t}$  events and 70000  $Z \rightarrow \nu\nu$  events (produced with MADGRAPH). The  $t\bar{t}$  and  $Z \rightarrow \nu\nu$  are then combined into the final training sample with equal total integrated weight. For the purposes of defining the training signal, a top candidate which could be tagged as a top quark is defined by a trijet combination if it satisfies the conditions mentioned in section 6.2.2. The top candidates are then separated into signal and background categories by matching them to the generator-level quarks from top decays. This matching requires that each of the three jets in the top candidate is matched to a generator-level quark originating from a single top quark within a cone of  $\Delta R < 0.4$ . Additionally, the total Lorentz vector of the top candidate must be matched to the

generator-level top quark within a cone of  $\Delta R < 0.6$ . Each individual top candidate is then weighted appropriately to flatten the  $p_T$  spectrum independently for signal and background for each MC sample in order to ensure low- $p_T$  events do not dominate the training. Finally, the weights of all background events are adjusted so that the integrated weights of signal and background match.

An extensive study has been done to get the optimal performance of the MVA tagger. The Receiver Operating Characteristic (ROC) curve is used to gauge the tagger performance. In the context of the MVA study, the ROC curve is defined in two ways, event-wise and object-wise. The event-wise ROC is the ratio of the efficiency over the fake rate whereas the object-wise ROC is defined as the ratio of the true positive rate (TPR) over the false positive rate (FPR). Efficiency and fake rate are already defined in section 5.2 and other two quantities are defined as follows:

$$\begin{aligned} \text{TPR} &= \frac{\text{Number of tagged top candidates matched to a generator-level top quark}}{\text{Number of top candidates matched to a generator-level top quark}}, \\ \text{FPR} &= \frac{\text{Number of tagged top candidates not matched to a generator-level top quark}}{\text{Number of top candidates not matched to a generator-level top quark}}, \end{aligned} \tag{A.1}$$

Here, 'tagged top' indicates the trijet combination (top candidate) with MVA discriminator value above a given threshold.

For final training, the number of trees and the maximum tree depth are tuned. The optimized list of specific features (input variables) used in the random forest training is mentioned in section 6.2.2. One of the powerful set of features is the selection of trijet rest frame variables in place of the similar lab frame quantities. The effect of using rest frame variables versus lab frame kinematic variables is shown clearly in the improvement to the ROC curve seen in Fig. A.1.

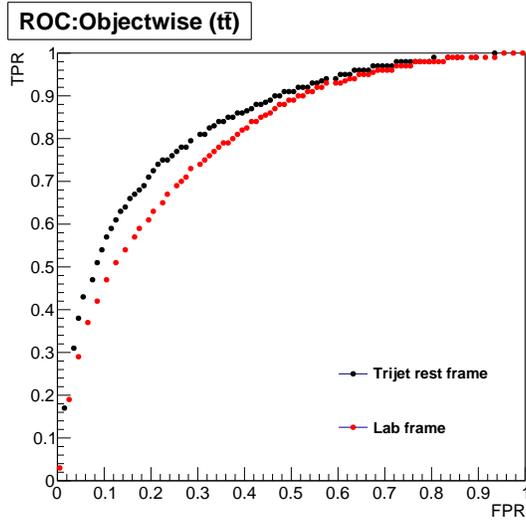


Figure A.1: The ROC curves for random forest with depth 14 trained using lab frame and rest frame variables.

## MVA optimisation and tuning study

To achieve the best performance of the tagger, a number of optimization studies based on the performance from the ROC curve are done.

- Depth optimization in training: the choice of maximum depth effects discriminator distribution. As the tagged top selection depends on the shape of the discriminator, it is important to set a proper depth. We vary the number of depths in the random forest training and check the impact on the MVA discriminator distribution and ROC performance for fixed input variables. Though the change in maximum depth has a small effect on ROC performance, training with a number of depths close to the number of input variables shows better performance compared to very high or very low number of depths.
- Input variables finalization: training with different combinations of input variables is done in order to obtain the minimum number of effective variables. Adding a lot of variables would increase the risk of bias. So we start with

the variables used in cut and count approach(Sec. 5.2) i.e., the mass of trijet candidate and the masses dijet combinations from that candidate. We also include different properties of jets used to form a top candidate. Removing spin correlation variables and jet charge,  $\eta$  does not affect the tagger performance whereas removing the b-tag CSV and quark-gluon discriminator values gives worse performance in the ROC curve.

