



BROWN

Brown University

The Department of Physics

SEARCH FOR PAIR PRODUCTION OF VECTOR-LIKE QUARKS
IN LEPTONIC FINAL STATES IN PROTON-PROTON COLLISIONS
AT 13 TeV AT THE CMS DETECTOR IN THE LHC

*A dissertation submitted to
the Department of Physics and Graduate School of Brown University
in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy in Physics*

by

Jess Wing Yan Wong

May 2022



© Copyright 2022 by Jess Wing Yan Wong

This dissertation by Wing Yan Wong is accepted in its present form
by the Department of Physics as satisfying the
dissertation requirement for the degree of Doctor of Philosophy

Date _____
Professor Meenakshi Narain, Advisor

Recommended to the Graduate Council

Date _____
Professor JiJi Fan, Reader

Date _____
Professor Jonathan Pober, Reader

Approved by the Graduate Council

Date _____
Professor Andrew G. Campbell,
Dean of the Graduate School

Wing Yan (Jess), Wong

Curriculum Vitae

Department of Physics
Barus and Holley Building
Brown University
182 Hope Street
Providence, RI 02912

wing_yan_wong@brown.edu
wwyjess@gmail.com

Education

- **Ph.D. Physics**, Brown University, August 2022
Advisor: Prof. Meenakshi Narain
- **M.Sc. Physics**, Brown University, 2019
- **B.Sc. Physics and Astronomy**, University of Hong Kong, 2017
Advisor: Dr. Yanjun Tu & Dr. Jeremy Jin Leong Lim

Research Experience

Search for Vector-Like Quarks in CMS, Jun 2019–present

- A search for pair production of heavy vector-like T or B quarks in proton-proton collisions at $\sqrt{s} = 13$ TeV was done using the full Run 2 data collected with the CMS detector at the CERN LHC from 2016 to 2018. Vector-like T quarks are predicted to decay to bW , tZ , and tH whereas vector-like B quarks are predicted to decay to tW , bZ , and bH . Pair production of vector-like T or B quarks would result in a wide range of final states. This project searches in 2017–2018 data in a multi-lepton channel, corresponding to final states with at least three leptons, and a same-sign dilepton channel, with exactly two leptons of the same electric charge in the final state. Both final states are rare for SM background processes. Full Run 2 result is obtained by combining this search with 2016 data reported in a published analysis and a single-lepton channel. Scanning over different branching ratios, the 95% CL upper limits on the production cross sections exclude T mass up to 1.54 TeV and B mass up to 1.56 TeV. The exclusion limit on T mass is the strongest to date in CERN.

Study of Muon ID Efficiencies dependence on Event Topology, Mar 2020–present

- The efficiencies of a muon passing a certain ID vary with the energy (p_T) and direction (η) of its motion. In addition, efficiencies may depend on the event topologies. A study was done to compare ID efficiencies in Monte Carlo (MC) simulations of Drell-Yan (DY) processes (for the conventional tag-and-probe measurement) and $t\bar{t}$ processes (common background in many analyses). The rates of a MC-truth muon passing a cut-based ID at tight working point are measured in the two MC simulations and compared. Developments of packages for data preparation and analysis are involved. The result shows that the efficiencies drop drastically beyond p_T of 30 GeV, from 0.96 down to 0.88, in $t\bar{t}$ events,

while efficiencies distribute like a gentle hill between 0.95 to 0.97 in DY events. Further study is proposed to understand the differences, and this may be used to optimize scale factor measurements in the future.

Development of ID Scale Factors for Electrons and Muons, Aug 2019–Sep 2020

- The efficiencies of a lepton passing a certain ID are different in data and in MC simulations. To account for this effect, the scale factors, which are the ratios of efficiency in data to that in MC, are calculated and applied to scale MC yields. Efficiencies are measured by the tag-and-probe method. In Drell-Yan events with opposite-signed same-flavour dilepton in the final state, the dilepton invariant mass will be close to Z mass. Thus, one of the two leptons can be selected as a tag with some tight selections. With the tag selected, its corresponding partner, the probe, is found by criteria such as same lepton flavour, opposite electric charge, and invariant mass near Z . With no ID or kinematic selections applied, the collection of probes is unbiased and can be used to count the rate of passing an ID. Lepton scale factors were measured for reconstructed leptons passing IDs, which are used officially in CMS. Scale factors were also measured for leptons of some IDs passing mini-Isolation working points, which are used in analyses.

Interference Study for a Heavy Higgs Search at ATLAS (Senior Project), Jun 2016–May 2017

- The HKU ATLAS group searched for a heavy Higgs boson in 13 TeV proton-proton ($p-p$) collisions at ATLAS in $t\bar{t}b\bar{b}$ final states. An extended Higgs sector is introduced in the Two Higgs Doublet Model (2-HDM). This project was composed of two parts: I) signal modelling for $t\bar{t}H, H \rightarrow b\bar{b}$, II) interference study on $t\bar{t}b\bar{b}$ final states. Both parts made use of Monte Carlo event generation to simulate $p-p$ collisions and produced samples of thousands of events. In I) signal modelling, the distributions of transverse momentum of the neutral heavy Higgs boson and collision energy looked different at $m_H = 150$ or 200 GeV. The distributions were almost the same for higher mass points, from 300 to 1000 GeV. More studies will be needed to explain the observation. In II) interference study, the distributions of the invariant mass of decay products from heavy Higgs bosons were well explained by the theoretical cross-section times branching ratio. The only abnormality was that the number of events with charged Higgs production at $m_H = 400$ GeV was almost twice that at $m_H = 500, 700$ GeV. Investigations to understand the relation between the value of cross-section $\times$ branching ratio and the width of heavy Higgs boson mass were proposed at the end of the project.

Strong Gravitational Lensing of RXJ2248 (Undergraduate RA), Dec 2014–May 2016

- A mass model for CLASH cluster RXJ2248 was built using strong gravitational lensing analysis. The mass model that best fitted the lensing and delensing of background galaxies was found by WSLAP (a IDL+Fortran+Python-based package) using data images from MUSE and Hubble Frontier Field. The predicted semi-parametric mass model was tested for its accuracy by comparing predictions of magnitude magnification, relensed positions, and source redshifts to measurement. The work was published in 2016.

Teaching Experience

- Teaching Assistant**, Brown University

Course	Instructor	Term
Foundations of Electromagnetism and Modern Physics (PHYS0060) Introduction to Relativity, Waves and Quantum Physics (PHYS0160)	Prof. Meenakshi Narain	2019 Spring
Foundations of Mechanics (PHYS0050) Analytical Mechanics (PHYS0070)	Prof. Greg Landsberg	2018 Fall
Astronomy (PHYS0220)	Prof. Jonathan Pober	2018 Spring
Astronomy and Astrophysics (PHYS270)	Prof. David Cutts	2017 Fall

Awards & Fellowships

- Physics Merit Fellowship 2021–2022
- Mr. Armin and Mrs. Lillian Kitchell (Enrichment) Scholarships 2016–2017
- 3rd Best Presenter in Research Colloquium for Science Undergraduate Students 2016–2017
- Overseas Research Fellowship (heavy Higgs search at ATLAS in CERN) Jun–Aug 2016
- Best Poster Presenters Nov 2017
- HKU Foundation Scholarships for Outstanding Students 2013–2014
- Science Entrance Scholarship 2013–2014
- Dr. P.M. Hui Memorial Scholarship 2013–2014

List of Publications

- Chan, Brian MY, Tom Broadhurst, Jeremy Lim, Jess Wong, Jose M. Diego, and Dan Coe. “Redshift Determinations from a Self-consistent Grid-based Lens Model for the Hubble Frontiers Field Cluster RXC J2248. 7- 4431 (AS1063).” The Astrophysical Journal 888, no. 1 (2020): 35.
- Diego, Jose M., Tom Broadhurst, Jess Wong, Joseph Silk, Jeremy Lim, Wei Zheng, Daniel Lam, and Holland Ford. “A free-form mass model of the Hubble Frontier Fields cluster AS1063 (RXC J2248. 7-4431) with over one hundred constraints.” Monthly Notices of the Royal Astronomical Society 459, no. 4 (2016): 3447-3459.

Talks & Posters

- **APS APRIL MEETING, AMERICAN PHYSICAL SOCIETY 04/18/2021**

Search for Vector-Like-Quarks T/B in Same-Sign Dilepton and Multi-lepton Final States at $\sqrt{s} = 13$ TeV, Presented, U. Heintz, J. Hogan, M. Narain, S. Sagir, R. Syarif, E. Usai, W.Y. Wong, CMS collaboration

- **GORDON RESEARCH SEMINAR/CONFERENCE ON PARTICLE PHYSICS, THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY 06/29/2019**

Search for Vector-Like-Quarks T/B in Leptonic Final States at $\sqrt{s} = 13$ TeV, Presented, U. Heintz, J. Hogan, M. Narain, Y. Duh, S. Sagir, R. Syarif, E. Usai, W.Y. Wong, CMS collaboration

- **GORDON RESEARCH SEMINAR/CONFERENCE ON PARTICLE PHYSICS, THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY 06/24/2017**

Interference Study on $t\bar{t}b\bar{b}$ Final State by Monte Carlo Event Generation for a Heavy Higgs Search at the ATLAS Experiment at the LHC, Presented, Nicola Orlando, Yanjun Tu, Wing Yan Wong, ATLAS collaboration.

Skills & Workshops

- **Computing:** C++, Python, CMS software/ATLAS software/ROOT, PySpark dataframe, IDL, and some knowledge on Fortran. Also trained for advanced use of MS Word, Excel, PowerPoint, etc. Some workshops attended are:
 - **Machine Learning Graduate Course**, LPC at Fermilab, Spring 2022. A remote-only class taught by Prof. Sergei Gleyzer of the University of Alabama (UA).
 - **CMS Data Analysis School**, LPC at Fermilab, 01/13/2020–01/17/2020. A one-week intensive program with hands-on exercises led by experts in CMS.
 - **Hands-on Advanced Tutorial Sessions at the LPC (HATS@LPC)**, LPC at Fermilab, 06/03/2019–06/04/2019. Hands-on tutorials on handling jets at CMS.
 - **C++ Course**, LPC at Fermilab, 05/20/2019–05/24/2019. A C++ course taught by Prof. Glenn Downing of UT Austin.
- **Statistics:** Handling of errors, data presentation, and some experience with machine learning (classification and deep neural network).
- **Language:** Proficiency in English and Chinese; Fluency in spoken English, Cantonese, and Putonghua.
- **Communication/Collaboration:** effective communication

Acknowledgment

The past five years, in pursuit of my PhD in Physics, have been uneasy, but rewarding. Here, I want to thank my parents, Tak Sang Wong and Pui Fong Leung, and my sister, Ho Yan Wong, for unconditionally supporting my decisions and for their love and care. I would like to express my appreciation to Alfred Chi Yung Ho, for his company and cheers in part of the journey. I am also grateful for the encouragement I received from my friends and physics buddies from Hong Kong. You gave me the strength to carry on in the face of hardships and doubts.

I want to take this opportunity to express my gratitude to Antimo Russo and the late Evelyn Yip. I am blessed to have known you as my tango teachers, my life mentors, and my beloved brother and sister. The day we met was a turning point in my life, and I shall forever be thankful.

I owe a very important debt of gratitude to Gabriel Gallardo for his company in my darkest days, for introducing me to the field of high energy experiment, for his advice, and more. I am more than lucky to have you as my friend.

Furthermore, I would like to offer my special thanks to my good friends and housemates, Yangrui Hu and Chenyu Zhang, for their company and support and for lightening up my days. It has been a fun and memorable time.

I must thank my undergraduate advisors, Dr. Jeremy Lim and Dr. Yanjun Tu, for the opportunities and guidance they have offered, which allowed me to embark on this overseas PhD journey. My heartfelt gratitude goes to my PhD supervisor, Prof. Meenakshi Narain, who has been supportive and enlightening. It is my honour to be a member of your research group. Thank you to the readers, Prof. JiJi Fan and Prof. Johnathan Pober, for your time and patience. I am deeply grateful to our collaborator, Dr. Julie Hogan, who has been encouraging and helpful.

I want to thank everyone in the High Energy Experiment group at Brown for the friendly and cooperative environment: Atheer Alzahrani, Gabriele Benelli, Bjorn Burkle, Xavier Coubez, Prof. David Cutts, Mary Hadley, Prof. Ulrich Heintz, Nicholas Hinton, Andrew Kent, Andrei Korotkov, Martin Kwok, Taeun Kwon, Ted Laird, Prof. Greg Landsberg, Otto Lau, Jangbae Lee, Daniel Li, Michael Lukasik, Jingyu Luo, Prof. Meenakshi Narain, Jessica Nelson, Nikolas Pervan, Trevor Russell, Jiwoo Seo, Xiaohe Shen, Farrah Simpson, Eric Spencer, Rizki Syarif, Lindsey Tensen, Emanuele Usai, Xuli Yan, David Yu, and Wenyu Zhang. My thanks also go to the conveners, colleagues, and staff in CERN and Fermilab for the successful collaborations.

To close, I want to thank all faculties, administrative staff, and fellow students in the Department of Physics at Brown University. It was my pleasure to be a part of this big family. I appreciate all the help and kindness I have received from you.

Contents

Curriculum Vitae	iv
Acknowledgments	viii
1 Introduction	1
2 The Standard Model	4
2.1 The Elementary Particles	5
2.2 The Quantum Field Theory	7
2.2.1 Standard Model without Higgs	7
2.2.2 Higgs Mechanism and Electroweak Symmetry Breaking	10
2.2.3 Leptons and Yukawa Coupling	13
2.2.4 Quarks and Quantum Chromodynamics	14
3 Vector-Like Quarks	16
3.1 SM Extension	16
3.2 Motivations for Searches	19
4 The CMS Experiment at the LHC	22

4.1	The Large Hadron Collider (LHC)	22
4.2	The Compact Muon Solenoid (CMS) Detector	27
4.2.1	Coordinate System	28
4.2.2	The Inner Tracker	29
4.2.3	The Electromagnetic Calorimeter (ECAL)	32
4.2.4	The Hadron Calorimeter (HCAL)	34
4.2.5	The Muon Chambers	37
4.3	Triggering System	41
4.4	Physics Objects Reconstruction	42
4.4.1	Primary Vertex and Pileup	43
4.4.2	Tracks	44
4.4.3	Electrons and Photons	44
4.4.4	Muons	45
4.4.5	Jets and Missing Transverse Momentum	45
5	Search for Pair Production of T/B in	
	Leptonic Final States	46
5.1	Search Channels	47
5.1.1	Multi-Lepton (3L+) Final State	48
5.1.2	Same-sign Dilepton (SS2L) Final State	48
5.2	Data and Monte Carlo	48
5.3	Physics Object Selection	53

5.3.1	Leptons	54
5.3.2	Jets and Missing Transverse Momentum	60
5.4	Event Selection	61
5.4.1	Triggers for the Analysis	61
5.4.2	Base Offline Event Selection	62
5.4.3	Control Region and Signal Region	64
5.5	Background Estimation	66
5.5.1	Corrections on MC	66
5.5.2	Matrix Method for Non-prompt Background Estimation	67
5.5.3	Prompt Rates	69
5.5.4	Fake Rates	72
5.5.5	Fake Rates Cross Check	77
5.5.6	3L+ Channel: η -Dependence of Muon Fake Rate	79
5.5.7	SS2L Channel: Fake Rate Scan for Systematics	82
5.5.8	SS2L Channel: Background from Charge Misidentification	84
5.6	Nonprompt Background Closure Tests with $t\bar{t}$ MC	86
5.6.1	Closure tests in Multi-lepton Channel	86
5.6.2	3+L Channel: Cut-and-Count Fake Rate Cross Check	86
5.6.3	3+L Channel: CR2 to SR Fake Rate Compatibility	88
5.6.4	Closure tests in SS2L Channel	95
5.6.5	SS2L Channel: Cut-and-Count Fake Rate	95

5.6.6	SS2L Channel: Test for Matrix Method	96
5.7	Systematic Uncertainties	99
5.8	Kinematic Distributions	103
5.8.1	3+L Channel: Control Region CR2	103
5.8.2	SS2L Channel: Control Regions	105
5.8.3	3+L Channel: Signal Region	108
5.8.4	SS2L Channel: Signal Region	110
5.9	3+L Channel: Optimization and Final Selection	112
5.10	Limits	117
5.10.1	Limits in 3+L Channel	118
5.10.2	Limits in SS2L Channel	122
5.11	Statistical Treatment and Tests	125
5.11.1	Statistical Tests in 3+L Channel	126
5.11.2	3+L Channel: Nuisance pulls	126
5.11.3	3+L Channel: Impacts	126
5.11.4	3+L Channel: Bias Test	129
5.11.5	3+L Channel: Goodness of Fit	129
5.11.6	Statistical Tests in SS2L Channel	131
5.11.7	SS2L Channel: Nuisance pulls	131
5.11.8	SS2L Channel: Impacts	132
5.11.9	SS2L Channel: Bias Test	134

5.11.10 SS2L Channel: Goodness of Fit	134
5.12 Combined Result: 1 + 2 + 3 Leptons in full Run 2	140
5.12.1 Limits	140
6 Summary	144
Appendix A Photon and Muon Identification/Isolation Performances	147
A.1 Photons	147
A.2 Muons	148
A.3 Tag-and-Probe Method	149
A.4 Photon Loose ID	150
A.5 Muons in Different Event Topologies	153
A.5.1 Framework Validation and Development	154
A.5.2 MC Truth Efficiencies	155
References	159

List of Tables

3.1	A table comparing the representations of SM quarks and VLQs. Their quantum numbers, electric charges, and Lagrangian terms in unitary gauge are shown. The four VLQ fields, vertically aligned in the same rows, are arranged to form different multiplets.	17
5.1	2017 data sample names and the specific processing group for locating the files in the CMS database.	50
5.2	2018 data sample names and the specific processing group for locating the files in the CMS database.	51
5.3	2017 and 2018 background MC Samples from RunIIFall17MiniAODv2 and RunIIAutumn18MiniAOD campaigns. The third column lists the cross-section of each process [92] and the last two columns indicate whether the processes are used in the same-sign dilepton and/or the multi-lepton channel.	51
5.4	2017 and 2018 signal MC Samples from RunIIFall17MiniAODv2 and RunIIAutumn18MiniAOD campaigns respectively. The last column lists the cross-section of each process [105].	52
5.5	JEC/JER versions used in the analysis	60
5.6	HLT trigger paths considered in the multi-lepton and same-sign dilepton analysis. .	61

5.7	Unscaled event counts after pre-selections in four random subsets of 2018 MC samples. Trigger cut in this table accepts more HLT paths than the ones adopted by this analysis. The lower part of the table shows the event counts in the same subset which passed the analysis triggers in Table 5.6 according to the flavour categories. Signal samples listed here have equal branching fractions for each of the three decay modes.	63
5.8	Background yields of events in 2017 samples with two same-sign tight leptons and that with an additional cut of exactly two tight leptons. Uncertainties on nonprompt background yields are evaluated as shifted histograms, and thus no value is quoted.	63
5.9	Background yields of events in 2018 samples with two same-sign tight leptons and that with an additional cut of exactly two tight leptons. Uncertainties on nonprompt background yield are evaluated as shifted histograms, and thus no number is quoted in this table.	63
5.10	Table summarizing event selections in multi-lepton and same-sign dilepton channels.	65
5.11	Electron and muon prompt rates for different p_T ranges in 2017 data. The values in table are applied in the SS2L channel and, for electron only, the multi-lepton channel. An η -averaged muon prompt rate of 0.928 ± 0.0150 is applied in the multi-lepton channel instead.	70
5.12	Electron and muon prompt rates for different p_T ranges in 2018 data. The values in table are applied in the SS2L channel and, for electron only, the multi-lepton channel. An η -averaged muon prompt rate of 0.931 ± 0.01059 is applied in the multi-lepton channel instead.	71
5.13	$ \eta $ -dependent μ fake rates measured in 2017 CR2 by fitting distributions of leptons' p_T while fixing e fake rate at 0.1	79
5.14	$ \eta $ -dependent μ fake rates measured in 2018 CR2 by fitting distributions of leptons' p_T while fixing e fake rate at 0.08	79

5.15	Nonprompt background yields from 2017 samples in each flavour category predicted with $ \eta $ -dependent and flat μ fake rates. The percentage changes in yields are calculated.	82
5.16	Nonprompt background yields from 2018 samples in each flavour category predicted with $ \eta $ -dependent and flat μ fake rates. The percentage changes in yields are calculated.	82
5.17	A table showing the applied 3+L fake rates and the best-fitted fake rates from the scan in SS2L low H_T^{lep} CR for each lepton flavour in 2017 and 2018 respectively. The last column shows the shifts in fake rates for a one-sided systematic shape uncertainty.	84
5.18	CMID rates measured in 2017 and 2018 samples by minimizing the negative log likelihood.	86
5.19	Fake rates measured by cut-and-count in 2017 $t\bar{t}$ sample with three or more leptons.	87
5.20	Fake rates measured by cut-and-count in 2018 $t\bar{t}$ sample with three or more leptons.	87
5.21	Observed and predicted yields of $t\bar{t}$ MC events with two prompt and one nonprompt leptons. Predicted yield is obtained by applying cut-and-count fake rates in Table 5.19 and 5.20	94
5.22	Fake rates measured by cut-and-count in 2017 $t\bar{t}$ sample with two same-sign leptons for each origin of the nonprompt lepton.	96
5.23	Fake rates measured by cut-and-count in 2018 $t\bar{t}$ sample with two same-sign leptons for each origin of the nonprompt lepton.	96
5.24	Summary of systematic uncertainties grouped according to their correlations across channels. Ranges indicate representative values across different lepton flavour categories. Uncertainties that effect different background sources are labelled as “MC” for all simulations, “OS” for opposite-sign dilepton events used to derive CMID background, and “NP” for nonprompt background. Uncertainties on signal are labelled as “Accept.” when they are scaled for acceptance-only.	100

5.25	Systematic uncertainties on the data-driven nonprompt background from 2017 data in multi-lepton channel.	101
5.26	Systematic uncertainties on the data-driven nonprompt background from 2018 data in multi-lepton channel.	102
5.27	Systematic uncertainties on data-driven backgrounds in SS2L channel for 2017 and 2018 data.	102
5.28	Yields in distributions of leptons' p_T in the multi-lepton channel CR2 in 2017 and the associated rate uncertainties on NP background.	103
5.29	Yields in distributions of leptons' p_T in the multi-lepton channel CR2 in 2018 and the associated rate uncertainties on NP background.	103
5.30	Summary of final selections applied on top of SR (defined in Table 5.10) for finding limits and performing statistical tests in the following sections.	117
5.31	List of uncertainties' correlation between different years.	117
5.32	The exclusion limits on T mass measured in the multi-lepton channel for the singlet and doublet scenarios in the 2016 analysis [21], in 2017 and 2018, and in full Run 2.	120
5.33	The exclusion limits on T mass measured in the SS2L channel for the singlet and doublet scenarios in the 2016 analysis [21], in 2017 and 2018, and in full Run 2.	124
5.34	Summary of systematic uncertainty sources for the various analysis channels, grouped according to channel. The second column shows uncertainties for the single-lepton channel, evaluated with 2016–2018 data combined. The third and fourth columns show uncertainties for the other channels, evaluated using 2017 or 2018 data. Ranges indicate representative values across different lepton flavour categories. Uncertainties that affect different sources are indicated as “MC” for all simulation (including signal), “OS” for opposite-sign dilepton events used to derive CMID background, and “NP” for nonprompt lepton background.	141
5.35	Expected and observed lower limits on the T and B quark masses.	143

6.1	The exclusion limits on VLQ masses in pair production searches reported by the CMS and ATLAS collaborations over the years.	145
A.1	Requirements of cut-based photon ID at loose working point.	150

List of Figures

2.1	Particles described in the SM and their observed properties [29] with masses updated.	4
2.2	Schematic diagrams showing categorizations of the SM particle with two different sets of subordinate level fermion categories determined by electric charge (top) or by generations (bottom).	5
3.1	Feynman diagrams of $T\bar{T}$ and $B\bar{B}$ pair productions and their decay modes [21]. . . .	19
3.2	Feynman diagrams of $X_{5/3}\bar{X}_{5/3}$ pair productions and their decay [40].	19
3.3	Feynman diagrams of possible single production of a VLQ [35].	21
4.1	A schematic diagram of the CERN accelerator complex, indicating the machines' names, the year built they were built, and their sizes when applicable [46].	23
4.2	Cross section of the LHC dipole magnet [51].	24
4.3	Cross section of the LHC quadrupole magnet [52].	24
4.4	The current LHC layout and planned installations for future upgrades to HL-LHC [53].	25
4.5	A schematic diagram of the paths of proton beams in the LHC [54]. The stars indicate where proton-proton collisions occur as the two beam paths cross.	26
4.6	Plots showing peak instantaneous luminosities (top) and integrated luminosities (bottom) recorded at the CMS since the first run in 2010 [55].	27
4.7	A cutaway view of the CMS detector, showing the layers of sub-detectors [57]. . . .	28

4.8	An illustration showing the coordinate system with (z, η, ϕ) [63].	29
4.9	Schematic diagrams of the pixel detector before and after phase 1 upgrade during Run 2 [64].	30
4.10	The r - z distributions of sensor modules in a quarter of the silicon tracker at the end of Run 2 [66, 67]. The pixel detector modules are shown in green, while the single- and double-sided strip modules are shown in red and blue respectively.	31
4.11	The hit efficiencies of the pixel detector against instantaneous luminosity before (left) and after (right) the Phase-1 upgrade [64].	31
4.12	Layout of the ECAL in CMS showing the arrangements and positions of different substructures (left) and an example of a detector module in the ECAL barrel holding PbWO_4 crystals (right) [61].	33
4.13	Layout of supercrystals in a Dee in the ECAL endcaps (left) [61] and the ECAL's η -coverage (right) [68].	33
4.14	A schematic diagram showing a quarter of the HCAL in Run 2 [71]. The coloured patches show the groupings of detector modules.	34
4.15	A schematic diagram showing the structure of a HB wedge, made of stainless steel and brass plates with slots for scintillator trays (left) [74] and a sketch of segmentations in scintillator trays (right) [61].	35
4.16	A schematic diagram showing the arrangement of absorbers and scintillators in the HE [75].	36
4.17	Layout of the DT detectors in a wheel in the barrel [76]. There are four stations per sector, spaced apart by steel return yokes (in gray). Each blue box is a drift chamber.	39
4.18	A schematic diagram showing an example of muon detection in the DT chambers (left) [80] and a schematic diagram of the cross section of a DT chamber (right) [61].	39

4.19	Layout CSC chambers in the endcaps in ME1 with three rings and ME2–4 with two rings (left) [81] updated for Phase-1 and the cutaway view of a cathode strip chambers (right) [61].	40
4.20	A schematic diagram showing the positions of muon chambers in a quarter of the CMS detector [77]. Thin blue boxes show where the RPCs are located.	40
4.21	A schematic diagram showing the cross section of a double-gap RPC module in the muon detector [82].	41
4.22	A schematic diagram showing how SM particles interact with different sub-detectors in the CMS detector [85].	43
5.1	Scale factors of electrons passing $I_{\text{mini}} < 0.1$ (top) and < 0.4 (bottom) in bins of η and p_T measured in 2017 data and simulations.	55
5.2	Scale factors of electrons passing $I_{\text{mini}} < 0.1$ (top) and < 0.4 (bottom) in bins of η and p_T measured in 2018 data and simulations.	56
5.3	Scale factors of muons passing $I_{\text{mini}} < 0.1$ (top) and < 0.4 (bottom) in bins of η and p_T measured for 2017 data and simulations.	58
5.4	Scale factors of muons passing $I_{\text{mini}} < 0.1$ (top) and < 0.4 (bottom) in bins of η and p_T measured for 2018 data and simulations.	59
5.5	Measured electron prompt rates as a function of η (left) and p_T (right) in 2017 data.	69
5.6	Measured muon prompt rates as a function of η (left) and p_T (right) in 2017 data.	70
5.7	Measured electron prompt rates as a function of η (left) and p_T (right) in 2018 data.	71
5.8	Measured muon prompt rates as a function of η (left) and p_T (right) in 2018 data.	71

5.9	Distributions of leptons' p_T in the multi-lepton channel CR2 in each flavour category. Data from 2017 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV scaled by a factor of 10 (50) using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with fake rates measured in CR2. Statistical and systematic uncertainties are shown by the hatched region.	73
5.10	Distributions of leptons' p_T in the multi-lepton channel CR2 in each flavour category. Data from 2018 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV scaled by a factor of 10 (50) using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with fake rates measured in CR2. Statistical and systematic uncertainties are shown by the hatched region.	73
5.11	Measured χ^2 in 2017 between the predicted and observed number of events in the multi-lepton channel CR2 summed over all flavour categories, as a function of electron (f_e) and muon (f_μ) nonprompt rates. The intersection of the two black lines indicates the position of the minimum, and the markers indicate bins which deviate from the minimum χ^2 by less than 2.3.	75
5.12	Distributions of Gaussian probability marginalized for electron (left) and muon (right) fake rates in CR2 using 2017 data. The fitted values of sigma are the uncertainties on measured fake rates.	75
5.13	Measured χ^2 in 2018 between the predicted and observed number of events in the multi-lepton channel CR2 summed over all flavour categories, as a function of electron (f_e) and muon (f_μ) nonprompt rates. The intersection of the two black lines indicates the position of the minimum, and the markers indicate bins which deviate from the minimum χ^2 by less than 2.28.	76
5.14	Distributions of Gaussian probability marginalized for electron (left) and muon (right) fake rates in CR2 using 2018 data. The fitted values of sigma are the uncertainties on measured fake rates.	76

5.15	Measured χ^2 in 2017 between the predicted and observed number of events in the multi-lepton channel CR1 summed over all flavour categories, as a function of electron (f_e) and muon (f_μ) nonprompt rates. The intersection of the two black lines indicates the position of the minimum, and the markers indicate bins which deviate from the minimum χ^2 by less than 2.3.	78
5.16	Measured χ^2 in 2018 between the predicted and observed number of events in the multi-lepton channel CR1 summed over all flavour categories, as a function of electron (f_e) and muon (f_μ) nonprompt rates. The intersection of the two black lines indicates the position of the minimum, and the markers indicate bins which deviate from the minimum χ^2 by less than 2.3.	78
5.17	Distributions of leptons' p_T in the multi-lepton channel CR2 in each flavour category. Data from 2017 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV scaled by a factor of 10 (50) using solid (dashed) lines, and background using filled histograms. Shapes of scaled $T\bar{T}$ distributions are easier to read. Nonprompt background is estimated with measured flat (left) and $ \eta $ -dependent (right) muon fake rates. Statistical and systematic uncertainties are shown by the hatched region. .	80
5.18	Distributions of leptons' p_T in the multi-lepton channel CR2 in each flavour category. Data from 2018 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV scaled by a factor of 10 (50) using solid (dashed) lines, and background using filled histograms. Shapes of scaled $T\bar{T}$ distributions are easier to read. Nonprompt background is estimated with measured flat (left) and $ \eta $ -dependent (right) muon fake rates. Statistical and systematic uncertainties are shown by the hatched region. .	81
5.19	Distributions of $H_T^{\text{lep}} < 1000$ GeV in the SS2L channel in ee and $\mu\mu$ categories. Data from 2017 (top) and 2018 (bottom) is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with the shifted fake rates. Statistical and systematic uncertainties are shown by the hatched region.	83

5.20	Measured electron CMID rate in $p_T < 100$ GeV (top), 100–200 GeV (mid) and > 200 GeV (bottom) in 2017 (left) and 2018 (right) as a function of $ \eta $. The markers at $1.44 < \eta < 1.56$ are dummy holders for the barrel-endcap gap in ECAL and will be ignored.	85
5.21	Measured $\chi^2_{t\bar{t}}$ in 2017 $t\bar{t}$ sample between the “predicted” and “observed” number of events in the multi-lepton channel CR2 summed over all flavour categories, as a function of electron (f_e) and muon (f_μ) nonprompt rates. The intersection of the two black lines indicates the position of the minimum.	90
5.22	Measured $\chi^2_{t\bar{t}}$ in 2017 $t\bar{t}$ sample between the “predicted” and “observed” number of events in the multi-lepton channel SR summed over all flavour categories, as a function of electron (f_e) and muon (f_μ) nonprompt rates. The intersection of the two black lines indicates the position of the minimum.	90
5.23	Distributions of leptons’ p_T in the multi-lepton channel CR2 in each flavour category. “Observed” event yield from 2017 $t\bar{t}$ sample is shown using black markers and the “predicted” event yield derived with the measured fake rates of $(f_e, f_\mu) = (0.08, 0.14)$ is shown using filled histograms. Statistical uncertainties on “predicted” are shown by the hatched region.	91
5.24	Distributions of leptons’ p_T in the multi-lepton channel CR2 in each flavour category. “Observed” event yield from 2018 $t\bar{t}$ sample is shown using black markers and the “predicted” event yield derived with the measured fake rates of $(f_e, f_\mu) = (0.08, 0.12)$ is shown using filled histograms. Statistical uncertainties on “predicted” are shown by the hatched region.	91
5.25	Measured $\chi^2_{t\bar{t}}$ in 2018 $t\bar{t}$ sample between the “predicted” and “observed” number of events in the multi-lepton channel CR2 summed over all flavour categories, as a function of electron (f_e) and muon (f_μ) nonprompt rates. The intersection of the two black lines indicates the position of the minimum.	92

5.26	Measured $\chi^2_{t\bar{t}}$ in 2018 $t\bar{t}$ sample between the “predicted” and “observed” number of events in the multi-lepton channel SR summed over all flavour categories, as a function of electron (f_e) and muon (f_μ) nonprompt rates. The intersection of the two black lines indicates the position of the minimum.	92
5.27	Distributions of leptons’ p_T in the multi-lepton channel CR2 in each flavour category. “Observed” event yield from 2018 $t\bar{t}$ sample is shown using black markers and the “predicted” event yield derived with the measured fake rates of $(f_e, f_\mu) = (0.08, 0.13)$ is shown using filled histograms. Statistical uncertainties on “predicted” are shown by the hatched region.	93
5.28	Distributions of leptons’ p_T in the multi-lepton channel SR in each flavour category. “Observed” event yield from 2018 $t\bar{t}$ sample is shown using black markers and the “predicted” event yield derived with the measured fake rates of $(f_e, f_\mu) = (0.08, 0.11)$ is shown using filled histograms. Statistical uncertainties on “predicted” are shown by the hatched region.	93
5.29	Distributions of H_T^{lep} in the SS2L channel for nonprompt electrons decaying from b quarks (top left), c quarks (top right), and light quarks (bottom left), and electrons matched to a “fake” GEN particle with the wrong flavour (bottom right). “Observed” event yield from 2017 $t\bar{t}$ sample is shown using black markers and the “predicted” event yield derived with the cut-and-count fake rates in Table 5.22 is shown using red solid lines. Statistical uncertainties are shown by the error bars.	97
5.30	Distributions of H_T^{lep} in the SS2L channel for nonprompt muons decaying from b quarks (top left), c quarks (top right), and light quarks (mid left), muons matched to a “fake” GEN particle with the wrong flavour (mid right), and muons not matched to any GEN particle (bottom). “Observed” event yield from 2017 $t\bar{t}$ sample is shown using black markers and the “predicted” event yield derived with the cut-and-count fake rates in Table 5.22 is shown using red solid lines. Statistical uncertainties are shown by the error bars.	97

5.31	Distributions of H_T^{lep} in the SS2L channel for nonprompt electrons decaying from b quarks (top left), c quarks (top right), and light quarks (mid left), electrons matched to a “fake” GEN particle with the wrong flavour (mid right), and electrons not matched to any GEN particle (bottom). “Observed” event yield from 2018 $t\bar{t}$ sample is shown using black markers and the “predicted” event yield derived with the cut-and-count fake rates in Table 5.23 is shown using red solid lines. Statistical uncertainties are shown by the error bars.	98
5.32	Distributions of H_T^{lep} in the SS2L channel for nonprompt muons decaying from b quarks (top left), c quarks (top right), light quarks (mid left), muons matched to a “fake” GEN particle with the wrong flavour (mid right), and muons not matched to any GEN particle (bottom). “Observed” event yield from 2018 $t\bar{t}$ sample is shown using black markers and the “predicted” event yield derived with the cut-and-count fake rates in Table 5.23 is shown using red solid lines. Statistical uncertainties are shown by the error bars.	98
5.33	Distributions of leptons’ p_T (top left), S_T (top right), AK4 jets’ p_T (bottom left) and b-tagged AK4 jet multiplicity (bottom right) in the multi-lepton channel CR2 in each flavour category. Data from 2017 is shown using black markers, scaled $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with measured fake rates (Section 5.5.4). Statistical and systematic uncertainties are shown by the hatched region.	104
5.34	Distributions of leptons’ p_T (top left), S_T (top right), AK4 jets’ p_T (bottom left) and b-tagged AK4 jet multiplicity (bottom right) in the multi-lepton channel CR2 in each flavour category. Data from 2018 is shown using black markers, scaled $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with measured fake rates (Section 5.5.4). Statistical and systematic uncertainties are shown by the hatched region.	104

5.35	Distribution of H_T^{lep} in the SS2L channel low H_T^{lep} CR in each flavour category. Data from 2017 (left) and 2018 (right) is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Statistical and systematic uncertainties are shown by the hatched region.	105
5.36	Distribution of H_T^{lep} in the SS2L channel nJet = 3 CR in each flavour category. Data from 2017 (left) and 2018 (right) is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Statistical and systematic uncertainties are shown by the hatched region.	106
5.37	Distributions of leading lepton p_T (left) and leading jet p_T (right) in the SS2L channel nJet = 3 CR in each flavour category. Data from 2017 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Statistical and systematic uncertainties are shown by the hatched region.	107
5.38	Distributions of leading lepton p_T (left) and leading jet p_T (right) in the SS2L channel nJet = 3 CR in each flavour category. Data from 2018 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Statistical and systematic uncertainties are shown by the hatched region.	108
5.39	Distributions of leptons' p_T (top left), S_T (top right), AK4 jets' p_T (bottom left) and b-tagged AK4 jet multiplicity (bottom right) in the multi-lepton channel SR in each flavour category. Data from 2017 is shown using black markers, scaled $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with measured fake rates (Section 5.5.4). Statistical and systematic uncertainties are shown by the hatched region.	109

5.40	Distributions of leptons' p_T (top left), S_T (top right), AK4 jets' p_T (bottom left) and b-tagged AK4 jet multiplicity (bottom right) in the multi-lepton channel SR in each flavour category. Data from 2018 is shown using black markers, scaled $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with measured fake rates (Section 5.5.4). Statistical and systematic uncertainties are shown by the hatched region.	109
5.41	Distribution of H_T^{lep} in the SS2L channel SR in each flavour category. Data from 2017 (left) and 2018 (right) is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Statistical and systematic uncertainties are shown by the hatched region.	110
5.42	Distribution of leading lepton p_T (left) and leading jet p_T (right) in the SS2L channel SR in each flavour category. Data from 2017 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Statistical and systematic uncertainties are shown by the hatched region.	111
5.43	Distribution of leading lepton p_T (left) and leading jet p_T (right) in the SS2L channel SR in each flavour category. Data from 2018 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Statistical and systematic uncertainties are shown by the hatched region.	112
5.44	Expected exclusion limit on $T\bar{T}$ (left) and $B\bar{B}$ (right) production cross sections in 2017 in the multi-lepton channel as a function of cut values on p_T of the leading b-tagged jet, for the singlet (top), doublet (middle), and $T \rightarrow tZ/B \rightarrow tW$ -only (bottom) scenarios.	113
5.45	Expected exclusion limit on $T\bar{T}$ (left) and $B\bar{B}$ (right) production cross sections in 2017 in the multi-lepton channel as a function of cut values on relative p_T of the closest lepton-jet pair, for the singlet (top), doublet (middle), and $T \rightarrow tZ/B \rightarrow tW$ -only (bottom) scenarios.	114

5.46	Expected exclusion limit on $T\bar{T}$ (left) and $B\bar{B}$ (right) production cross sections in 2018 in the multi-lepton channel as a function of cut values on p_T of the leading b-tagged jet, for the singlet (top), doublet (middle), and $T \rightarrow tZ/B \rightarrow tW$ -only (bottom) scenarios.	115
5.47	Expected exclusion limit on $T\bar{T}$ (left) and $B\bar{B}$ (right) production cross sections in 2018 in the multi-lepton channel as a function of cut values on relative p_T of the closest lepton-jet pair, for the singlet (top), doublet (middle), and $T \rightarrow tZ/B \rightarrow tW$ -only (bottom) scenarios.	116
5.48	Distribution of S_T in the multi-lepton channel SR with final selections in each flavour category. Data from 2017 and 2018 is shown using black markers and $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines. Estimated background after fitting to Run 2 data (Section 5.12) is shown with filled histograms. Statistical and systematic uncertainties are shown by the hatched region.	118
5.49	The 95% CL upper limits on $T\bar{T}$ production cross sections as a function of T masses for singlet (top), doublet (bottom left), and $T \rightarrow tZ$ only (bottom right) scenarios in 2017 and 2018 in 3+L channel.	119
5.50	The 95% CL upper limits on $B\bar{B}$ production cross sections as a function of B masses for singlet (top), doublet (bottom left), and $B \rightarrow tW$ only (bottom right) scenarios in 2017 and 2018 in 3+L channel.	120
5.51	The 95% CL upper limits on $T\bar{T}$ production cross sections as a function of T masses for singlet (top), doublet (bottom left), and $T \rightarrow tZ$ only (bottom right) scenarios in the full Run 2 in 3+L channel.	121
5.52	The 95% CL upper limits on $B\bar{B}$ production cross sections as a function of B masses for singlet (top), doublet (bottom left), and $B \rightarrow tW$ only (bottom right) scenarios in the full Run 2 in 3+L channel.	121

5.53	Distribution of H_T^{lep} in the SS2L channel SR in each flavour category. Data from 2017 and 2018 is shown using black markers, $T\bar{T}$ signals using solid lines, and background using filled histograms. Statistical and systematic uncertainties in the background prediction before performing the fit to data are shown by the hatched region.	122
5.54	The 95% CL upper limits on $T\bar{T}$ production cross sections as a function of T masses for singlet (left) and doublet (right) scenarios in 2017 and 2018 in SS2L channel. . .	123
5.55	The 95% CL upper limits on $B\bar{B}$ production cross sections as a function of B masses for singlet (left) and doublet (right) scenarios in 2017 and 2018 in SS2L channel. . .	123
5.56	The 95% CL upper limits on $T\bar{T}$ production cross sections as a function of T masses for singlet (left) and doublet (right) scenarios in the full Run 2 in SS2L channel. . . .	124
5.57	The 95% CL upper limits on $B\bar{B}$ production cross sections as a function of B masses for singlet (left) and doublet (right) scenarios in the full Run 2 in SS2L channel. . . .	124
5.58	Normalized (unitless) nuisance pulls and constraints of the maximum likelihood fits in the multi-lepton channel SR in 2017 and 2018 (top) and the full Run 2 (bottom). For each nuisance parameter in x -axis, the marker shows the difference in its value before and after the fit, normalized by its value before the fit, and the error bar shows the uncertainty after the fit. The blue markers indicate results from the fit with background-only model, fixing signal strength to zero, and the red markers denote the result for a fit with background+signal model with a floating r . For prefit, the difference is always zero, shown with a solid black line, and the normalized uncertainty range of 1 to -1 is shown with shaded area.	127
5.59	The top 30 nuisance parameter impacts from maximum likelihood fits in the multi-lepton channel SR in 2017 and 2018 (top) and the full Run 2 (bottom). The unitless fitted signal strength is shown on the top right corner of each plot.	128

5.60	Distributions of signal strength pulls when injecting $r = 0$ for mass points 1000 GeV in $T\bar{T}$ (left) and $B\bar{B}$ (right) analysis in the multi-lepton channel in 2017+2018 combined (top) and 2016+2017+2018 combined (bottom).	129
5.61	Distributions of the saturated χ^2 metric in toys (markers) generated with background-only model and data (red line) in maximum likelihood fit in the multi-lepton CR1 in 2017 and 2018.	130
5.62	Distribution of S_T in the multi-lepton channel CR1 in each flavour category. Data from 2017 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV scaled by a factor of 10 (50) using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with fake rates measured in CR2. Statistical and systematic uncertainties are shown by the hatched region.	130
5.63	Distribution of S_T in the multi-lepton channel CR1 in each flavour category. Data from 2018 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV scaled by a factor of 10 (50) using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with fake rates measured in CR2. Statistical and systematic uncertainties are shown by the hatched region.	131
5.64	Normalized (unitless) nuisance pulls and constraints of the maximum likelihood fits in the SS2L channel SR in 2017 and 2018 (top) and the full Run 2 (bottom).	132
5.65	The top 30 nuisance parameter impacts from maximum likelihood fits in the SS2L channel SR in 2017 and 2018 (top) and the full Run 2 (bottom).	133
5.66	Distributions of signal strength pulls when injecting $r = 0$ for mass points 1000 GeV in $T\bar{T}$ (left) and $B\bar{B}$ (right) analysis in the SS2L channel in 2017+2018 combined (top) and 2016+2017+2018 combined (bottom).	134
5.67	Distributions of the saturated χ^2 metric in toys (markers) generated with background-only model and data (red line) in maximum likelihood fit in the SS2L channel control region of $400 < H_T^{\text{lep}} < 1000$ GeV in 2017, 2018, and 2017+2018.	135

5.68	Distribution of $400 < H_T^{\text{lep}} < 1000$ GeV in the SS2L channel low H_T^{lep} CR in each flavour category. Data from 2017 (top) and 2018 (bottom) is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with fake rates measured in CR2. Statistical and systematic uncertainties are shown by the hatched region.	135
5.69	Distributions of the saturated χ^2 metric in toys (markers) generated with background-only model and data (red line) in maximum likelihood fit in the SS2L channel control region of $400 < H_T^{\text{lep}} < 1000$ GeV in 2018 in ee (left), $e\mu$ (mid), and $\mu\mu$ (right) categories independently.	136
5.70	Distributions of H_T^{lep} with pre-fit (top) and post-fit (bottom) backgrounds in the SS2L channel control region of $400 < H_T^{\text{lep}} < 1000$ GeV in ee (left), $e\mu$ (middle), and $\mu\mu$ (right) categories. Data from 2018 is shown using black markers, background using blue solid lines, and upward error using green solid lines. Statistical uncertainties are shown by the yellow filled region.	136
5.71	Distributions of the saturated χ^2 metric in toys (markers) generated with background-only model and data (red line) in maximum likelihood fit in the SS2L channel control region of $400 < H_T^{\text{lep}} < 700$ GeV in 2018.	137
5.72	Distributions of the saturated χ^2 metric in toys (black markers) generated with background-only model and data (red line) in analysis of 2018 in combined (left) and $e\mu$ (right) categories in nJet = 3 control region.	137
5.73	Distribution of H_T^{lep} in the SS2L channel nJet = 3 CR in each flavour category. Data from 2018 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Statistical and systematic uncertainties are shown by the hatched region.	138

5.74	Distributions of the saturated χ^2 metric in toys (markers) generated with background-only model and data (red line) in maximum likelihood fit in the SS2L channel control region of $700 < H_T^{\text{lep}} < 1000$ GeV in 2018 with nJet = 4 to 11, from left to right and top to bottom.	139
5.75	Expected 95% CL upper limits on $T\bar{T}$ production for the singlet (left) and doublet (right) benchmarks.	142
5.76	Expected 95% CL upper limits on $B\bar{B}$ production for the singlet (left) and doublet (right) benchmarks.	142
5.77	The 95% CL expected (left) and observed (right) lower limits on the T quark masses for many possible branching fraction combinations.	143
5.78	The 95% CL expected (left) and observed (right) lower limits on the B quark masses for many possible branching fraction combinations.	143
A.1	A diagram illustrating the tag and probe leptons from a Z mother.	149
A.2	Fits to distributions of dilepton masses for passing and failing probe in bin $35 < p_T < 50$ GeV and $-1.44 < \eta < -0.8$. 2017 UL data is fitted by nominal (top left), altSig (bottom left), and altBkg (bottom right) fit functions respectively. DY MC fitted by MC-altSig fit function (top right).	151
A.3	Fits to distributions of dilepton masses for passing and failing probe in bin $20 < p_T < 35$ GeV and $0.0 < \eta < 0.8$. 2017 UL data is fitted by nominal (top left), altSig (bottom left), and altBkg (bottom right) fit functions respectively. DY MC fitted by MC-altSig fit function (top right).	151
A.4	The measured data efficiencies and scale factors as a function of p_T (left) and η (right) in 2016 UL (top), 2017 UL (middle), and 2018 UL (bottom).	152
A.5	Measured efficiencies of cut-based tight muons passing mini-Isolation < 0.1 as a function of p_T and η in 2017 data by spark_tnp (left) and TnPUtills (right). . .	154

A.6	Measured efficiencies of cut-based tight muons passing mini-Isolation < 0.1 as a function of p_T and η in DY MC events for 2017 by <code>spark_tnp</code> (left) and <code>TnPUtils</code> (right).	155
A.7	Measured MC truth efficiencies of GEN muons passing cut-based tight ID as a function of p_T and $ \eta $ in 2018 MC samples of $t\bar{t}$ (left) and Drell-Yan (right) processes.	156
A.8	Measured MC truth efficiencies of GEN muons passing cut-based tight ID as a function of p_T in $ \eta < 0.9$ (top left), $0.9 < \eta < 1.2$ (top right), $1.2 < \eta < 2.1$ (bottom left), and $1.2 < \eta < 2.4$ (bottom right) in 2018 for $t\bar{t}$ and Drell-Yan processes.	157
A.9	Measured MC truth efficiencies of GEN muons passing cut-based tight ID as a function of p_T in 2018 DY MC events for different dZ ranges.	158

Chapter 1

Introduction

What are ‘things’ made of? It is a simple question that a child would ask. Yet it is a complicated question to answer, which has driven many scholarly developments and breakthroughs. A biologist may explain the definitions of life, from kingdoms and phyla all the way to genus and species. A chemist may answer with chemical compositions. Physicists, however, try to find an answer as close to the absolute fundamentality as possible. The question becomes — what is ‘everything’ made of?

Among the ancient Greek philosophers, Plato was the first to relate the word ‘element’ to the origin of matter in traceable written records [1, 2]. This puts in place the idea of fundamental particles. The word ‘element’ was defined, at that time, as a body that is not divisible into other bodies that are different in form. Particle physics is the modern field of physics that studies elementary particles, which are the smallest, simplest, and most fundamental constituents of matter.

Experimental particle physics has a long history that dates back to the late 1800s [3]. The discovery of electrons with J. J. Thomson’s cathode ray experiment is one of the earliest particle physics experiments. In the century that followed, experimentalists found photons, then muons, neutrinos, and their anti-partners. Often, these elementary particles were observed in bound states called hadrons which include baryons (e.g., protons, neutrons, and Λ) and mesons (e.g., pions and kaons). In 1964, the quark model, suggesting that hadrons are composed of quarks, was proposed. One way to find proof was to probe the insides of hadrons, which was among the motivations for building large accelerators, such as SLAC in the late 1960s [4, 5]. Despite doubts and challenges, quarks (but not hadrons) were eventually acknowledged as a family of fundamental particles after the discovery of the J/ψ meson (a bound state of two charm quarks) at Brookhaven Cosmotron and SLAC [6, 7].

The theory of fundamental particles, the Standard Model (SM), which included the weak interactions via vector bosons, was formulated in 1978. The model gained credibility after the discovery of W and Z bosons with masses close to predictions, at a proton-antiproton collider (SPS) at CERN in 1983 [8, 9]. The SM was a success in describing quarks, leptons, bosons, and their interactions. However, there was one missing puzzle in the SM — the Higgs boson. The related Higgs mechanism is important as it explains the mass of other fundamental particles. In 2012, the discovery of a Higgs boson was announced by both the CMS and ATLAS collaborations at CERN after four years of proton-proton collisions in the Large Hadron Collider (LHC) [10, 11, 12]. With this ground-breaking news, all elementary particles in the SM have been experimentally observed.

However, this is still far from the end of the exploration. While the SM best describes observations among existing theories, there are several motivations for extending the theory “Beyond the Standard Model” (BSM). One concern is the very different scale of the electroweak and gravitational sectors, known as the “hierarchy problem” [13]. It leads to the need for fine-tuning, or the “naturalness problem”, given the lightness of the observed Higgs mass. Another drive is to find the sector for dark matter, as astronomical observations support its existence. It is also many scientists’ hope that, in BSM, gravity can be included in the picture to complete a “Theory of Everything” (ToE). Physicists never cease to explore the unknown and understand the truth of nature.

This dissertation will describe a BSM search performed at the CMS experiment at CERN. It is an analysis looking for vector-like quarks (VLQs), which are new particles introduced in some BSM theories. The analysis aims at finding the rare occurrences, if VLQs are ever produced, in proton-proton collisions recorded at the CMS from 2016 to 2018. This analysis focuses on events with either two charged leptons of the same electric charge or at least three charged leptons from VLQ decays. Since these final states are relatively infrequent for SM processes, the small background yield makes it advantageous for observing new physics.

Pair productions of vector-like T and B quarks have been searched for previously at CMS and ATLAS in proton-proton collisions with a center of mass energy ($\sqrt{s}$) of 7 TeV [14, 15, 16], 8 TeV [17, 18, 19, 20], and 13 TeV [21, 22, 23, 24, 25, 26]. Combining these publications, T and B quarks are excluded for masses below 1.31–1.42 TeV and 1.22–1.58 TeV respectively, depending on the

branching ratios of each decay mode. The search presented in this dissertation uses the complete dataset at $\sqrt{s} = 13$ TeV and finds more stringent exclusion limits.

Chapter 2

The Standard Model

The Standard Model (SM), which combines the Glashow-Weinberg-Salam model and quantum chromodynamics, was introduced in the 1970s. The SM is a quantum field theory with gauge group $SU(3) \times SU(2) \times U(1)$ that can describe all the fundamental particles that have been observed so far, including six flavours of quarks and their anti-particle partners, six types of leptons and their anti-particle partners, four types of gauge bosons, and a Higgs boson¹. An overview of particles in the SM is shown in Figure 2.1.

This chapter is written with the following references: [3, 27, 28].

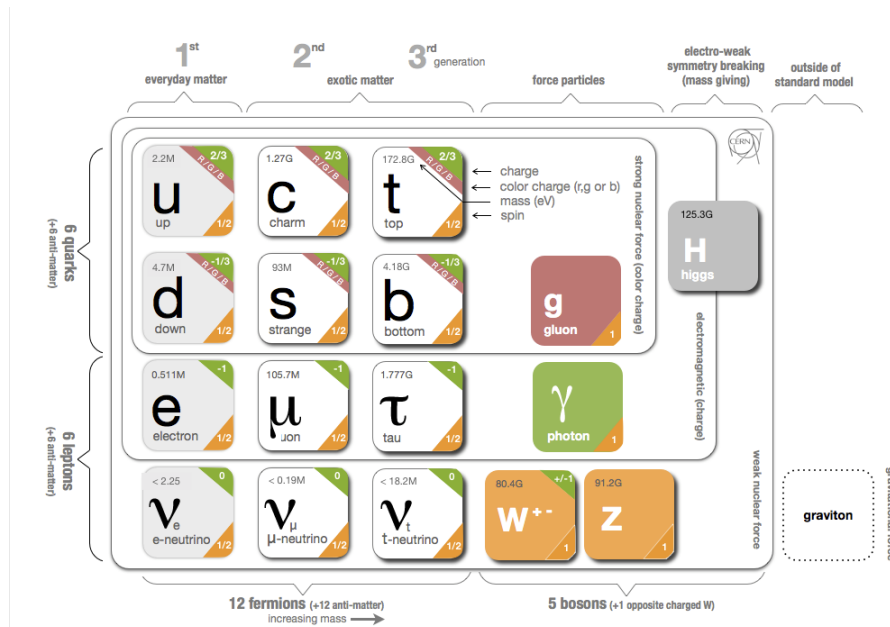


Figure 2.1: Particles described in the SM and their observed properties [29] with masses updated.

¹While some new physics theories propose that Higgs boson may be a composite state particle, there is no experimental evidence so far. Thus, we define the Higgs boson as fundamental particle unless proven otherwise.

2.1 The Elementary Particles

The 17 types of observed elementary particles are displayed in Figure 2.1. The three boxes, from the smallest to the largest, circle fermions participating in the strong, electromagnetic, and weak interactions respectively. The box for the weak interaction is the largest outermost one and circles all fermions. The curly braces at the periphery of the figure show various categories that particles can divide into according to their properties. The categorizations are illustrated in Figure 2.2 in which categories are written in coloured boxes, and decision criteria are labelled beside connecting arrows.

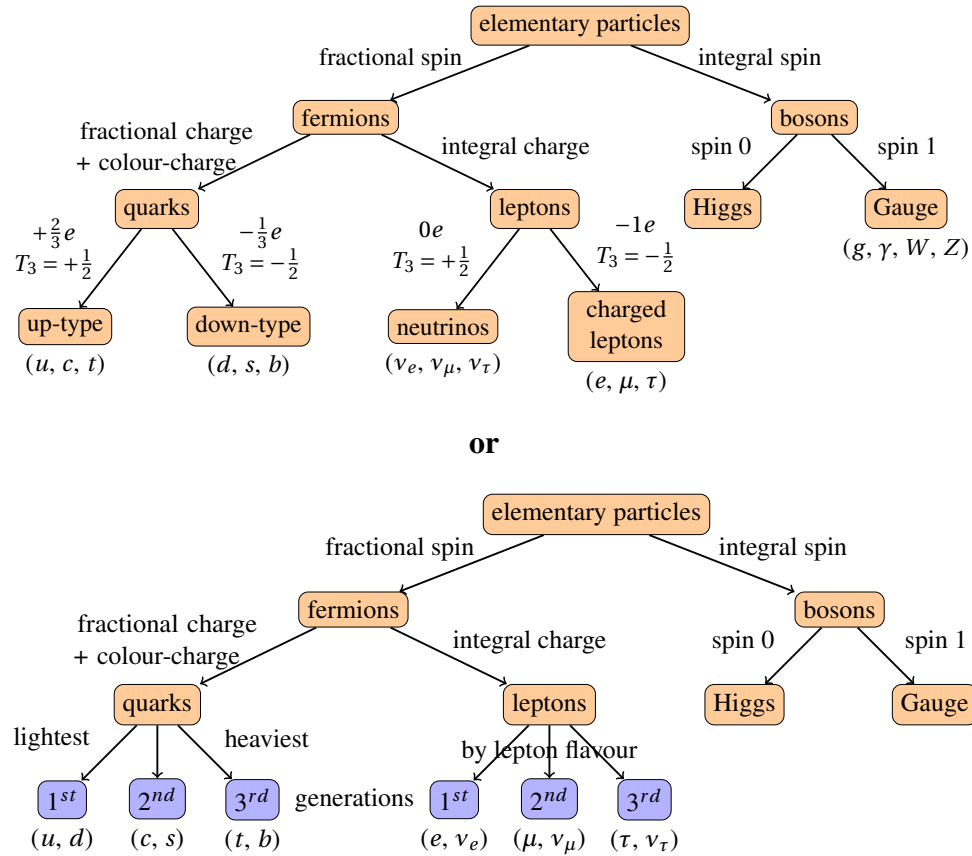


Figure 2.2: Schematic diagrams showing categorizations of the SM particle with two different sets of subordinate level fermion categories determined by electric charge (top) or by generations (bottom).

Particles are divided by their spins into bosons with integral spin and fermions with fractional spin, according to quantum field theory definitions. Ignoring particles' colour multiplicities in this paragraph, there are five bosons and twelve spin- $\frac{1}{2}$ fermions (and twelve anti-particle partners). The five bosons are split into the Higgs boson with spin-0 and four gauge bosons with spin-1. The twelve fermions are separated into six quarks with fractional electric charge and six leptons with

integral electric charge, for electric charges in units of $e = 1.6 \times 10^{-19}$ C. The colour-charged quarks can be sub-divided into up- and down-type, by their electric charges of $+\frac{2}{3}$ and $-\frac{1}{3}$ and/or by their weak isospins of $T_3 = \pm\frac{1}{2}$ in the left-handed (LH) sector, or sub-divided into the first, second, and third generations with different mass ranges. The leptons, allowed in electroweak gauge group only, are separated into charged leptons with charge -1 , and neutrinos with charge 0 , or into the first, second, and third generations by the lepton flavour quantum numbers. Similarly, LH leptons can be categorized by weak isospins into charged leptons with $T_3 = -\frac{1}{2}$ and neutrinos with $T_3 = +\frac{1}{2}$.

The six flavours of quarks are up (u), down (d), strange (s), charm (c), bottom (b), and top (t), in ascending order of mass. The t quark, with a mass of 172.76 GeV [30], is the heaviest particle in the SM. Quarks participate in all three of electromagnetic, strong, and weak interactions. Each quark is colour-charged with one of red, green, or blue. Often arranged into upper and lower rows, the up-type quarks are u , c , and t , while the down-type quarks are d , s , and b . The LH up-type quarks carry a weak isospin of $+\frac{1}{2}$, and the LH down-type quarks carry a weak isospin of $-\frac{1}{2}$. Arranging into three columns, the three generations contain u and d , c and s , and t and b respectively.

Each quark has an anti-particle partner with an anti-colour charge (anti-red, anti-green, and anti-blue) and an electric charge of the opposite sign ($-\frac{2}{3}$ or $+\frac{1}{3}$). Combinations of up-type and down-type quarks and anti-quarks form colourless bound states with integral electric charges, which can exist as free isolated particles. Mesons are formed by a quark and an anti-quark with opposite colour charges. Baryons are formed by three quarks or three anti-quarks, each of different colour charges. For example, a proton is composed of two up-type quarks and one down-type quark, with an electric charge of $\frac{2}{3} + \frac{2}{3} - \frac{1}{3} = 1$.

The six types of leptons are tau (τ), muon (μ), electron (e), tau neutrino (ν_τ), muon neutrino (ν_μ), and electron neutrino (ν_e). Though not explained by the SM, neutrinos have non-zero mass according to experiments. The lightest neutrino observed has a mass of < 2.25 eV [30]. Leptons participate in the weak interaction, but not in the strong interaction, as they have no colour charge. Only charged leptons, τ , μ , and e , are involved in the electromagnetic interaction. The three charged leptons correspond to three different lepton flavour quantum numbers. There is one neutrino for each lepton flavour: ν_τ , ν_μ , and ν_e . Leptons are also arranged into three generations by lepton flavour. Each

lepton has an anti-particle partner with the opposite charge (+1 or 0).

Spin-one gauge bosons are mediators of forces. Only three of the four fundamental forces of nature are described within the SM: strong, electromagnetic, and weak forces. The gravitational force, mediated by the spin-2 graviton, is beyond the SM. Massless photons are electromagnetic force carriers, while massive $W^\pm$ and Z bosons are weak force carriers. Massless gluons are strong force carriers. In the strong interaction, also known as quantum chromodynamics (QCD), colour-charged gluons and quarks interact. Unlike quarks, gluons have colour charges that are linear combinations of colour/anti-colour pairs. There are eight linearly independent combinations of the nine possible non-colourless colour/anti-colour pairs, such as red/anti-blue. Thus, there are eight different gluons in the SM.

Finally, there is one and only one type of scalar boson in the SM, called Higgs boson. It was the last elementary particle being observed in the SM, only discovered a decade ago with a mass of 125 GeV. The spin-0 Higgs boson is a crucial inclusion in the SM for particles to acquire mass through the Higgs mechanism.

2.2 The Quantum Field Theory

Under quantum field theory (QFT), a Lagrangian of a gauge group can be written, given the constraints of mass dimensions of four or less, gauge symmetries, and Lorentz invariance. The terms in the complete Lagrangian would tell us the masses of particles, and the fields and interactions that are allowed. Recall that the gauge group of the Standard Model is $SU(3) \times SU(2) \times U(1)$, under the symmetry of which, no mass terms of the gauge bosons nor leptons would be directly allowed in the Lagrangian. This is obviously in conflict with our observations in experiments and is resolved by introducing the Higgs field and the Higgs Mechanism.

2.2.1 Standard Model without Higgs

In the Standard Model, the Lagrangian is required to follow $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetry in which C is the colour charge, L is the left-handedness of fermions with weak isospin, and Y is the hypercharge. Without the Higgs field, the Lagrangian is symmetric if and only if there is no mass

term for gauge bosons and fermions (Equation 2.1).

$SU(3)_C$ is the strong interaction sector and involves only colour-charged particles: quarks and gluons. Since mass terms for quarks $m_q \bar{\psi} \psi$ are invariant under $SU(3)_C$, and gluons are indeed massless, the Lagrangian terms for strong interaction sector can be written without the Higgs field.

$$\begin{aligned} \mathcal{L}_{\text{SM}} = & -\frac{1}{4} G_{\mu\nu}^a G_{\mu\nu}^{a\mu\nu} - \frac{1}{4} F_{\mu\nu}^a F_{\mu\nu}^{a\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \bar{L}_i i D_\mu \gamma^\mu L_i \\ & + \bar{e}_{R_i} i D_\mu \gamma^\mu e_{R_i} + \bar{Q}_i i D_\mu \gamma^\mu Q_i + \bar{u}_{R_i} i D_\mu \gamma^\mu u_{R_i} + \bar{d}_{R_i} i D_\mu \gamma^\mu d_{R_i} \end{aligned}$$

$$\text{where } \left\{ \begin{array}{l} G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_3 f^{abc} G_\mu^b G_\nu^c \\ \text{for } a \in \{1, 2, \dots, 8\}, \text{ commutator of the generators} = [T^a, T^b] = i f^{abc} T_c \\ \text{and } g_3 = SU(3)_C \text{ coupling constant} \\ F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g_2 \epsilon^{abc} A_\mu^b A_\nu^c \\ \text{for } a \in \{1, 2, 3\}, \text{ commutator of the generators} = [T^a, T^b] = i \epsilon^{abc} T_c \\ \text{and } g_2 = SU(2)_L \text{ coupling constant} \\ B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \\ D_\mu = \partial_\mu - i g_3 T_a G_\mu^a - i g_2 T_a F_\mu^a - i g_1 \frac{Y}{2} B_\mu \\ \text{for commutator of the generators} = [Y, Y] = 0 \\ \text{and } g_1 = U(1)_Y \text{ coupling constant} \end{array} \right.$$

for gauge fields G_μ 's, A_μ 's and B_μ of the three gauge groups.

Given $i = \{1, 2, 3\}$ for the three generations and L or R for left-handed or right-handed.

$$\left\{ \begin{array}{l} L_1 = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L \quad e_{R_1} = e_R^- \quad Q_1 = \begin{pmatrix} u \\ d \end{pmatrix}_L \quad u_{R_1} = u_R \quad d_{R_1} = d_R \\ L_2 = \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L, \quad e_{R_2} = \mu_R^-, \quad Q_2 = \begin{pmatrix} c \\ s \end{pmatrix}_L, \quad u_{R_2} = c_R, \quad d_{R_2} = s_R \\ L_3 = \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L \quad e_{R_3} = \tau_R^- \quad Q_3 = \begin{pmatrix} t \\ b \end{pmatrix}_L \quad u_{R_3} = t_R \quad d_{R_3} = b_R \end{array} \right. \quad (2.1)$$

In the electroweak interaction sector, the Lagrangian should be invariant under $SU(2)_L \times U(1)_Y$ gauge transformation:

$$\begin{aligned} \text{fermions} & \begin{cases} \psi_L \rightarrow e^{i\alpha_a(x)T^a + i\beta(x)Y} \psi_L \\ \psi_R \rightarrow e^{i\beta(x)Y} \psi_R \end{cases} \\ \text{gauge bosons} & \begin{cases} A_\mu^a(x) \rightarrow A_\mu^a(x) - \frac{1}{g_2} \partial_\mu \alpha^a(x) - \epsilon^{abc} A_\mu^b(x) \alpha^c(x) \\ B_\mu(x) \rightarrow B_\mu(x) - \frac{1}{g_1} \partial_\mu \beta(x) \end{cases} \end{aligned} \quad (2.2)$$

where g_2 and g_1 are coupling constants, T^a and Y are generators, α and β are real parameters, and A_μ^a and B_μ are gauge fields for $SU(2)_L$ and $U(1)_Y$ respectively. x is the spacetime coordinates.

In the Lagrangian, each particle field should have a kinetic term and a mass term. By definition,

$$\begin{aligned} \text{for a fermion} & \begin{cases} \text{kinetic term} = i\bar{\psi} D_\mu \gamma^\mu \psi \\ \text{mass term} = -m\bar{\psi} \psi \end{cases} \\ \text{for a gauge boson} & \begin{cases} \text{kinetic term} = -\frac{1}{4} F_{\mu\nu}^a F_a^{\mu\nu} \\ \text{mass term} = -\frac{m^2}{2} A_\mu A^\mu \end{cases} \end{aligned} \quad (2.3)$$

where the covariant derivative $D_\mu = \partial_\mu - i g_2 T_a A_\mu^a - i \frac{g_1}{2} Y B_\mu$ for left-handed leptons and $D_\mu = \partial_\mu - i g_1 \frac{Y}{2} B_\mu$ for right-handed leptons. In Equation 2.3, for any vector field A_μ , its field strength tensor is $F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f^{abc} A_\mu^b A_\nu^c$.

One can perform the gauge transformation on each of the terms and shall find that the mass terms for W , Z , leptons and quarks are not invariant under $SU(2)_L \times U(1)_Y$, which means that these particles are massless under this theory. However, experiments have observed non-zero masses for these particles. One way to generate W , Z , and fermion masses while keeping $SU(2)_L \times U(1)_Y$ as part of the SM gauge group is to introduce the Higgs mechanism, which was attested with the discovery of the Higgs boson in 2012 [10, 11, 12]. The Higgs mechanism and mass generation are explained in the following sections.

2.2.2 Higgs Mechanism and Electroweak Symmetry Breaking

For $W^\pm$ and Z bosons to acquire mass, a complex scalar doublet field i.e. the Higgs field is added. Focusing on the electroweak sector, which is the $SU(2) \times U(1)$ part of the gauge group, the Higgs field ϕ has a representation of $(2, -\frac{1}{2})$. The standard form of Lagrangian of a scalar field ϕ in a non-abelian gauge group contains a kinematic term for ϕ , the potential of ϕ , and a kinematic term for each of the gauge fields associated with the $SU(2) \times U(1)$ gauge group.

$$\mathcal{L} = -(D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi) - \frac{1}{4} F^{a\mu\nu} F_{\mu\nu}^a - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} \quad (2.4)$$

Mass term of a gauge field has the form shown in Equation 2.3, which is seemingly absent in Equation 2.4. However, there are massive $W^\pm$ and Z bosons in this sector. In fact, Equation 2.4 is just a starting point. To arrive at the final Lagrangian of the SM with all the mass and interaction terms, one needs to work with the Higgs field. In QFT, mass terms appear after some algebra manipulations involving the Higgs field, concluded as the Higgs Mechanism. This is why people say that the Higgs field gives mass to particles.

Except at extremely high temperatures in the primordial environment, the Higgs field has a non-zero vacuum expectation value (VEV), which leads to the spontaneous breaking of the gauge group. Finding the VEV of a scalar field ϕ is equivalent to finding the minimum point of its potential $V(\phi)$. The famous Mexican-hat potential of Higgs is given in Equation 2.5 in its most generalized form

$$V(\phi) = \frac{1}{4} \lambda \left(\phi^\dagger \phi - \frac{1}{2} v^2 \right)^2 \quad (2.5)$$

The complex doublet field ϕ has a simple expression of VEV with only one non-zero component:

$$\langle 0 | \phi(x) | 0 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} v \\ 0 \end{pmatrix} \quad (2.6)$$

Algebraically, spontaneous symmetry breaking is to do a Taylor expansion of the Lagrangian around the VEV, i.e, substituting Equation 2.7 into the Lagrangian. $H(x)$ in Equation 2.7 is a real scalar field

associated with the observed Higgs boson.

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} v + H(x) \\ 0 \end{pmatrix} \quad (2.7)$$

To visualize this process, imagine a ball at the centre of the Mexican hat, where it is symmetric in every direction. With a small perturbation, it rolls down in a random direction to a point at the bottom of the circular valley. Now, at any of these minimum points, the ball only sees symmetry in one direction. In QFT language, the $SU(2) \times U(1)$ symmetry is broken to a $U(1)$. This process is called the electroweak symmetry breaking (EWSB).

It is the most trivial to expand the potential,

$$\begin{aligned} V(\phi) &= \frac{1}{4} \lambda \left(\frac{1}{2} (v + H)^2 - \frac{1}{2} v^2 \right)^2 \\ &= \frac{1}{4} \lambda v^2 H^2 + \frac{1}{4} \lambda v H^3 + \frac{1}{16} \lambda H^4 \end{aligned} \quad (2.8)$$

The first term in Equation 2.8 is the mass term of the field H . Thus, the mass of the Higgs boson is $m_H = \sqrt{\frac{1}{2} \lambda v^2}$.

The kinematic term of the scalar field is $-(D^\mu \phi)^\dagger (D_\mu \phi)$ where D_μ is the symbol for covariant derivative. For a gauge-invariant term in the Lagrangian, the covariant derivative is defined as

$$(D_\mu \phi)_i = \partial_\mu \phi_i - i \left[g_2 A_\mu^a T^a + g_1 B_\mu Y \right]_i^j \phi_j \quad (2.9)$$

where $T^a = \frac{1}{2} \sigma^a$ for $a \in 1, 2, 3$ and $Y = -\frac{1}{2} I$. All the indices that are not specified on the left hand side of the equations are summed over all possible values ($\sum_{j \in 1, 2}$ and $\sum_{a \in 1, 2, 3}$). Expanding the kinematic term at VEV, there is a kinematic term for the Higgs boson and a term involving the square of a 2×2 matrix of generators in Equation 2.10. This latter term with matrix is what will produce the gauge bosons' mass terms.

$$\begin{aligned} &-(D^\mu \phi)^\dagger (D_\mu \phi) \\ &= -\frac{1}{2} \partial^\mu H \partial_\mu H - \frac{1}{8} (v + H)^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix}^\dagger \begin{pmatrix} g_2 A_\mu^3 - g_1 B_\mu & g_2 (A_\mu^1 - i A_m^2 u) \\ g_2 (A_\mu^1 + i A_m^2 u) & -g_2 A_\mu^3 - g_1 B_\mu \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \end{aligned} \quad (2.10)$$

In a clean unitary gauge, the lower component of the scalar field doublet is zero, as a result, only the top left element of the squared matrix in Equation 2.10 is kept. However, before working out the matrix elements, it is more convenient for writing and interpretation to first define some new variables. In Equation 2.11, a weak mixing angle is defined to express the ratio of coupling constants g_1 and g_2 , and rewrite the mixing of gauge fields A 's and B into four new gauge fields $W^\pm$, Z and A .

$$\begin{aligned} \text{weak mixing angle} &: \theta_w = \tan^{-1} \left(\frac{g_1}{g_2} \right) \\ \text{gauge fields} &: \begin{cases} W_\mu^\pm = -\frac{1}{\sqrt{2}}(A_\mu^1 \mp A_\mu^2) \\ Z_\mu = c_w A_\mu^3 - s_w B_\mu \\ A_\mu = s_w A_\mu^3 + c_w B_\mu \end{cases} \quad \text{for } s_w = \sin \theta_w \text{ and } c_w = \cos \theta_w \end{aligned} \quad (2.11)$$

then,

$$\begin{pmatrix} g_2 A_\mu^3 - g_1 B_\mu & g_2(A_\mu^1 - i A_\mu^2 u) \\ g_2(A_\mu^1 + i A_\mu^2 u) & -g_2 A_\mu^3 - g_1 B_\mu \end{pmatrix} = \begin{pmatrix} \frac{1}{c_w} Z_\mu & \sqrt{2} W_\mu^+ \\ \sqrt{2} W_\mu^- & \dots \end{pmatrix}.$$

The bottom right element in the matrix is a bit complicated to write, but it is not needed for figuring out the Lagrangian terms. After the change of variables, there are two mass terms in the Lagrangian from the interactions of gauge fields W and Z with the Higgs field. Note that H is the real scalar field for the Higgs boson and must not be included in the mass constant. Thus, v is factored out. Then, Z boson has a mass of $m_Z = \frac{g_2 v}{2c_w}$, and $W^\pm$ has a mass of $m_W = \frac{g_2 v}{2}$ as shown in Equation 2.12. There is one more gauge field A_μ which does not have a mass term, i.e., it is massless. It is identified as the gauge field in the unbroken $U(1)$ of electromagnetism.

$$\mathcal{L}_{\text{mass}} = - \left(\frac{g_2 v}{2} \right)^2 (1 + v^{-1} H)^2 W^{+\mu} W_\mu^- - \frac{1}{2} \left(\frac{g_2 v}{2c_w} \right)^2 (1 + v^{-1} H)^2 Z^\mu Z_\mu \quad (2.12)$$

The field strength tensors $F_{\mu\nu}$ and $B_{\mu\nu}$ in Equation 2.4 are independent of the added complex scalar field. However, since gauge fields A 's and B are re-arranged into $W^\pm$, Z , and A in the Higgs mechanism, change of variables is applied to the definitions of $F_{\mu\nu}$ and $B_{\mu\nu}$ to obtain the kinematic

terms for gauge fields $W^\pm$, Z , and A . The full Lagrangian for $SU(2) \times U(1)$ is given in Equation 2.13.

$$\begin{aligned}
\mathcal{L} = & -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{4}Z_{\mu\nu}Z^{\mu\nu} - D^\dagger{}^\mu W^{-\nu} D_\mu W_\nu^+ + D^\dagger{}^\mu W^\nu D_\nu W_\mu^+ \\
& + ie(F^{\mu\nu} + \cot\theta_w Z^{\mu\nu}) W_\mu^+ W_\mu^- - \frac{1}{2} \left(\frac{e^2}{\sin^2\theta_w} \right) (W^{+\mu} W_\mu^- W^{+\nu} W_\nu^- - W^{+\mu} W_\mu^+ W^{-\nu} W_\nu^-) \\
& - (m_W^2 W^{+\mu} W_\mu^- + \frac{1}{2} m_Z^2 Z^\mu Z_\mu) (1 + v^{-1} H)^2 \\
& - \frac{1}{2} \partial^\mu H \partial_\mu H - \frac{1}{2} m_H^2 H^2 - \frac{1}{2} m_H^2 v^{-1} H^3 - \frac{1}{8} m_H^2 v^{-2} H^4
\end{aligned}$$

$$\text{where } \begin{cases} D_\mu = \partial_\mu - ie(A_\mu + \cot\theta_w Z_\mu) \\ F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \quad ; \quad Z_{\mu\nu} = \partial_\mu Z_\nu - \partial_\nu Z_\mu \\ \theta_w = \tan^{-1} \left(\frac{g_1}{g_2} \right) \quad ; \quad e = g_2 \sin\theta_w \\ m_H = \sqrt{\frac{\lambda}{2}} v \quad ; \quad m_Z = \frac{g_2 v}{c_w} \quad ; \quad m_W = \frac{g_2 v}{2} \end{cases} \quad (2.13)$$

2.2.3 Leptons and Yukawa Coupling

Mass terms of W and Z bosons are allowed in the Lagrangian with the Higgs mechanism. Yet the electroweak sector is incomplete without lepton mass terms. Since each generation is a copy of the same representations in the Lagrangian, the mass terms of the second and third generations have the same form as those of the first generation leptons. The derivation below omits the generation index and assumes that there is only one generation.

In the electroweak $SU(2) \times U(1)$ gauge group, there are two left-handed Weyl fields per lepton generation with representations of $(2, -\frac{1}{2})$ and $(1, +1)$ respectively. The former is a doublet field, and the latter is a singlet field, which are labelled as $L_1 = \begin{pmatrix} \nu_e \\ e \end{pmatrix}$ and $\bar{e}_1$ for the first generation. Note that $\bar{e}_1$ is the hermitian conjugate of e_{R_1} defined in Equation 2.1. Physically, $\bar{e}_1$ corresponds to the left-handed charged anti-leptons and e_{R_1} the right-handed charged leptons. The superscript for the electric charge on the lower element of L_1 is dropped for convenience. Their kinematic terms follow the same form as shown in Equation 2.3. Note that gauge fields are present in the definition of

covariant derivative, which becomes the interaction terms of leptons and electroweak gauge bosons in the SM after the change of variables defined in Equation 2.11.

Mass terms for these lepton fields arise in the Lagrangian through Yukawa coupling, which is the interaction of Weyl fields and the Higgs field as shown in Equation 2.14.

$$\mathcal{L}_{\text{Yuk}} = -y_1 \epsilon^{ij} \phi_i L_{1j} \bar{e}_1 + h.c. \quad (2.14)$$

With EWSB, the Lagrangian term is Taylor expanded at the VEV defined in Equation 2.7. The non-zero terms are the mixing of e and $\bar{e}_1$ and its hermitian conjugate. It is useful to re-arrange the Weyl fields into a Dirac field ε and Majorana field $\mathcal{N}$ as shown in Equation 2.15. Then, it becomes obvious that the Yukawa term is a mass term for the Dirac field with mass $m_{e1} = \frac{y_1 v}{\sqrt{2}}$ as shown in Equation 2.16.

$$\begin{aligned} \varepsilon &= \begin{pmatrix} e \\ \bar{e}_1^\dagger \end{pmatrix} \\ \mathcal{N} &= \begin{pmatrix} \nu \\ \nu^\dagger \end{pmatrix} \end{aligned} \quad (2.15)$$

$$\mathcal{L}_{\text{Yuk}} = -\frac{1}{\sqrt{2}} y_1 (v + H) (e \bar{e}_1 + \bar{e}_1^\dagger e^\dagger) = -m_{e1} \bar{\varepsilon} \varepsilon (1 + v^{-1} H) \quad (2.16)$$

Similar to W and Z , with spontaneous symmetry breaking, the Higgs field gives mass to charged leptons. For the Yukawa coupling of all three generations, after diagonalization, there are only the mass terms of the three charged leptons $m_{e_i} = \frac{y_i v}{\sqrt{2}}$ for $i \in 1, 2, 3$. The different masses of charged leptons are therefore determined by the Yukawa coupling constants for each generation. However, neutrinos remain massless in this model. The observation that neutrinos do in fact have mass is one of the motivations for extending the SM.

2.2.4 Quarks and Quantum Chromodynamics

The Lagrangian terms for quarks in quantum chromodynamics (QCD) can be derived similarly to those for leptons. For QCD, the full SM gauge group $SU(3) \times SU(2) \times U(1)$ is considered. In each

generation, there are three Weyl fields associated to quarks with representations $(3, 2, \frac{1}{6})$, $(\bar{3}, 1, -\frac{2}{3})$, and $(\bar{3}, 1, \frac{1}{3})$. For the i -th generation, these fields are labelled as Q_i , $\bar{u}_i$, and $\bar{d}_i$ respectively. Again, note that $\bar{u}_i$ and $\bar{d}_i$ are hermitian conjugates (left-handed anti-quarks) of the right-handed quarks u_{R_i} and d_{R_i} in Equation 2.1. The kinematic terms for QCD are the last three terms in Equation 2.1 and their hermitian conjugates. The covariant derivative for each field is different based on their representations (e.g., a $SU(2)$ singlet has no coupling term in its covariant derivative), and it is necessary to distinguish between gauge fields and generators for $SU(3)$ and $SU(2)$. Nevertheless, the gauge fields in covariant derivatives give rise to interaction terms of quarks and gauge bosons in the SM Lagrangian.

Mass terms of quarks arise from Yukawa coupling of the three Weyl fields after EWSB. The Lagrangian for the i -th generation only is shown in Equation 2.18, where Q_i are doublets defined in Equation 2.1. After multiplying the matrices, one can see that the couplings are always between up-type quark fields and $\bar{u}_i$, or down-type quark fields and $\bar{d}_i$. Similar to what was done for leptons, new Dirac fields $\mathcal{U}_i$ and $\mathcal{D}_i$ are defined in Equation 2.17, such that mass terms become apparent. For a compact display of the equation, u_i is used to represent the upper element of Q_i defined in Equation 2.1, while d_i is used for the lower element. The masses of up-type quarks and down-type quarks are then $m_{u_i} = \frac{y'_i v}{\sqrt{2}}$ and $m_{d_i} = \frac{y''_i v}{\sqrt{2}}$ for the i -th generation.

$$\begin{aligned}\mathcal{U}_i &= \begin{pmatrix} u_i \\ \bar{u}_i^\dagger \end{pmatrix} \\ \mathcal{D}_i &= \begin{pmatrix} d_i \\ \bar{d}_i^\dagger \end{pmatrix}\end{aligned}\tag{2.17}$$

$$\begin{aligned}\mathcal{L}_{\text{Yuk}} &= -y'_i \epsilon^{ab} \phi_a Q_{ib} \bar{d}_i - y''_i \phi_a^\dagger Q_{ia} \bar{u}_i + h.c. \\ &= -y'_i \frac{1}{\sqrt{2}} (v + H) d_i \bar{d}_i - y''_i \frac{1}{\sqrt{2}} (v + H) u_i \bar{u}_i + h.c. \\ &= -\frac{y'_i v}{\sqrt{2}} (1 + v^{-1} H) \bar{\mathcal{D}}^i \mathcal{D}_i - \frac{y''_i v}{\sqrt{2}} (1 + v^{-1} H) \bar{\mathcal{U}}^i \mathcal{U}_i\end{aligned}\tag{2.18}$$

Chapter 3

Vector-Like Quarks

The Standard Model is a successful theory, but not a full picture of nature. What is the force carrier of gravitational force that interacts with all massive particles? Why do particle masses vary over so many orders of magnitude, or in other words, why are coupling constants so different between different generations? What is dark matter, abundantly observed in the universe, made of? Are neutrinos massless, as given in the SM, or massive, as supported by the observation of neutrino oscillations? There is a handful of evidence showing that there is something new beyond the Standard Model (BSM).

Vector-like quarks (VLQs) are a type of new particle introduced in a number of BSM theories tackling the hierarchy problem, such as the Little Higgs model [31, 32] and the Composite Higgs model [33, 34]. The VLQ sector is a simple extension of the SM that has not been excluded by experiments.

This chapter is written with the following references: [35, 36, 37, 38, 39].

3.1 SM Extension

Naively, an easy way to extend the SM would be to increase members in one of the sectors. However, fourth-generation quarks have been excluded by experiments. Prohibition of higher generations is suspected to be a result of the chirality of couplings between fermions and W boson. This inspires the investigation of a new fermion sector, similar to quarks but non-chiral. These are vector-like quarks.

Like the SM quarks, VLQs are spin- $\frac{1}{2}$ with fractional electric charges. They are also colour-charged, i.e., they participate in the strong interaction. However, VLQs have the same representations for both left-handed and right-handed components in the gauge group. Such a requirement has a large impact: the gauge theory for this sector is not chiral, and gauge-invariant mass terms are allowed in the Lagrangian without the need of Higgs mechanism. In renormalizable theories, only seven representations can be written for VLQs. They are listed in Table 3.1, grouped by their $SU(2)$ quantum number. The multiplets are formed by a total of four new fields: $X_{5/3}$, $T_{2/3}$, $B_{-1/3}$, and $Y_{-4/3}$ where the subscripts denote their electric charges.

Table 3.1: A table comparing the representations of SM quarks and VLQs. Their quantum numbers, electric charges, and Lagrangian terms in unitary gauge are shown. The four VLQ fields, vertically aligned in the same rows, are arranged to form different multiplets.

ψ	SM	VLQs as $SU(2)$ Multiplets						
	(chiral) quarks	Singlet		Doublet			Triplets	
	$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}$	$\begin{pmatrix} T \\ B \end{pmatrix}$	$\begin{pmatrix} X \\ T \end{pmatrix}$	$\begin{pmatrix} T \\ B \end{pmatrix}$	$\begin{pmatrix} B \\ Y \end{pmatrix}$	$\begin{pmatrix} X \\ T \\ B \end{pmatrix}$	$\begin{pmatrix} T \\ B \\ Y \end{pmatrix}$	
$SU(2)_L$	$2_{(L)} / 1_{(R)}$	1		2			3	
$U(1)_Y$	$\frac{1}{6}_{(L)} / \frac{2}{3}_{(u_R)} / -\frac{1}{3}_{(d_R)}$	$\frac{2}{3}$	$-\frac{1}{3}$	$\frac{7}{6}$	$\frac{1}{6}$	$-\frac{5}{6}$	$\frac{2}{3}$	$-\frac{1}{3}$
T_3	$\frac{1}{2}_{(u_L)} / -\frac{1}{2}_{(d_L)} / 0_{(R)}$	0		$\frac{1}{2}$ $-1/2$			1 0 -1	
Q_{EM}	$\frac{2}{3}_{(u)} / -\frac{1}{3}_{(d)}$	$\frac{2}{3}$	$-\frac{1}{3}$	$\frac{5}{3}$ $2/3$	$2/3$ $-1/3$	$-1/3$ $-4/3$	$\frac{5}{3}$ $2/3$ $-1/3$	$2/3$ $-1/3$ $-4/3$
$-\mathcal{L}_{\text{Yuk}}$	$y''_{IJ} \overline{Q}_I(\phi) u_{RJ}$ $y'_{IJ} \overline{Q}_I(\epsilon\phi) d_{RJ}$	$\lambda_1 \overline{Q}_3(\phi) T_R$ $\lambda_2 \overline{Q}_3(\epsilon\phi) B_R$	$\lambda_3 \overline{\psi}_{(XT),L}(\epsilon\phi) t_R$ $\lambda_4 \overline{\psi}_{(TB),L}(\phi) t_R + \lambda_5 \overline{\psi}_{(TB),L}(\epsilon\phi) b_R$ $\lambda_6 \overline{\psi}_{(BY),L}(\phi) b_R$	$\lambda_7 \overline{Q}_3(\sigma^a \phi) \psi^a_{(XTB),R}$ $\lambda_8 \overline{Q}_3(\sigma^a \epsilon\phi) \psi^a_{(XTB),R}$				

In a simplified scenario, which is considered for this dissertation, VLQs' interactions with the SM quarks through the Higgs field are limited to the third generation (Q_3 , t_R , b_R). This simplification is favored both experimentally for up-type VLQ and theoretically. For the up-type fermions, precision measurements on SM model parameters have highly constrained the mixing of VLQ with lighter generations. On the theory side, given the hierarchy of Yukawa coupling strength between generations, VLQ mixing with the third generation is expected to be dominant.

Direct mixing of VLQs and SM fermions can be ruled out by rotating fields. Among the four VLQs, X and Y are mass eigenstates, while T and B are rearranged into doublets with the SM t and b quarks. The final mass terms are given in Equation 3.1. With unitary rotation matrices $V_{L,R}^{t,b}$ defined in Equation 3.2, the 2×2 mass matrices M^t and M^b are diagonalized by performing $V_L^{t,b} M^{t,b} (V_R^{t,b})^\dagger$. Thus, masses of t , b and VLQs can be related by the four rotation angles $\theta_{L,R}^{t,b}$. In the doublet mass eigenstates $V_{L,R}^t T'_{L,R}$ and $V_{L,R}^b B'_{L,R}$, each element is the sum of field components of the VLQ and SM quark, such as $\cos\theta_L^t t_L - \sin\theta_L^t T_L$. Since the rotation angles describe the mixing of T and t , and B and b in the mass eigenstates, they are also called mixing angles.

$$\mathcal{L}_m = -\overline{T'_L} M^t T'_R - \overline{B'_L} M^b B'_R - \overline{Y_{L,R}} M^Y Y_{L,R} - \overline{X_{L,R}} M_X X_{L,R}$$

where $T'_{L,R} = \begin{pmatrix} t_{L,R} \\ T_{L,R} \end{pmatrix}$ and $B'_{L,R} = \begin{pmatrix} b_{L,R} \\ B_{L,R} \end{pmatrix}$ (3.1)

$$V_{L,R}^{t,b} = \begin{pmatrix} \cos\theta_{L,R}^{t,b} & -\sin\theta_{L,R}^{t,b} \\ \sin\theta_{L,R}^{t,b} & \cos\theta_{L,R}^{t,b} \end{pmatrix} \quad (3.2)$$

Just like any other fermion in the SM, each new representation adds a kinematic term of the form shown in Equation 2.3. The gauge fields and generators in the covariant derivative depend on the chosen representations. By rewriting the Lagrangian in terms of mass eigenstates in Equation 3.1 and gauge fields in Equation 2.11, it can be shown that VLQs couple to the electroweak gauge bosons with either a SM third-generation quark or another VLQ where electric charge differs by 1. Since VLQs are heavy, with masses excluded up to the TeV scale by experiments, this coupling is interpreted as the decay of the VLQs. The analysis presented in this dissertation searches for T , which only decays to tH , tZ , or bW and B , which only decays to bH , bZ , or tW . The branching fractions vary for different representations and models. The Feynman diagrams of pair production of $T\overline{T}$ and $B\overline{B}$ and their possible decays are shown in Figure 3.1. It is also possible to extend the analysis to search for $X_{5/3}$, which decays to tW only. The Feynman diagram of such production processes in the LHC is shown in Figure 3.2.

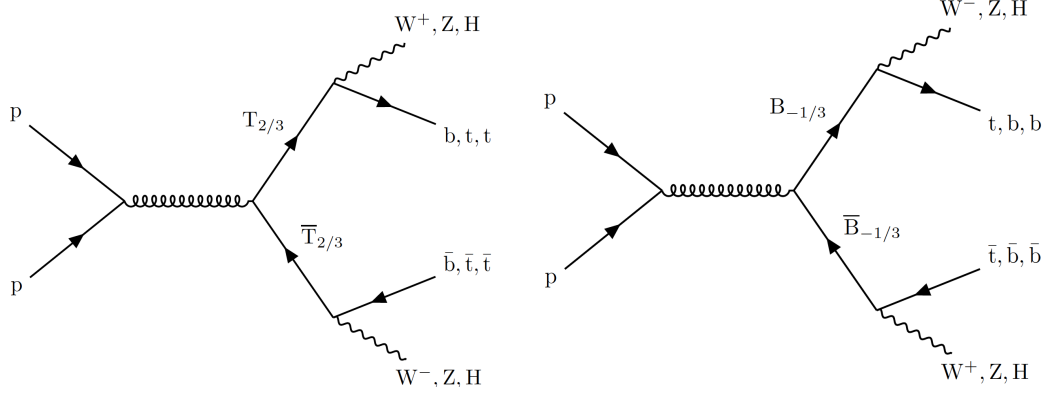


Figure 3.1: Feynman diagrams of $T\bar{T}$ and $B\bar{B}$ pair productions and their decay modes [21].

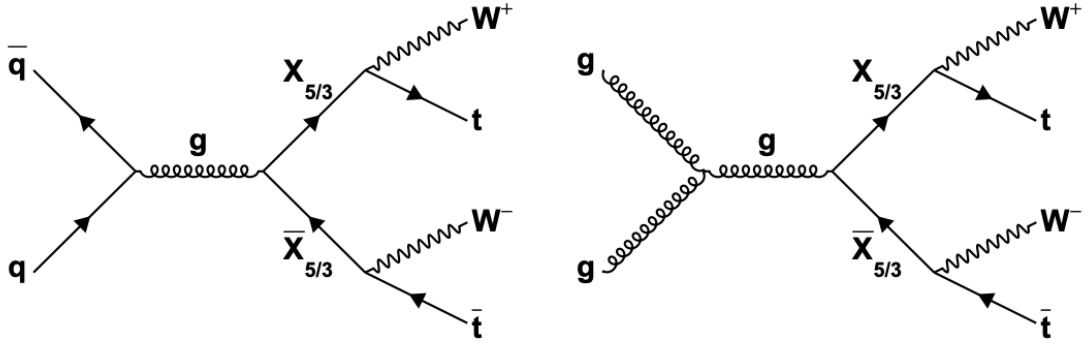


Figure 3.2: Feynman diagrams of $X_{5/3}\bar{X}_{5/3}$ pair productions and their decay [40].

3.2 Motivations for Searches

VLQs are popular search targets in high energy physics experiments. There are several reasons why physicists are particularly interested in these hypothetical new particles. The list below first describes the major theoretical drive, followed by various experiments which can test and constraint the hypothesis. The variety of experiments provides greater chances of finding VLQs, if they are real.

1. Higgs Boson Mass

- Under the SM, there are diverging corrections in the calculation of Higgs mass. In quite a number of models introducing the VLQs, Higgs mass is stabilized by adding corrections that cancel out the leading top quark quantum loop corrections.

2. Constraints from Flavour-Changing-Neutral-Currents (FCNC) Searches

- FCNC is greatly suppressed in the SM, occurring only within loop diagrams, and has not yet been observed experimentally. However, when one or more VLQs are present, the Zqq coupling is modified. The Z boson can couple to a VLQ and a SM quark, resulting in a change of quark flavour quantum number at a vertex, thus giving rise to FCNC. Results of FCNC searches in high energy experiments can therefore constrain the VLQ model by comparing with the SM.

3. Constraints from Measurements of CKM Matrix Elements

- In the SM, there is only one 3×3 Cabibbo–Kobayashi–Maskawa (CKM) matrix for the left-handed sector, which describes the quark mixing through a W boson [41]. Parametrization of the CKM matrix. If VLQs exist, the CKM matrix will be modified, and there would be a second CKM matrix for the right-handed sector which introduces a new CP-violating KM phases. Precision measurements of CKM matrix elements can be used to constrain VLQ properties by comparing with the SM.

4. Constraints from Measurements of SM Couplings

- Presence of one or more VLQ introduces new Feynman diagrams and alters the Lagrangian coupling terms with SM quarks that mix with VLQs. As a result, precision measurements complement VLQ searches by observing deviations from the SM and setting constraints.