- Comparison between Scikit-learn and openCV: it is also checked that the random forest gives similar ROC curves obtained using Scikit-learn and openCV package which in turn enable us to use the training result obtained from Scikit-learn in openCV integrated into the analysis framework to evaluate the resolved part of the top tagger.

Two tagged tops should not share a jet in their constituents. So the overlap between tagged candidates is resolved favoring the candidate having the highest MVA discriminator value. This is based on a study that showed that the purity increases with the increase in MVA discriminator value. We define purity as the ratio of the number of tagged tops (or reconstructed tops) matched to generator level top to the total number of tagged tops.

## **MVA performance**

The random forest algorithm is a strong improvement over the existing cut-based algorithm. The exact performance of the random forest algorithm is dependent on the discriminator cut chosen. The discriminator for the final training is shown in Fig. A.2. Here, the discriminator is not shown for all top candidates but instead is shown only for tops that do not share a jet with another top candidate. Though the final discriminator threshold is chosen as 0.85 based on signal sensitivity study, the

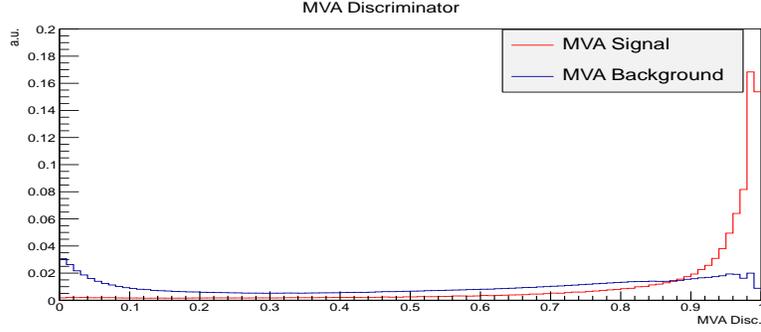


Figure A.2: The discriminator for signal and background from the final trained random forest separated between signal- and background-like candidates.

results from MVA tagger are shown in term of tagging efficiency and fake rate for the discriminator value of 0.7.

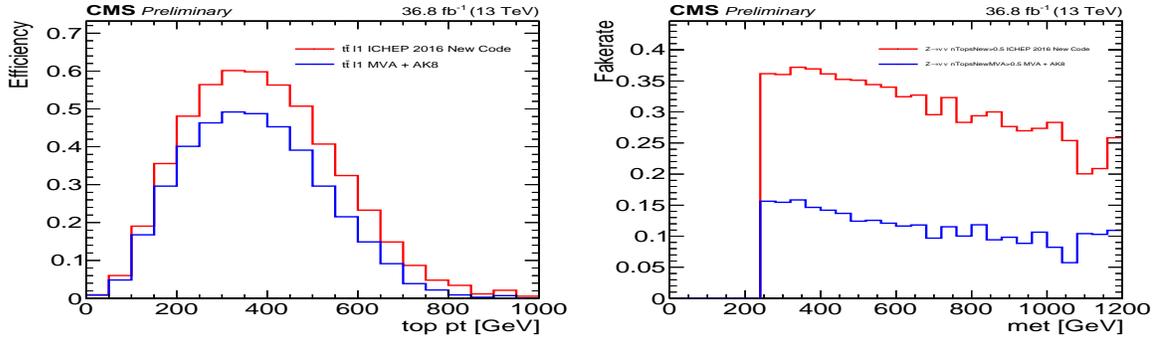


Figure A.3: The efficiency (left) and fake rate (right) of the resolved top tagging algorithm in term of generated top quark  $p_T$  and event  $p_T^{miss}$  respectively. Blue lines represent the MVA algorithm and the red one indicates original cut-based algorithm. The efficiency and fake rate are measured according to the equation 5.3 in simulated semileptonic  $t\bar{t}$  and Z+jets events respectively. The legend 'ICHEP 2016 new code' means cut-based algorithm runs on 2016 ICHEP samples and 'MVA + AK8' indicates the new algorithm which is combined with AK8 tagger but the results presented here are from MVA tagger alone and with discriminator threshold of 0.7.

# Bibliography

- [1] F. Pedregosa *et al.*, Journal of Machine Learning Research 12 2825 (2011).
- [2] Itseez, “Open Source Computer Vision Library” (2015).