5. Easily Tested by Model-Independent Searches

- The properties of VLQs are mainly constrained by a few model parameters, which can be measured and compared regardless of the model.

6. Single Production for Testing Higher VLQ Masses

- VLQs can be produced singly or in pairs. Figure 3.3 shows Feynman diagrams of single productions of a VLQ. Single productions are more model-dependent, but the production processes dominate at high VLQ mass. This allows separate analyses exploring the high mass search region of VLQs.

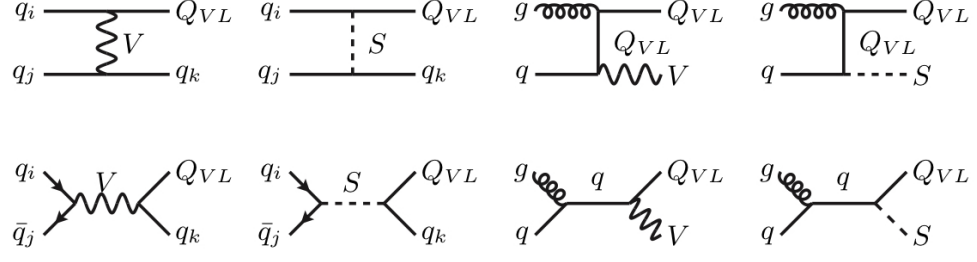


Figure 3.3: Feynman diagrams of possible single production of a VLQ [35].

Chapter 4

The CMS Experiment at the LHC

4.1 The Large Hadron Collider (LHC)

The Large Hadron Collider (LHC) is the world's largest particle accelerator, built and operated by the European Council for Nuclear Research (CERN) on the French-Swiss border [42]. It is a ring structure with a circumference of 27 km, situated 100 m underground. During the latest operating period between 2016 and 2018 (Run 2), it produced the most powerful proton-proton collisions to date with a centre-of-mass energy $\sqrt{s}$ of 13 TeV [43].

The high collision energy is achieved through successive stages in the CERN accelerator complex, illustrated in Figure 4.1. Proton bunches from ionized hydrogen gas, are first boosted to 50 MeV in Linac by alternating electric fields in a linear chain of conductors [44, 45]. They are then accelerated to 1.4 GeV in the 157 m Booster. The Booster is a synchrotron in which particles circulate and gain speed from pulsating electric fields. The circular path is maintained by magnetic fields which increase with the speed of the particles. Proton bunches are further accelerated to 2.6 GeV in the 628 m Proton Synchrotron (PS) where beams of proton bunches are formed with 25 ns (7.5 m) bunch spacings. Proton beams leave the PS and enter the Super Proton Synchrotron (SPS), where their energy is increased to 450 GeV, before they are injected into the LHC. In the LHC, beam energy is finally raised to 6.5 TeV.

CERN's Accelerator Complex

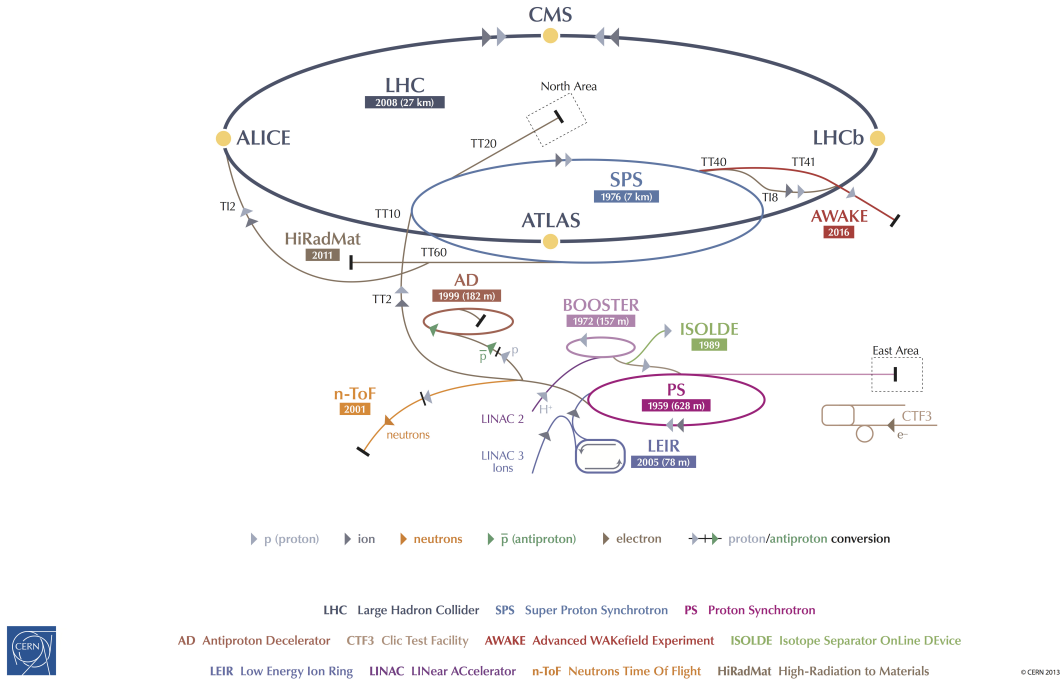


Figure 4.1: A schematic diagram of the CERN accelerator complex, indicating the machines' names, the year built they were built, and their sizes when applicable [46].

The LHC was built in the same tunnel as the Large Electron Positron Collider (LEP) after its deconstruction and started running in 2008 [47]. The LHC ring is composed of eight curves and eight straight sections, connected alternately. A majority of the eight curves are filled by 1232 superconducting main dipole magnets, each 15 m long with a cross-sectional diameter of 4 m, contributing to two thirds of the LHC ring [44, 48]. Figure 4.2 shows the cross section of a main dipole magnet. In the iron yoke of each dipole magnet, there are two apertures 194 mm apart, known as a “2-in-1” magnet, allowing beams to circulate in counter directions while fitting in the limited space underground. The magnetic field strength needed to bend the energetic beam is 8.3 T, which is provided by Nb-Ti superconductors cooled to 1.9 K (-271.3 °C) by helium in superfluid phase [49, 50]. These superconducting magnets are relatively economical to produce, but more sensitive to quenches compared to other superconductors studied during LHC’s design. The remaining curved sections are filled by 392 main quadrupole magnets, each of 3.25 m long, contributing to 1.2 km of the LHC ring. Figure 4.3 shows the cross section of a main quadrupole magnet.

LHC DIPOLE : STANDARD CROSS-SECTION

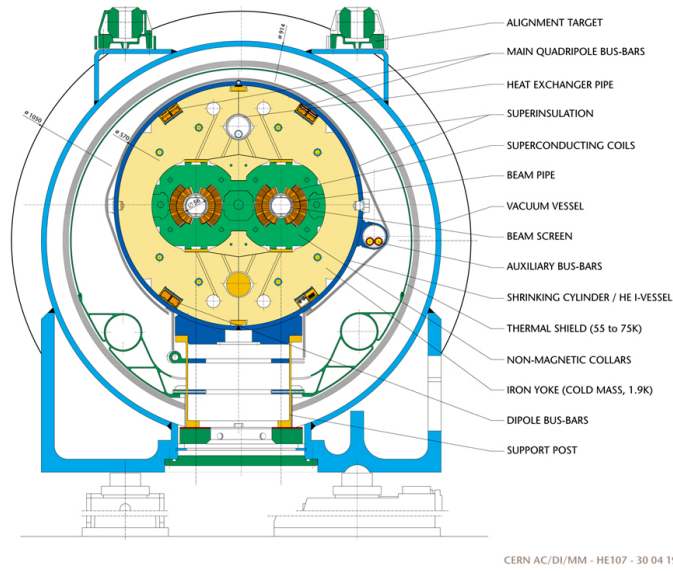


Figure 4.2: Cross section of the LHC dipole magnet [51].

LHC quadrupole cross section

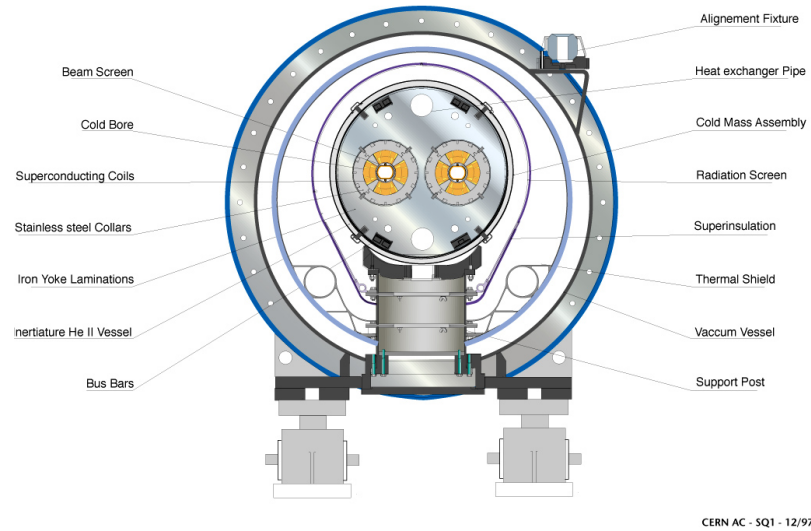


Figure 4.3: Cross section of the LHC quadrupole magnet [52].

Connected alternately to the eight curves in the LHC, the eight straight sections, each ~ 528 m long, are named Point 1 to 8 as shown in Figure 4.4. Each Point, with access to the surface, houses either a detector or facilities for beam-control. There are four detectors: A Toroidal LHC Apparatus (ATLAS) on Point 1, A Large Ion Collider Experiment (ALICE) on Point 2, Compact Muon Solenoid

(CMS) on Point 5, and Large Hadron Collider beauty (LHCb) on Point 8. ATLAS and CMS are general purpose detectors while the other two are designed for heavy-ion (ALICE) and flavour physics experiments (LHCb). Point 3, 4, 6, and 7 host utilities for collimating off-momentum protons, accelerating protons, dumping beam, and cleaning beam halo respectively. Figure 4.4 also shows the construction work needed for future upgrade to High Luminosity LHC (HL-LHC).

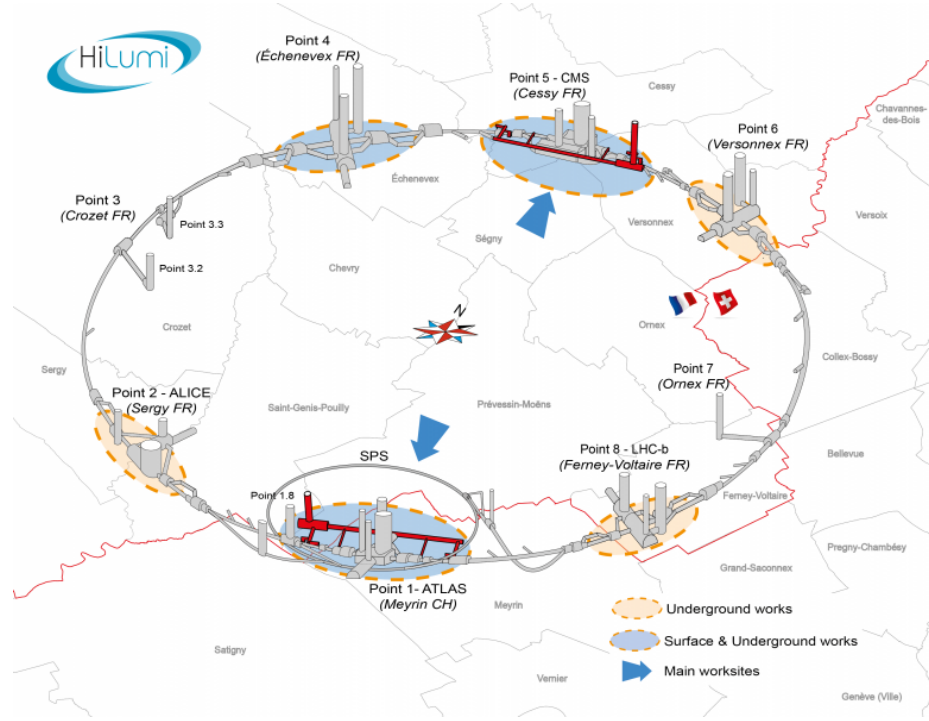


Figure 4.4: The current LHC layout and planned installations for future upgrades to HL-LHC [53].

The two beam pipes in the LHC ring stay separated, but cross at four points where collisions can occur, as illustrated in Figure 4.5. Proton beams from the SPS are injected through the two pipes on each side of ATLAS, in opposite directions. The crossing angles of the beams are controlled such that they circulate and accelerate without meeting and collide at a regulated rate when the desired beam energy is reached. The two beams are made to collide at the centre of a detector with centre-of-mass energy of 13 TeV, twice the energy of each individual beam.

While the high collision energy of the LHC allows a greater variety of production processes to occur, a large number of proton-proton interactions must be generated to collect sufficient statistics for rare events. The number of occurrences of a specific process (N) in a second is given by Equation 4.1, where σ is the cross section of the process and L the instantaneous luminosity [49]. Thus, the

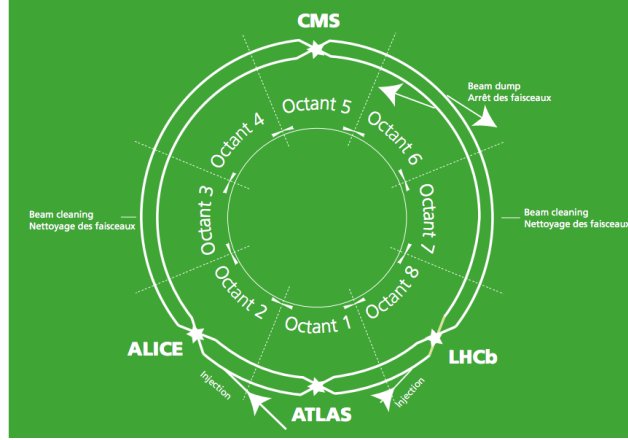


Figure 4.5: A schematic diagram of the paths of proton beams in the LHC [54]. The stars indicate where proton-proton collisions occur as the two beam paths cross.

integrated luminosity in units of inverse femtobarn ($1 \text{ fb}^{-1} = 10^{43} \text{ m}^{-2}$) is usually reported as a measure of the number of events generated in the LHC.

$$N = \sigma L \quad (4.1)$$

$$L = \frac{N_p^2 n_b f_{rev} \gamma_{rel}}{4\pi \epsilon_n \beta^*} \left[1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*} \right)^2 \right]^{-\frac{1}{2}} \quad (4.2)$$

The instantaneous luminosity, when two beams of protons cross paths in the LHC, is defined by Equation 4.2. In Run 2 in the LHC, the number of proton bunches per beam is $n_b = 2808$, the initial number of protons per bunch $N_p = 1.2 \times 10^{11}$, the revolution frequency $f_{rev} = 11245 \text{ s}^{-1}$ for the LHC's size and speed-of-light, the transverse emittance $\epsilon \approx 3.8 \text{ } \mu\text{m}$ when normalized for variations due to acceleration, and the beta function at the collision point $\beta^* \approx 60 \text{ cm}$ for the transverse beam size [42, 44]. The relativistic gamma factor γ_{rel} approaches unity near the speed-of-light. The inverse square root factor at the end of the equation is the geometric luminosity reduction factor, which accounts for the crossing angle θ_c at the collision point and other effects. In the reduction factor, σ_z is the root-mean-square bunch length at the collision point, and σ^* the root-mean-square transverse bunch size. In Run 2, LHC reached a peak instantaneous luminosity of $2 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ [43]. The integrated luminosity collected by CMS and ATLAS in Run 2 was about 150 fb^{-1} [55, 56]. Figure 4.6 shows the improvements over the years in luminosity, both instantaneous and integrated.

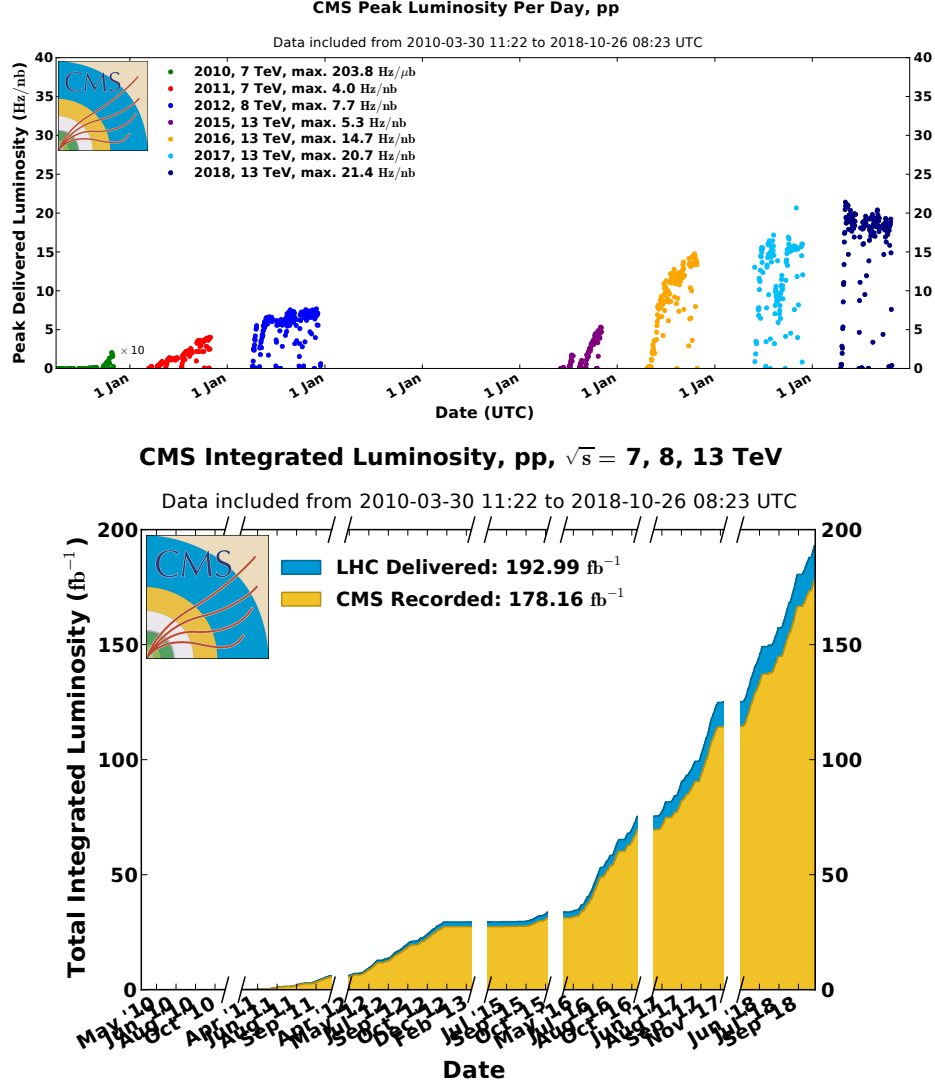


Figure 4.6: Plots showing peak instantaneous luminosities (top) and integrated luminosities (bottom) recorded at the CMS since the first run in 2010 [55].

4.2 The Compact Muon Solenoid (CMS) Detector

The Compact Muon Solenoid (CMS) is a cylindrical general-purpose detector composed of concentric layers, four of which are sub-detectors, as shown in Figure 4.7. CMS, 15 m in height, 21 m in length, and 14,000 tonnes in weight, is “compact” with half the size but double the weight of ATLAS, the only other general purpose detector at the LHC [57, 58]. The CMS detector is designed to accurately identify and measure the properties of muons, which are common in BSM signatures. To identify charges and measure the motions of decay particles, a superconducting solenoid, 6 m high and

12.5 m long, is installed at the centre of the detector, producing a magnetic field of 3.8 T, which is much more powerful than that of ATLAS [59, 60, 61]. Within the solenoid, there are inner tracker, electromagnetic calorimeter, and hadron calorimeter, while the outermost muon system wraps around the solenoid.

The CMS detector probes for new physics from hadron decays or heavy-ion collisions. During Run 2, CMS recorded data from 151 fb^{-1} of proton-proton collisions at 13 TeV, $2397 \mu\text{b}^{-1}$ of lead-lead collisions at 5.02 TeV, and 39.2 pb^{-1} of proton-lead collisions at 8.16 TeV.

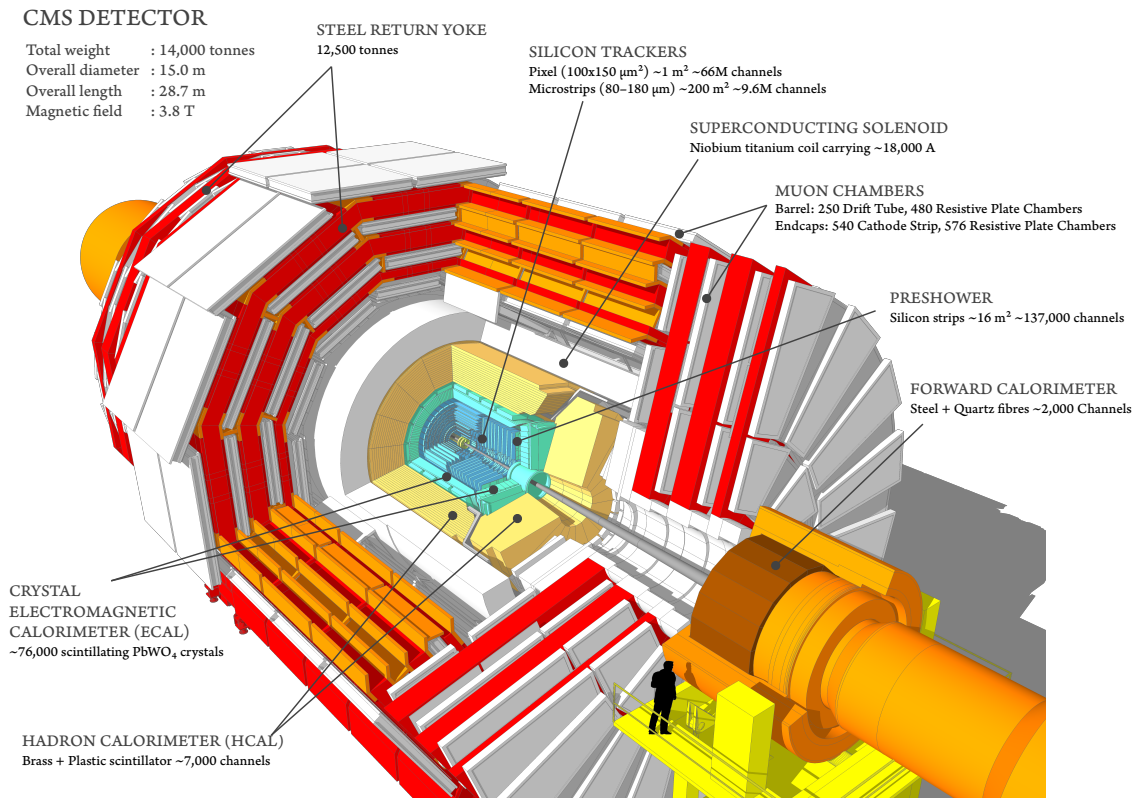


Figure 4.7: A cutaway view of the CMS detector, showing the layers of sub-detectors [57].

4.2.1 Coordinate System

For convenience, particle detectors adopt a coordinate system with the origin set at the nominal collision point. The y -axis points upward, and the x -axis points towards the centre of the LHC ring, while the z -axis aligns with the beam pipe [61]. The x - y plane is the transverse plane in which all particle momenta are captured by the detector, and conservation law is applicable. The direction

of motion of a particle is described by its pseudorapidity and azimuthal angle. The pseudorapidity η is defined in Equation 4.3, where θ is the angle between the vector of interest and the positive z -axis [62], as illustrated in Figure 4.8. A particle traveling perpendicular to the beam line has $\eta = 0$. The more boosted (“forward”) the particle, the larger the absolute value of η . The azimuthal angle ϕ is the angle between the vector of interest and the positive x -axis in the transverse plane. This coordinate system is adopted for its advantage of being invariant under Lorentz boosts along the z -axis. Separations between two vectors are often described by ΔR , defined in Equation 4.4.

$$\eta = -\ln \left[\tan \frac{\theta}{2} \right] \quad (4.3)$$

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \quad (4.4)$$

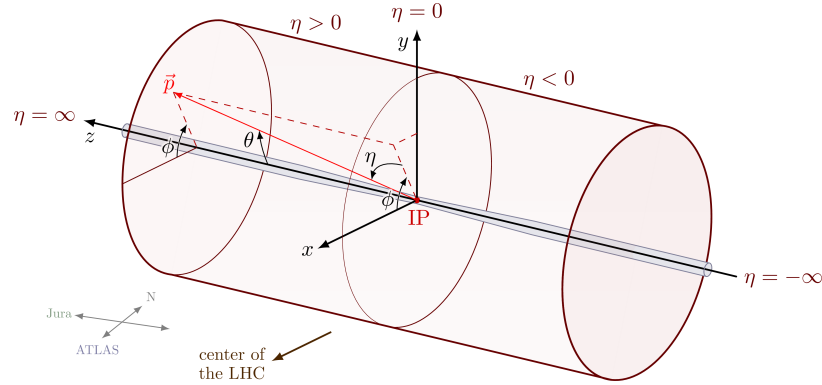


Figure 4.8: An illustration showing the coordinate system with (z, η, ϕ) [63].

4.2.2 The Inner Tracker

The inner tracker, as its name implies, is the innermost layer in the CMS detector. Despite being only 2.5 m in diameter and 5.8 m in length, it is the world’s largest silicon tracker, in terms of volume and silicon sensors’ coverage ($> 200 \text{ m}^2$) [61]. When a charged particle penetrates the p-n junction in a sensor, it induces ionizations and generates a current which is collected by readout chips (ROC). The recorded particle hits can be used to reconstruct tracks for particle identification and locate secondary vertices, which are signatures of heavy-flavour particle decays. There are two major types of sensors, pixels and strips, which form the inner pixel detector and the outer silicon strip detector respectively.

The pixel detector, lying just a few centimeters from the collision point, is crucial for detecting secondary vertices, which are formed by decays of heavy-flavour particles, such as b quarks and τ leptons. Its readouts also provide seeds for track reconstruction. Like most other sub-detectors, the pixel detector consists of concentric barrels and two forward detectors to provide a large acceptance region. In the middle of Run 2 (2017), the pixel detector underwent its Phase-1 upgrade to accommodate the enhanced luminosity of the LHC. Figure 4.9 shows the layout of the pixel detector before and after the upgrade.

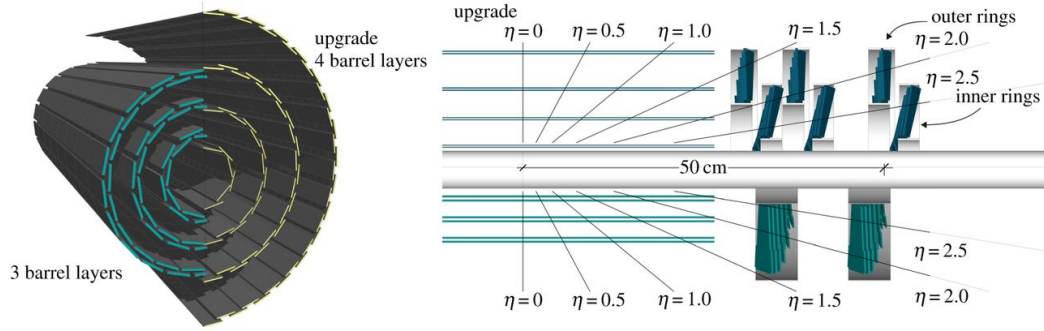


Figure 4.9: Schematic diagrams of the pixel detector before and after phase 1 upgrade during Run 2 [64].

There are now four layers instead of three in the barrel region. The innermost pixel layer is closer to the beam line, only 3.3 cm away, prompting the need for a smaller beam pipe. In the two forward regions (end-caps), the disk structure is modified to contain two rings, and an extra layer of disk is added. The layout of detector modules after the Phase-1 upgrade is shown in Figure 4.10. This upgrade adds 58 million to the 66 million pixels of the same size ($100 \times 150 \mu\text{m}$), resulting in a total of $16.2 \times 64.8 \text{ mm}^2$ of active area [65]. Each charged particle in the acceptance region of $|\eta| < 2.5$ leaves at most four hits in the pixel detector. With the change from 3-hit to 4-hit reconstruction in addition to new ROCs with increased bandwidth and improved radiation tolerance, the upgraded pixel detector has higher tracking efficiency, lower data loss, and better resolution. The performances of the pixel detector before and after the Phase-1 upgrade are shown in Figure 4.11.

The strip detector, lying beyond the pixel detector, receives much lower particle flux, and thus can be constructed with silicon strips while keeping the occupancy at a few percent. It is built with a total of 9.3 million strips, of length 80-180 μm , summing up to 198 m^2 of active area. The strip detector is divided into four parts, according to their locations and types of sensor modules used.

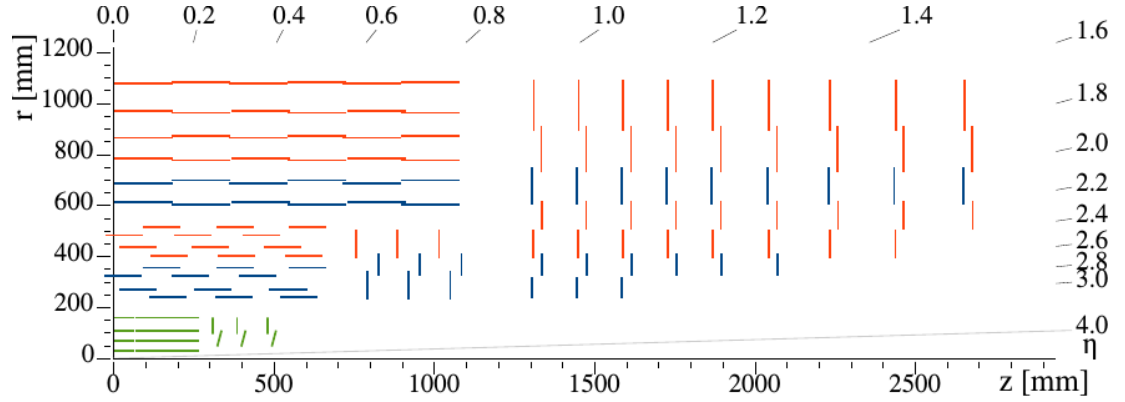


Figure 4.10: The r - z distributions of sensor modules in a quarter of the silicon tracker at the end of Run 2 [66, 67]. The pixel detector modules are shown in green, while the single- and double-sided strip modules are shown in red and blue respectively.

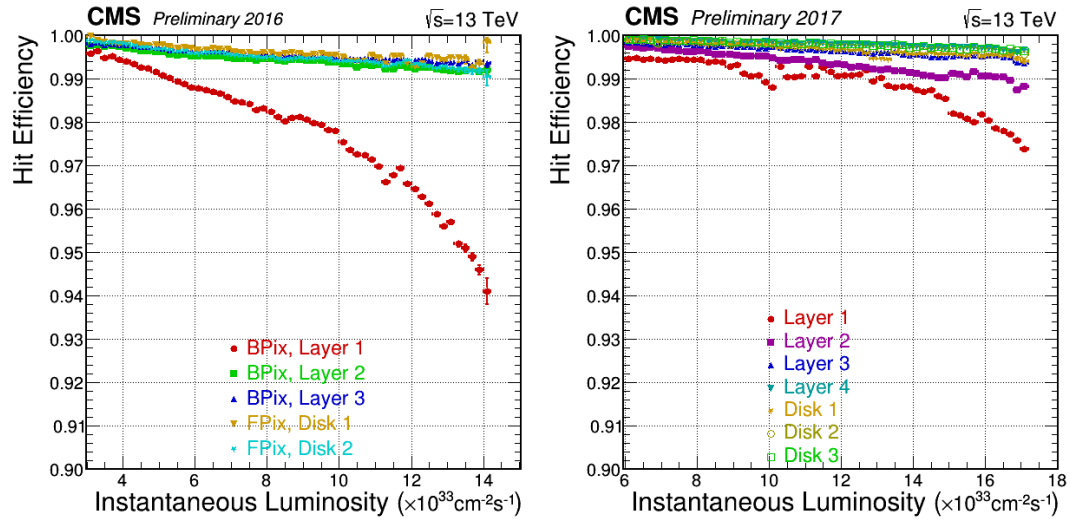


Figure 4.11: The hit efficiencies of the pixel detector against instantaneous luminosity before (left) and after (right) the Phase-1 upgrade [64].

There are ten layers in the barrel region, four inner ones form the Tracker Inner Barrel (TIB), and the six outer ones form the Tracker Outer Barrel (TOB). There are three disks at both ends of TIB, each with three concentric rings of sensors, forming the Track Inner Disks (TID). Finally, on each side in the forward region, there are Tracker Endcaps (TEC), made up of nine disks. The TEC disks are formed by four to seven rings. One can refer to Figure 4.10 for visualization. Located closer to the collision point, the TIB, TID, and the four inner rings of TEC use silicon sensors strips that are $320\text{ }\mu\text{m}$ thick. Further out, in the TOB and the three outer rings of TEC, flux is reduced. Thus, longer yet thicker strips, of thickness $500\text{ }\mu\text{m}$, are used for a high signal-to-noise ratio. Some layers

and rings use double-sided sensor modules which are two strip sensors glued back-to-back with a rotation angle of 100 mrad in between, indicated in blue in Figure 4.10. There was no hardware change on the strip detector during Phase-1 upgrade, but the cooling system has improved, lowering the operating temperature from 4 to -20 °C. This mitigates radiation damage.

4.2.3 The Electromagnetic Calorimeter (ECAL)

The electromagnetic calorimeter (ECAL) is a homogeneous crystal calorimeter which measures the energy of electromagnetic particles, mainly electrons and photons, with high resolution. This is particularly useful for the mass reconstruction of particles decaying to photons or leptons, such as the Higgs boson. The hermetic CMS ECAL, the largest in the history of high energy physics experiments, is made up of 75848 lead tungstate (PbWO_4) scintillating crystals, 61200 in the barrel and 7324 in each endcap [61]. For a better performance in the LHC environment, the calorimeter should have higher granularity, fast response, and be radiation-hard. This is achieved by using PbWO_4 crystals, which have a high density of 8.28 g/cm^3 , a short Molière radius of 2.2 cm, and a short scintillation decay time of 25 ns (for $\sim 80\%$ of the light) [64].

In the ECAL barrel (EB), the PbWO_4 crystals are arranged in 36 supermodules, each made up of four modules with varying shapes depending on the η position, as shown in Figure 4.12. The four types of modules are assemblies of 400-500 PbWO_4 crystals packed in submodules. Crystals in a submodule are separated by 0.35 mm, and submodules are separated by 0.5 mm in a module. Each crystal is tapered in shape, with cross section of $22 \times 22 \text{ mm}^2$ in front and $26 \times 26 \text{ mm}^2$ at the rear. Crystals are placed at an angle of 3° from their position vectors (from nominal collision point to crystals) to avoid particles travelling in parallel to the cracks in between. The EB covers an acceptance region of $|\eta| < 1.479$, and light signals in crystals are collected and amplified by a connected photodetector, avalanche photodiodes (APDs), which work well under the LHC and CMS environment.

In the ECAL endcaps (EE), PbWO_4 crystals are mounted to four semicircular disks, known as Dees. In a Dee, crystals are arranged in full or partial supercrystals, which are units of 5×5 bounded crystals, as shown in Figure 4.13. All crystals in EE are identical, with a length of 220 mm and a cross section of $28.62 \times 28.62 \text{ mm}^2$ in front and $30 \times 30 \text{ mm}^2$ at the rear. They point 1300 mm

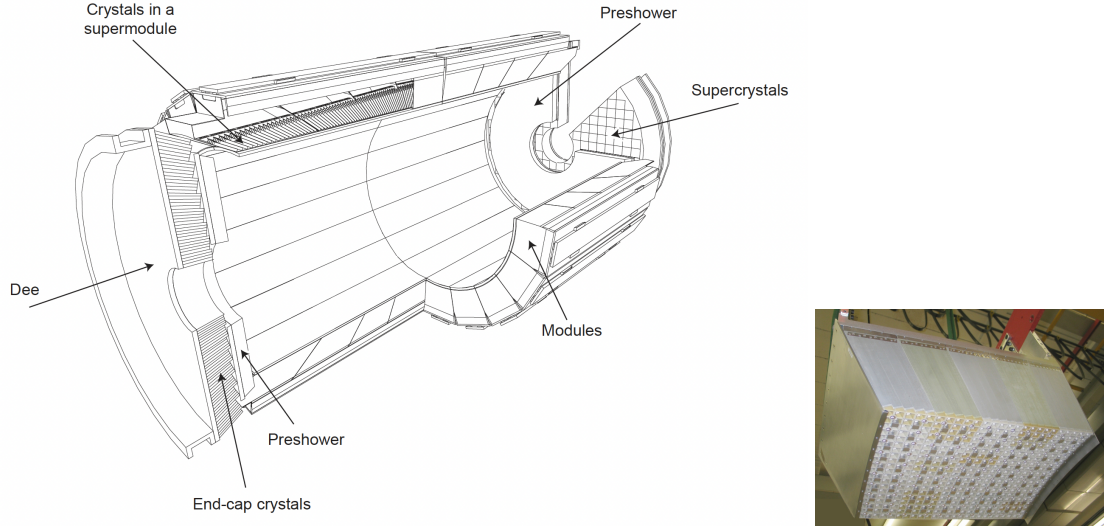


Figure 4.12: Layout of the ECAL in CMS showing the arrangements and positions of different substructures (left) and an example of a detector module in the ECAL barrel holding PbWO_4 crystals (right) [61].

off the collision point to create an angle of $2\text{--}8^\circ$ from their position vectors, again to avoid parallel alignment with trajectories. The EE covers the region of $1.479 < |\eta| < 3.0$ with read-outs by vacuum phototriodes (VPTs). In front of the endcaps, there are preshower detectors made of tungsten absorbers and silicon strips. The preshower detectors, covering only $1.653 < |\eta| < 2.6$, help to discriminate between photons and π^0 mesons.

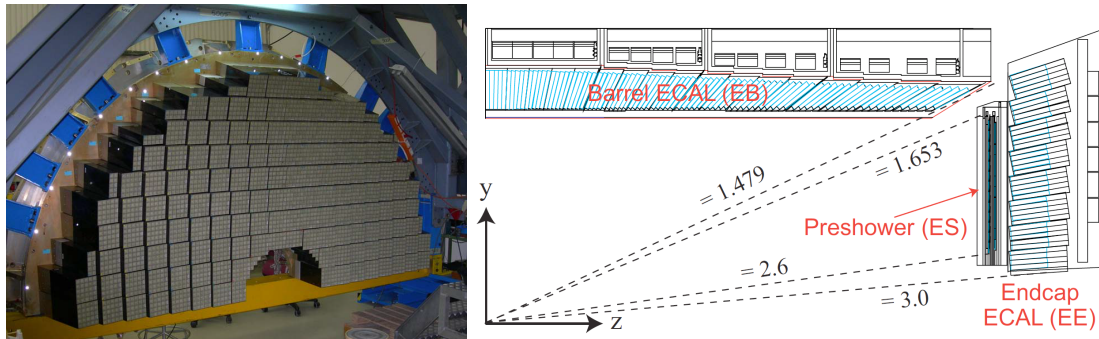


Figure 4.13: Layout of supercrystals in a Dee in the ECAL endcaps (left) [61] and the ECAL's η -coverage (right) [68].

Both light yields in crystals and internal gains in photodetectors (APDs and VPTs) are temperature-dependent. The ECAL operates at around 18°C by water cooling for an optimized performance [69]. Under the high instantaneous and integrated luminosity in the LHC, radiation damage received by the ECAL is more severe and light yields of PbWO_4 crystals are lowered. For a stable readout-energy

conversion, ECAL measurements are corrected, every 40 minutes with a laser monitoring system, every 2 days by channel-to-channel calibrations and every year by statistical calibrations [70].

4.2.4 The Hadron Calorimeter (HCAL)

The hadron calorimeter (HCAL) measures the energy of jets, which are hadron showers from hadronization of QCD particles. Since the total momentum in transverse plane is zero and conserved, combining the precise energy measurements in ECAL and HCAL, as well as information from other sub-detectors, gives a measurement of the transverse momentum of non-interacting particles, such as neutrinos in SM and dark matter in BSM. This is reported as the missing transverse momentum p_T^{miss} , which is the magnitude of the negative vector sum of all other measured objects. The HCAL is composed of four substructures: the concentric HCAL barrel (HB), the endcaps (HE), an outer calorimeter (HO) outside the solenoid, and a forward calorimeter (HF) on the two far sides beyond the muon detectors, as shown in Figure 4.14. The HCAL is hermetic, covering $|\eta|$ up to 5.2, to avoid hadrons from escaping the HCAL, which is important for finding the p_T^{miss} [61].

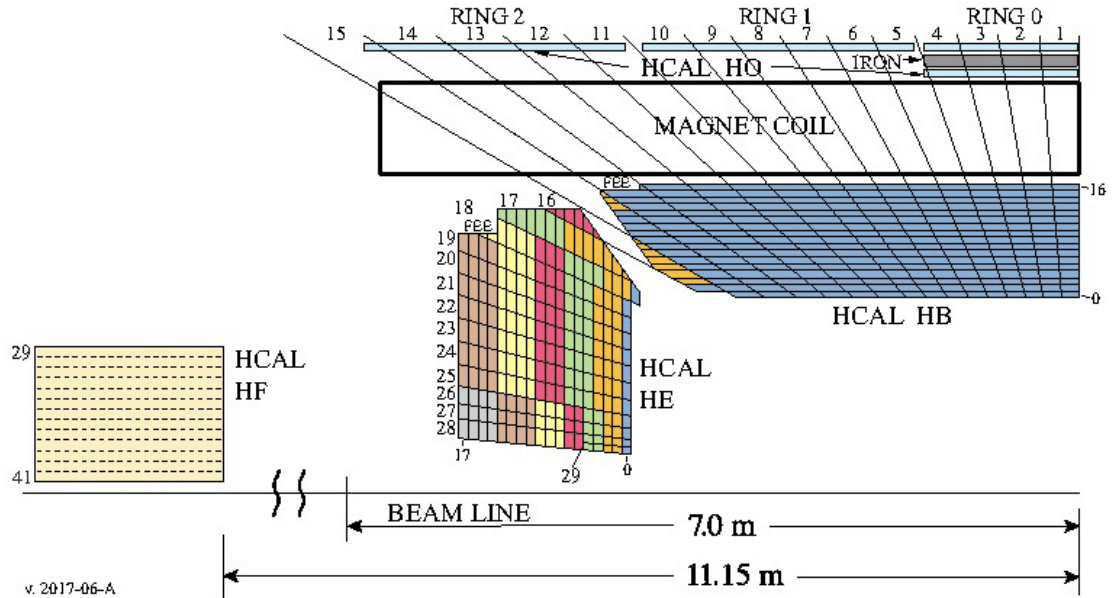


Figure 4.14: A schematic diagram showing a quarter of the HCAL in Run 2 [71]. The coloured patches show the groupings of detector modules.

The HCAL is a sampling calorimeter, which uses different materials for energy absorption, shower production, and scintillation in alternating layers, unlike the homogeneous ECAL. Both the HB and HE use brass absorbers and plastic scintillators, but with different geometric arrangements. These non-magnetic C26000 brass alloys have a high density of 8.53 g/cm^3 and a short radiation length of 1.49 cm, and thus are suitable for “stopping” hadrons showers with limited thickness inside the solenoid. The HO utilizes the solenoids’ iron yoke as absorbers with one or two layer(s) of plastic scintillators installed [71]. The HF uses steel absorbers and fiber scintillators with quartz core and acrylic clad [72].

The HCAL Barrel (HB)

The HB, which covers $|\eta| < 1.3$, is constructed from 36 brass-scintillator wedges. The wedges are lined up into two identical half-barrel rings of 18 wedges, with a minimal separation of $< 2 \text{ mm}$. Each wedge, spreading over an azimuthal angle ϕ of 20° , contains a 40 mm thick stainless steel front plate, eight 50.5 mm thick brass plates, six 56.5 mm thick brass plates, and a 75 mm thick stainless steel back plate staggered in order, in a fashion that leaves slots for insertion of scintillators as shown on the left in Figure 4.15. With scintillator slots before, in between, and after the plates, there are a total of 17 layers of read-outs in the radial direction [73]. In each layer in a wedge, four mechanical units of scintillators, called scintillator trays or magatiles, are inserted, one on each side and two in the central slot. This supports ϕ -segmentations of four sectors per wedge. A scintillator tray consists of a maximum of 16 read-out towers, as shown on the right in Figure 4.15, which allows η -segmentations of 16 sectors per wedge. Thus, each readout from the HB covers an area of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$.

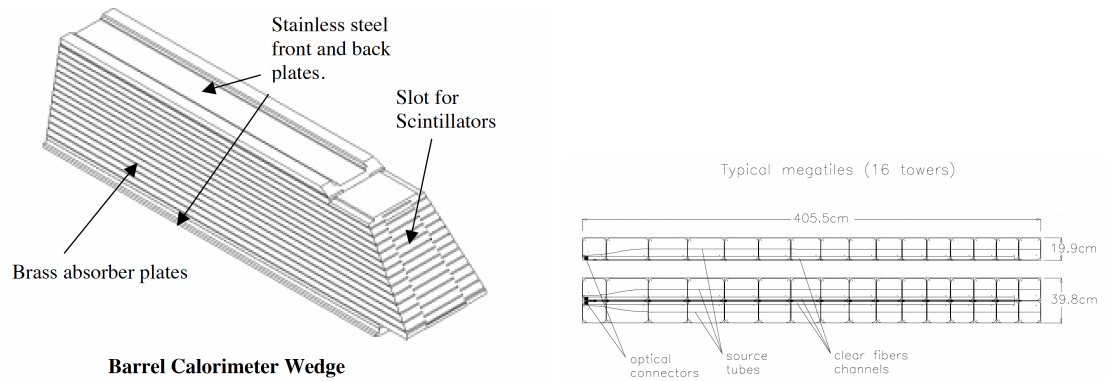


Figure 4.15: A schematic diagram showing the structure of a HB wedge, made of stainless steel and brass plates with slots for scintillator trays (left) [74] and a sketch of segmentations in scintillator trays (right) [61].

The HCAL Endcaps (HE)

The HE covers $1.3 < |\eta| < 3$ where roughly one third of the decay particles are captured [75]. It is made up of the same brass-scintillator materials as HB, but in a different geometry. Staggering brass sector plates as shown in Figure 4.16, the HE has an absorber thickness of 79 mm between two 9-mm slots for scintillators. This design does not require any additional support structure and ensures that there is no dead zone. The HE is rotationally symmetric in ϕ with 18 folds [73]. The depths and η -segmentations of HE, designed to suit its purpose, can be seen in Figure 4.14. There are 19 layers and a maximum of 14 sectors in $|\eta|$ per quarter of the detector. The cut-away block on the top left corner of HE is where the front-end electronics lies, and an extra layer is added before layer-0 of tower 18 to ensure sufficient absorbing thickness. The segmentations in HE have a size of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for $|\eta| < 1.6$ and $\Delta\eta \times \Delta\phi \approx 0.17 \times 0.17$ for $|\eta| > 1.6$. During the Phase-1 upgrade, photodetectors in HE were replaced by silicon photomultipliers (SiPMs) which improved the η -segmentations and detection performance [64].

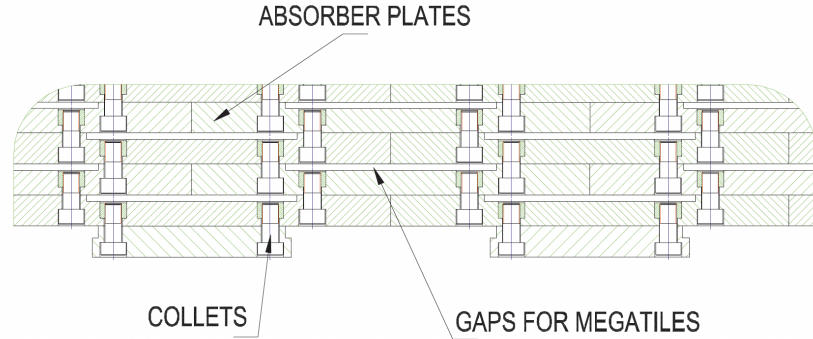


Figure 4.16: A schematic diagram showing the arrangement of absorbers and scintillators in the HE [75].

The HCAL Outer Calorimeter (HO)

The HO, which covers $|\eta| < 1.3$ like HB, is a tail catcher installed beyond the magnet coil to pick up shower energy that penetrates through the HB. Limited within the solenoid, the combined absorber thickness of HB does not provide sufficient stopping power for all jets, and roughly 5% of jets with energy > 100 GeV will escape [61]. The HO is designed to make use of the existing iron yoke for magnetic field return as an extended absorption layer. There are only five η sectors in total, corresponding to the five rings of return yoke. Each ring, 19.5 cm thick, is placed parallel to the

beam line, and thus the absorption depth is larger at higher $|\eta|$. For the four rings at higher $|\eta|$, a single layer of scintillators tiles is installed between the iron ring and the muon detector. For the central ring, to compensate for the smaller absorption depth, two layers of scintillators are placed on either side of the iron ring. The granularity $\Delta\eta \times \Delta\phi$ of the HO matches roughly to the HB, for the convenience of reconstructing energy “towers”.

The HCAL Forward Calorimeter (HF)

The HF, lying closest to the beam line, extends the HCAL detection to $2.9 < |\eta| < 5.0$, making the HCAL more hermetic [64]. To fit mechanically with other sub-detectors, such a high pseudorapidity part is placed at $\Delta z = 11.15$ m from the collision point, essentially sticking out on both sides of the main detector cylindrical body (Figure 4.7). Due to the proximity to the beam line, the HF experiences over seven times more radiation than other HCAL substructures [61]. Within the HF, the closer a scintillator is to the beam line, the higher the instantaneous luminosity it experiences. To build a radiation-hard forward calorimeter that will survive for at least ten years of operation, quartz-fibers are used. They are aligned, 5 mm apart, parallel to the beam line in between steel absorbers. When a charged particle passes through the quartz-fibers, which are active mediums, it radiates Cherenkov light. The Cherenkov signals are grouped, collected, and amplified by photomultiplier tubes (PMT) at the back of the HF. During the Phase-1 upgrade, improved PMTs are installed to reduce the efficiency of fake hits. The grouping of readouts gives the HF a granularity of $\Delta\eta \times \Delta\phi = 0.175 \times 0.175$ for $|\eta| < 4.7$, similar to the larger segmentations of the HE for $|\eta| > 1.6$. For the remaining high $|\eta| > 4.7$ which is strongly radiated, HF has a larger segmentation of $\Delta\eta \times \Delta\phi = 0.175 \times 0.35$ [76].

4.2.5 The Muon Chambers

At the beginning of Section 4.2, it was already mentioned that CMS puts a lot of emphasis on muon detection. To realize the full potential of muon final states in analyses, the CMS muon system is designed to identify muons and measure their kinematics with high accuracy in a wide range of $|\eta| < 2.4$ and transverse momentum $p_T > 2\text{--}4$ GeV [77, 78]. The muon chambers are built as the outermost concentric cylindrical layer of sub-detectors because of muons’ large penetration depth in metals. Most SM particles are terminated before reaching the front plate of the muon chambers, but

muons, as well as neutrinos, will penetrate through the muon chambers. Like most sub-detectors, the muon chambers form a barrel and two endcaps, using three types of gas-ionization detectors interleaved with the steel return yokes [61, 76, 79]. The gaseous detectors are drift tubes (DT) in the barrel, cathode strip chambers (CSCs) in the endcaps, and resistive plate chambers (RPCs) with a fast response for triggers, which will be explained later in Section 4.3.

In the barrel, the muon detectors are arranged in five identical wheels to cover $|\eta| < 1.2$, following the geometry of the solenoid. There are twelve azimuthal sectors per wheel and four layers of detector stations per sector, as shown in Figure 4.17. Each station is built from mechanical units called drift chambers. In the three inner stations, one chamber fills a station, so there are a total of $5 \times 12 = 60$ drift chambers per layer. In the outermost layer, the top and bottom stations are filled with two chambers, and thus a total of $5 \times (12 + 2) = 70$ drift chambers. Each chamber, as shown in Figure 4.18, contains three 4-layers drift cells (“super-layers”) aligned such that wires for signal currents run perpendicular to the z -direction for r - z positioning in the middle super-layer, while the others have wires parallel to the beam line for r - ϕ positions. The outermost stations have chambers with only two super-layers, both for r - ϕ positions. The bottom r - ϕ super-layers are separated from the rest by a honeycomb plate for better angular resolution. In each super-layer, four layers of drift cells are staggered with a shift of half-cell, which adds timing information for more accurate detection. There are two layers of RPCs in the two inner stations and one layer in the two outer stations, as shown in thin blue boxes in Figure 4.20.

In the endcaps, the muon detectors cover $0.9 < |\eta| < 2.4$, overlapping with the barrel. In this region, the muon system receives higher radiation flux compared to the barrel, and the magnetic field is not uniform. Thus, DTs are not a good choice of detectors in the endcaps. CSCs with better resistance to radiation and faster response are used instead. There are again four layers of CSC stations (ME1–4), with two to three rings, interleaved with disks of steel return yokes, as shown in Figure 4.19 and 4.20. The two endcaps contain a total of $(36 \times 3 + (18 + 36) \times 3) \times 2 = 540$ cathode strip chambers. Each trapezium chamber holds six planes of azimuthally running anode wires, alternating with seven layers of radially running cathode strips with a width of constant $\Delta\phi$, as shown on the right of Figure 4.19. The azimuthal arrangement of wires can determine the r coordinate, while the ϕ coordinate is estimated by analysing the signal current. The largest CSCs have a size of $\sim 3.4 \times 1.5 \text{ m}^2$. The entire

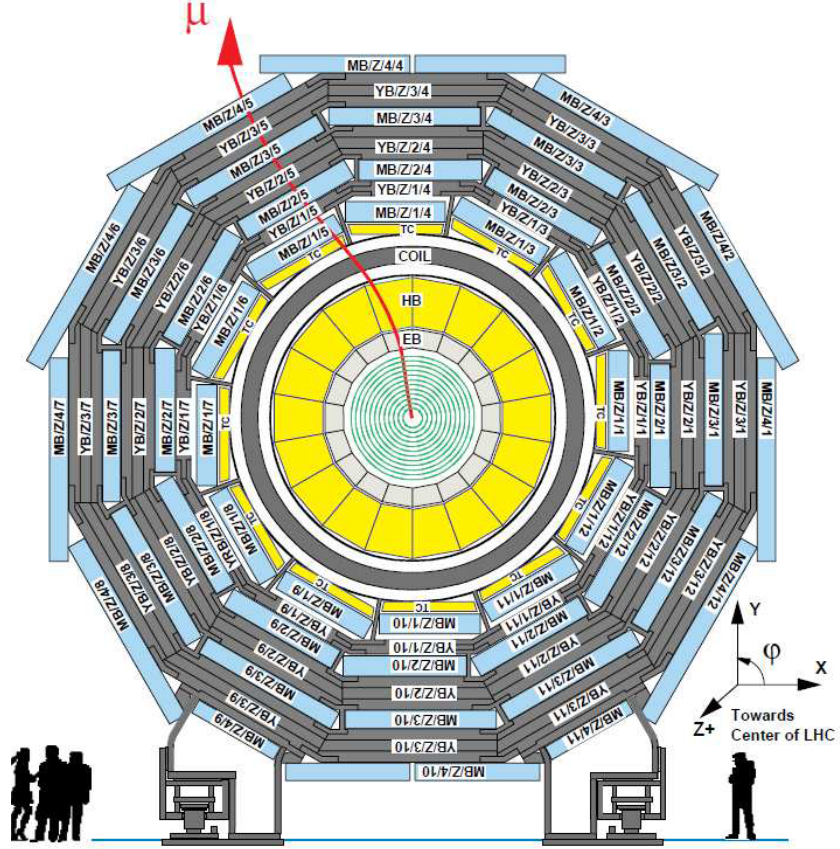


Figure 4.17: Layout of the DT detectors in a wheel in the barrel [76]. There are four stations per sector, spaced apart by steel return yokes (in gray). Each blue box is a drift chamber.

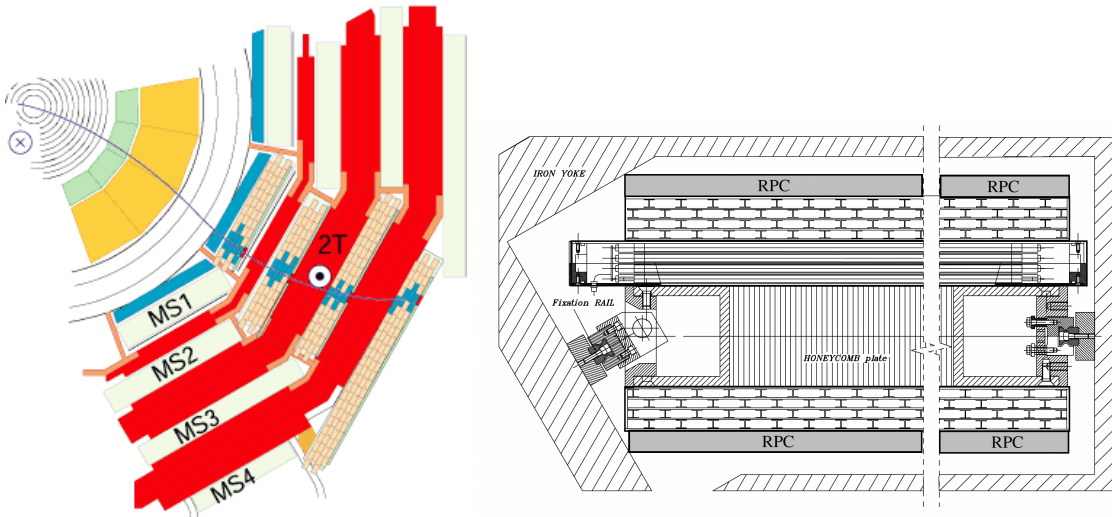


Figure 4.18: A schematic diagram showing an example of muon detection in the DT chambers (left) [80] and a schematic diagram of the cross section of a DT chamber (right) [61].

sub-detector contains 266112 cathode strips and 210816 channels of anode wires, supporting precise measurements of muons [77]. There are again four stations of RPCs in the endcaps as shown in Figure 4.20.

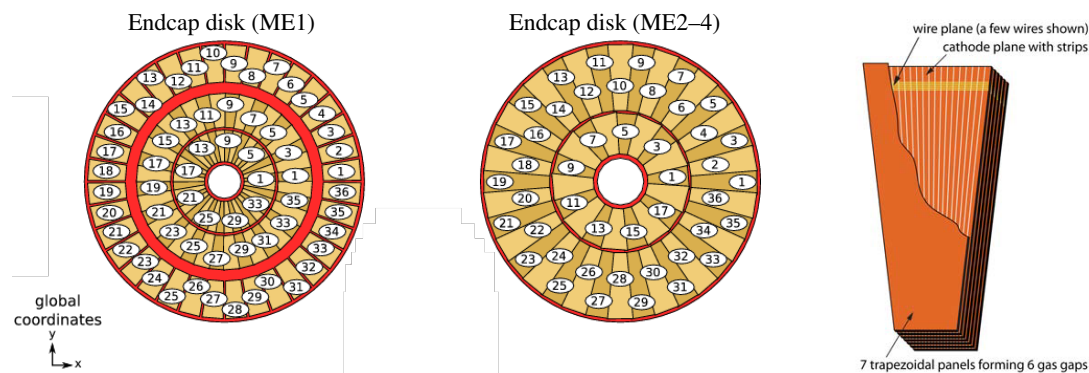


Figure 4.19: Layout CSC chambers in the endcaps in ME1 with three rings and ME2-4 with two rings (left) [81] updated for Phase-1 and the cutaway view of a cathode strip chambers (right) [61].

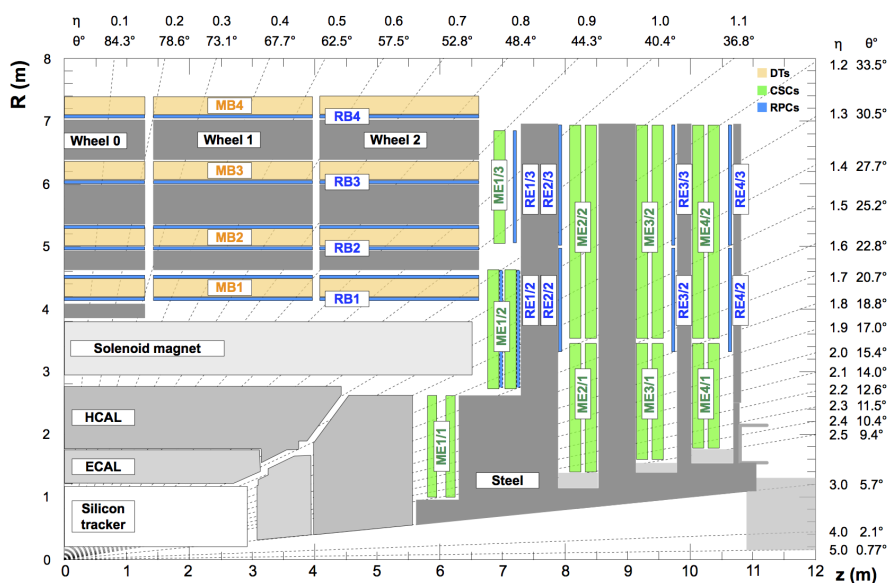


Figure 4.20: A schematic diagram showing the positions of muon chambers in a quarter of the CMS detector [77]. Thin blue boxes show where the RPCs are located.

The RPCs, distributed in both the barrel and endcaps, cover $|\eta| < 1.9$ only. There are a total of 480 RPCs modules in the barrel and 576 modules in the endcaps. Though mainly installed for triggering, they also measure a coarse position of muons. The RPCs consist of two gas-gap sandwiching readout strips. Each gas gap is a thin box-like module with two resistive plates separated by spacers. As

shown in Figure 4.21, when a muon passes through the PRC, it ionizes the gas. Under the voltage setup, the resulting avalanche electrons penetrate the resistive plates freely and are collected by the read-out strips. The muon's momentum and position can be determined by analysing the avalanche pattern. These RPC detectors have a fast response time, shorter than the 25 ns time gap between two bunch crossings, which can correctly tag the detected muons to a bunch crossing and serve as a trigger.

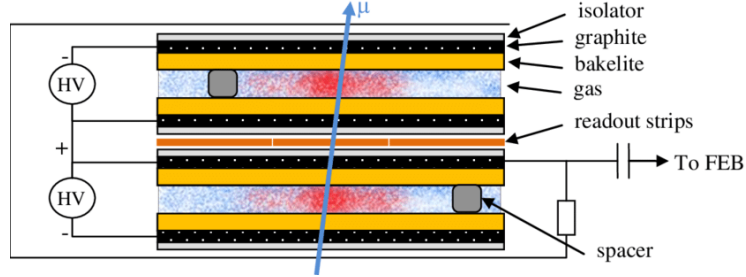


Figure 4.21: A schematic diagram showing the cross section of a double-gap RPC module in the muon detector [82].

4.3 Triggering System

In Run 2, LHC proton beams have a bunch spacing of 25 ns, corresponding to a bunch crossing frequency of 40 MHz, with about 55 simultaneous interactions per bunch crossing (also known as pileup) [61, 83, 84]. However, due to limitations in computing power, the input rate of offline data storage is roughly 1 kHz. To reduce the rate from 40 MHz down to 1kHz while keeping the more interesting rare events, a two-level trigger system is adopted by the CMS. The two sequential steps are Level-1 (L1) Trigger and High-Level Trigger (HLT).

Events are first filtered by L1 Trigger, reducing the data stream rate to approximately 100 kHz, with a short processing time $< 4 \mu\text{s}$. The hardware-based L1 Trigger makes use of the segmented readouts from HCAL, ECAL, and the muon chambers. These pieces of simplified coarse information are called trigger primitives (TP). No track information is used since track reconstruction is time-consuming. With energy deposition towers in calorimeters and hit patterns in the muon chambers, L1 Trigger filter events based on quickly reconstructed crude objects, such as muons, (indistinguishable) electrons/photons, taus, jets, and p_T^{miss} , with energy above some defined thresholds. L1 Triggers are

mainly designed to remove inelastic p - p interactions and $b\bar{b}$ processes which have production cross sections at much higher order of magnitudes.

Events with L1-accept are then further slimmed by HLT, reducing the rate to 1 kHz. With around 26000 cores, the average computing time is 260 ms. The software-based HLT makes use of the full detector readouts and performs simplified regional reconstructions. There are hundreds of HLT paths, corresponding to different selection sequences. Typically, the sequences follow four steps with increasing computational complexity to minimize the average computing time. First, by requiring events to pass a specific L1 trigger, the rate is cut down by a factor of ten with < 1 ms. Second, selections are applied on calorimetric towers, which can be reconstructed quickly in roughly 50 ms. This Level-2 (L2) calorimetric selections reduce the rate down by at least a factor of five. At Level-3 (L3), tracks are reconstructed, only within a region around the calorimetric tower, to facilitate the distinction between electrons and photons, identification of jets from b-quark decays (b-tagging), identification of hadronic decays of τ leptons etc. This step reduces the rate by a factor of four in about 100 ms. Finally, a simplified version of Particle Flow (PF), which is the offline reconstruction explained in Section 4.4, is performed in nearly 1000 ms. Selections are applied on PF objects, i.e., particle candidates, to complete the HLT.

4.4 Physics Objects Reconstruction

Elementary particles each present a different signature to the CMS detector. An overview of the unique signatures for each type of detectable final state particles is shown in Figure 4.22. All charged particles from an interaction leave hits in the inner tracker. These hits are linked to find possible trajectories, called tracks, which have curvatures from which their momentum can be measured, given the 3.8 T solenoid magnetic field [85]. Electric charges can be determined from the direction of the curves. Tracks are extrapolated backward to find intersections which are possible interaction points, called vertices. Neighbouring calorimetric towers are grouped into clusters in certain η - ϕ regions according to the energy distributions of showers from bremsstrahlung of electrons, pair production of photons, or hadronization of QCD particles. Electrons and photons have all their energy deposited in the ECAL, except for some very energetic ones which punch through and leave

some energy in HCAL. On the other hand, hadrons with $3 < p_T < 300$ GeV leave a small fraction of energy in the ECAL and are fully stopped in HCAL [86]. Energy fractions in the two calorimeters and geometries of shower energy are checked to distinguish different particles and determine their energy and direction of motion. Muons interact weakly with the CMS detector and are mainly reconstructed by matching tracks in the inner tracker to that in muon chambers. Some particles, such as neutrinos in the SM, are non-interactive and escape the CMS detector with no trace, leaving a non-zero vector sum of p_T of detected objects. The quantity p_T^{miss} represents the total energy of these escaped particles. The described reconstruction and related quality checks are done by the Particle Flow (PF) algorithm. Reconstructed final state PF objects include vertices, electrons, photons, muons, jets, and missing transverse momentum.

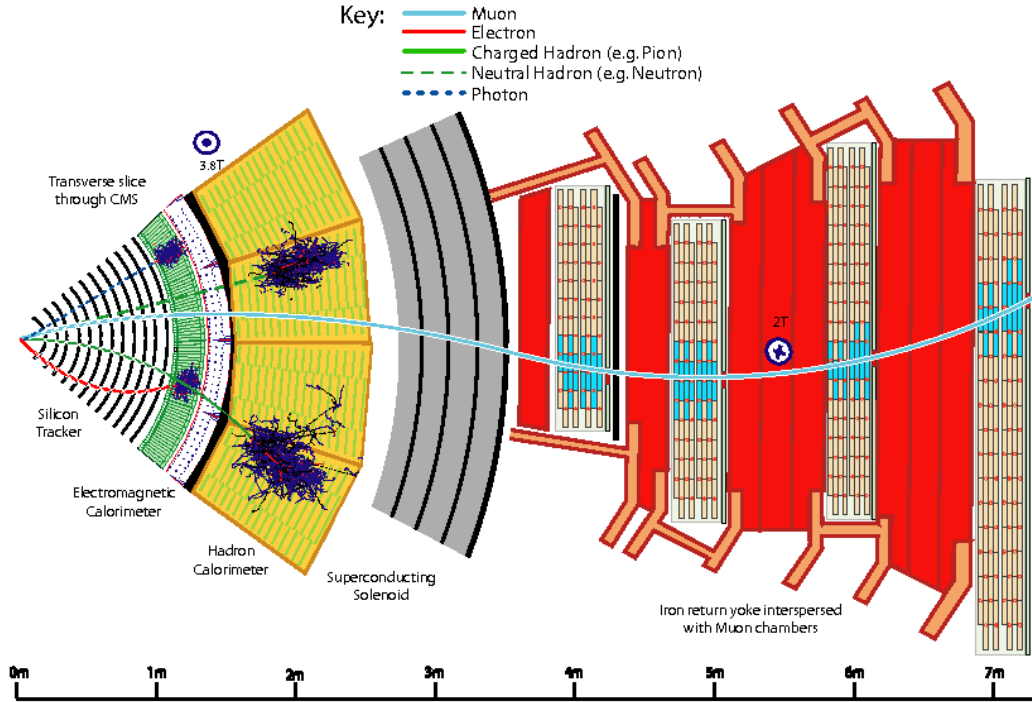


Figure 4.22: A schematic diagram showing how SM particles interact with different sub-detectors in the CMS detector [85].

4.4.1 Primary Vertex and Pileup

In Run 2, there are about 55 simultaneous interactions per bunch crossing [83]. Vertices in an event are reconstructed from all tracks that are compatible with the LHC beam spot in the x - y

plane. The vertex with the highest $\sum p_T^2$ is selected as the primary vertex, corresponding to the interaction of interest in an event, and the remaining ones are referred to as pileup vertices. Accurate determination of vertices is important for distinguishing objects from pileup background and for identifying secondary vertices, which correspond to the decays of particles with long lifetimes [87].

4.4.2 Tracks

Tracks are connected hits in the inner tracker. To find valid tracks among the numerous possible connections, tracks are built starting from a few tightly selected hits, called seeds, which are determined using the new Cellular-Automaton (CA) techniques after the Phase-1 upgrade of the pixel detector. Track seeds are composed of a pair of hits from two consecutive layers in the pixel detector that agrees with the vertex, in addition to one or two compatible hits from the remaining layers. The use of CA seeding techniques has improved track reconstruction. Based on Kalman Filtering (KF), seeds are associated to relevant hits in all tracker layers with expected patterns in the CMS environment. A track must connect at least eight hits with no more than one layer of missing hits. There are many combinations of hits, and tracking is not fast. For more efficient computation, iterative tracking first reconstructs the fewer high-confidence tracks with tighter requirements and repeats in the reduced set of unassigned hits with looser cuts.

4.4.3 Electrons and Photons

Reconstruction of electrons and photons uses readouts from the inner tracker, ECAL, and HCAL. When they transverse through the ECAL, electrons emit bremsstrahlung photons, while photons decay to electron-positron pairs which also emit bremsstrahlung photons, leaving energy showers in ECAL. The shower shape is checked against expectations. For example, an isolated prompt electron or photon should have a single-core shower. To collect all energy from the shower, “superclusters” are defined. Seeding at the local maximum of an energy tower, neighbouring crystals are clustered if they have decreasing energy and have not yet been assigned to any cluster. Due to the magnetic push, showers spread out along ϕ -direction, and thus clusters are grouped to form superclusters. In an ECAL-driven approach, superclusters are used to seed tracks to find the associated tracks efficiently. HCAL is checked for little energy deposition. Measurement of electron energy combines

the momentum of track and the shower energy from ECAL, while that of photon energy just uses ECAL shower energy.

4.4.4 Muons

Reconstruction of muons is based on hits in the inner tracker and muon chambers. Hits in the muons chambers, including DTs, CSCs, and RPCs, are fitted to recognizable patterns and grouped into stand-alone track segments. Hits in the inner tracker are joined to form the inner tracks. Global muons are reconstructed “outside-in”, by matching a stand-alone track to an inner track. Reversely, tracker muons are reconstructed “inside-out”, by extrapolating inner tracks and finding at least one track segment along. Momenta of muons are determined by the curvature of inner tracks when $p_T < 200$ GeV, and by tracks refits with the best goodness of fit otherwise.

4.4.5 Jets and Missing Transverse Momentum

Reconstruction of a jet uses readouts from the inner tracker and calorimeters. QCD particles, or partons (quarks and gluons), decay and undergo fragmentation and hadronization to form showers of partons. Calorimetric energy towers are clustered by the anti- k_T algorithms with a cone size of $\Delta R < 0.4$ for normal jets and $\Delta R < 0.8$ for more energetic large-R jets. Depending on the distance parameters, the reconstructed objects are called AK4 jets and AK8 jets respectively. The momentum vector of a jet is the vector sum of all constituent partons within the cone of a cluster.

Jets are tagged by the flavour of the quarks from which they originate. The main physics principle of flavour tagging is the secondary vertex formed by different decay lengths (i.e., long lifetimes). For example, in a process known as b-tagging, jets from b quark decays can be identified by one of the several machine learning algorithms.

Non-interactive particles, such as neutrinos, are observed as the missing transverse momentum (p_T^{miss}). By conservation of momentum, p_T^{miss} is defined as the magnitude of the negative vector sum of the transverse momenta of all objects in an event. Other high-level objects considered are H_T , which is the scalar sum of p_T of AK4 jets, and S_T , which is the scalar sum of H_T , p_T^{miss} , and p_T of leptons.

Chapter 5

Search for Pair Production of T/B in Leptonic Final States

A Higgs boson with mass near 125 GeV was discovered at the LHC a decade ago and completed the Standard Model (SM) [10, 11, 12]. While experiments have proven that the SM is a successful theory, there are still observations or questions that the SM cannot fully address, and thus a theory beyond the SM (BSM) is needed. Some BSM theories are developed to cancel the diverging corrections in Higgs mass calculation so as to avoid fine-tuning for the lightness of the Higgs boson. Among these theories, several introduced the hypothetical VLQs as an extension to the SM. Examples of such theories include Little Higgs [31, 32] and Composite Higgs [33, 34].

VLQs are introduced in Chapter 3. They are non-chiral fermions and do not gain mass through Yukawa coupling like SM fermions. They are often called top quark partners as they only mix with the third generation quarks in the SM in a simplified scenario. Interactions with other generations are suppressed by flavour-changing processes and the large t -quark Yukawa coupling. There are four types of VLQ: $X_{5/3}$, $T_{2/3}$, $B_{-1/3}$, $Y_{-4/3}$.

T and B quarks have an electric charge of $\frac{2}{3}$ and $-\frac{1}{3}$ respectively, same as their Standard Model partners, but are predicted to be a lot heavier, with masses in TeV scale. Given the heaviness, T quarks can decay into bW , tZ or tH while B quarks can decay into tW , bZ or bH . The T and B quarks can be written as singlets (T) and (B), doublets (T, B) or doublets and triplets including X and Y in the Lagrangian of the SM gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$. The branching fractions vary for different multiplets of the VLQs. In the asymptotic limit of VLQ masses tend to infinity,

singlet T (B) has a 50% chance of decaying to bW (tW) while the remaining two decay modes each has a branching ratio of 25%, and doublet T (B) has 50% chance each decaying to tH and tZ (bH and bZ), but never to bW (tW).

In previous pair production searches at $\sqrt{s} = 7$ TeV [14, 15, 16], 8 TeV [17, 18, 19, 20] by the CMS and ATLAS Collaborations, T and B quarks were excluded for masses below 0.72–0.95 TeV and 0.58–0.81 TeV respectively at 95% C.L., considering different branching fractions.

The CMS searched for pair production of vector-like T and B quarks with leptons in the final states in 13 TeV proton-proton collisions in the LHC. This dissertation reports the analysis on data collected from 2017 to 2018. An equivalent search analyzing 2016 data was published and excluded T and B masses below 1.14–1.3 TeV and 0.91–1.24 TeV [21]. The majority of the search strategies were inherited from the 2016 analysis.

The result from this analysis is combined with the equivalent 2016 search to obtain a result for the full Run 2 data. Considering all lepton search channels, the exclusion limits are improved from that of 2016 and the strongest exclusion on T mass is found, comparing to the T mass limit of 1.31–1.42 TeV from other pair production searches at 13 TeV by the CMS and ATLAS Collaborations [22, 23, 24, 25, 26].

Note that, in this chapter, all distributions and event yields of VLQ signals are obtained for branching ratios in singlet scenario unless otherwise specified.

5.1 Search Channels

This analysis searches in 2017–2018 data in two different final states with leptons: the multi-lepton channel with three or more leptons and the same-sign dilepton channel with exactly two leptons. The combined full Run 2 result is merged with a single-lepton channel with exactly one lepton for completeness.

5.1.1 Multi-Lepton (3L+) Final State

The multi-lepton channel requires events to have three or more tight leptons (defined in Section 5.3.1) in both data and simulations. These events have the highest signal efficiency in $T \rightarrow tZ$ and $B \rightarrow tW$ decays. Selected events are split into four categories by flavours of the three leptons with the highest p_T . Data events with three or more loose leptons are used to derive data-driven background.

5.1.2 Same-sign Dilepton (SS2L) Final State

Complementary to the multi-lepton channel, this analysis also searches for VLQ pairs in final states with same-sign (SS) dilepton following the strategies of a $X_{5/3}$ search in 2016 data [88, 89]. This channel is more sensitive to decay modes of $T \rightarrow tH$, where H can decay into a pair of W , and $B \rightarrow tW$. There are three categories in this channel, determined by the lepton flavours of the dilepton pair. Data events with at least two tight electrons which have reconstructed dilepton mass within 10 GeV from Z mass are used to measure the charge misidentification (CMID) rates. Data events with two or more loose leptons and events with two opposite-signed tight leptons are used to derive nonprompt and CMID backgrounds respectively.

5.2 Data and Monte Carlo

The analysis looks at the Run 2 data, proton-proton ($p-p$) collisions at $\sqrt{s} = 13$ TeV with bunch spacing of 25 ns, collected at the CMS in the LHC during 2017 and 2018 with an integrated luminosity of 41.53 fb^{-1} and 59.74 fb^{-1} [90, 91]. The lists of data samples are shown in Table 5.1 and Table 5.2. Background processes are predicted using Monte Carlo (MC) simulations. Samples of MC simulations are generated with the same bunch spacing in the RunIIFall17MiniAODv2 and RunIIAutumn18MiniAOD campaigns, which are names assigned for groups of samples generated at different times of the year to match the LHC running conditions. MC samples are listed in Table 5.3 together with their production cross sections [92]. Data and MC samples were processed in the CMSSW framework with version 10.2.16UL and analyzed with version 10.2.10.

Data events are filtered to keep only certified events recorded in the following JSON files:

- Cert_294927-306462_13TeV_EOY2017ReReco_Collisions17_JSON.txt
- Cert_314472-325175_13TeV_17SeptEarlyReReco2018ABC_PromptEraD_Collisions18_JSON.txt

Unique event IDs were checked to remove any overlap among the three dilepton dataset DoubleEG (EGamma), MuonEG and DoubleMuon. For example, an event with one electron and two muons may be stored in both MuonEG and DoubleMuon datasets.

For simplification, some background processes are grouped, namely the diboson processes (VV), triboson processes (VVV), $t\bar{t}$ plus vector boson processes (TTV) or $t\bar{t}$ plus boson processes (TTX). In the same-sign dilepton channel, the VV processes include W^+W^+ , WZ , and ZZ , the VVV processes include WWZ and WZZ , while the TTX processes include $t\bar{t}W$, $t\bar{t}Z$, and $t\bar{t}H$. The $t\bar{t}t\bar{t}$ process is a group on its own. There are also backgrounds from nonprompt (NP) sources and charge misidentification (CMID), which are derived by data-driven methods. In the multi-lepton channel, the group VV consists of WZ and ZZ samples, the group VVV contains WWW , WWZ , WZZ , and ZZZ samples, and the group TTV consists of $t\bar{t}W$ and $t\bar{t}Z$. There is also a contribution of background from nonprompt sources derived by data-driven method (ddbkg).

MC samples are produced by different matrix-element generators, listed in the second column in Table 5.3 and Table 5.4. For simulations of the 2017–2018 conditions, NNPDF_3.1 parton-distribution-function (PDF) sets at next-to-next-to-leading order (next-to-NLO) [93] are used for SM background processes and PDF4LHC15 PDF set at NLO is used for VLQ signal samples [94].

The SM $t\bar{t}$, $t\bar{t}H$, ZZ , and the 2018 WZ samples are simulated by the POWHEG v2 generator [95, 96, 97, 98]. The MADGRAPH5_aMC@NLO generator [99] version 2.4.2 is used to simulate the other SM processes. W^+W^+ is produced at leading order (LO) while the VVV, TTV, Drell-Yan, $t\bar{t}t\bar{t}$, and the 2017 WZ samples are all generated at NLO. Among them, $t\bar{t}W$ and WZ are simulated with the FxFx [100] matching scheme. Additionally, $t\bar{t}W$ sample is produced with MadSpin interface [101].

$T\bar{T}$ and $B\bar{B}$ signal MC simulations are produced by generator MADGRAPH5_aMC@NLO 2.4.2.

Signal samples at ten different mass points from 900 to 1800 GeV were considered. They are listed in Table 5.4 together with their mass-dependent production cross sections at next-to-NLO computed by the Top++2.0 program [102]. The generator assumes equal branching ratios of 33% each for $T \rightarrow bW$, tZ , and tH , and $B \rightarrow tW$, bZ , and bH . In the analysis, signal samples are scaled with branching ratios corresponding to the scenario under study.

For all MC simulations, parton showering and the underlying kinematics are simulated by the PYTHIA 8 [103] generator versions 8.226 and 8.230 for 2017 and 2018 running conditions. All samples undergo detector simulations by GEANT4 [104]. In the LHC, multiple simultaneous interactions (“pileup”) were detected at the same time from the current or previous bunch crossings. This condition must be simulated in MC samples and later re-weighted to match the mean number of interactions per bunch crossing observed in data.

Table 5.1: 2017 data sample names and the specific processing group for locating the files in the CMS database.

Dataset	Reconstruction Group
DoubleEG	Run2017B-31Mar2018-v1
DoubleEG	Run2017C-31Mar2018-v1
DoubleEG	Run2017D-31Mar2018-v1
DoubleEG	Run2017E-31Mar2018-v1
DoubleEG	Run2017F-31Mar2018-v1
DoubleMuon	Run2017B-31Mar2018-v1
DoubleMuon	Run2017C-31Mar2018-v1
DoubleMuon	Run2017D-31Mar2018-v1
DoubleMuon	Run2017E-31Mar2018-v1
DoubleMuon	Run2017F-31Mar2018-v1
MuonEG	Run2017B-31Mar2018-v1
MuonEG	Run2017C-31Mar2018-v1
MuonEG	Run2017D-31Mar2018-v1
MuonEG	Run2017E-31Mar2018-v1
MuonEG	Run2017F-31Mar2018-v1
Total int. lumi	41.53 fb ⁻¹

Table 5.2: 2018 data sample names and the specific processing group for locating the files in the CMS database.

Dataset	Reconstruction Group
EGamma	Run2018A-17Sep2018-v2
EGamma	Run2018B-17Sep2018-v1
EGamma	Run2018C-17Sep2018-v1
EGamma	Run2018D-22Jan2019-v2
DoubleMuon	Run2018A-17Sep2018-v2
DoubleMuon	Run2018B-17Sep2018-v1
DoubleMuon	Run2018C-17Sep2018-v1
DoubleMuon	Run2018D-PromptReco-v2
MuonEG	Run2018A-17Sep2018-v1
MuonEG	Run2018B-17Sep2018-v1
MuonEG	Run2018C-17Sep2018-v1
MuonEG	Run2018D-PromptReco-v2
Total int. lumi	59.7 fb ⁻¹

Table 5.3: 2017 and 2018 background MC Samples from RunIIFall17MiniAODv2 and RunIIAutumn18MiniAOD campaigns. The third column lists the cross-section of each process [92] and the last two columns indicate whether the processes are used in the same-sign dilepton and/or the multi-lepton channel.

MC for Background	Generators	Cross Section(pb)	SS2L	3L
WpWpJJ	madgraph-pythia8	0.03711	✓	
WZTo3LNu	amcatnloFXFX-pythia8 (2017)	4.42965	✓	✓
	powheg-pythia8 (2018)			
ZZTo4L	powheg-pythia8	1.256	✓	✓
WWW_4F	amcatnlo-pythia8	0.2086		✓
WWZ_4F	amcatnlo-pythia8	0.1651	✓	✓
WZZTo3L1Nu2Q	amcatnlo-pythia8	0.00256		✓
WZZ_ZTo2L	amcatnlo-pythia8	0.000567		✓
ZZZ	amcatnlo-pythia8	0.01398		✓
TTWJetsToLNu	amcatnloFXFX-madspin-pythia8	0.2043	✓	✓
TTZToLLNuNu_M-10	amcatnlo-pythia8	0.2529	✓	✓
ttHTToNonbb_M125	powheg-pythia8	0.215	✓	
TTTT	amcatnlo-pythia8	0.009103	✓	
MC for Checks	Generators	Cross Section (pb)	SS2L	3L
TTTo2L2Nu	powheg-pythia8	831.76×0.105	✓	✓
TTToSemiLeptonic	powheg-pythia8	831.76×0.438	✓	
DYJetsToLL_M-50	amcatnloFXFX-pythia8	6077.22	✓	

Table 5.4: 2017 and 2018 signal MC Samples from RunIIFall17MiniAODv2 and RunIIAutumn18MiniAOD campaigns respectively. The last column lists the cross-section of each process [105].

Signal	Generators	Cross Section (fb)
TprimeTprime_M-900	madgraph-pythia8	90.3
TprimeTprime_M-1000	madgraph-pythia8	44.0
TprimeTprime_M-1100	madgraph-pythia8	22.4
TprimeTprime_M-1200	madgraph-pythia8	11.8
TprimeTprime_M-1300	madgraph-pythia8	6.39
TprimeTprime_M-1400	madgraph-pythia8	3.54
TprimeTprime_M-1500	madgraph-pythia8	2.00
TprimeTprime_M-1600	madgraph-pythia8	1.148
TprimeTprime_M-1700	madgraph-pythia8	0.666
TprimeTprime_M-1800	madgraph-pythia8	0.391
BprimeBprime_M-900	madgraph-pythia8	90.3
BprimeBprime_M-1000	madgraph-pythia8	44.0
BprimeBprime_M-1100	madgraph-pythia8	22.4
BprimeBprime_M-1200	madgraph-pythia8	11.8
BprimeBprime_M-1300	madgraph-pythia8	6.39
BprimeBprime_M-1400	madgraph-pythia8	3.54
BprimeBprime_M-1500	madgraph-pythia8	2.00
BprimeBprime_M-1600	madgraph-pythia8	1.148
BprimeBprime_M-1700	madgraph-pythia8	0.666
BprimeBprime_M-1800	madgraph-pythia8	0.391

5.3 Physics Object Selection

In both data and MC samples, events and physics objects are checked against various requirements to ensure their qualities before selecting events for the analysis. The definitions of accepted events and physics objects are given in this section. Particle Flow (PF) objects defined in Section 4.4 are used.

All data and MC events are required to pass several quality filters. There are filters dedicated to remove events with anomalous detector signals due to electronics noise in the HCAL (“HBHENoiseFilter” and “HBHENoiseIsolatedFilter”) or the calibration and bias of ECAL readouts (“ecalBadCalibFilter” and “EcalDeadCellTriggerPrimitiveFilter”). There is also a filter designed to handle events induced from beam halo [106] (“globalSuperTightHalo2016Filter”) and another filter handling events with poorly reconstructed muons in PF (“BadPFMuonFilter”). Acting on data samples only, there is a filter for events with anomalous readouts from supercrystals in EE.

Accepted events must have a well-reconstructed vertex. The primary vertex (PV) is the reconstructed vertex with the highest $\sum p_T^2$ which corresponds to the interesting interaction that will be studied. The PV must pass the “goodVertices” filter, together with cuts on its coordinates $z < 24.0$ cm and $\rho < 2.0$ cm. For good estimations of PV properties, the quality of vertex fitting is controlled by the requirement of degrees of freedom ≥ 4 .

Pileup (PU) refers to the other interactions that are not from the PV [107]. To match the detector simulation of MC samples to the actual PU environment in the LHC, PU is recorded [108]. In MC simulations, distributions of the number of true interactions are stored without any selection. In data, events are filtered by JSON files listed in Section 5.2 before calculating the distributions of PU using the instantaneous luminosity, revolution frequency, and total inelastic p - p cross section. MC samples are reweighted to correct for the differences in PU distributions between data and MC samples. The uncertainty due to pileup reweighting is calculated by setting the minimum bias cross section to 69.2 mb and varying it by $\pm 4.6\%$.

Final state physics objects reconstructed by the PF algorithm are further required to pass some quality control sequences known as IDs. There are usually two to three working points of stringency for each type of ID.

5.3.1 Leptons

In this analysis, reconstructed PF electrons are checked with a multivariate (MVA) ID with no isolation (mvaEleID-Fall17-noIso-V2) and a mini-isolation ID. Two working points of the MVA ID are considered: wp90 with an efficiency (or passing rate) of 90% and wpL with an efficiency of 98%. Mini-isolation I_{mini} is a measure of how isolated a lepton is. It considers all hadron and photon candidates within a cone around the lepton with a p_T -dependent radius $\mathcal{R}$ defined in Equation 5.1. I_{mini} is defined as the ratio of the scalar sum of p_T of all candidates in the cone, corrected for PU by the effective area method [109], to p_T of the lepton. Setting a maximum cut on I_{mini} removes leptons produced in jets while using a p_T -dependent cone size maintains a relatively high efficiency in boosted events, such as decays of the heavy VLQs.

$$\mathcal{R} = \frac{10 \text{ GeV}}{\min(\max(p_T, 50 \text{ GeV}), 200 \text{ GeV})} \quad (5.1)$$

A “tight” electron is defined as passing the MVA ID at wp90 and $I_{\text{mini}} < 0.1$, with a combined efficiency of roughly 86–89%, depending on p_T and η . A “loose” electron is one that passes the MVA ID at wpL and $I_{\text{mini}} < 0.4$, with an efficiency of approximately 97%.

To account for the inefficiencies of electron identification, three scale factors are applied for the MVA ID, mini-isolation, and reconstruction respectively. Scale factors are the ratios of efficiencies measured in data and MC. The official electron MVA ID and reconstruction scale factors provided by the EGamma physics object group (EGamma POG) of the CMS collaboration [110] are directly applied. For mini-isolation, scale factors are applied in the MC for the efficiencies of a MVA wp90 (wpL) electron passing the $I_{\text{mini}} < 0.1$ (0.4) according to its p_T and η , so that the MC efficiencies matches the data. Scale factors are measured in data and MC using the tag-and-probe method [111, 112]. The approved result is shown in Figures 5.1 and 5.2 [113]. The Mini-isolation scale factors are applied with a flat systematic uncertainty of 1.5%.

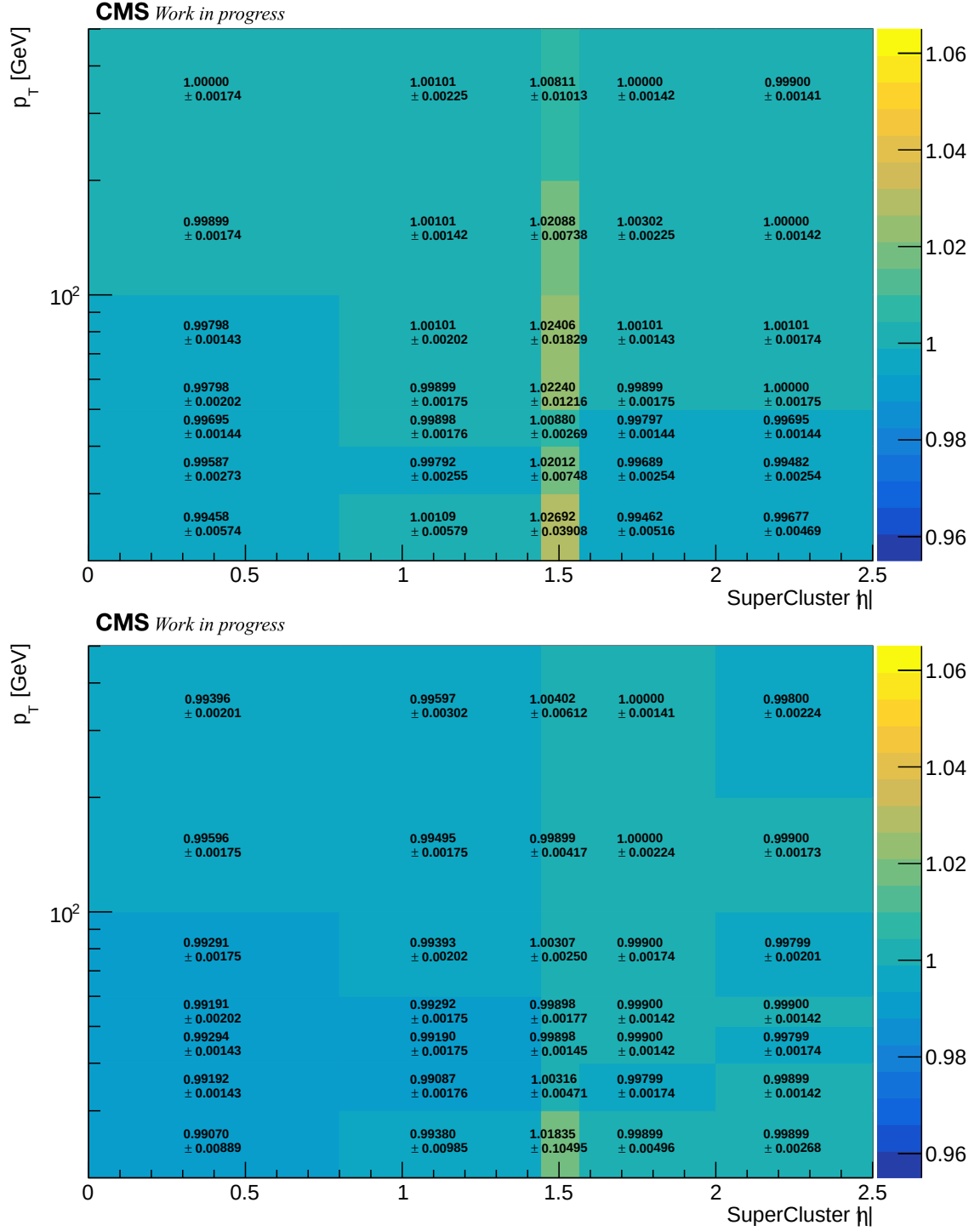


Figure 5.1: Scale factors of electrons passing $I_{\text{mini}} < 0.1$ (top) and < 0.4 (bottom) in bins of η and p_T measured in 2017 data and simulations.

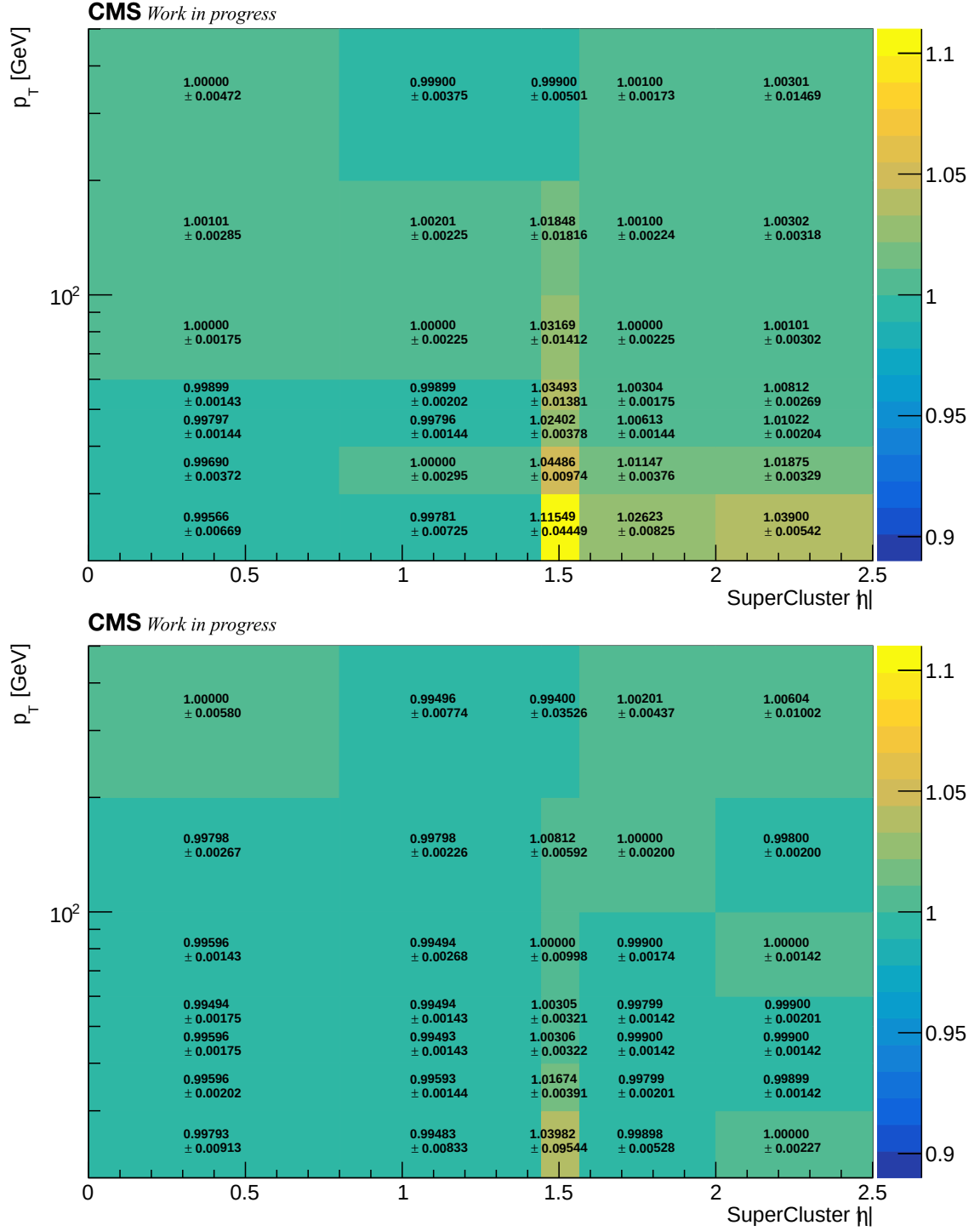


Figure 5.2: Scale factors of electrons passing $I_{\text{mini}} < 0.1$ (top) and < 0.4 (bottom) in bins of η and p_T measured in 2018 data and simulations.

PF muons are first checked against the following criteria for a valid reconstruction:

- Reconstructed as a global muon in PF algorithm
- The fitting of tracks and hits with $\chi^2/\text{ndof} < 10$
- Number of hits in muon chamber > 0
- Number of muon chambers with matched segments > 1
- $d_{xy} < 2$ mm between the muon track and the PV
- $d_z < 5$ mm between the muon track and the PV
- Number of hits in pixel detector > 0
- Number of hits in inner track > 5

The qualified PF global muons are then checked with a series of cuts recommended by CMS Muon POG, known as cut-based ID, and the mini-isolation. Muons are “tight” if they pass Muon POG’s cut-based ID at tight working point and $I_{\text{mini}} < 0.1$, with a combined efficiency of approximately 88–98%, while the definition of “loose” muons combines Muon POG’s cut-based ID at Loose working point and $I_{\text{mini}} < 0.4$, with an efficiency of roughly 99%.

Due to the well-established muon system in the CMS, inefficiencies due to reconstruction is very close to zero and no correction is needed. However, similar to electrons, correction scale factors are applied on muons for identification and isolation requirements. The official scale factors of cut-based ID measured by the Muon POG [114, 115] are applied. The scale factors on mini-isolation are also measured in data by tag-and-probe method. The results were approved by the Muon POG and shown in Figures 5.3 and 5.4 [116]. The isolation scale factors are applied with a flat systematic uncertainty of 1%.

All lepton collections are slimmed by selections of $p_T > 30$ GeV and $|\eta| < 2.4$ before being considered in the analysis.

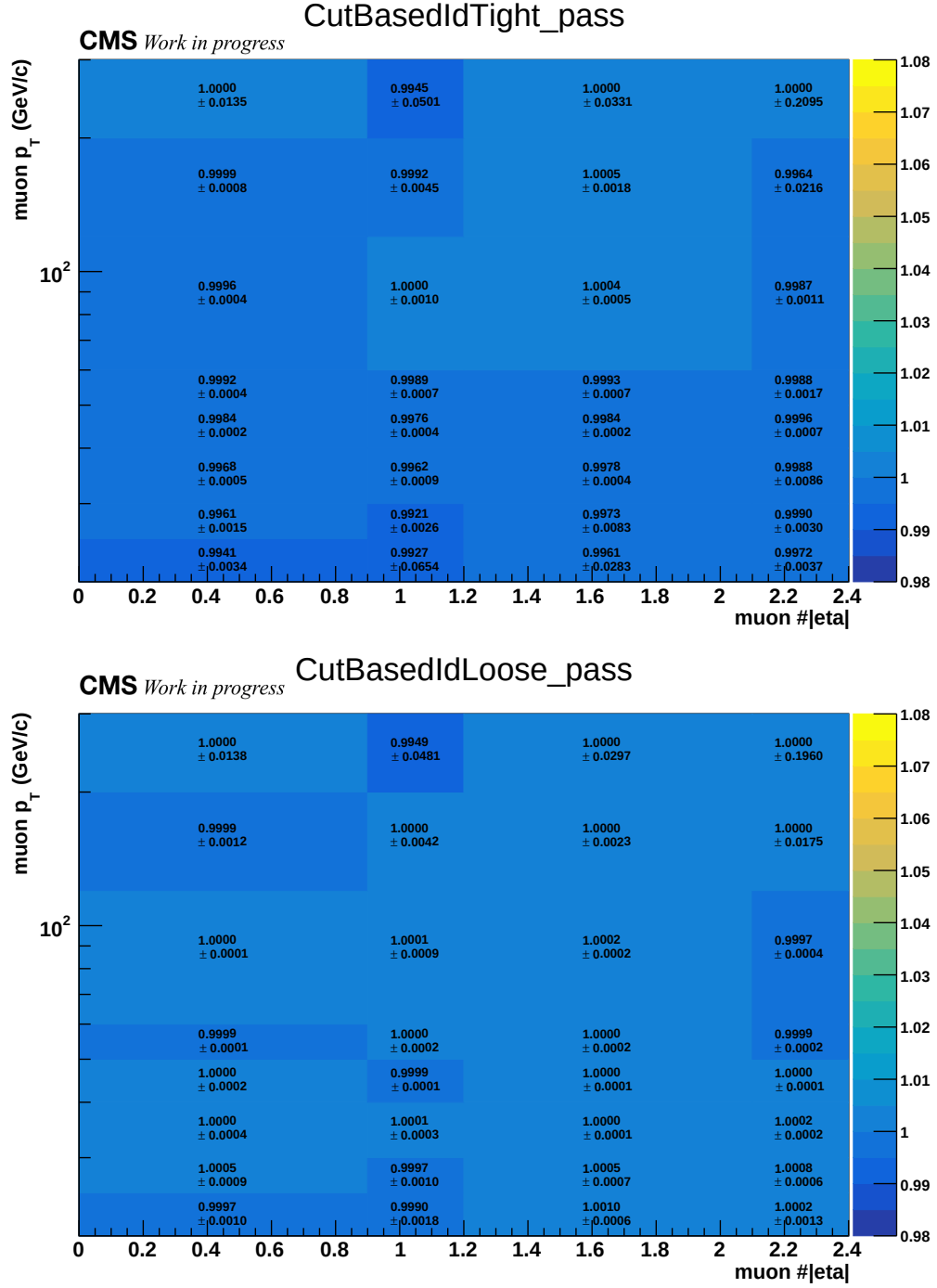


Figure 5.3: Scale factors of muons passing $I_{\text{mini}} < 0.1$ (top) and < 0.4 (bottom) in bins of η and p_T measured for 2017 data and simulations.

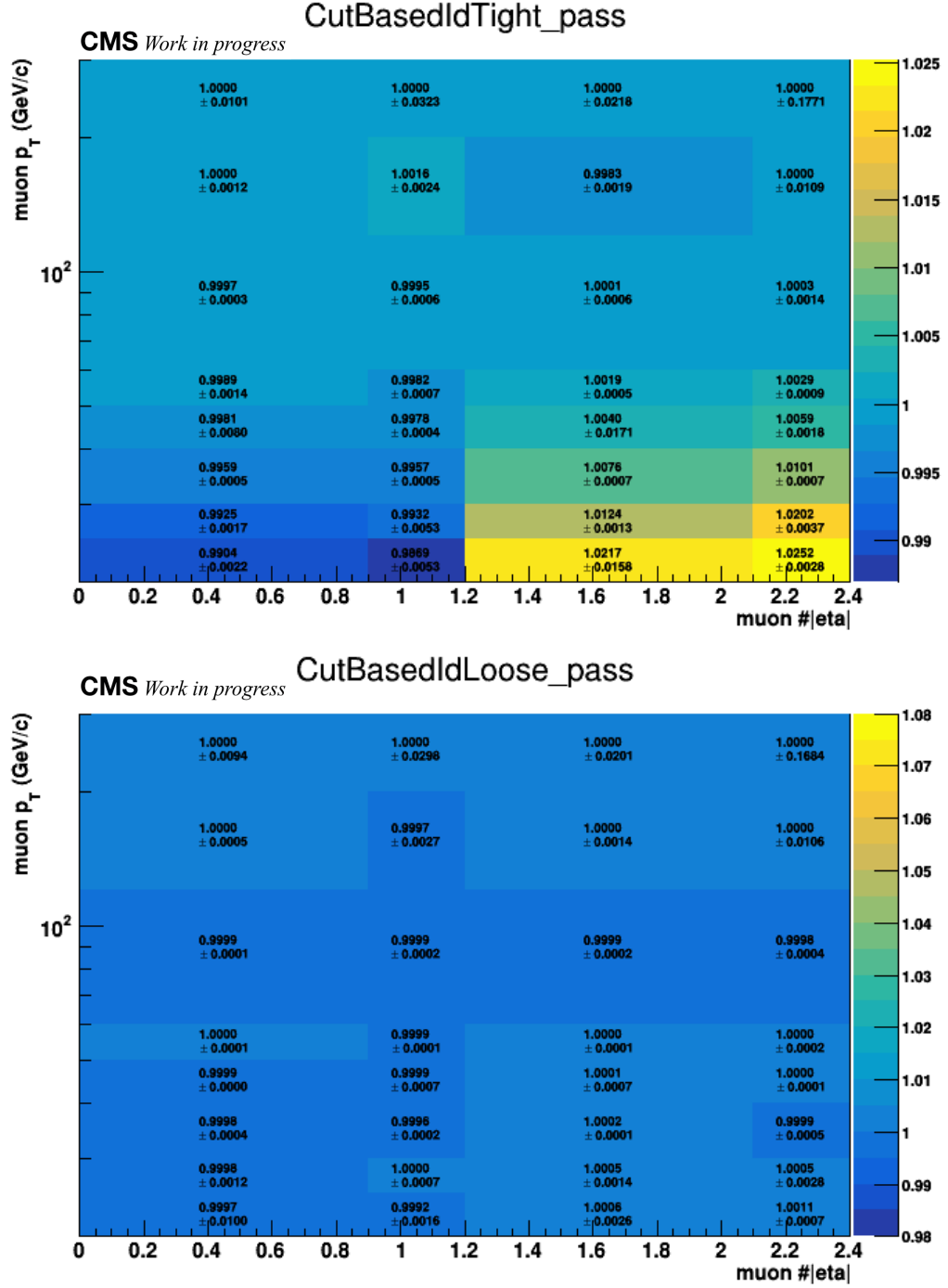


Figure 5.4: Scale factors of muons passing $I_{\text{mini}} < 0.1$ (top) and < 0.4 (bottom) in bins of η and p_T measured for 2018 data and simulations.

5.3.2 Jets and Missing Transverse Momentum

This analysis considers PF jets with pileup contributions removed by the Charge-Hadron-Subtraction (CHS) algorithm [107]. Jet energy depositions in ECAL and HCAL are clustered by the anti- k_T algorithm with a size parameter of 0.4 (AK4). AK4 jets are “cleaned” by removing leptons with separation $\Delta R < 0.4$. Cleaned AK4 jets are required to pass the cut-based jet ID at tight working point recommended by the CMS JetMET group [117]. The jet collection in the analysis keeps only cleaned AK4 jets with $p_T > 30$ GeV and $|\eta| < 2.4$. For measured energies of AK4 CHS jets, “L1FastJet”, “L2Relative” and “L3Absolute” corrections are applied in both data and MC samples, to correct for pileups, and the η - and p_T -dependence of detector responses [118, 119, 120]. “L2L3Residual” corrections are applied on data to address small residual differences in responses between data and MC. Scale factors are applied for jet energy correction (JEC) and jet energy resolution (JER) using the hybrid method. Table 5.5 lists the JEC and JER versions used [121, 122].

Table 5.5: JEC/JER versions used in the analysis

Era	JEC	JER
2017	Fall17_17Nov2017_V32	Fall17_V3
2018	Autumn18_V19	Autumn18_V7

Flavour tagging, which predicts the likeliness of a jet originating from a particular hadron type, is an important feature of jets. In this analysis, since each VLQ decays to a third generation quark with a boson, jets from b quarks are expected. The b-tagging of a AK4 jet is determined by a multivariate discriminator, called DeepJet [123, 124], at medium working point, which has a mis-identification rate of 1% against jets from light partons and a corresponding tagging efficiency near 82% for jets with $p_T > 30$ GeV [125]. The corresponding scale factors are applied with the jet-by-jet b-tagging status update method [126, 127].

Missing transverse momentum p_T^{miss} is the magnitude of the negative vector sum of p_T of all objects in an event. The PF reconstructed p_T^{miss} , or MET, are updated by “Type-1” corrections to propagate the changes due to jet energy corrections, which are derived using all the jets in an event with have $p_T > 15$ GeV, electromagnetic energy fraction < 0.9 and no overlap with any muon. Other quantities considered in the analysis are the scalar sum of AK4 jets’ p_T , labelled H_T , and the scalar sum of H_T , p_T^{miss} and leptons’ p_T , labelled S_T .

5.4 Event Selection

5.4.1 Triggers for the Analysis

For both search channels in this analysis, events are required to have at least two leptons. Data collected at the CMS with dilepton triggers listed in Table 5.6 are considered.

Table 5.6: HLT trigger paths considered in the multi-lepton and same-sign dilepton analysis.

Category	Trigger path
Multi-Lepton Channel	
eee	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v
$ee\mu$	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v
$e\mu\mu$	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v
$\mu\mu\mu$	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8_v HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8_v
SS2L Channel	
ee	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v
$e\mu$	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v
$\mu\mu$	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8_v

A SUSY multi-lepton analysis using a similar set of triggers have measured the trigger efficiencies and corresponding scale factors [128]. The SUSY analysis measured average trigger efficiencies of approximately 93–94% in 2017 data and 96–98% in 2018 data. The SUSY group used a di-electron trigger without DZ filtering, but both the measured efficiencies and scale factors for DZ filtering are very close to unity. They also used a di-muon trigger with a lower mass requirement, but scale factors are measured by tag-and-probe method which requires di-muon masses in a Z mass window way above the trigger mass criteria. Thus, their measured scale factors should be applicable to our analysis. According to the referenced literature, uncertainty of 1% is applied for each lepton leg (except for leading leg of di-electron trigger) and a 0.3% for DZ filtering. For the leading leg of di-electron trigger, upward and downward uncertainties are applied bin by bin, according to values reported in the referenced literature. Given that the SUSY group used a different definition of “tight” leptons, after comparing between our own calculation and SUSY’s measurement of scale factors

of cross-flavour trigger efficiencies, an extra uncertainty of 2% is added to each lepton leg. This resolution is agreed by the contact person from EGamma and Muon POGs.

5.4.2 Base Offline Event Selection

Leptons considered in this analysis have $p_T > 30$ GeV and are produced within $|\eta| < 2.4$. These requirements are suitable for the triggers described in Section 5.4.1. Lepton IDs are defined in Section 5.3.1. This analysis considers only jets with $p_T > 30$ GeV and $|\eta| < 2.4$, as described in Section 5.3.2.

In multi-lepton channel, events are required to have at least one AK4 jet, as expected from VLQ decays, and $p_T^{\text{miss}} > 20$ GeV, for neutrinos from leptonic decays of W bosons. Although the p_T^{miss} requirement throws away some events with Z boson(s), processes with W bosons in final state have higher cross-sections. In addition, events are also required to have at least one b-tagged AK4 jet, for b from decays of T , and invariant masses of same-flavour opposite-sign (OS) lepton pairs $M_{\text{OS}\ell\ell} > 20$ GeV (when present), to reject low mass resonances. Requirements of b-tagging follow the descriptions in section 5.3.2. Table 5.7 shows the cut-flow of pre-selections in four subsets of background processes and signal events, at several mass points, generated in 2018 running conditions. Note that the p_T^{miss} cut has very little effect on signals but is significant in background processes.

In SS2L channel, event selection criteria is modified from the 2016 $X_{5/3}$ search [88, 89]. Besides the cut on lepton multiplicity, events are required to have at least three AK4 jets and a leading lepton with $p_T > 40$ GeV. This channel also requires that invariant mass of the same-sign dilepton pair $M_{\text{SS}\ell\ell} > 20$ GeV, to reject decays from quarkonia. SS2L channel vetoes di-electron events with an invariant mass M_{ee} within the range of $M_Z \pm 10$ GeV to reject Z boson decays. Another associated veto considers all possible dilepton pairs which combines one lepton from the SS leptons and a loose lepton in an event, and rejects the event if any of the invariant masses falls within $M_Z \pm 10$ GeV. Table 5.8 and 5.9 show the yields of each background source when requiring the presence of a SS lepton pair in an event and after further requiring that there is no tight lepton other than the SS lepton pair.

Table 5.7: Unscaled event counts after pre-selections in four random subsets of 2018 MC samples. Trigger cut in this table accepts more HLT paths than the ones adopted by this analysis. The lower part of the table shows the event counts in the same subset which passed the analysis triggers in Table 5.6 according to the flavour categories. Signal samples listed here have equal branching fractions for each of the three decay modes.

Samples	T1000	T1200	T1500	T1800	ttW	ttZ	ZZZ	ZZ
No selection	2962	2865	2772	2556	81776	76037	15472	88194
Trigger	2962	2865	2772	2556	81776	76037	15472	88194
Primary Vertex	2962	2865	2772	2556	81776	76037	15472	88194
Pass filters	2944	2838	2743	2510	81654	75937	15450	88164
Min Loose Leptons	182	160	191	198	3133	10052	1051	11829
Min jet multiplicity	182	160	191	198	3065	9978	891	5024
Max jet multiplicity	182	160	191	198	3065	9978	891	5024
Leading jet pt	182	160	191	198	3065	9978	891	5024
MET	182	157	188	196	2946	9395	723	3809
Selected Triggers	93	84	96	101	1010	4223	347	1640
$M_{\text{OS}\ell\ell}$	90	83	94	100	999	4145	345	1590
b-tagged jet multiplicity	81	79	85	90	853	3554	86	107

Table 5.8: Background yields of events in 2017 samples with two same-sign tight leptons and that with an additional cut of exactly two tight leptons. Uncertainties on nonprompt background yields are evaluated as shifted histograms, and thus no value is quoted.

Sample	SSDL	nLep = 2	% decrease
TTZ	357.58 ± 46.10	130.09 ± 16.78	63.6
TTW	350.90 ± 66.81	299.28 ± 56.99	14.7
TTH	124.55 ± 40.39	95.25 ± 30.88	23.5
TTTT	11.56 ± 2.92	8.48 ± 2.15	26.6
WZ	6372.64 ± 1390.20	3091.71 ± 674.52	51.5
ZZ	1360.55 ± 287.50	410.45 ± 86.74	69.8
WpWp	173.41 ± 131.10	173.25 ± 130.97	0.1
WWZ	62.20 ± 18.51	24.97 ± 7.47	59.9
NonPrompt	$12137.79 \pm \sigma$	$9653.80 \pm \sigma$	20.5
ChargeMisID	14969.95 ± 5164.6	14969.95 ± 5164.6	0.0

Table 5.9: Background yields of events in 2018 samples with two same-sign tight leptons and that with an additional cut of exactly two tight leptons. Uncertainties on nonprompt background yield are evaluated as shifted histograms, and thus no number is quoted in this table.

Sample	SSDL	nLep = 2	% decrease
TTZ	552.01 ± 75.66	205.35 ± 28.16	62.8
TTW	529.53 ± 101.66	452.48 ± 86.87	14.6
TTH	181.72 ± 99.41	138.91 ± 75.99	23.6
TTTT	17.75 ± 5.00	13.21 ± 3.72	25.6
WZ	8269.24 ± 1855.20	4039.36 ± 906.39	51.2
ZZ	1993.93 ± 966.07	610.45 ± 295.78	69.4
WpWp	269.29 ± 208.34	269.11 ± 208.21	0.1
WWZ	91.09 ± 79.97	34.36 ± 30.18	62.3
NonPrompt	$14828.05 \pm \sigma$	$11352.91 \pm \sigma$	23.4
ChargeMisID	16910.38 ± 10146.2	16910.38 ± 10146.2	0.0

5.4.3 Control Region and Signal Region

The multi-lepton channel is most sensitive to signal decays of $T\bar{T} \rightarrow tZ + tZ/tH/bW$ or $B\bar{B} \rightarrow tW + tW/bZ/bH$. In these decay modes, a minimum of three jets are produced. Thus, the signal region (SR) is defined to have at least three AK4 jets that passed the requirements in Section 5.3.2. A control region (CR), which does not overlap with the signal region, is needed to measure the lepton fake rates, as described in Section 5.5.4, and check for data-background agreement. Two control regions are defined, labelled CR2 and CR1, based on the jet multiplicity. CR2 has events with exactly two AK4 jets and is used for measurement of lepton fake rates. CR1 has exactly one AK4 jet and is used for validation. All the aforementioned selections are applied on control regions, including at least one b-tagged AK4 jets.

In SS2L channel, the most sensitive signal decays are $T\bar{T} \rightarrow tH + tH/tZ/bW$ and $B\bar{B} \rightarrow tW + tW/bZ/bH$, which produce a minimum of four jets when there are only two leptons in the event. So the signal region (SR) is defined to have exactly two tight leptons with same-sign charges and at least four AK4 jets. The discriminating variable in SS2L channel is H_T^{lep} , defined as the scalar sum of all AK4 jets' p_T and the leptons' p_T . A control region is again needed for validations. Two control regions are defined: a CR with all SR cuts but limiting to the lower H_T^{lep} region where signal is almost absent and another CR with exactly three AK4 jets in which full H_T^{lep} distributions are utilized.

Event selections for both channels are summarized in Table 5.10 below.

Concerning the HEM issues in 2018 [129], η - ϕ distributions of electrons and jets together with distributions of down-scaled jet p_T in MC samples were studied. For multi-lepton channel, no electron veto or correction is required. For SS2L channel, electrons in $-2.5 < \eta < -1.479$ and $-1.55 < \phi < -0.9$ are vetoed in 2018 data and MC samples, and no correction is applied on jets.

Table 5.10: Table summarizing event selections in multi-lepton and same-sign dilepton channels.

multi-lepton	SS2L
Object Reconstruction	
Lepton $p_T > 30$ GeV and $ \eta < 2.4$	
AK4 jets $p_T > 30$ GeV and $ \eta < 2.4$	
Offline Selection	
• ≥ 3 loose leptons • $p_T^{\text{miss}} > 20$ GeV • ≥ 1 AK4 jets • ≥ 1 b-tagged AK4 jet • if OS pairs are present, $\min(M_{\text{OS}\ell\ell}) > 20$ GeV	• ≥ 2 loose leptons and ≤ 2 tight leptons • Leading lepton $p_T > 40$ GeV • ≥ 3 AK4 jets • veto $76.1 < M_{ee} < 106.1$ GeV (ee only) • veto $76.1 < M(\ell + \ell_{\text{loose}}) < 106.1$ GeV (ℓ from SS pair + any loose leptons) • $M_{\text{SS}\ell\ell} > 20$ GeV
Signal Region	
• ≥ 3 tight leptons • ≥ 3 AK4 jets	• exactly 2 SS tight leptons • ≥ 4 AK4 jets • $H_T^{\text{lep}} > 400$ GeV
Control Regions	
1. CR1 (for cross-check) • exactly 3 leptons • exactly 1 AK4 jet 2. CR2 (for nonprompt rates) • exactly 3 leptons • exactly 2 AK4 jet	1. low H_T^{lep} • ≥ 4 AK4 jets • $H_T^{\text{lep}} < 1000$ GeV 2. nJet = 3 • exactly 3 AK4 jets

5.5 Background Estimation

There are three types of background:

- Prompt background: This background type includes events from Standard Model processes with the same final states that are defined in Section 5.1.1 and 5.1.2. They are predicted by Monte Carlo simulations. There are diboson processes (VV), triboson processes (VVV), top quark pairs with gauge or Higgs bosons (TTV/TTX) and four tops (TTTT). The details of MC samples used in each group for each of the channels are given in Section 5.2.
- Non-prompt (NP) background: This includes events with fakes, such as nonprompt leptons that passed tight lepton ID, jets mis-reconstructed as a lepton, and energetic leptons decaying from jets that passed the isolation requirements. To predict the contribution from this type of background, data events with leptons passing loose ID are used. The measurement of prompt and nonprompt rates and the data-driven background prediction will be described in the Sections 5.5.2–5.5.4.
- Charge misidentification (CMID) background: This is a background source **unique to SS2L channel**. It includes events that passed the requirement of same-sign charges due to wrongly reconstructed lepton charges. CMID background is predicted using data events with opposite-signed charges. The measurement of CMID rates and the data-driven prediction will be described in Section 5.5.8.

5.5.1 Corrections on MC

On all MC simulations, the following corrections are applied:

- Pileup reweighting (described in Section 5.3)
- Scale factors of Lepton IDs, lepton isolations, electron reconstructions, and triggers
- Prefiring weights in 2017 samples to correct for the shifted ECAL pulses which affect the timing calibration [84]

5.5.2 Matrix Method for Non-prompt Background Estimation

All leptons passing the loose ID can be described in two ways: tight/loose (passing loose ID, but not tight ID) and prompt/nonprompt. Prompt leptons can be tight or loose. So do nonprompt leptons. An event can be categorized according to the number of tight and loose leptons or by the number of prompt and nonprompt leptons. Since the total number of event is the same, by defining the fractions of prompt or nonprompt leptons that pass the tight ID as prompt rate (p) and fake rate (f) respectively, the number of events in a certain prompt/nonprompt category can be related to the number of events in different tight/loose categories. Assuming that p and f are well estimated in the parameter space of interest, the relationship can be generalized to Equation 5.2, where $\mathcal{M}$ is the relation matrix, subscripts p or f denote a prompt or a nonprompt lepton in the event, and subscripts t or l denote a tight or a loose lepton respectively [130].

$$\begin{pmatrix} N_{t...t} \\ N_{t...l} \\ \vdots \\ N_{l...l} \end{pmatrix} = \mathcal{M} \begin{pmatrix} N_{p...p} \\ N_{p...f} \\ \vdots \\ N_{f...f} \end{pmatrix} = \mathcal{M} \mathcal{M}^{-1} \begin{pmatrix} N_{t...t} \\ N_{t...l} \\ \vdots \\ N_{l...l} \end{pmatrix} \quad (5.2)$$

For dilepton events, there are three possible combinations of leptons: tight-tight, tight-loose, and loose-loose / prompt-prompt, prompt-nonprompt, and nonprompt-nonprompt. In the simplified case with same flavour leptons ($ee/\mu\mu$), the relation matrix $\mathcal{M}$ is a 3×3 matrix which can be easily written, as shown in Equation 5.3, in terms of prompt rate p and fake rate f .

$$\mathcal{M} = \begin{pmatrix} p^2 & pf & f^2 \\ 2p(1-p) & f(1-p) + p(1-f) & (1-f)^2 \\ (1-p)^2 & (1-p)(1-f) & (1-f)^2 \end{pmatrix} \quad (5.3)$$

From the first row of the relation matrix, corresponding to events with two tight leptons, as required in SS2L channel in this analysis, one can tell that the N_{pp} term is the real events in which production processes (both SM and/or VLQs) actually decay to two leptons. The remaining terms, $pf \cdot N_{pf} + f^2 \cdot N_{ff}$, contribute to the nonprompt background which are events passing SS2L selection due to

the presence of one or more nonprompt lepton. See Equation 5.4.

$$N_{tt} = \overbrace{p^2 \cdot N_{pp}}^{real} + \overbrace{pf \cdot N_{pf} + f^2 \cdot N_{ff}}^{fake} \quad (5.4)$$

In multi-lepton channel, the relation matrix $\mathcal{M}$ is a 4×4 matrix. The four rows correspond to N_{ppp} , N_{ppf} , N_{pff} , and N_{fff} / N_{ttt} , N_{ttl} , N_{tll} , and N_{lll} . Similarly, in the simplified case with same flavour leptons ($eee/\mu\mu\mu$), the relation matrix can be expressed in terms of prompt rate p and fake rate f as shown in Equation 5.5.

$$\mathcal{M} = \begin{pmatrix} p^3 & p^2 f & p f^2 & f^3 \\ 3p^2(1-p) & p^2(1-p) + 2p(1-p)f & 2pf(1-f) + (1-p)f^2 & 3f^2(1-f) \\ 3p(1-p)^2 & 2p(1-p)(1-f) + (1-p)^2 f & p(1-f)^2 + 2(1-p)f(1-f) & 3f(1-f)^2 \\ (1-p)^3 & (1-p)^2(1-f) & (1-p)(1-f)^2 & (1-f)^3 \end{pmatrix} \quad (5.5)$$

Among events with three tight leptons (i.e., the first row of the matrix), as required in multi-lepton channel, only the contribution from N_{ppp} is real. The remaining terms are the fakes, i.e., nonprompt background. See Equation 5.6.

$$N_{ttt} = \overbrace{p^3 \cdot N_{ppp}}^{real} + \overbrace{p^2 f \cdot N_{ppf} + p f^2 \cdot N_{pff} + f^3 \cdot N_{fff}}^{fake} \quad (5.6)$$

In Equation 5.4 and Equation 5.6, the left-hand sides show the total number of events expected (N_{tt} or N_{ttt}), which is equivalent to the total background. The real part is estimated from MC samples of SM processes. To find the fakes, besides prompt rate p and fake rate f , one also needs to know the numbers of events with one or more nonprompt leptons, e.g., N_{pf} , N_{ppf} , which are not directly measurable. Fortunately, taking the inverse of $\mathcal{M}$, $N_{...f}$'s can be expressed in terms of $N_{t...t}$, $N_{t...l}$'s, and $N_{l...l}$ as shown in Equation 5.2. While $N_{t...t}$ is the unknown being evaluated, $N_{...l}$'s can be found by directly counting in data. Then, to complete Equation 5.4 and Equation 5.6, one needs to determine the prompt rates and fake rates. Their measurements are explained in the coming Section 5.5.3 and 5.5.4.

Note that Equation 5.4 and Equation 5.6 are simplified for the case of same-flavour leptons. The actual equations applied in the analysis are varied for the different combinations of electrons and muons [130].

5.5.3 Prompt Rates

Prompt rate is the efficiency of a prompt loose lepton passing the “tight” requirements defined in Section 5.3.1, and is essential for the estimation of nonprompt background as described in Section 5.5.2. Prompt rates for electrons and muons are measured separately using tag-and-probe method in Drell-Yan events in data. Events are required to pass one of the dilepton triggers listed in Section 5.4.1 and contain a tag lepton that passed the “tight” selections with $p_T > 30$ GeV and $|\eta| < 2.4$. Event yields are obtained by simply counting the number of event before and after applying the cuts under study. Prompt rate is the count of probes passing “tight” divided by the total number of probes passing “loose” in qualified tag-and-probe events.

The distributions of electron and muon prompt rates measured in 2017 samples are shown in Figures 5.5 and 5.6. The distributions of muon prompt rates are relatively flat, and thus it is acceptable to apply the η -averaged value 0.928 ± 0.0150 on all muons in multi-lepton channel. However, getting rid of legacies from the preceding 2016 analysis, p_T -binned muon prompt rates in Table 5.11 are applied in SS2L channel. On the other hand, electron prompt rates show a greater dependence on p_T . Therefore, different prompt rates have been assigned according to the electron p_T , as listed in Table 5.11, in both channels.

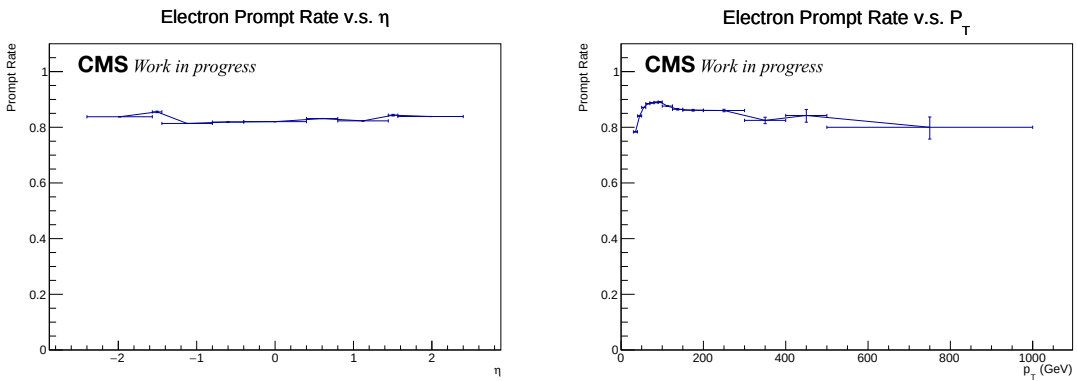


Figure 5.5: Measured electron prompt rates as a function of η (left) and p_T (right) in 2017 data.

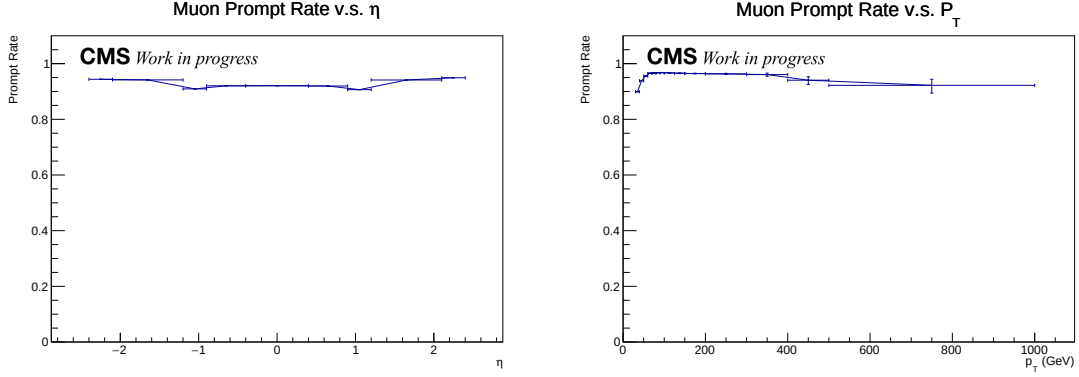


Figure 5.6: Measured muon prompt rates as a function of η (left) and p_T (right) in 2017 data.

Table 5.11: Electron and muon prompt rates for different p_T ranges in 2017 data. The values in table are applied in the SS2L channel and, for electron only, the multi-lepton channel. An η -averaged muon prompt rate of 0.928 ± 0.0150 is applied in the multi-lepton channel instead.

p_T (GeV)	Electron	Muon
30 –40	0.784 ± 0.0002	0.899 ± 0.0001
40 –50	0.841 ± 0.0002	0.939 ± 0.0001
50 –60	0.871 ± 0.0003	0.955 ± 0.0002
60 –70	0.885 ± 0.0006	0.966 ± 0.0003
70 –80	0.888 ± 0.0010	0.967 ± 0.0005
80 –90	0.889 ± 0.0014	0.968 ± 0.0006
90 –100	0.890 ± 0.0018	0.968 ± 0.0008
100–125	0.876 ± 0.0017	0.967 ± 0.0008
125–150	0.865 ± 0.0026	0.966 ± 0.0012
150–200	0.861 ± 0.0030	0.964 ± 0.0014
200–300	0.860 ± 0.0044	0.963 ± 0.0022
300–400	0.825 ± 0.0116	0.961 ± 0.0056
400–500	0.842 ± 0.0228	0.941 ± 0.0137
500–1000	0.800 ± 0.0397	0.922 ± 0.0251

The distributions of electron and muon prompt rates measured in 2018 samples are shown in Figures 5.7 and 5.8. Similarly, η -averaged muon prompt rate 0.931 ± 0.01059 is used in multi-lepton channel, and p_T -binned muon prompt rates in Table 5.12 are applied in same-sign dilepton channel. For electrons, p_T -binned prompt rates listed in Table 5.12 are applied in both channels.

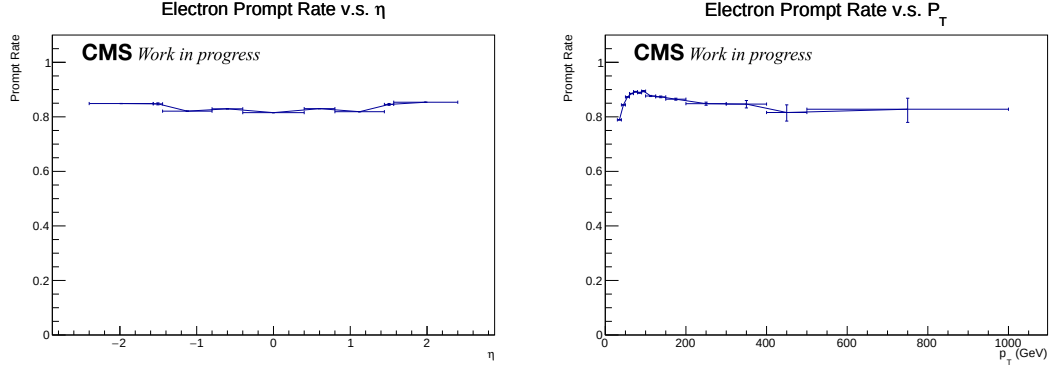


Figure 5.7: Measured electron prompt rates as a function of η (left) and p_T (right) in 2018 data.

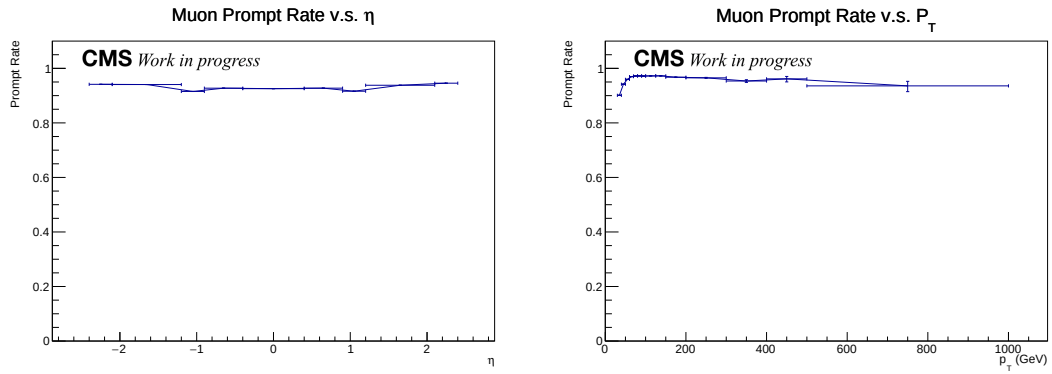


Figure 5.8: Measured muon prompt rates as a function of η (left) and p_T (right) in 2018 data.

Table 5.12: Electron and muon prompt rates for different p_T ranges in 2018 data. The values in table are applied in the SS2L channel and, for electron only, the multi-lepton channel. An η -averaged muon prompt rate of 0.931 ± 0.01059 is applied in the multi-lepton channel instead.

p_T (GeV)	Electron	Muon
30–40	0.789 ± 0.0003	0.901 ± 0.0001
40–50	0.844 ± 0.0002	0.942 ± 0.0001
50–60	0.873 ± 0.0004	0.960 ± 0.0001
60–70	0.885 ± 0.0007	0.970 ± 0.0002
70–80	0.892 ± 0.0012	0.972 ± 0.0003
80–90	0.888 ± 0.0017	0.972 ± 0.0005
90–100	0.894 ± 0.0022	0.972 ± 0.0007
100–125	0.877 ± 0.0020	0.972 ± 0.0006
125–150	0.873 ± 0.0031	0.973 ± 0.0009
150–200	0.865 ± 0.0036	0.968 ± 0.0011
200–300	0.848 ± 0.0057	0.965 ± 0.0018
300–400	0.847 ± 0.0136	0.954 ± 0.0049
400–500	0.816 ± 0.0298	0.961 ± 0.0101
500–1000	0.828 ± 0.0445	0.936 ± 0.0191

5.5.4 Fake Rates

After measuring prompt rates in Section 5.5.3, one more variable is needed to find N_{tt} with Equation 5.4 and N_{ttt} with Equation 5.6. Once $N_{t...t}$ is determined, the number of fakes can be calculated from the last few terms in Equation 5.4 and Equation 5.6. This variable in need, fake rate f , is estimated using a data-driven method [21, 131] in which nonprompt background is fitted to data with prompt background subtracted in the control region CR2 of the multi-lepton channel. The measured fake rates are applied to both same-sign dilepton and multi-lepton channel.

The data-driven method is based on two assumptions. First, it assumes that fake rate only depends on lepton flavour, i.e., electron fake rate and muon fake rate are constants. Second, it is assumed that data roughly equals to the sum of MC (prompt) background and nonprompt background in multi-lepton channel, which is true in control region where little signal is expected. With these assumptions, the best-fitted fake rate can be found by comparing nonprompt background, estimated by the matrix method, to data subtracted by prompt background from MC samples. To keep the signal region blinded, the fit is performed in the control region CR2 in multi-lepton channel with exactly two AK4 jets and three leptons that passed “loose” requirements.

In CR2, in bin i of each leptons’ p_T distribution (Figures 5.9–5.10), the number of data entries (N_{data}^i) is counted, and the prompt background yields are estimated with MC (N_{MC}^i). Then, nonprompt background in bin i (N_{NP}^i) is calculated using the matrix method for some values of electron fake rate f_e and muon fake rate f_μ , e.g., $f_e = 0.01$ and $f_\mu = 0.0$. To quantify how well an estimated nonprompt background fits to data subtracted by prompt background, we compute the χ^2 defined in Equation 5.7.

$$\chi^2(p, f) = \sum_{i \in \text{bins}} \frac{(N_{\text{data}}^i - (N_{\text{NP}}^i(p, f) + N_{\text{MC}}^i))^2}{N_{\text{NP}}^i(p, f) + N_{\text{MC}}^i} \quad (5.7)$$

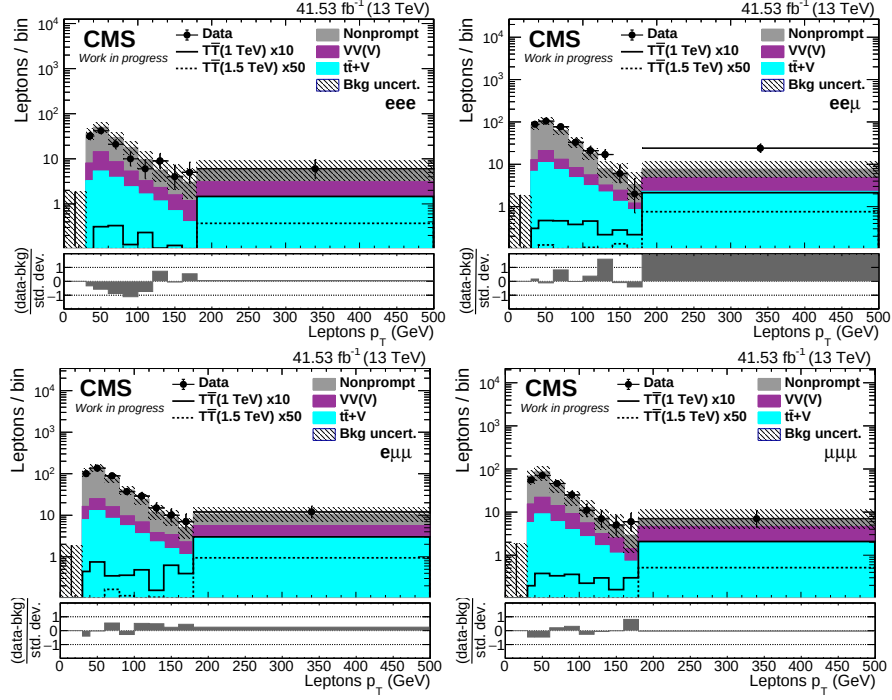


Figure 5.9: Distributions of leptons' p_T in the multi-lepton channel CR2 in each flavour category. Data from 2017 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV scaled by a factor of 10 (50) using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with fake rates measured in CR2. Statistical and systematic uncertainties are shown by the hatched region.

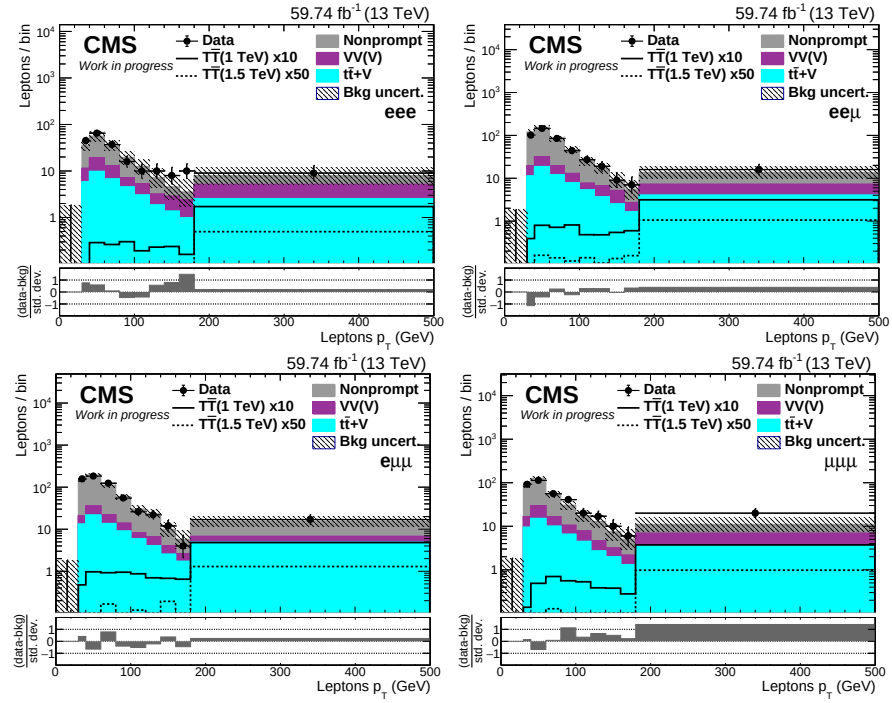


Figure 5.10: Distributions of leptons' p_T in the multi-lepton channel CR2 in each flavour category. Data from 2018 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV scaled by a factor of 10 (50) using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with fake rates measured in CR2. Statistical and systematic uncertainties are shown by the hatched region.

Starting from $(f_e = 0.0, f_\mu = 0.0)$, the computation of χ^2 is repeated after each incrementation of the electron fake rate or muon fake rate by 0.01, until they both reach 0.5. Eventually, a total of 51×51 χ^2 values are computed, for electron and muon fake rate each incrementing from 0.0 to 0.5 at every 0.01 interval independently. The best-fitted electron and muon fake rates are those that correspond to the minimum χ^2 .

Given the complexities in varying Equation 5.6 for different combinations of lepton flavours, fits are performed separately for each flavour category. The four 51×51 -binned χ^2 distributions are then added together to find the overall minimum, where the best-fitted electron and muon fake rates can be found.

Uncertainties are determined by the marginalized Gaussian probability distributions. Using the relation that probability $P(f_\mu, f_e) \propto \exp(-\chi^2/2)$, a 2D Gaussian probability distribution can be converted from the χ^2 distribution. Then, by marginalization, i.e, summing the distribution along one of the axes, the 2D distribution collapses to a 1D Gaussian in which the 1-sigma uncertainty is extracted for the corresponding fake rate. These measured uncertainties are directly included as part of the systematics in multi-lepton channel.

In 2017 samples, the computed χ^2 distribution is shown in Figure 5.11, and the 1D Gaussian distributions are shown in Figure 5.12. The minimum χ^2 value is 35.4, corresponding to electron and muon fake rates of 0.1 ± 0.01 and 0.15 ± 0.01 .

In 2018 samples, the computed χ^2 distribution is shown in Figure 5.13, and the 1D Gaussian distributions are shown in Figure 5.14. The minimum χ^2 value is 24.9 corresponding to electron and muon fake rates of 0.08 ± 0.01 and 0.16 ± 0.01 .

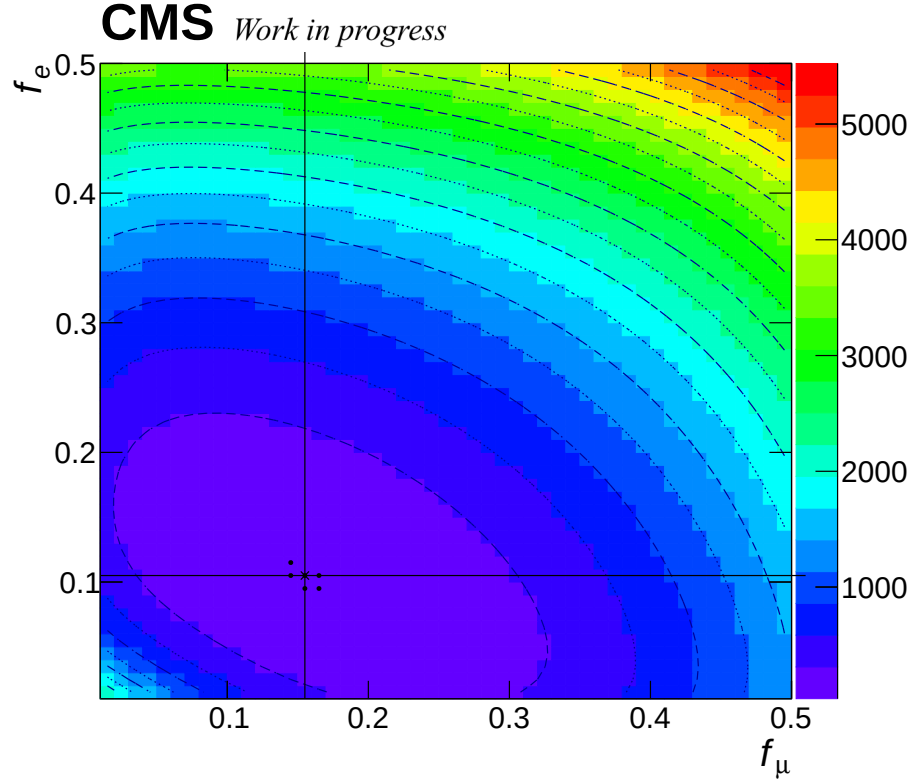


Figure 5.11: Measured χ^2 in 2017 between the predicted and observed number of events in the multi-lepton channel CR2 summed over all flavour categories, as a function of electron (f_e) and muon (f_μ) nonprompt rates. The intersection of the two black lines indicates the position of the minimum, and the markers indicate bins which deviate from the minimum χ^2 by less than 2.3.

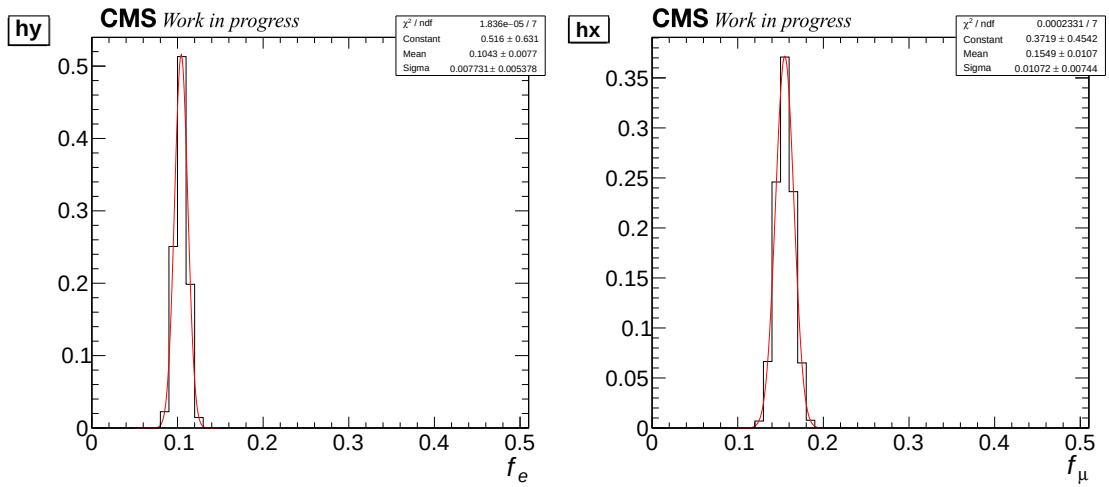


Figure 5.12: Distributions of Gaussian probability marginalized for electron (left) and muon (right) fake rates in CR2 using 2017 data. The fitted values of sigma are the uncertainties on measured fake rates.

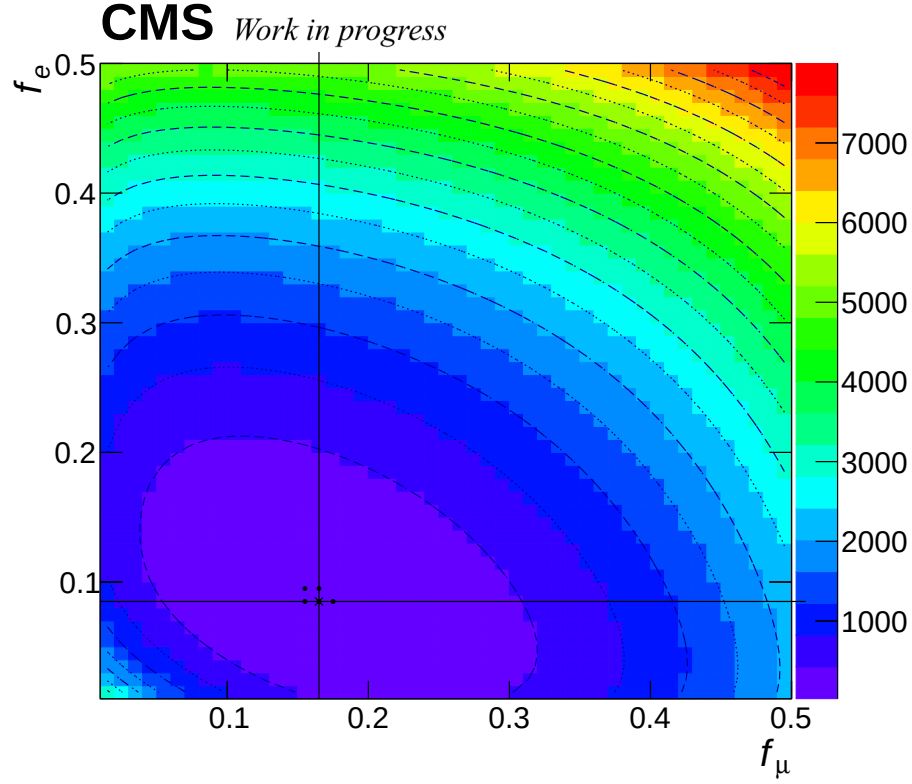


Figure 5.13: Measured χ^2 in 2018 between the predicted and observed number of events in the multi-lepton channel CR2 summed over all flavour categories, as a function of electron (f_e) and muon (f_μ) nonprompt rates. The intersection of the two black lines indicates the position of the minimum, and the markers indicate bins which deviate from the minimum χ^2 by less than 2.28.

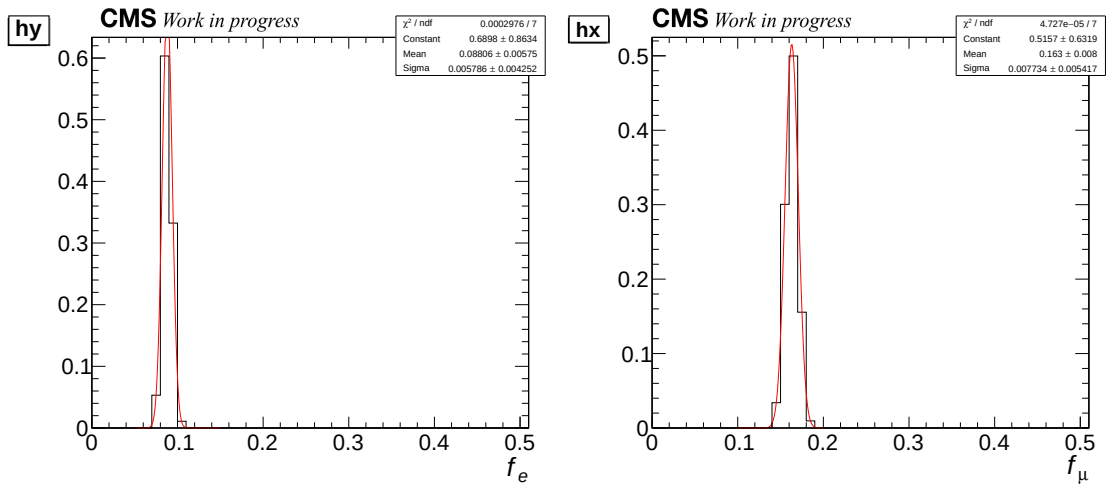


Figure 5.14: Distributions of Gaussian probability marginalized for electron (left) and muon (right) fake rates in CR2 using 2018 data. The fitted values of sigma are the uncertainties on measured fake rates.

5.5.5 Fake Rates Cross Check

A way to cross check with the measured fake rates in Section 5.5.4 is to repeat the measurement in a perpendicular region CR1. In CR1, events have exactly three leptons and exactly one AK4 jet. Again, nonprompt background for each set of lepton fake rates between 0.0 and 0.5 is estimated using the matrix method. The fitting performance is quantified by χ^2 defined in Equation 5.7. χ^2 is computed for each lepton fake rates.

Figure 5.15 shows the χ^2 distribution at different electron and muon fake rates measured in 2017 CR1 events. The point with minimum χ^2 of 52.9 is marked by the intersection of two straight black lines. The point corresponds to electron and muon fake rates of 0.12 ± 0.01 and 0.16 ± 0.01 . The measurements in CR1 and CR2 agree according to z-test. The test statistic of -1.41 and -0.71 are within the range of critical values from -1.96 to 1.96 for 95% confidence level.

Figure 5.16 shows the χ^2 distribution at different electron and muon fake rates measured in 2018 CR1 events. The point with minimum χ^2 of 44.5 is marked by the intersection of two straight black lines. The point corresponds to electron and muon fake rates of 0.12 ± 0.01 and 0.14 ± 0.01 . For muon fake rate, the measurements in CR1 and CR2 agree in z-test with a statistic of 1.41. Although the measurements of electron fake rate in CR1 and CR2 disagree in z-test with a statistic of -2.83, it is still fair to say that χ^2 scanning seems to work fine overall for measuring lepton fake rates.

Note that no systematics is assigned based on this cross check.

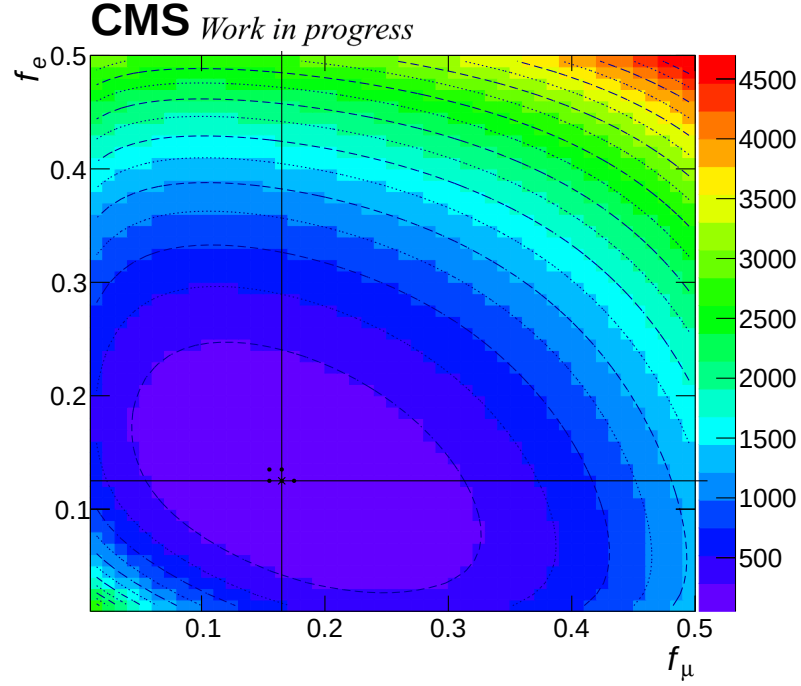


Figure 5.15: Measured χ^2 in 2017 between the predicted and observed number of events in the multi-lepton channel CR1 summed over all flavour categories, as a function of electron (f_e) and muon (f_μ) nonprompt rates. The intersection of the two black lines indicates the position of the minimum, and the markers indicate bins which deviate from the minimum χ^2 by less than 2.3.

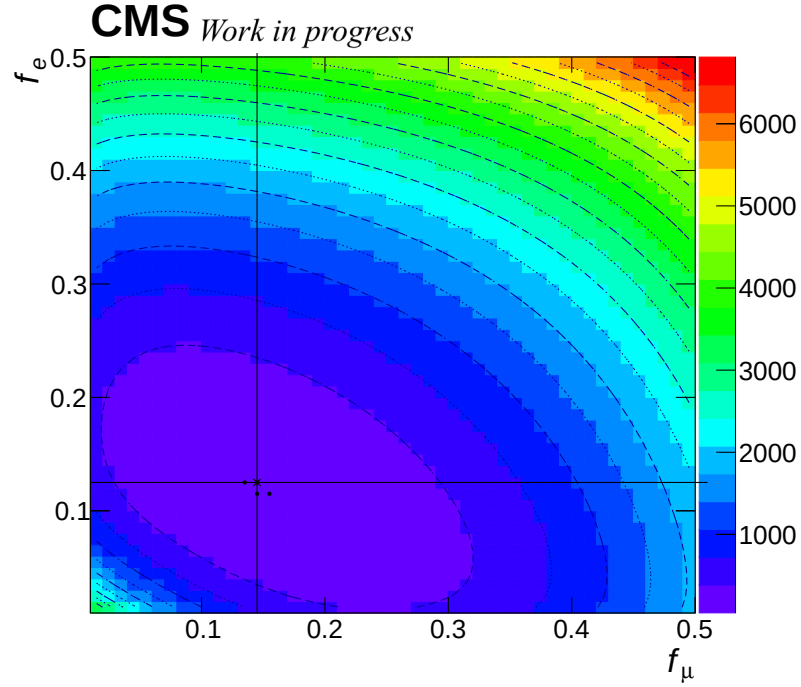


Figure 5.16: Measured χ^2 in 2018 between the predicted and observed number of events in the multi-lepton channel CR1 summed over all flavour categories, as a function of electron (f_e) and muon (f_μ) nonprompt rates. The intersection of the two black lines indicates the position of the minimum, and the markers indicate bins which deviate from the minimum χ^2 by less than 2.3.

5.5.6 3L+ Channel: η -Dependence of Muon Fake Rate

In the preceding 2016 analysis, fake rates are measured separately in SS2L channel with tag-and-probe method in same-flavour dilepton events [21, 89, 131]. The measurement shows that muon fake rate changes with rapidity (η). To investigate the η -dependence of muon fake rate, for each trilepton flavour, events are further categorized by the leading muon's rapidity to find the best-fitted muon fake rate in each defined $|\eta|$ range.

To uniquely divide events into $|\eta|$ groups, only the rapidity of the muon with highest p_T is considered. Since at least one muon is required for this study, the flavour category EEE is omitted. To test variations due to muon fake rates only, electron fake rate is fixed at measured value in Section 5.5.4, and then muon fake rate is scanned from 0 to 0.5 in events within an $|\eta|$ subgroup to find the minimum χ^2 . The best-fitted muon fake rates in different $|\eta|$ ranges and flavour categories are shown in Table 5.13–5.14 for measurement in 2017 and 2018 samples respectively.

Table 5.13: $|\eta|$ -dependent μ fake rates measured in 2017 CR2 by fitting distributions of leptons' p_T while fixing e fake rate at 0.1

$ \eta $	EEM	EMM	MMM	All	Sum
0.0-0.4	0.07 ± 0.067	0.22 ± 0.041	0.22 ± 0.029	0.19 ± 0.025	0.20 ± 0.023
0.4-0.9	0.18 ± 0.052	0.15 ± 0.033	0.16 ± 0.022	0.15 ± 0.020	0.16 ± 0.018
0.9-1.2	0.33 ± 0.044	0.18 ± 0.052	0.12 ± 0.037	0.18 ± 0.027	0.21 ± 0.027
1.2-2.4	0.19 ± 0.037	0.17 ± 0.030	0.12 ± 0.021	0.15 ± 0.017	0.15 ± 0.017

Table 5.14: $|\eta|$ -dependent μ fake rates measured in 2018 CR2 by fitting distributions of leptons' p_T while fixing e fake rate at 0.08

$ \eta $	EEM	EMM	MMM	All	Sum
0.0-0.4	0.10 ± 0.059	0.17 ± 0.029	0.19 ± 0.022	0.15 ± 0.019	0.18 ± 0.017
0.4-0.9	0.17 ± 0.042	0.16 ± 0.026	0.14 ± 0.018	0.15 ± 0.016	0.15 ± 0.015
0.9-1.2	0.16 ± 0.060	0.15 ± 0.030	0.22 ± 0.034	0.16 ± 0.023	0.18 ± 0.021
1.2-2.4	0.19 ± 0.029	0.15 ± 0.020	0.19 ± 0.020	0.17 ± 0.013	0.17 ± 0.013

Since muon fake rate is applied to each muon as a weight, the $|\eta|$ -dependent fake rates measured in EEM category best reflects the $|\eta|$ -dependence of fake rate of a single muon. When these $|\eta|$ -dependent fake rates measured in EEM are applied to all muons, the resulting distributions of leptons' p_T in CR2, displayed in Figure 5.17–5.18 for 2017 and 2018 respectively, show improved background prediction in a few bins only.

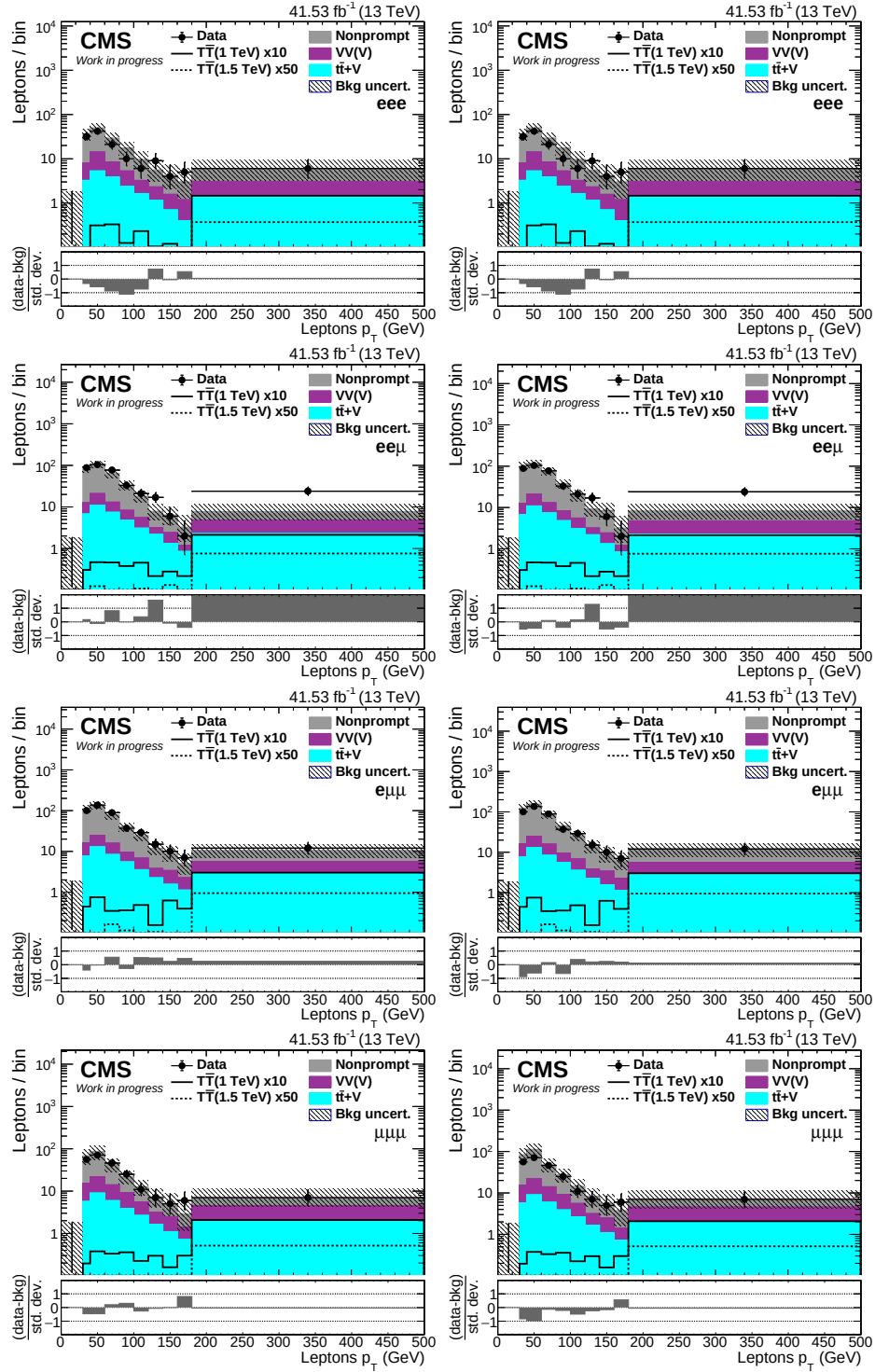


Figure 5.17: Distributions of leptons' p_T in the multi-lepton channel CR2 in each flavour category. Data from 2017 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV scaled by a factor of 10 (50) using solid (dashed) lines, and background using filled histograms. Shapes of scaled $T\bar{T}$ distributions are easier to read. Nonprompt background is estimated with measured flat (left) and $|\eta|$ -dependent (right) muon fake rates. Statistical and systematic uncertainties are shown by the hatched region.

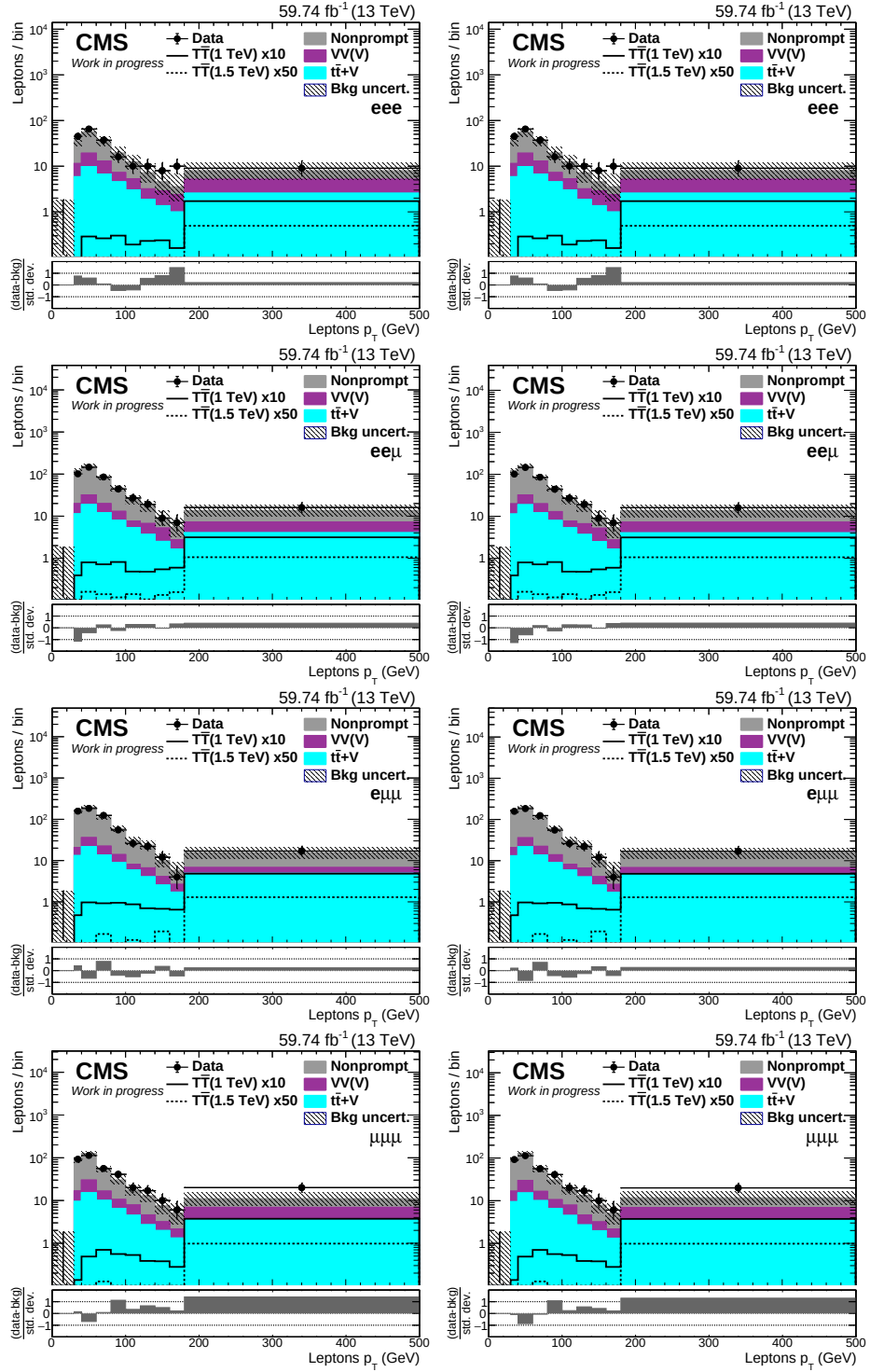


Figure 5.18: Distributions of leptons' p_T in the multi-lepton channel CR2 in each flavour category. Data from 2018 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV scaled by a factor of 10 (50) using solid (dashed) lines, and background using filled histograms. Shapes of scaled $T\bar{T}$ distributions are easier to read. Nonprompt background is estimated with measured flat (left) and $|\eta|$ -dependent (right) muon fake rates. Statistical and systematic uncertainties are shown by the hatched region.

While the change is barely observable in the plots, the yields are certainly different between constant and $|\eta|$ -dependent muon fake rates. To account for the changes due to $|\eta|$ -dependence of muon fake rate, the percentage differences in yields is considered as a source of systematics in the multi-lepton channel. The values are reported in Table 5.15 and 5.16.

Table 5.15: Nonprompt background yields from 2017 samples in each flavour category predicted with $|\eta|$ -dependent and flat μ fake rates. The percentage changes in yields are calculated.

	EEE	EEM	EMM	MMM	All
FR($ \eta $)	36.39 ± 13.0	96.55 ± 24.0	154.5 ± 36.0	70.89 ± 32.0	358.2 ± 82.0
flat FR	36.39 ± 13.0	87.73 ± 23.0	130.7 ± 32.0	51.11 ± 24.0	305.8 ± 73.0
$\Delta\%$	0.0%	10.1%	18.2%	38.7%	17.1%

Table 5.16: Nonprompt background yields from 2018 samples in each flavour category predicted with $|\eta|$ -dependent and flat μ fake rates. The percentage changes in yields are calculated.

	EEE	EEM	EMM	MMM	All
FR($ \eta $)	43.98 ± 28.0	121.5 ± 40.0	156.0 ± 41.0	76.94 ± 19.0	398.3 ± 110.0
flat FR	43.98 ± 28.0	120.2 ± 40.0	154.5 ± 41.0	75.9 ± 19.0	394.5 ± 110.0
$\Delta\%$	0.0%	1.1%	1.0%	1.4%	1.0%

5.5.7 SS2L Channel: Fake Rate Scan for Systematics

As aforementioned, fake rates are measured separately in a QCD-enriched control region in the SS2L channel with tag-and-probe method in the preceding 2016 analysis [21, 89, 131]. However, the same measurement in 2017 and 2018 data is unsuited to describe the nonprompt background in the SS2L channel SR and predicts a background that does not agree well with data in the CR.

In this analysis, fake rates measured in multi-lepton channel are applied to the SS2L channel instead. Taking the differences between the two channels into account, more conservative uncertainties are preferred. A separate estimation of systematics is developed for the 3+L fake rates in the SS2L channel.

The nonprompt background uncertainties in SS2L channel are estimated by a fake rate scan procedure which is similar to the 3+L fake rate measurement in Section 5.5.4. Scanning through fake rates in the range of $(\text{measuredvalue} \pm 0.1)$, nonprompt background is predicted for each step of 0.01, and the nonprompt yield is compared to $(\text{data} - \text{MCbackground} - \text{CMIDbackground})$. The minimum χ^2 (defined in Equation 5.7) corresponds to the best-fitted fake rates from the scan. The scan is

performed in the low H_T^{lep} CR ($H_T^{\text{lep}} < 1000$ GeV) of the SS2L channel. The best-fitted values are reported in Table 5.17. The distributions of $H_T^{\text{lep}} < 1000$ GeV obtained with the best-fitted fake rates from the scan in SS2L channel are shown in Figures 5.19.

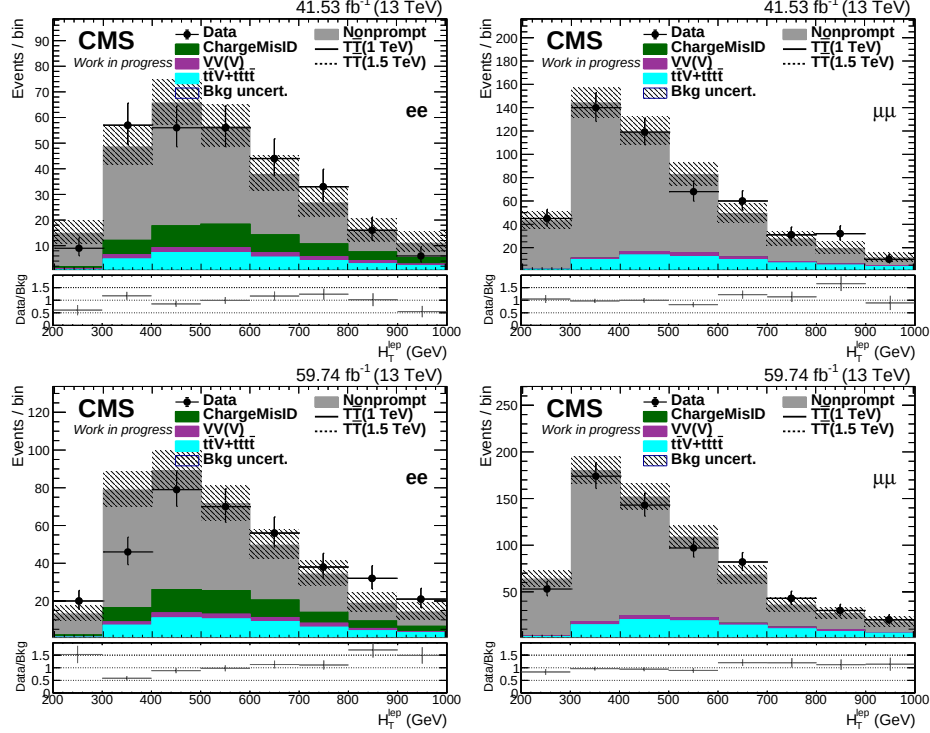


Figure 5.19: Distributions of $H_T^{\text{lep}} < 1000$ GeV in the SS2L channel in ee and $\mu\mu$ categories. Data from 2017 (top) and 2018 (bottom) is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with the shifted fake rates. Statistical and systematic uncertainties are shown by the hatched region.

The fake rate scans in SS2L and multi-lepton channels produce very different results for electrons, as expected from the observed distributions in control regions (Section 5.8.2). In SS2L channel, a very conservative one-sided shape uncertainty on nonprompt background is included by shifting the 3+L fake rates to the values from the scan in SS2L low H_T^{lep} CR, as detailed in Table 5.17. The templates for this fake rate systematic uncertainty in SS2L channel are produced by predicting nonprompt background events with the shifted fake rates.

Note that while it is fair to check for data-background agreement in the low H_T^{lep} CR, the CR is not suitable for directly measuring the nominal fake rates because distributions in the CR is not representative of the SR. Thus, the solution above is adopted.

Table 5.17: A table showing the applied 3+L fake rates and the best-fitted fake rates from the scan in SS2L low H_T^{lep} CR for each lepton flavour in 2017 and 2018 respectively. The last column shows the shifts in fake rates for a one-sided systematic shape uncertainty.

Lepton	year	Applied FR	FR from scan	uncertainty
e	2017	0.1	0.17	+0.07
	2018	0.08	0.16	+0.08
μ	2017	0.15	0.13	-0.02
	2018	0.16	0.14	-0.02

5.5.8 SS2L Channel: Background from Charge Misidentification

In SS2L channel, in addition to prompt and nonprompt background, there is also background from charge misidentification (CMID). CMID rates are dominated by electron's, which are an order of magnitude higher than muon's. The electron's CMID rates are measured in $Z \rightarrow ee$ events, by requiring $81.1 < M_{\ell\ell} < 101.1$ GeV in data. Instead of taking the ratio of same-sign to opposite-sign events as in the preceding 2016 $X_{5/3}$ search, the CMID rates are estimated using a likelihood fit [132], which considers the kinematics of both leptons and the scenario when both leptons' charges are mis-identified. Assuming constant CMID rate in a $p_T-|\eta|$ bin, for a di-electron event in which one electron falls in bin i and the other in bin j , the probability of observing a same-sign event is

$$p_{ij}^{\text{SS}} = (1 - \epsilon_i) \cdot \epsilon_j + \epsilon_i \cdot (1 - \epsilon_j) \quad (5.8)$$

where ϵ_i is the CMID efficiency in bin i . The total number of expected SS events is $n^{\text{SS}} = \sum_{ij} n_{ij} \cdot p_{ij}^{\text{SS}}$ where n_{ij} is the number of events with electron pair in bin i and bin j .

Expecting that observed SS events follow a binomial distribution, and that the probability is small,

$$P(n_{ij}^{\text{SS}} | \epsilon_i, \epsilon_j) = \left[C_{n_{ij}^{\text{SS}}}^n p^{n_{ij}^{\text{SS}}} (1-p)^{n-n_{ij}^{\text{SS}}} \right]_{ij} \xrightarrow{p \rightarrow 0} \frac{(n_{ij} \cdot p_{ij}^{\text{SS}})^{n_{ij}^{\text{SS}}} \cdot e^{-n_{ij} \cdot p_{ij}^{\text{SS}}}}{n_{ij}^{\text{SS}}!} \quad (5.9)$$

Then, the negative log likelihood is

$$-\ln \mathcal{L} = -\sum_{ij} \ln P(n_{ij}^{\text{SS}} | \epsilon_i, \epsilon_j) = -\sum_{ij} \left[n_{ij}^{\text{SS}} \ln(n_{ij} \cdot p_{ij}^{\text{SS}}) - n_{ij} \cdot p_{ij}^{\text{SS}} - \ln(n_{ij}^{\text{SS}}!) \right] \quad (5.10)$$

By minimizing Equation 5.10, the CMID rate is estimated for each p_T - $|\eta|$ bin. To simplify the calculation, and ensure sufficient statistics in each bin, there are only three p_T regions and five $|\eta|$ bins as shown in Table 5.18 together with the measured values. Figure 5.20 shows the distributions of measured CMID rates.

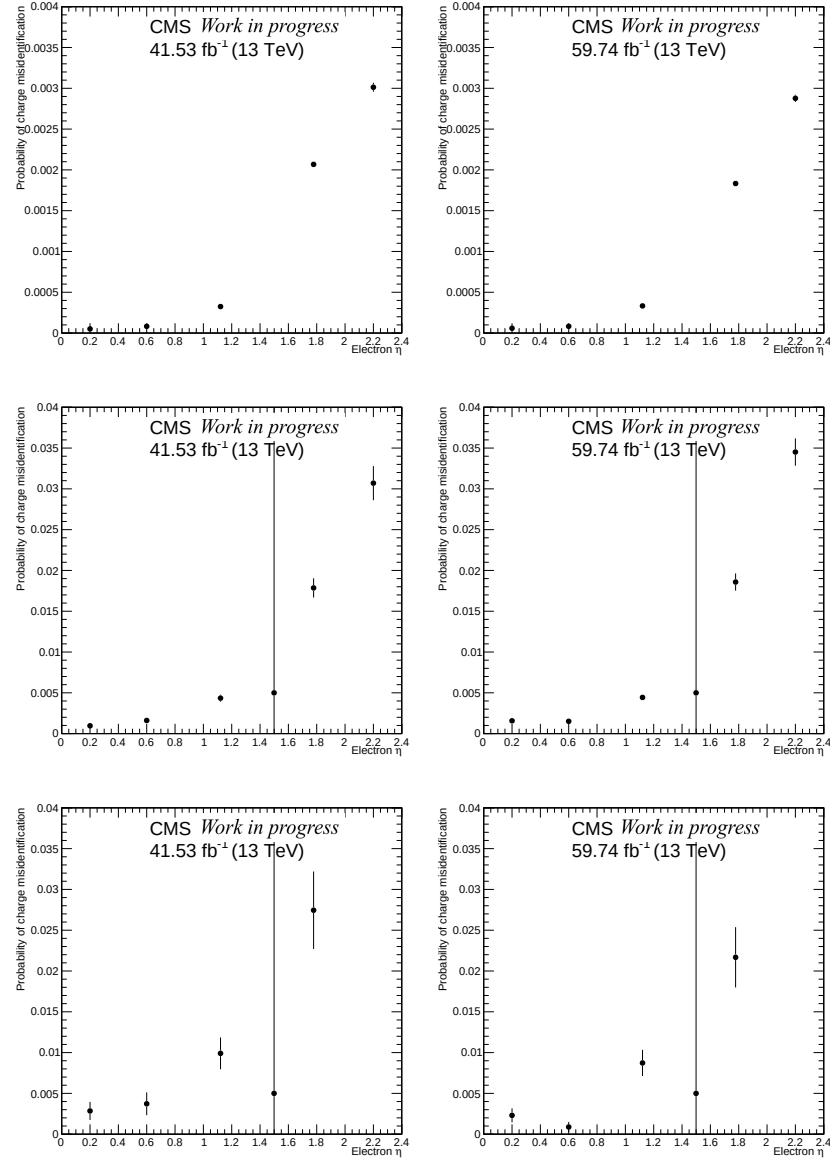


Figure 5.20: Measured electron CMID rate in $p_T < 100$ GeV (top), 100 – 200 GeV (mid) and > 200 GeV (bottom) in 2017 (left) and 2018 (right) as a function of $|\eta|$. The markers at $1.44 < |\eta| < 1.56$ are dummy holders for the barrel-endcap gap in ECAL and will be ignored.

Table 5.18: CMID rates measured in 2017 and 2018 samples by minimizing the negative log likelihood.

p_T (GeV)	$ \eta $	2017 CMID	2018 CMID
< 100	0.0–0.4	$5.08 \times 10^{-5} \pm 5.03 \times 10^{-6}$	$5.88 \times 10^{-5} \pm 4.98 \times 10^{-6}$
	0.4–0.8	$8.24 \times 10^{-5} \pm 6.36 \times 10^{-6}$	$8.25 \times 10^{-5} \pm 5.92 \times 10^{-6}$
	0.8–1.44	$3.25 \times 10^{-4} \pm 1.06 \times 10^{-5}$	$3.33 \times 10^{-4} \pm 1.00 \times 10^{-5}$
	1.56–2.0	$2.07 \times 10^{-3} \pm 3.19 \times 10^{-5}$	$1.83 \times 10^{-3} \pm 2.77 \times 10^{-5}$
	2.0–2.4	$3.01 \times 10^{-3} \pm 5.56 \times 10^{-5}$	$2.88 \times 10^{-3} \pm 4.40 \times 10^{-5}$
100–200	0.0–0.4	$9.59 \times 10^{-4} \pm 2.09 \times 10^{-4}$	$1.57 \times 10^{-3} \pm 2.25 \times 10^{-4}$
	0.4–0.8	$1.61 \times 10^{-3} \pm 2.86 \times 10^{-4}$	$1.51 \times 10^{-3} \pm 2.63 \times 10^{-4}$
	0.8–1.44	$4.35 \times 10^{-3} \pm 4.29 \times 10^{-4}$	$4.43 \times 10^{-3} \pm 3.36 \times 10^{-4}$
	1.56–2.4	$1.79 \times 10^{-2} \pm 1.18 \times 10^{-3}$	$1.86 \times 10^{-2} \pm 1.05 \times 10^{-3}$
	2.0–2.4	$3.07 \times 10^{-2} \pm 2.09 \times 10^{-3}$	$3.45 \times 10^{-2} \pm 1.67 \times 10^{-3}$
> 200	0.0–0.4	$2.85 \times 10^{-3} \pm 1.10 \times 10^{-3}$	$2.30 \times 10^{-3} \pm 8.57 \times 10^{-4}$
	0.4–0.8	$3.73 \times 10^{-3} \pm 1.40 \times 10^{-3}$	$8.80 \times 10^{-4} \pm 6.18 \times 10^{-4}$
	0.8–1.44	$9.90 \times 10^{-3} \pm 1.95 \times 10^{-3}$	$8.73 \times 10^{-3} \pm 1.60 \times 10^{-3}$
	1.56–2.4	$2.74 \times 10^{-2} \pm 4.75 \times 10^{-3}$	$2.17 \times 10^{-2} \pm 3.69 \times 10^{-3}$
	2.0–2.4	$5.41 \times 10^{-2} \pm 1.02 \times 10^{-2}$	$4.62 \times 10^{-2} \pm 6.89 \times 10^{-3}$

5.6 Nonprompt Background Closure Tests with $t\bar{t}$ MC

To make sure that the matrix method works and that nonprompt background is well estimated, a number of closure tests were performed using $t\bar{t} \rightarrow 2\ell + 2\nu$ MC sample. The simulated $t\bar{t}$ events are required to pass the triggers in Table 5.6 and all offline selections in Section 5.4.3 except that leptons only need to pass the “loose” requirements. In addition, events are required to have one and only one nonprompt lepton while the others must be prompt leptons. These kind of events is a major contribution to nonprompt background, which is derived from data in the analysis.

5.6.1 Closure tests in Multi-lepton Channel

For the multi-lepton channel, selected events are required to be labelled ppf , meaning that two of the trilepton are prompt, and the remaining one is nonprompt.

5.6.2 3+L Channel: Cut-and-Count Fake Rate Cross Check

The matrix method used to estimate nonprompt background depends greatly on the input value of prompt rate and fake rate. Prompt rate measurement in Section 5.5.3 is very straightforward with

relatively small error. So the prerequisite for the matrix method to work is dominantly having well estimated fake rates. There is a simple cross-check in Section 5.5.5 in which fake rates measured in two independent control regions nearly agree. Here, MC truth information is used to find fake rate in simulated $t\bar{t}$ trilepton events to cross-check with the data-driven scan.

In each selected ppf event, there is one and only one nonprompt lepton among the trilepton. The nonprompt lepton is categorized as tight or loose-not-tight for passing and failing the “tight” requirements (all leptons pass the “loose” requirements). Table 5.19 and 5.20 show the numbers of nonprompt leptons that are tight and loose-not-tight respectively. The ratio of $\frac{\text{tight}}{(\text{tight}) + (\text{loose-not-tight})}$ is calculated as a measurement of fake rates. This technique is known as cut-and-count for applying a cut and counting the number of entries passing and failing the applied cut.

In 2017 $t\bar{t}$ MC sample, the cut-and-count electron fake rate is $f_e = 0.09$, and the muon fake rate is $f_\mu = 0.15$. In 2018 $t\bar{t}$ MC sample, the cut-and-count fake rates are $f_e = 0.09$ and $f_\mu = 0.14$.

Recall the measurement in Section 5.5.4: $(f_e, f_\mu) = (0.10, 0.15)$ in 2017 and $(f_e, f_\mu) = (0.08, 0.16)$ in 2018. These values are very close to the ones from cut-and-count in $t\bar{t}$ MC, within ± 0.01 to 0.02.

Table 5.19: Fake rates measured by cut-and-count in 2017 $t\bar{t}$ sample with three or more leptons.

	nonprompt & tight	nonprompt & loose-not-tight	e fake rate
eee ppf	1629	15942	0.093
$ee\mu$ ppf	3827	37119	0.093
$e\mu\mu$ ppf	2365	21720	0.098
total	7821	74781	0.095
	nonprompt & tight	nonprompt & loose-not-tight	μ fake rate
$ee\mu$ ppf	1963	11248	0.015
$e\mu\mu$ ppf	4634	26354	0.015
$\mu\mu\mu$ ppf	2676	15168	0.015
total	9273	52770	0.015

Table 5.20: Fake rates measured by cut-and-count in 2018 $t\bar{t}$ sample with three or more leptons.

	nonprompt & tight	nonprompt & loose-not-tight	e fake rate
eee ppf	1429	14546	0.089
$ee\mu$ ppf	3505	34065	0.093
$e\mu\mu$ ppf	2044	20084	0.092
total	6978	68695	0.092
	nonprompt & tight	nonprompt & loose-not-tight	μ fake rate
$ee\mu$ ppf	1694	10417	0.014
$e\mu\mu$ ppf	4043	24199	0.014
$\mu\mu\mu$ ppf	2411	14521	0.014
total	8148	49137	0.014

5.6.3 3+L Channel: CR2 to SR Fake Rate Compatibility

Although Section 5.5.5 shows that the measured data-driven fake rates nearly agree in different control regions with different number of jets, it does not rule out the chance that fake rates might depend on hadronic activity.

In Section 5.5.4, fake rates are measured in CR2. To check that these values are also valid in SR, a mimicked χ^2 scan in $t\bar{t}$ MC samples is repeated in CR2 and SR.

Among the ppf events, if the nonprompt lepton passes the “tight” requirements, the event is labelled as “observed”. These “observed” events are treated as pseudo-data for the χ^2 scanning. By definition, weighting the number of loose-not-tight nonprompt leptons with fake rate gives the “predicted” number of tight nonprompt leptons. “Predicted” background is fitted to the “observed” pseudo-data. There is no need to consider prompt background.

Putting this in the framework of χ^2 scanning in Section 5.5.4, the two assumptions are:

1. Fake rate only depends on lepton flavour
2. “Observed” pseudo-data roughly equals to “predicted” nonprompt background

Removing prompt background from Equation 5.7 (or set $N_{MC} = 0$), $\chi_{t\bar{t}}^2$ is computed for all bins in the distributions of leptons’ p_T in $t\bar{t}$ MC samples, as defined in Equation 5.11.

$$\chi_{t\bar{t}}^2(p, f) = \sum_{i \in \text{bins}} \frac{(N_{\text{observed}}^i - N_{\text{predicted}}^i(p, f))^2}{N_{\text{predicted}}^i(p, f)} \quad (5.11)$$

Repeatedly computing $\chi_{t\bar{t}}^2$ for electron and muon fake rates incrementing independently, from 0.0 to 0.5 by steps of 0.01, a 51×51 -binned 2D $\chi_{t\bar{t}}^2$ distribution is produced using $t\bar{t}$ MC information. Similarly, the best-fitted electron and muon fake rates are those that correspond to the minimum $\chi_{t\bar{t}}^2$.

Figure 5.21 shows the distribution of $\chi_{t\bar{t}}^2$ at different electron and muon fake rates measured in CR2 of 2017 $t\bar{t}$ MC events. The minimum $\chi_{t\bar{t}}^2$ of 116.474 is marked with the intersection of two straight black lines. The point corresponds to electron and muon fake rates of 0.08 ± 0.01 and 0.14 ± 0.01 .

Figure 5.22 shows the $\chi^2_{t\bar{t}}$ distribution measured in SR of 2017 $t\bar{t}$ MC events. The minimum $\chi^2_{t\bar{t}}$ of 128.43 is marked with the intersection of two straight black lines. The point corresponds to electron and muon fake rates of 0.08 ± 0.01 and 0.12 ± 0.01 . The deviations between the two measurements are 0 for electron and 0.02 for muon. Although the two measurements agree within error, to be conservative, the deviations are considered in systematics. It is also reassuring to see that the two data-driven measurements are not too far from the cut-and-count values in Table 5.19. The “observed” and “predicted” distributions of leptons’ p_T in CR2 and SR in 2017 $t\bar{t}$ events are shown in Figures 5.23 and 5.24. The distributions for nonprompt leptons from “charm” and “light” origin, and muons from “fake”, have fluctuating or absent observed-predicted agreements (plotted in the lower panel) due to lack of statistics. The distribution of “fake” nonprompt electrons shows relatively good observed-predicted agreements at lower H_T^{lep} where there are sufficient statistics. The main contribution of nonprompt background is “bottom” nonprompts which show predicted distribution above that of observed due to the deviations in cut-and-count fake rates binned by η (applied) and that by H_T^{lep} (plotted).

For 2018 $t\bar{t}$ MC events, Figure 5.25 shows the distribution of $\chi^2_{t\bar{t}}$ at different electron and muon fake rates measured in CR2. The minimum $\chi^2_{t\bar{t}}$ of 96.99 is marked with the intersection of two straight black lines. The point corresponds to electron and muon fake rates of 0.08 ± 0.01 and 0.13 ± 0.01 . Figure 5.26 shows the $\chi^2_{t\bar{t}}$ distribution measured in SR. The minimum $\chi^2_{t\bar{t}}$ of 162.87 is marked with the intersection of two straight black lines. The point corresponds to electron and muon fake rates of 0.08 ± 0.01 and 0.11 ± 0.01 . The deviations between the two measurements are again 0 for electron and 0.01 for muon, which are included in systematics. Same as above, the two measurements agree within error and are close to the corresponding cut-and-count fake rates in Table 5.20. The “observed” and “predicted” distributions of leptons’ p_T in CR2 and SR in 2018 $t\bar{t}$ events are shown in Figures 5.27 and 5.28. These plots can be interpreted similarly to those in 2017.

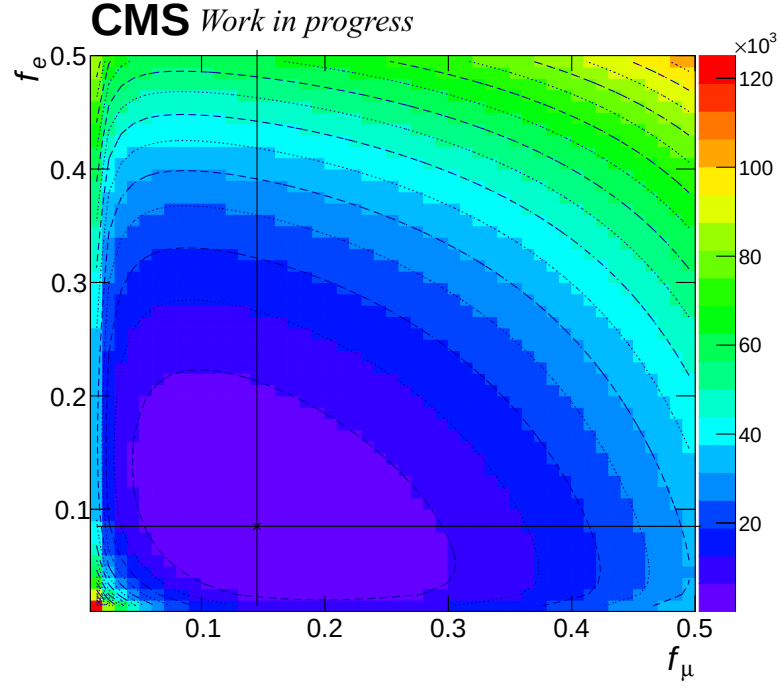


Figure 5.21: Measured χ^2_{CR2} in 2017 $t\bar{t}$ sample between the “predicted” and “observed” number of events in the multi-lepton channel CR2 summed over all flavour categories, as a function of electron (f_e) and muon (f_μ) nonprompt rates. The intersection of the two black lines indicates the position of the minimum.

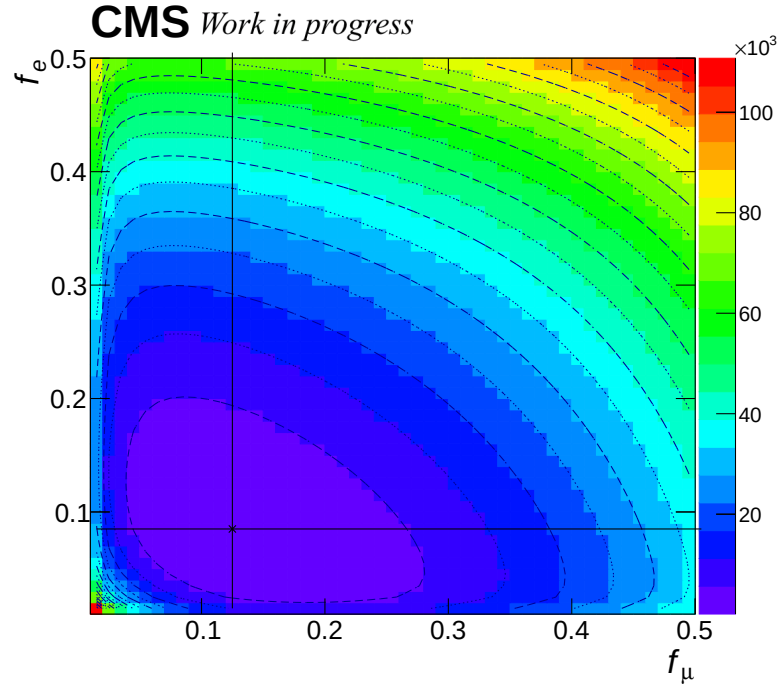


Figure 5.22: Measured χ^2_{SR} in 2017 $t\bar{t}$ sample between the “predicted” and “observed” number of events in the multi-lepton channel SR summed over all flavour categories, as a function of electron (f_e) and muon (f_μ) nonprompt rates. The intersection of the two black lines indicates the position of the minimum.

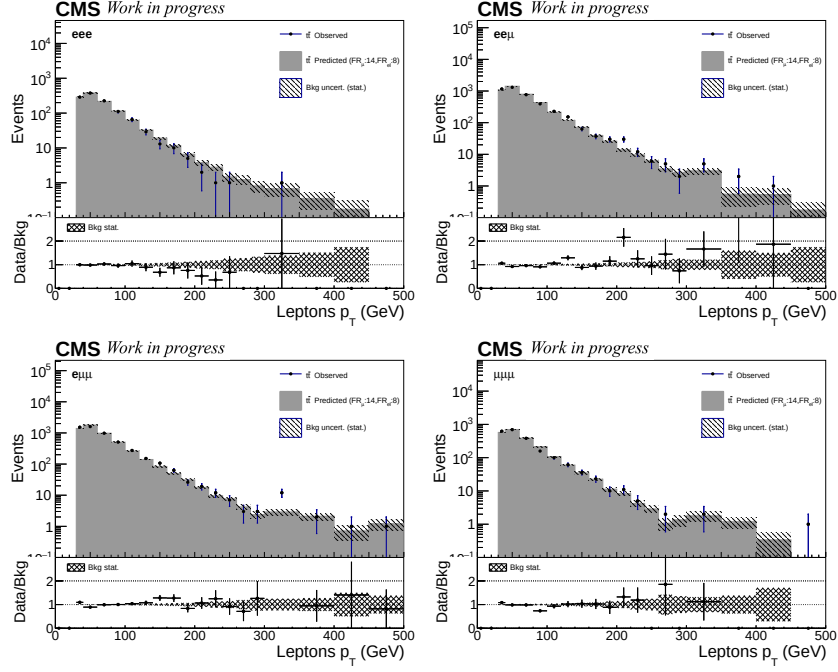


Figure 5.23: Distributions of leptons’ p_T in the multi-lepton channel CR2 in each flavour category. “Observed” event yield from 2017 $t\bar{t}$ sample is shown using black markers and the “predicted” event yield derived with the measured fake rates of $(f_e, f_\mu) = (0.08, 0.14)$ is shown using filled histograms. Statistical uncertainties on “predicted” are shown by the hatched region.

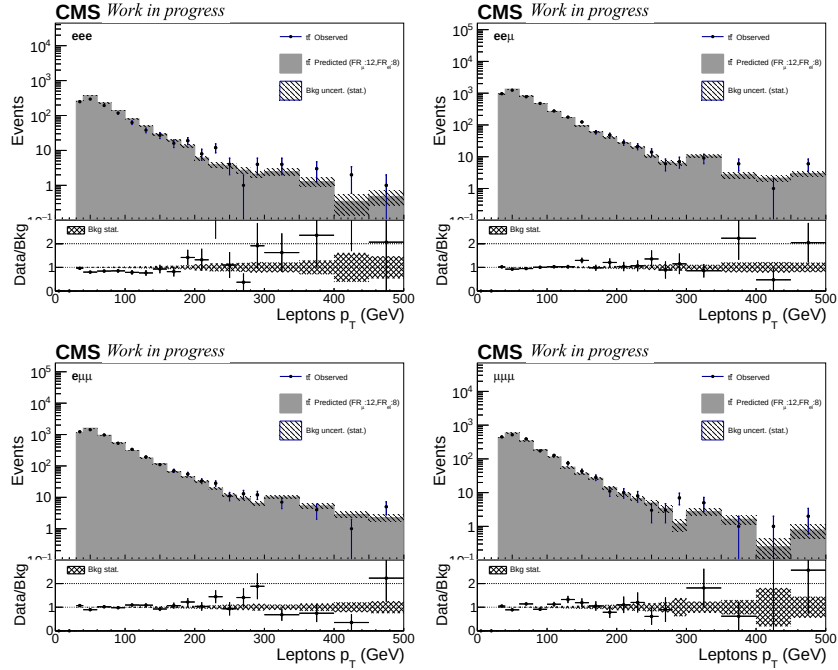


Figure 5.24: Distributions of leptons’ p_T in the multi-lepton channel CR2 in each flavour category. “Observed” event yield from 2018 $t\bar{t}$ sample is shown using black markers and the “predicted” event yield derived with the measured fake rates of $(f_e, f_\mu) = (0.08, 0.12)$ is shown using filled histograms. Statistical uncertainties on “predicted” are shown by the hatched region.

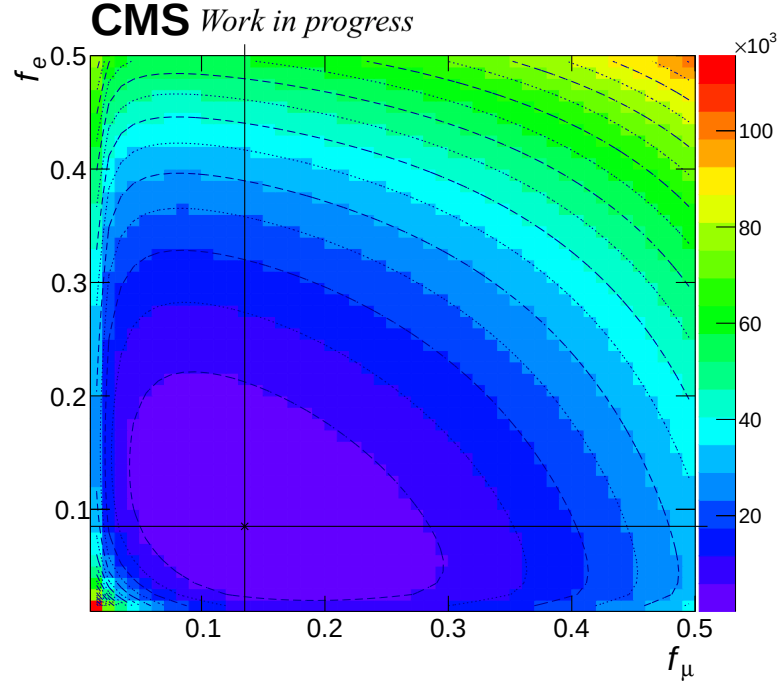


Figure 5.25: Measured χ^2_{tt} in 2018 $t\bar{t}$ sample between the “predicted” and “observed” number of events in the multi-lepton channel CR2 summed over all flavour categories, as a function of electron (f_e) and muon (f_μ) nonprompt rates. The intersection of the two black lines indicates the position of the minimum.

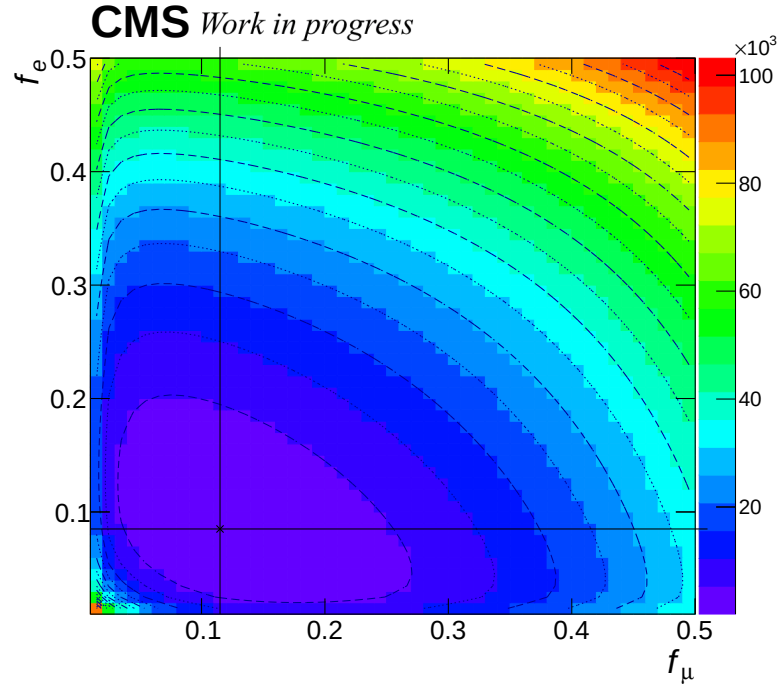


Figure 5.26: Measured $\chi^2_{t\bar{t}}$ in 2018 $t\bar{t}$ sample between the “predicted” and “observed” number of events in the multi-lepton channel SR summed over all flavour categories, as a function of electron (f_e) and muon (f_μ) nonprompt rates. The intersection of the two black lines indicates the position of the minimum.

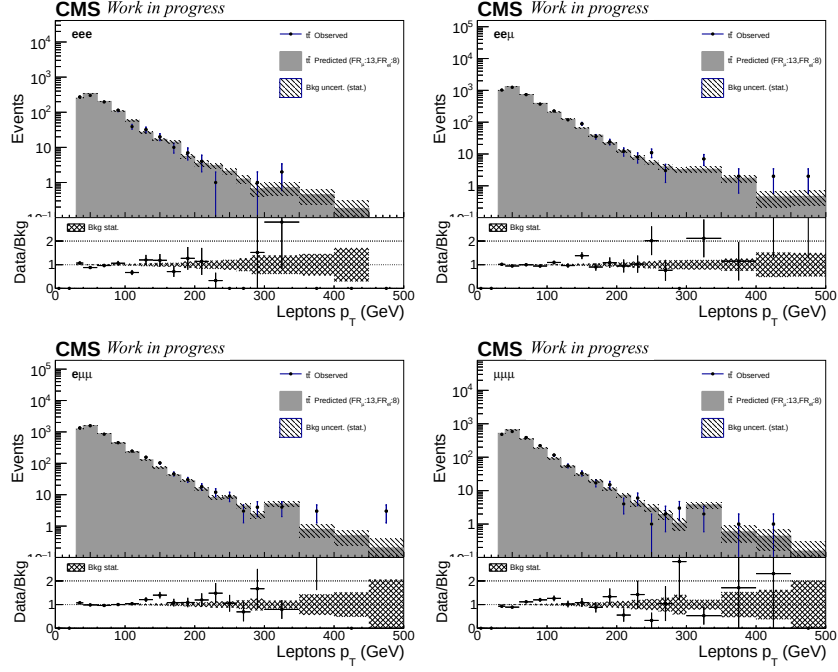


Figure 5.27: Distributions of leptons' p_T in the multi-lepton channel CR2 in each flavour category. “Observed” event yield from 2018 $t\bar{t}$ sample is shown using black markers and the “predicted” event yield derived with the measured fake rates of $(f_e, f_\mu) = (0.08, 0.13)$ is shown using filled histograms. Statistical uncertainties on “predicted” are shown by the hatched region.

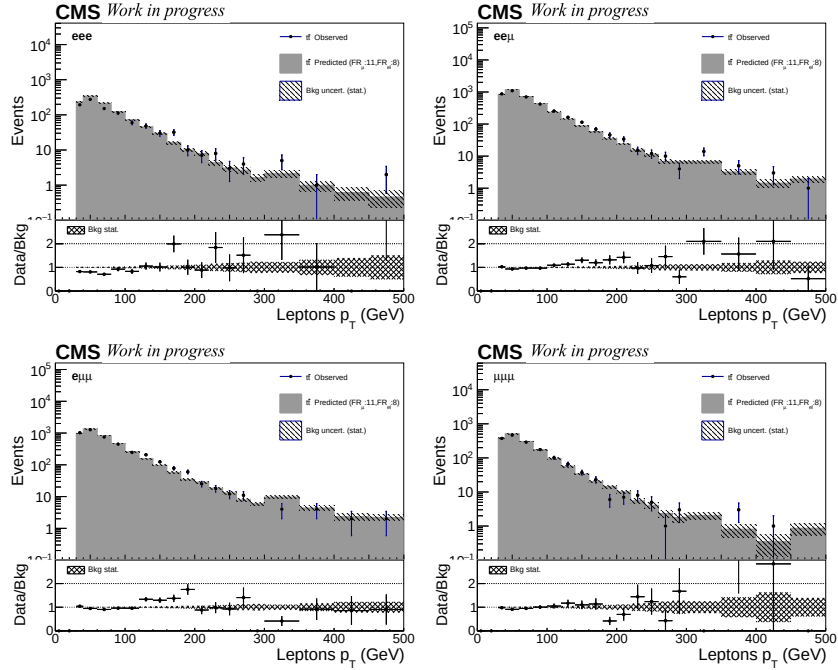


Figure 5.28: Distributions of leptons' p_T in the multi-lepton channel SR in each flavour category. “Observed” event yield from 2018 $t\bar{t}$ sample is shown using black markers and the “predicted” event yield derived with the measured fake rates of $(f_e, f_\mu) = (0.08, 0.11)$ is shown using filled histograms. Statistical uncertainties on “predicted” are shown by the hatched region.

Table 5.21 shows the number of “observed” and “predicted” events in each flavour category per year when applying the cut-and-count fake rates from Section 5.6.2. The deviations between “observed” and “predicted” events yields reported in the last column of the table is included as part of the “ ℓ NP rate” systematic uncertainty.

Table 5.21: Observed and predicted yields of $t\bar{t}$ MC events with two prompt and one nonprompt leptons. Predicted yield is obtained by applying cut-and-count fake rates in Table 5.19 and 5.20

ppf events	observed	predicted	$\Delta \left(\frac{\text{obs}-\text{pre}}{\text{pre}} \right)$
2017 all region			
eee	3492.0 ± 59.0	3592.0 ± 20.0	2.8%
$ee\mu$	12720.0 ± 110.0	13080.0 ± 43.0	2.8%
$e\mu\mu$	15460.0 ± 120.0	15900.0 ± 52.0	2.8%
$\mu\mu\mu$	5697.0 ± 75.0	6158.0 ± 34.0	7.5%
All	37370.0 ± 190.0	38730.0 ± 79.0	3.5%
2018 all region			
eee	3171.0 ± 56.0	3372.0 ± 20.0	6.0%
$ee\mu$	11880.0 ± 110.0	12150.0 ± 41.0	2.2%
$e\mu\mu$	14010.0 ± 120.0	14190.0 ± 47.0	1.3%
$\mu\mu\mu$	5448.0 ± 74.0	5605.0 ± 31.0	2.8%
All	34510.0 ± 190.0	35320.0 ± 73.0	2.3%

5.6.4 Closure tests in SS2L Channel

For SS2L channel, selected events for closure tests are required to be labelled pf , meaning that one of the dilepton is prompt, and the other is nonprompt. The simulated $t\bar{t}$ events are separated by the flavours of the generated (GEN) mother from which the nonprompt lepton originate. The origin of each lepton is classified with the following steps:

1. Match to a stable outgoing final state GEN particle with the same lepton flavour within a cone of $\Delta R < 0.1$ OR match to the closest GEN particle
2. IF matched GEN is found within a cone of $\Delta R < 0.3$, continue to the next step. ELSE origin = **“Unmatched”**
3. IF matched GEN is nonprompt, continue to the next step. ELSE discard events
4. IF matched GEN is of the same flavour, continue to the next step. ELSE origin = **“Fake”**
5. IF mother of matched GEN is a b quark, origin = **“Bottom”**

ELSE IF mother of matched GEN is a c quark, origin = **“Charm”**

ELSE origin = **“Light”**

5.6.5 SS2L Channel: Cut-and-Count Fake Rate

In each selected pf event, there is one and only one nonprompt lepton in the dilepton. The nonprompt lepton is categorized as tight or loose-not-tight for passing and failing the “tight” requirements. The ratio of $\frac{\text{tight}}{\text{tight}+\text{loose-not-tight}}$ is calculated as the cut-and-count fake rates, listed in Table 5.22 and 5.23. In 2017 $t\bar{t}$ MC sample, the overall cut-and-count electron fake rate is $f_e = 0.11$, and the muon fake rate is $f_\mu = 0.17$. In 2018 $t\bar{t}$ MC sample, the cut-and-count fake rates are $f_e = 0.10$, and $f_\mu = 0.16$.

Recall the measurement in Section 5.5.4: $(f_e, f_\mu) = (0.10, 0.15)$ in 2017 and $(f_e, f_\mu) = (0.08, 0.16)$ in 2018. These values are very close to the cut-and-count value SS2L in $t\bar{t}$ MC events, within ± 0.01 to 0.02.

Table 5.22: Fake rates measured by cut-and-count in 2017 $t\bar{t}$ sample with two same-sign leptons for each origin of the nonprompt lepton.

Origin	e fake rate	μ fake rate
Light Quarks	0.079 ± 0.003	0.147 ± 0.070
Charm Quarks	0.081 ± 0.006	0.246 ± 0.015
Bottom Quarks	0.118 ± 0.000	0.181 ± 0.001
Fakes	0.087 ± 0.001	0.041 ± 0.001
Unmatched	0.367 ± 0.010	0.017 ± 0.001
Overall	0.110 ± 0.000	0.169 ± 0.001

Table 5.23: Fake rates measured by cut-and-count in 2018 $t\bar{t}$ sample with two same-sign leptons for each origin of the nonprompt lepton.

Origin	e fake rate	μ fake rate
Light Quarks	0.072 ± 0.003	0.273 ± 0.178
Charm Quarks	0.068 ± 0.006	0.227 ± 0.015
Bottom Quarks	0.115 ± 0.000	0.172 ± 0.001
Fakes	0.086 ± 0.001	0.041 ± 0.001
Unmatched	0.337 ± 0.010	0.015 ± 0.001
Overall	0.108 ± 0.000	0.161 ± 0.001

Among the ppf events, if the nonprompt lepton passes the “tight” requirements, the event is labelled as “observed”. These “observed” events are treated as pseudo-data for the χ^2 scanning. By definition, weighting the number of loose-not-tight nonprompt leptons with fake rate gives the “predicted” number of tight nonprompt leptons. “Predicted” background is fitted to the “observed” pseudo-data. There is no need to consider prompt background.

5.6.6 SS2L Channel: Test for Matrix Method

Similar to multi-lepton channel, if the nonprompt lepton in the SS pair is tight, the event is labelled as “observed”. These “observed” events are treated as pseudo-data. The “predicted” events are derived by weighting the number of loose-not-tight events with $f(f-1)$ where f is the fake rate. The matrix method can be tested by comparing the distributions of “observed” and “predicted” events using $t\bar{t}$ MC sample with cut-and-count fake rates in Table 5.22–5.23 applied, independently for each origin of nonprompt leptons. The results are shown in Figures 5.29–5.32.

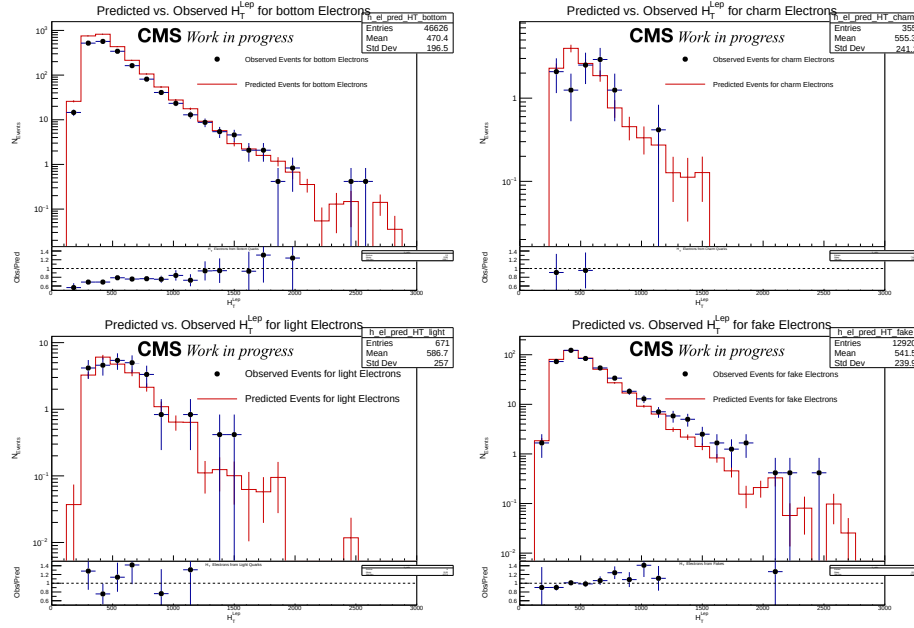


Figure 5.29: Distributions of H_T^{lep} in the SS2L channel for nonprompt electrons decaying from b quarks (top left), c quarks (top right), and light quarks (bottom left), and electrons matched to a "fake" GEN particle with the wrong flavour (bottom right). "Observed" event yield from 2017 $t\bar{t}$ sample is shown using black markers and the "predicted" event yield derived with the cut-and-count fake rates in Table 5.22 is shown using red solid lines. Statistical uncertainties are shown by the error bars.

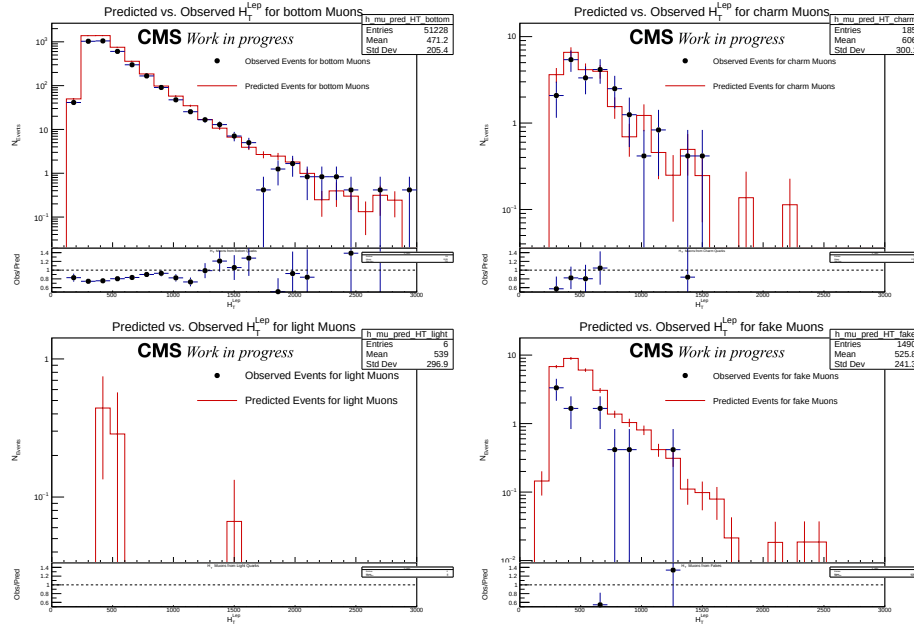


Figure 5.30: Distributions of H_T^{lep} in the SS2L channel for nonprompt muons decaying from b quarks (top left), c quarks (top right), and light quarks (mid left), muons matched to a "fake" GEN particle with the wrong flavour (mid right), and muons not matched to any GEN particle (bottom). "Observed" event yield from 2017 $t\bar{t}$ sample is shown using black markers and the "predicted" event yield derived with the cut-and-count fake rates in Table 5.22 is shown using red solid lines. Statistical uncertainties are shown by the error bars.

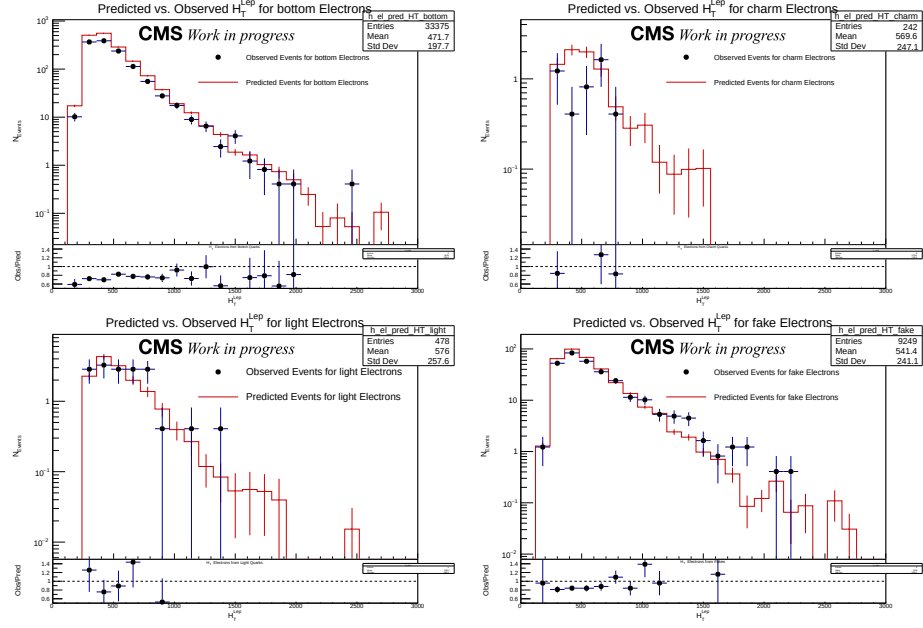


Figure 5.31: Distributions of H_T^{lep} in the SS2L channel for nonprompt electrons decaying from b quarks (top left), c quarks (top right), and light quarks (mid left), electrons matched to a “fake” GEN particle with the wrong flavour (mid right), and electrons not matched to any GEN particle (bottom). “Observed” event yield from 2018 $t\bar{t}$ sample is shown using black markers and the “predicted” event yield derived with the cut-and-count fake rates in Table 5.23 is shown using red solid lines. Statistical uncertainties are shown by the error bars.

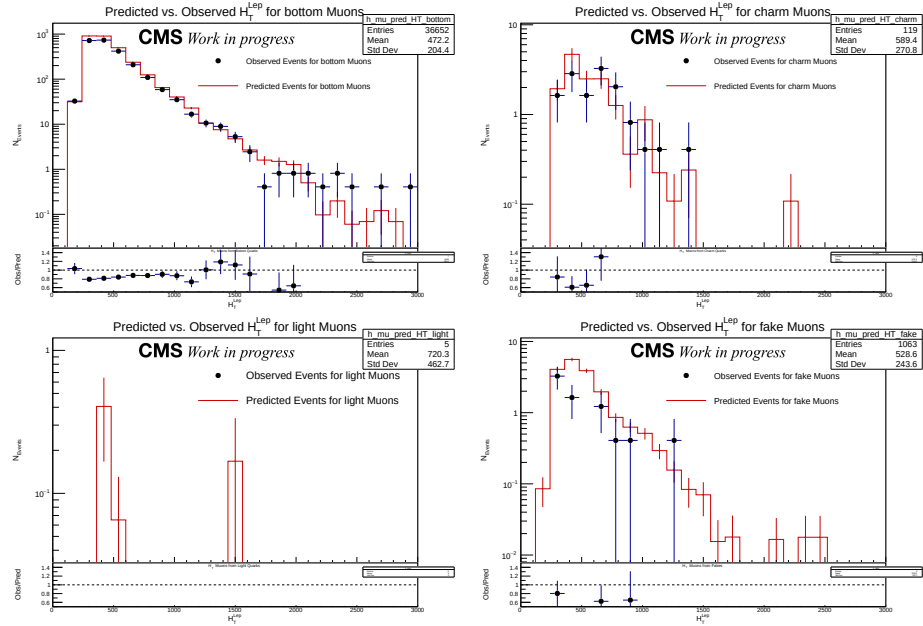


Figure 5.32: Distributions of H_T^{lep} in the SS2L channel for nonprompt muons decaying from b quarks (top left), c quarks (top right), light quarks (mid left), muons matched to a “fake” GEN particle with the wrong flavour (mid right), and muons not matched to any GEN particle (bottom). “Observed” event yield from 2018 $t\bar{t}$ sample is shown using black markers and the “predicted” event yield derived with the cut-and-count fake rates in Table 5.23 is shown using red solid lines. Statistical uncertainties are shown by the error bars.

5.7 Systematic Uncertainties

Various sources of systematic uncertainties and their applied values are summarized in Table 5.24. Sources of rate uncertainties are cited in the table when applicable. There is no uncertainty on muon tracking [133]. Shape uncertainties are described below:

- Electron ID scale factors: the uncertainty is applied in p_T - η bins according to values provided by the EGamma POG [110].
- Inelastic cross section (pileup): this uncertainty is obtained by applying the recomputed pileup weights after varying the total inelastic cross section of 69.2 mb by $\pm 4.6\%$
- L1 trigger timing (prefiring): the uncertainty on prefiring weights (2017 only) is applied in p_T - η bins based on the instructions from POG [134].
- Trigger efficiency: the uncertainties for all triggers are flat rates except for the leading leg of the di-electron trigger, which is applied in p_T - η bin [128]. A flat uncertainty of 2% is added to account for the different definitions of “tight” leptons between this and the cited analyses [128]. Uncertainties are treated as uncorrelated for each flavour category.
- Renormalization Scale: this uncertainty on the QCD renormalization μ_R and factorization μ_F scales in MC simulations is obtained by varying both scales up and down by a factor of two, including their symmetric shifts, in each bin. For signals, this uncertainty is scaled to represent only the effect of event selections in the analysis, i.e., acceptance-only.
- PDFs: the bin-by-bin upward and downward shifts are computed by the quadrature sum of Hessian uncertainties for all PDF variations in a set $\left(\sqrt{\sum_i (\text{yield}_{\text{PDF}_i} - \text{yield}_{\text{nom}})^2} \right)$. The shifted histograms are created by varying the nominal histograms with this uncertainty. In 2017–2018 simulations of background processes, the NNPDF3.1 PDF set with 100 variations is considered. In 2017–2018 signal samples, the default NNPDF3.1 PDF set creates many events with negative weights, and thus the PDF4LHC15 PDF set is used instead. The uncertainty for signals is scaled to represent only the effect of event selections in the analysis, i.e., acceptance-only.

- Jet energy scale and resolution (JEC and JER): these uncertainties on AK4 jets are varied according to the method prescribed by the POG [121, 122] and propagated to distributions of p_T^{miss} and other variables calculated with jets (e.g., S_T , H_T^{lep}).
- DeepJet b-tagging: the uncertainties for b or c quark tagging and light quark or gluon misidentification are applied separately.

Table 5.24: Summary of systematic uncertainties grouped according to their correlations across channels. Ranges indicate representative values across different lepton flavour categories. Uncertainties that effect different background sources are labelled as “MC” for all simulations, “OS” for opposite-sign dilepton events used to derive CMID background, and “NP” for nonprompt background. Uncertainties on signal are labelled as “Accept.” when they are scaled for acceptance-only.

Source	Uncertainty		Samples	
	2017	2018	Signal	Bkgd.
<i>Correlated across all channels</i>				
Luminosity [135]	2.3%	2.5%	Yes	MC
μ ID + Iso [114, 115, 116]	2.5-7.5%	2.5-7.5%	Yes	MC
e Reco + Iso [113, 136]	1.4-4.2%	1.8-5.4%	Yes	MC
e ID [110]	$\pm\sigma(p_T, \eta)$		Yes	MC
Pileup	$\sigma_{\text{inel.}} \pm 4.6\%$		Yes	MC
Trigger prefiring	$\pm\sigma$		Yes	MC
Jet energy scale	$\pm\sigma(p_T, \eta)$		Yes	MC
Jet energy res.	$\pm\sigma(\eta)$		Yes	MC
Ren./Fact. scale	env($\times 2, \times 0.5$)		Accept.	MC
PDF	$\pm\sigma$		Accept.	MC
<i>SS dilepton channel</i>				
ee trigger [128]	$\pm\sigma(p_T, \eta)$		Yes	MC
$e\mu$ trigger [128]	$\pm\sigma(p_T, \eta)$		Yes	MC
$\mu\mu$ trigger [128]	$\pm\sigma(p_T, \eta)$		Yes	MC
Charge misid.	34.5%	60.0%	No	OS
e/μ prompt rate	$\pm\sigma(p_T)$		No	NP
e NP rate	NP rate+0.07	NP rate+0.08	No	NP
μ NP rate	NP rate-0.02		No	NP
<i>Multi-lepton channel</i>				
ee trigger [128]	$\pm\sigma(p_T, \eta)$		Yes	MC
$e\mu$ trigger [128]	$\pm\sigma(p_T, \eta)$		Yes	MC
$\mu\mu$ trigger [128]	$\pm\sigma(p_T, \eta)$		Yes	MC
e prompt rate	1.0-6.6%	0.9-7.3%	No	NP
μ prompt rate	1.0-6.6%	1.2-5.5%	No	NP
ℓ NP rate	3.9-23.0%	1.4-24.6%	No	NP
μ NP rate in $\eta(\mu)$	0-38.7%	1.0-1.4%	No	NP
e/μ NP rate	$\pm\sqrt{\sigma_{\text{gaus}}^2 + \sigma_{\Delta(f_{\text{SR}}, f_{\text{CR2}})}^2}$		No	NP
b tag: b/c	$\pm\sigma(p_T)$		Yes	MC
b tag: light	$\pm\sigma(p_T)$		Yes	MC

More details on systematic uncertainties on nonprompt background in multi-lepton channel are given below in Table 5.25 and 5.26. In multi-lepton channel, uncertainties from prompt rate (Section 5.5.3) are estimated by comparing yields. Separately for each flavour, the prompt rate is replaced by unity to obtain the shifted event yields in SR. The percentage changes in yields are applied as rate uncertainties.

Uncertainties due to muon fake rate η -dependence are estimated by the differences in yields between the η -dependent and constant muon fake rates. Details can be found in Section 5.5.6 in Table 5.15 and 5.16.

Uncertainties from lepton nonprompt rate measured using the matrix method (described in Section 5.5.2 and 5.5.4) are estimated by the ratio of data-background deviation to nonprompt background yield. The ratios are measured in both $t\bar{t}$ MC events (Table 5.21) and CR2 in data (Table 5.28–5.29). The “ ℓ NP rate” uncertainty is the quadratic sum of the two ratios, applied as rate uncertainties. There is another source of systematic uncertainties on nonprompt rate which is lepton-flavour dependent. “ e/μ NP rate” is a shape uncertainty obtained by shifting the electron and muon nonprompt rates up and down. The size of the shift is determined by the quadrature sum of measurement error from marginalizing the χ^2 scan in Section 5.5.4 and the deviation between nonprompt rates measured in SR and CR2 in $t\bar{t}$ MC samples in Section 5.6.3.

Table 5.25: Systematic uncertainties on the data-driven nonprompt background from 2017 data in multi-lepton channel.

Sources	Uncertainty (2017)					Measurement
	eee	$ee\mu$	$e\mu\mu$	$\mu\mu\mu$	All	
e prompt rate	6.6%	2.8%	1.0%	0.0%	2.0%	Change in yield when p_e set to 1
μ prompt rate	0.0%	1.0%	1.5%	6.6%	2.1%	Change in yield when p_μ set to 1
μ NP rate η	0.0%	10.1%	18.2%	38.7%	17.1%	Apply f_μ as function of η
ℓ NP rate	23.1%	16.0%	3.9%	13.5%	3.5%	$\frac{ \text{data-tot. bkg} }{\text{NP bkg}}$ in $t\bar{t}$ MC and data CR2
e/μ NP rate	shift NP rates by $\pm$ quad. sum of scan err. in data CR2 and $\Delta(f_{\text{SR}}, f_{\text{CR2}})$ in $t\bar{t}$ MC					

Details of systematic uncertainties on nonprompt and CMID backgrounds in SS2L channel are given below in Table 5.27. In SS2L channel, uncertainties from prompt rates are applied as shape uncertainties from shifting prompt rates up and down by the measurement errors in Section 5.5.3 for each lepton flavour. Uncertainties on nonprompt rates are one-sided shape uncertainties obtained by shifting the nonprompt rates to the best-fitted value from χ^2 scan in low H_T^{lep} CR (Section 5.5.7).

Table 5.26: Systematic uncertainties on the data-driven nonprompt background from 2018 data in multi-lepton channel.

Sources	Uncertainty (2018)					Measurement
	eee	$ee\mu$	$e\mu\mu$	$\mu\mu\mu$	All	
e prompt rate	7.1%	2.6%	0.8%	0.0%	1.9%	Change in yield when p_e set to 1
μ prompt rate	0.0%	1.2%	1.6%	5.5%	2.1%	Change in yield when p_μ set to 1
μ NP rate η	0.0%	1.1%	1.0%	1.4%	1.0%	Apply f_μ as function of η
ℓ NP rate	24.5%	5.6%	1.4%	8.2%	3.2%	$\frac{ \text{data-tot. bkg} }{\text{NP bkg}}$ in $t\bar{t}$ MC and data CR2
e/μ NP rate	shift NP rates by $\pm$ quad. sum of scan err. in data CR2 and $\Delta(f_{\text{SR}}, f_{\text{CR2}})$ in $t\bar{t}$ MC					

Uncertainties on CMID are contributed by two parts, the uncertainty on the negative log likelihood fit and the differences in topology of Drell-Yan events in which CMID rates are measured and $t\bar{t}$ events with opposite-sign dilepton in which CMID rates are applied. The uncertainty from event topologies are evaluated by counting the CMID rates using MC truth information. Each tight electrons with $p_T > 30$ GeV is matched to a GEN electron within a cone of $\Delta R < 0.1$. The reconstructed charge is compared with the MC truth information. The charge misidentification in Drell-Yan and $t\bar{t}$ MC samples are obtained by counting the fractions of leptons with reconstructed charge different from the matched GEN particles.

Table 5.27: Systematic uncertainties on data-driven backgrounds in SS2L channel for 2017 and 2018 data.

	2017	2018
e/μ prompt rate	shift prompt rate by measurement err.	
e NP rate	shift f_e by +0.07	shift f_e by +0.08
μ NP rate	shift f_μ by -0.02	
Charge mis-ID	34.5%	60.0%

5.8 Kinematic Distributions

5.8.1 3+L Channel: Control Region CR2

Table 5.28 and 5.29 recorded the event yields in the multi-lepton channel CR2. The values of $\frac{|\text{data}-\text{tot. bkg}|}{\text{NP bkg}}$ reported in the table are used to derive part of the “ ℓ NP rate” systematic uncertainty.

Table 5.28: Yields in distributions of leptons’ p_T in the multi-lepton channel CR2 in 2017 and the associated rate uncertainties on NP background.

	EEE	EEM	EMM	MMM	All
Data	135.0 ± 12.0	372.0 ± 19.0	435.0 ± 21.0	234.0 ± 15.0	1176.0 ± 34.0
NP background	113.8 ± 31.0	259.2 ± 56.0	337.8 ± 72.0	176.6 ± 81.0	887.4 ± 180.0
Total Background	161.1 ± 31.0	331.0 ± 57.0	425.9 ± 73.0	253.8 ± 81.0	1172.0 ± 190.0
$\frac{ \text{data}-\text{tot. bkg} }{\text{NP bkg}}$	22.9%	15.8%	2.7%	11.2%	0.5%

Table 5.29: Yields in distributions of leptons’ p_T in the multi-lepton channel CR2 in 2018 and the associated rate uncertainties on NP background.

	EEE	EEM	EMM	MMM	All
Data	210.0 ± 14.0	453.0 ± 21.0	600.0 ± 24.0	375.0 ± 19.0	1638.0 ± 40.0
NP background	113.7 ± 34.0	358.0 ± 46.0	481.1 ± 59.0	257.1 ± 51.0	1210.0 ± 150.0
Total background	183.1 ± 34.0	471.7 ± 47.0	601.8 ± 60.0	355.2 ± 52.0	1611.0 ± 160.0
$\frac{ \text{data}-\text{tot. bkg} }{\text{NP bkg}}$	23.7%	-5.2%	-0.4%	7.7%	2.2%

Figures 5.33–5.34 show the unblinded distributions of some kinematic variables in the multi-lepton channel CR2 combining all lepton flavour categories, with the measured prompt rates and fake rates applied. These distributions show that data agrees with background in general. $T\bar{T}$ distributions in plots are scaled up so that they are easier to read and comparable to background.

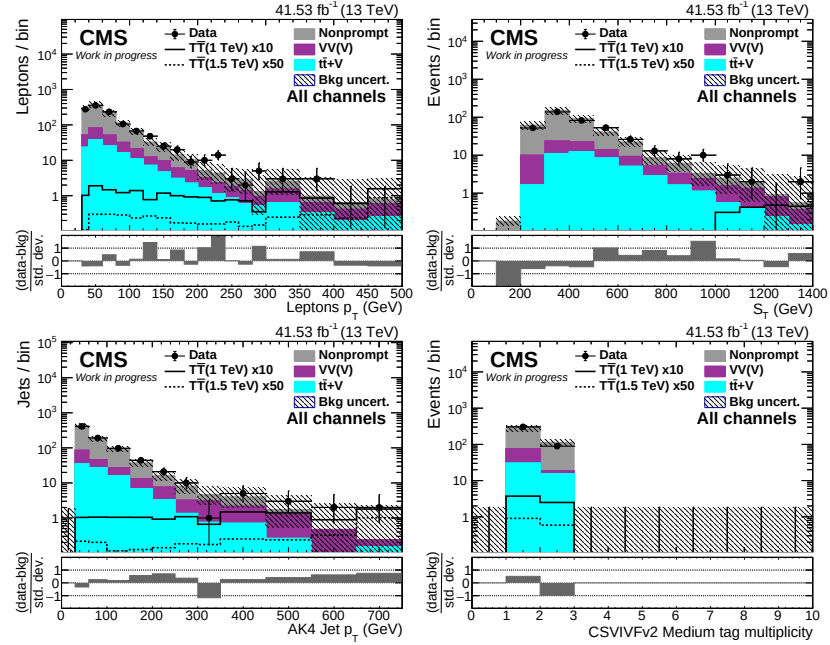


Figure 5.33: Distributions of leptons' p_T (top left), S_T (top right), AK4 jets' p_T (bottom left) and b-tagged AK4 jet multiplicity (bottom right) in the multi-lepton channel CR2 in each flavour category. Data from 2017 is shown using black markers, scaled $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with measured fake rates (Section 5.5.4). Statistical and systematic uncertainties are shown by the hatched region.

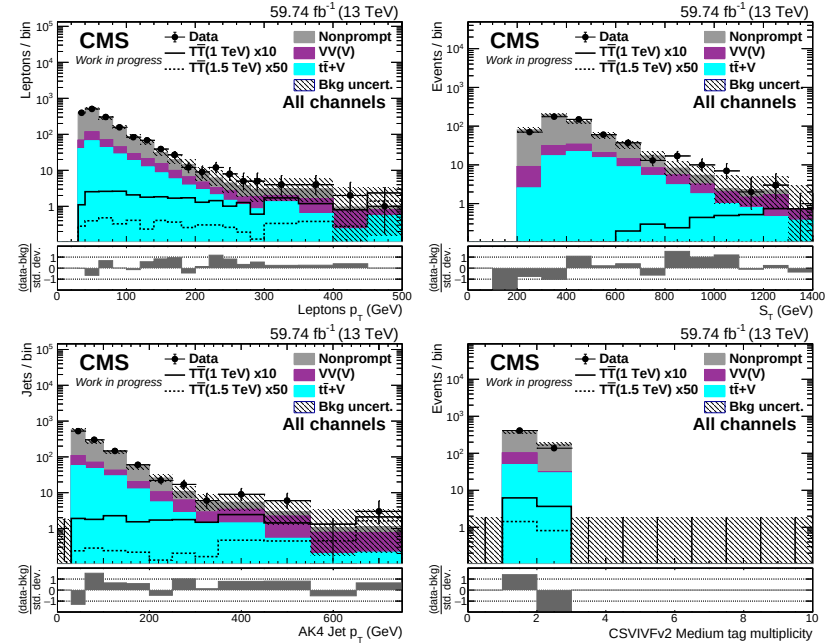


Figure 5.34: Distributions of leptons' p_T (top left), S_T (top right), AK4 jets' p_T (bottom left) and b-tagged AK4 jet multiplicity (bottom right) in the multi-lepton channel CR2 in each flavour category. Data from 2018 is shown using black markers, scaled $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with measured fake rates (Section 5.5.4). Statistical and systematic uncertainties are shown by the hatched region.

5.8.2 SS2L Channel: Control Regions

Figure 5.35 shows the unblinded distribution of the discriminant H_T^{lep} in each flavour category in the SS2L channel low H_T^{lep} control region ($H_T^{\text{lep}} < 1000$ GeV), with measured prompt rates, fake rates and charge misidentification rates applied. Figure 5.36 shows the distributions of H_T^{lep} in the control region with exactly 3 AK4 jets ($n_{\text{Jet}} = 3$). Figures 5.37 –5.38 show distributions of some kinematic variables in different flavour categories in the $n_{\text{Jet}} = 3$ control region. It is observed in these distributions that data, and predicted background agree within ~ 1 standard deviation overall.

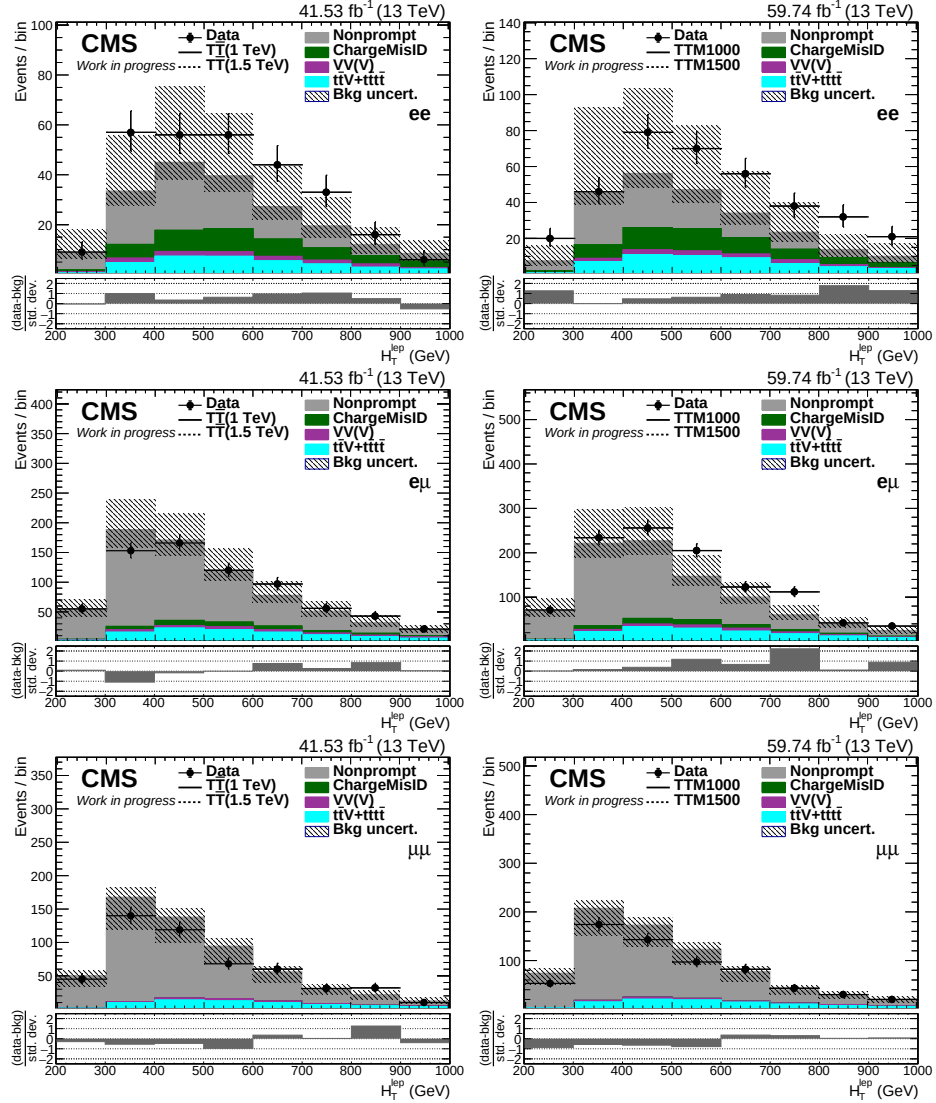


Figure 5.35: Distribution of H_T^{lep} in the SS2L channel low H_T^{lep} CR in each flavour category. Data from 2017 (left) and 2018 (right) is shown using black markers, TT with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Statistical and systematic uncertainties are shown by the hatched region.

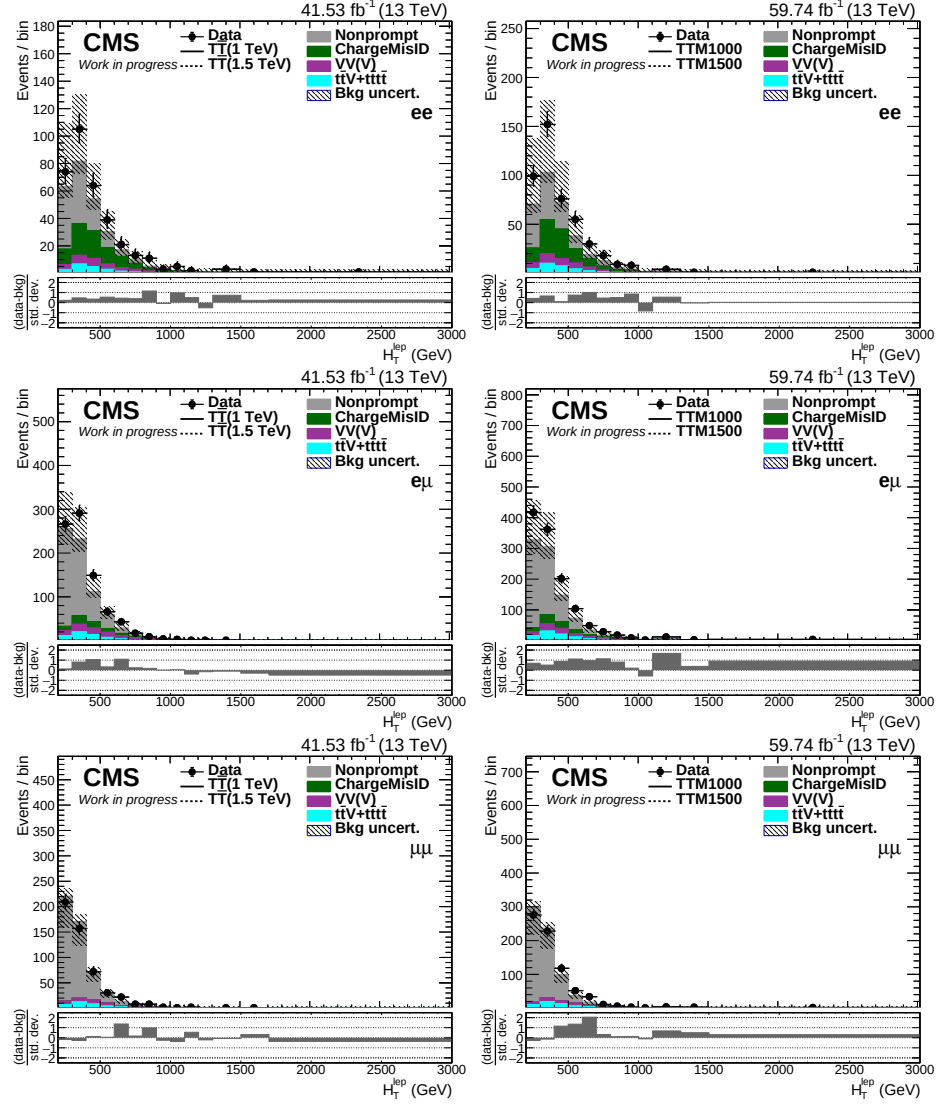


Figure 5.36: Distribution of H_T^{lep} in the SS2L channel $n\text{Jet} = 3$ CR in each flavour category. Data from 2017 (left) and 2018 (right) is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Statistical and systematic uncertainties are shown by the hatched region.

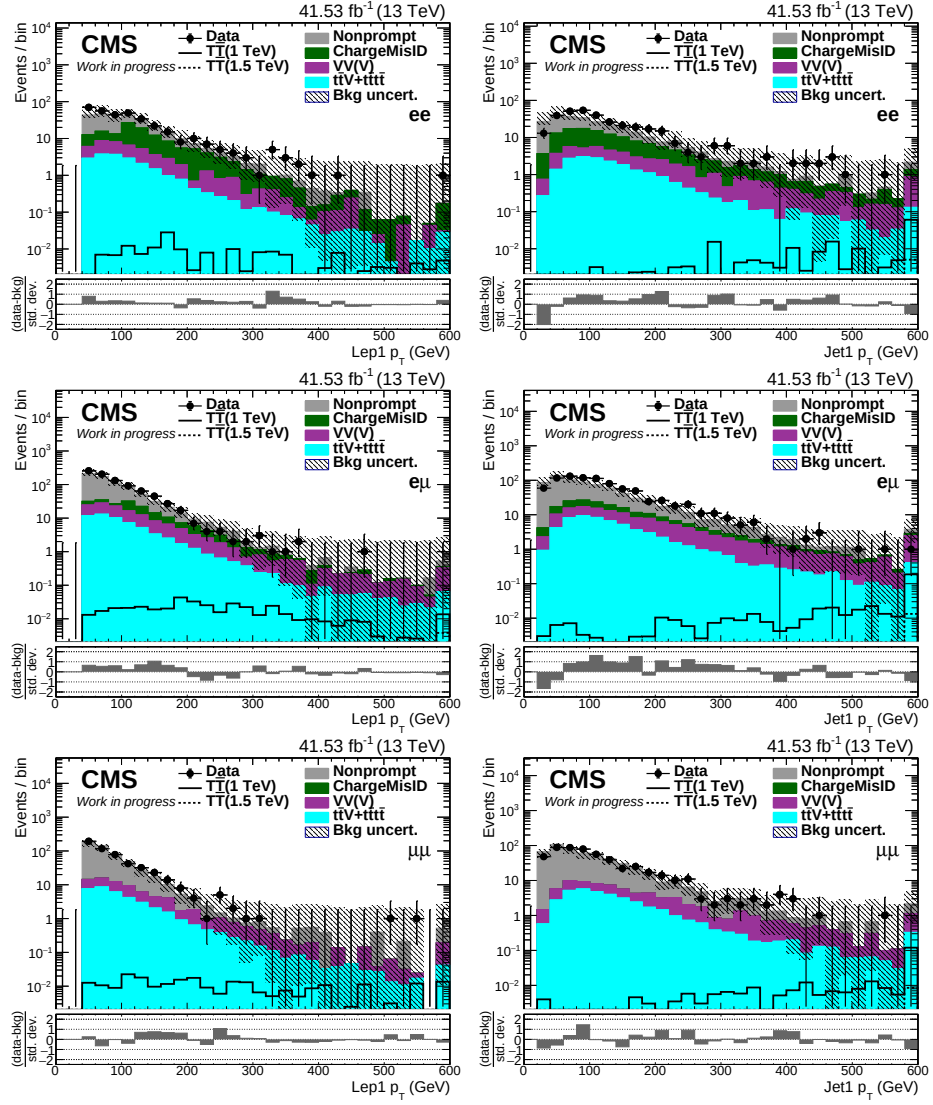


Figure 5.37: Distributions of leading lepton p_T (left) and leading jet p_T (right) in the SS2L channel $n_{\text{Jet}} = 3$ CR in each flavour category. Data from 2017 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Statistical and systematic uncertainties are shown by the hatched region.

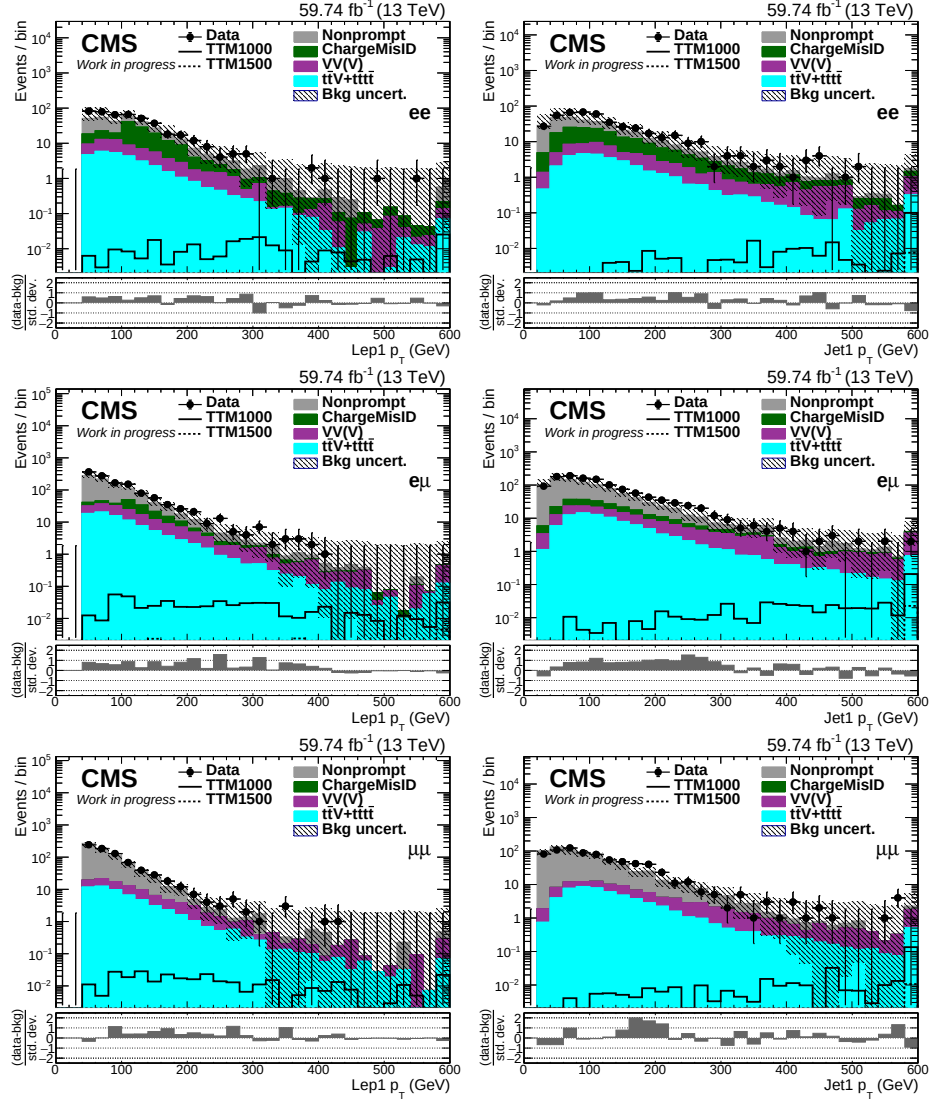


Figure 5.38: Distributions of leading lepton p_T (left) and leading jet p_T (right) in the SS2L channel $n\text{Jet} = 3$ CR in each flavour category. Data from 2018 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Statistical and systematic uncertainties are shown by the hatched region.

5.8.3 3+L Channel: Signal Region

Figures 5.39–5.40 show the unblinded distributions of some kinematic variables in the multi-lepton channel SR combining all lepton flavour categories, with the measured prompt rates and fake rates applied. Similar to plots for the control region, $T\bar{T}$ distributions in plots are scaled up so that they are easier to read and comparable to background.

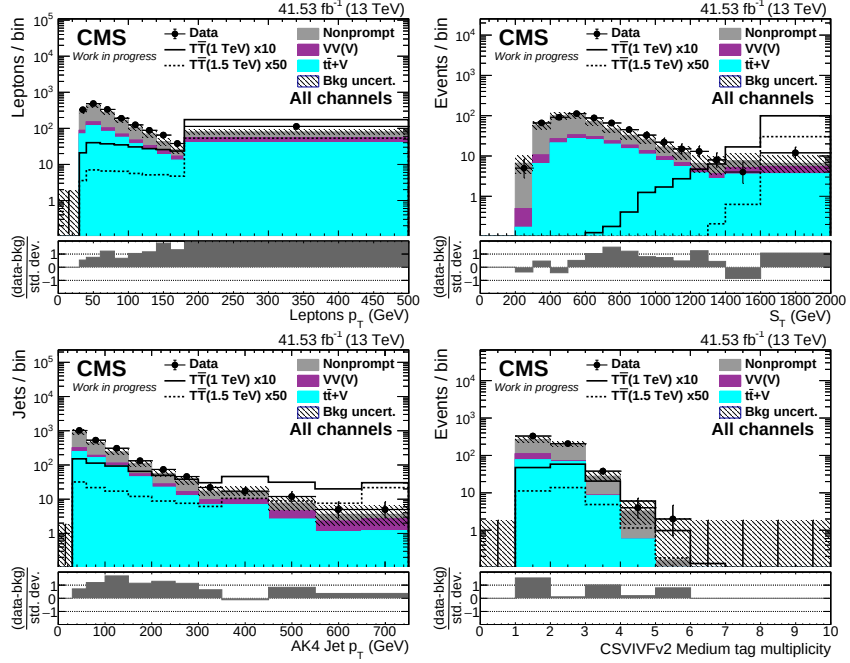


Figure 5.39: Distributions of leptons' p_T (top left), S_T (top right), AK4 jets' p_T (bottom left) and b-tagged AK4 jet multiplicity (bottom right) in the multi-lepton channel SR in each flavour category. Data from 2017 is shown using black markers, scaled $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with measured fake rates (Section 5.5.4). Statistical and systematic uncertainties are shown by the hatched region.

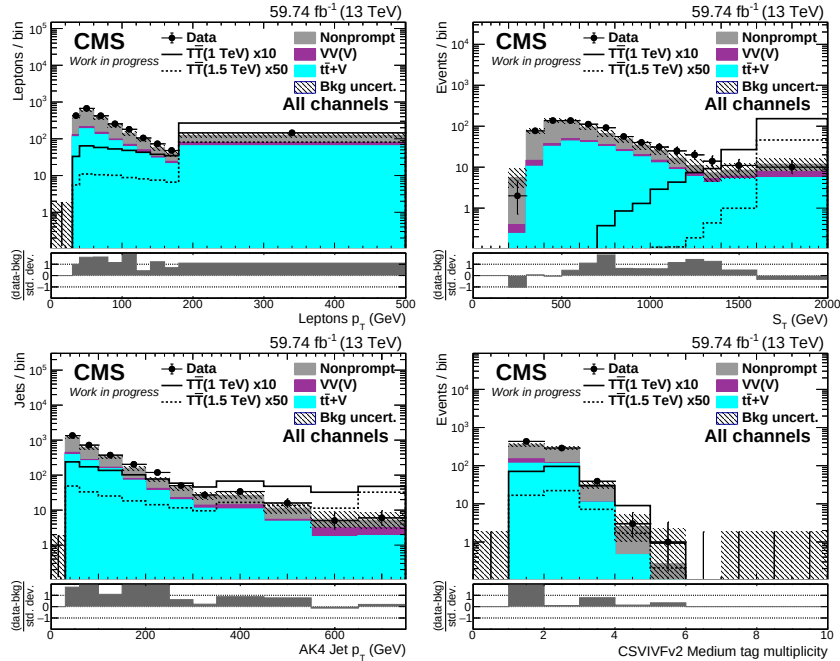


Figure 5.40: Distributions of leptons' p_T (top left), S_T (top right), AK4 jets' p_T (bottom left) and b-tagged AK4 jet multiplicity (bottom right) in the multi-lepton channel SR in each flavour category. Data from 2018 is shown using black markers, scaled $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with measured fake rates (Section 5.5.4). Statistical and systematic uncertainties are shown by the hatched region.

5.8.4 SS2L Channel: Signal Region

Figures 5.41–5.43 show the unblinded distributions of various kinematic variables in the SS2L channel SR in each flavour category, with the measured prompt rates, fake rates and charge misidentification rates applied.

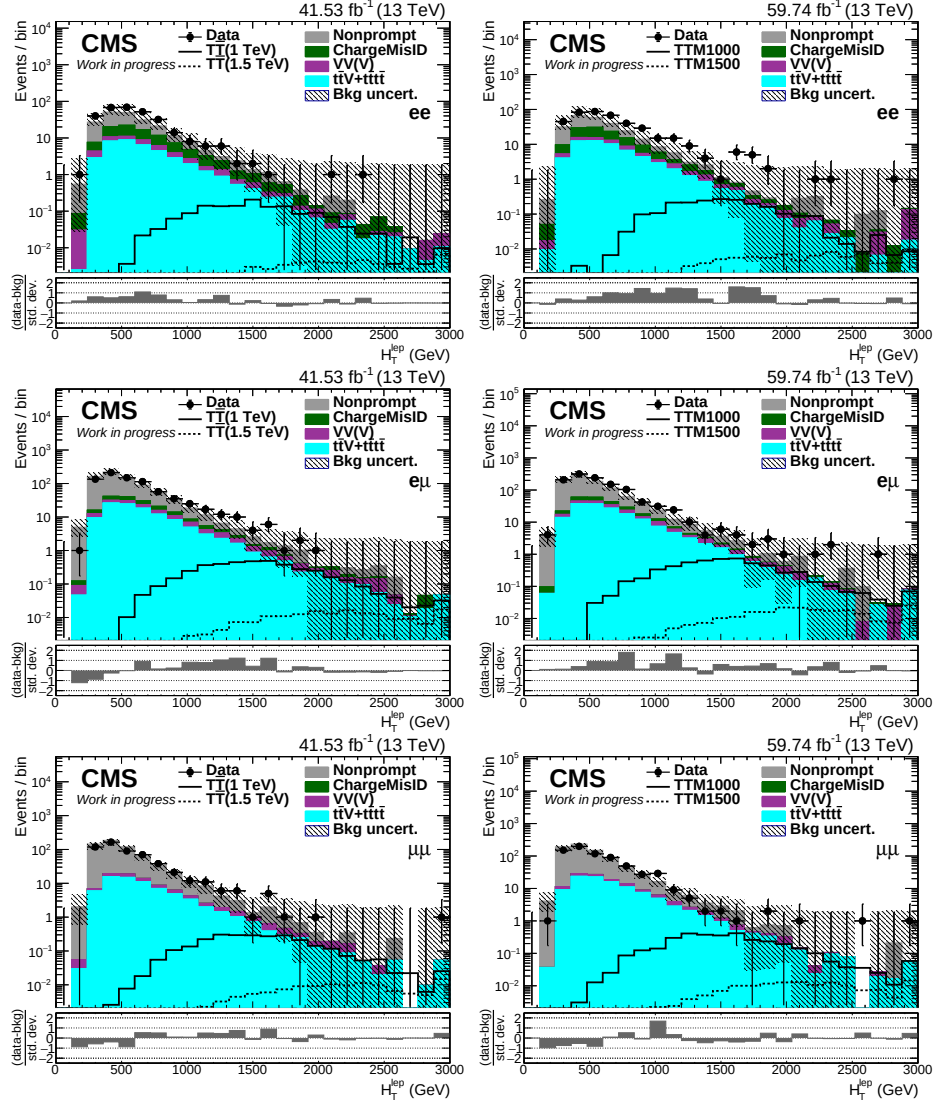


Figure 5.41: Distribution of H_T^{lep} in the SS2L channel SR in each flavour category. Data from 2017 (left) and 2018 (right) is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Statistical and systematic uncertainties are shown by the hatched region.

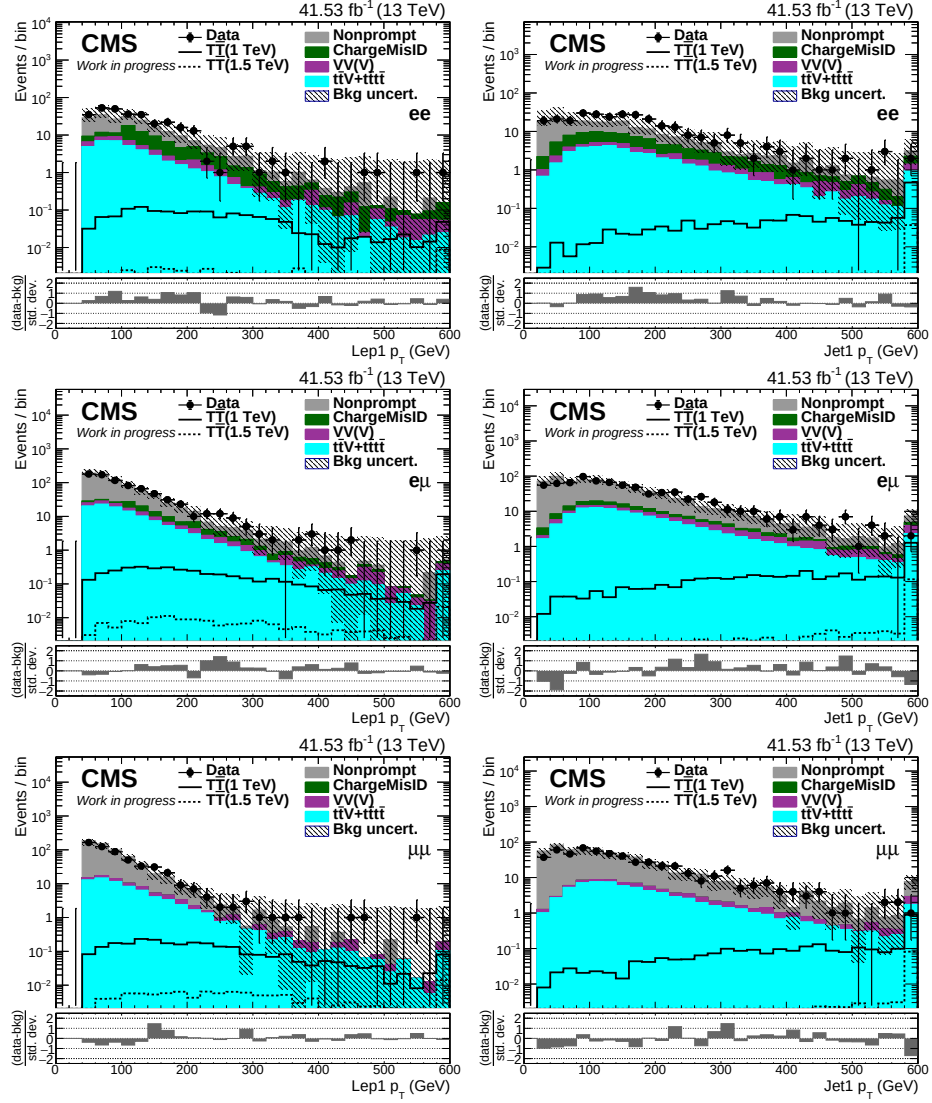


Figure 5.42: Distribution of leading lepton p_T (left) and leading jet p_T (right) in the SS2L channel SR in each flavour category. Data from 2017 is shown using black markers, Tt with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Statistical and systematic uncertainties are shown by the hatched region.

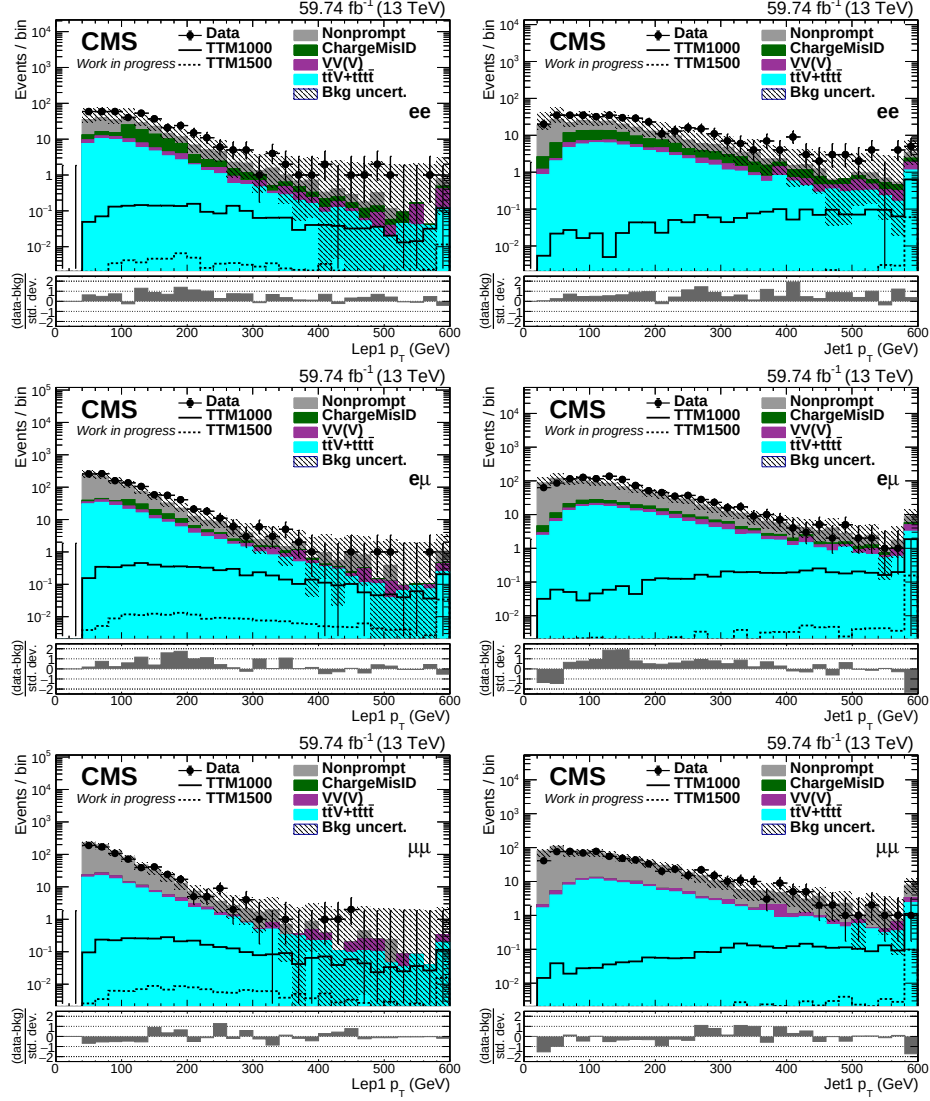


Figure 5.43: Distribution of leading lepton p_T (left) and leading jet p_T (right) in the SS2L channel SR in each flavour category. Data from 2018 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Statistical and systematic uncertainties are shown by the hatched region.

5.9 3+L Channel: Optimization and Final Selection

On top of event selections described in Section 5.4.2, some additional cuts may improve the exclusion limits. In multi-lepton channel, to investigate how various selection cuts affect the exclusion limits, the expected limits on $T\bar{T}$ (left) and $B\bar{B}$ (right) production cross sections are calculated for different sets of cuts added. To visualize the result, the expected limits are plotted as a function of the values at which the studied variable is cut. A number of kinematic variables were explored, and observable

changes are seen only in selections on the leading b-tagged AK4 jet's p_T and the relative p_T between the lepton-jet pair with minimum separation, which is the transverse component of a lepton's $\vec{p}_T$ relative to the closest jet's $\vec{p}_T$. Since VLQs are heavy, their decay products would be more boosted. Thus, a larger leading b-tagged AK4 jet's p_T and relative p_T flavour VLQ events. The measured expected limits on signal cross sections as a function of cut values are shown in Figures 5.44–5.47 for each of on the two variables.

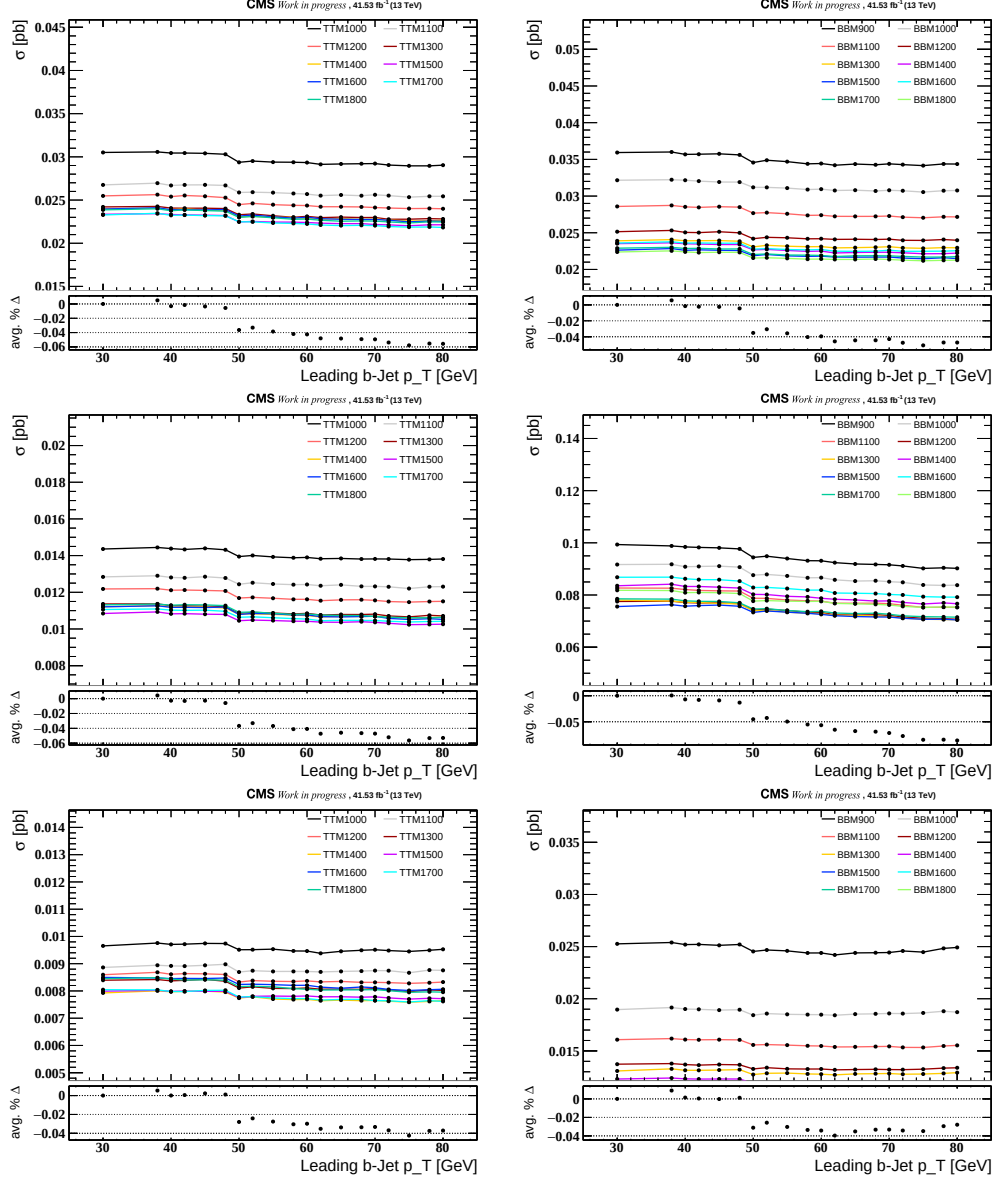


Figure 5.44: Expected exclusion limit on $T\bar{T}$ (left) and $B\bar{B}$ (right) production cross sections in 2017 in the multi-lepton channel as a function of cut values on p_T of the leading b-tagged jet, for the singlet (top), doublet (middle), and $T \rightarrow tZ/B \rightarrow tW$ -only (bottom) scenarios.

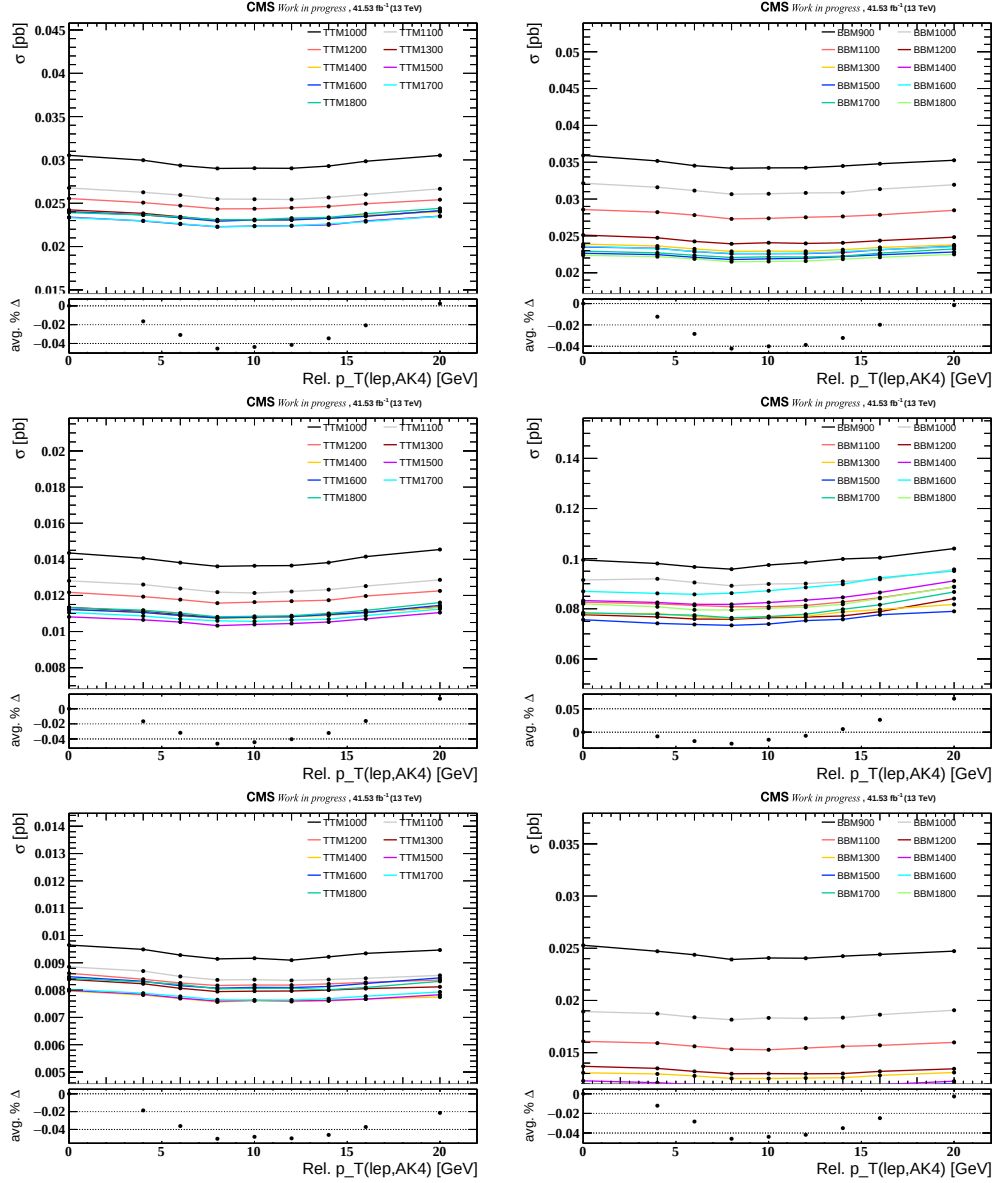


Figure 5.45: Expected exclusion limit on $T\bar{T}$ (left) and $B\bar{B}$ (right) production cross sections in 2017 in the multi-lepton channel as a function of cut values on relative p_T of the closest lepton-jet pair, for the singlet (top), doublet (middle), and $T \rightarrow tZ/B \rightarrow tW$ -only (bottom) scenarios.

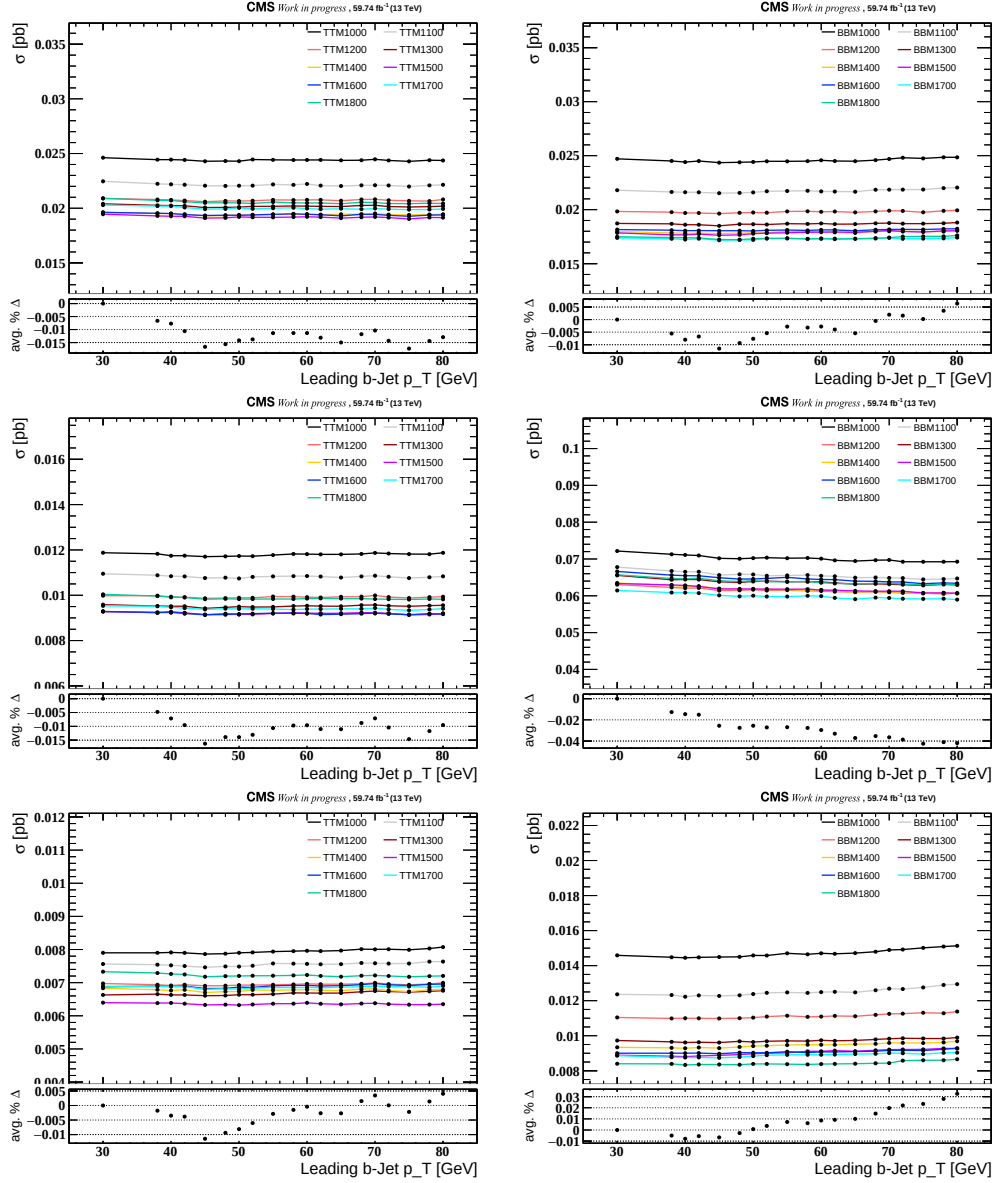


Figure 5.46: Expected exclusion limit on $T\bar{T}$ (left) and $B\bar{B}$ (right) production cross sections in 2018 in the multi-lepton channel as a function of cut values on p_T of the leading b-tagged jet, for the singlet (top), doublet (middle), and $T \rightarrow tZ/B \rightarrow tW$ -only (bottom) scenarios.

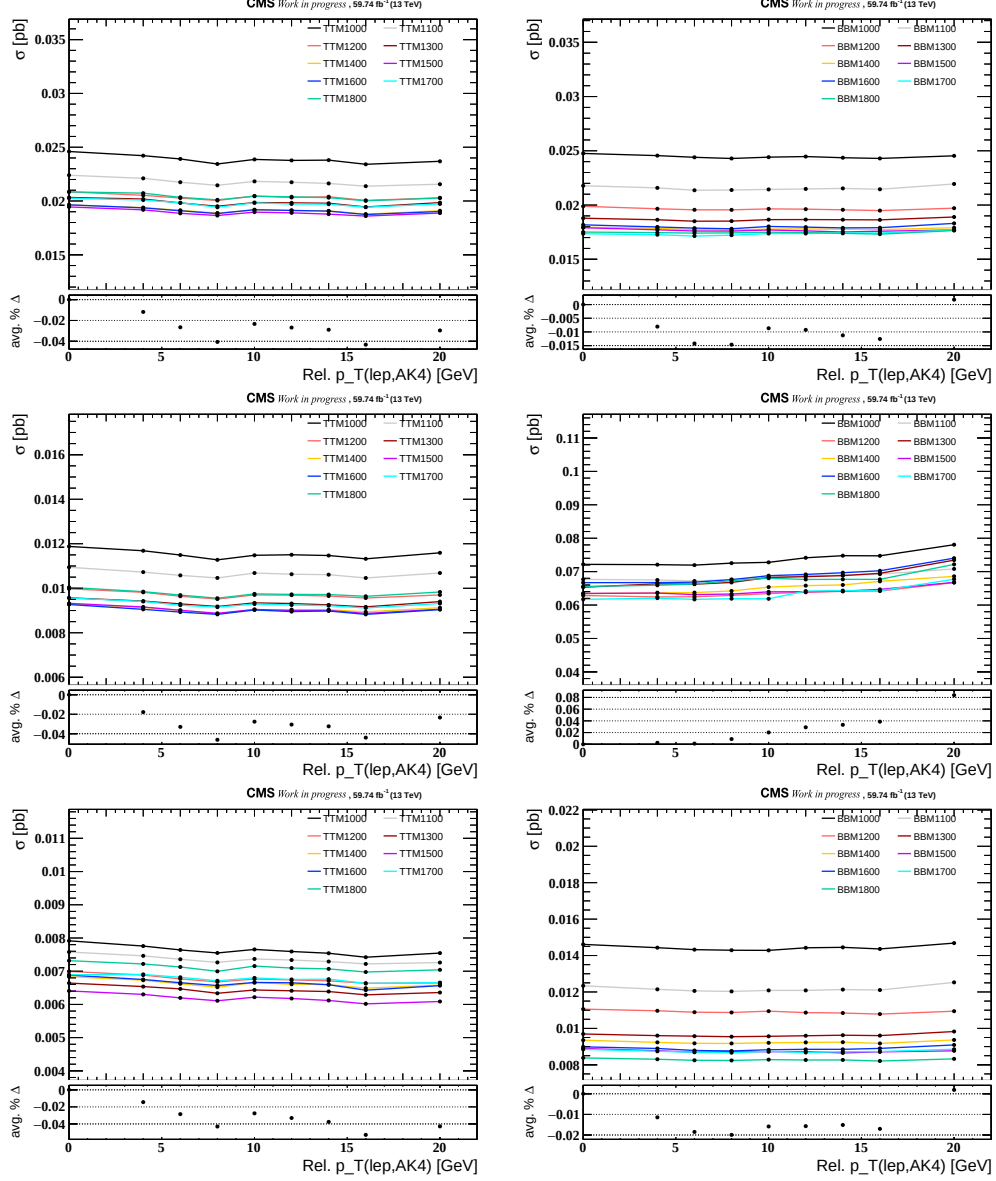


Figure 5.47: Expected exclusion limit on $T\bar{T}$ (left) and $B\bar{B}$ (right) production cross sections in 2018 in the multi-lepton channel as a function of cut values on relative p_T of the closest lepton-jet pair, for the singlet (top), doublet (middle), and $T \rightarrow tZ/B \rightarrow tW$ -only (bottom) scenarios.

Considering all mass points and scenarios studied, final selections are defined with $p_T^{\text{bjet1}} > 50$ GeV and $p_T^{\text{rel}}(\text{lep, jet}) > 8$ GeV for 2017, and $p_T^{\text{bjet1}} > 45$ GeV and $p_T^{\text{rel}}(\text{lep, jet}) > 8$ GeV for 2018, despite a small improvement of only 2%. Due to lack of statistics, bins at $S_T < 300$ GeV are omitted in exclusion limit measurement in Section 5.10 and statistical tests in Section 5.11.1. In SS2L channel, there are no additional selection except $H_T^{\text{lep}} > 400$ GeV to remove bins with insufficient statistics. Final selections are summarized in Table 5.30.

Table 5.30: Summary of final selections applied on top of SR (defined in Table 5.10) for finding limits and performing statistical tests in the following sections.

Final Selections	
multi-lepton	SS2L
• $S_T > 300$ GeV distributions • leading b-Jet $p_T > 50(45)$ GeV in 2017(2018) • $p_T^{\text{rel}}(\text{lep}, \text{jet}) > 8$ GeV	• $H_T^{\text{lep}} > 400$ GeV distributions

5.10 Limits

To find the exclusion limits, distributions of the discriminating variable in SR with final selections are used for shape-based likelihood fits. Using the Higgs Combine program [137], we calculate the 95% CL observed and expected upper limits on the production cross sections of $T\bar{T}$ and $B\bar{B}$ at each simulated mass points between 0.9 and 1.8 TeV (Table 5.4) using the “Asymptotic CLs” method. With “autoMCstats” method in Higgs Combine, statistical uncertainties are treated by the Barlow-Beeston lite method [138, 139] for event yields above 1.0. Otherwise, Poisson PDF is used. Systematic uncertainties in Table 5.24 enters the Higgs Combine as nuisance parameters. Normalization-only uncertainties are configured as log-normal priors, and shape uncertainties are treated with templates morphing with Gaussian priors. Signal cross section is configured as a flat prior. Correlations of nuisances across the years are listed in Table 5.31.

Table 5.31: List of uncertainties’ correlation between different years.

Systematic	correlated?
Luminosity	Yes
Electron ID	No
Electron ISO+RECO	Yes
Muon ID+ISO	Yes
Trigger	Yes
Prompt Rate	No
NP Rate	No
Charge mis-ID	No
JEC/JER	No
PDF	Yes
Renormalization Scale	Yes
Pileup	Yes
b tagging	Yes
b mis-tagging	Yes
Prefiring (2017 only)	2017-only

5.10.1 Limits in 3+L Channel

In multi-lepton channel, the discriminant used for likelihood fits is S_T , which is the scalar sum of AK4 jets' p_T , leptons' p_T and p_T^{miss} . The distributions must pass all selection criteria for SR defined in Section 5.4.3 and the final selections defined in Section 5.9.

The S_T distributions passing the final selections are passed on to the Higgs Combine for a shape-based likelihood fit. Figure 5.48 shows the background distributions in S_T , for 2017 and 2018, after a likelihood fit to all leptonic channels and full Run 2 data, which will be explained in Section 5.12. Data and $T\bar{T}$ signal at VLQ mass of 1.0 (1.5) TeV (before the fit) are overlaid in the same figure.

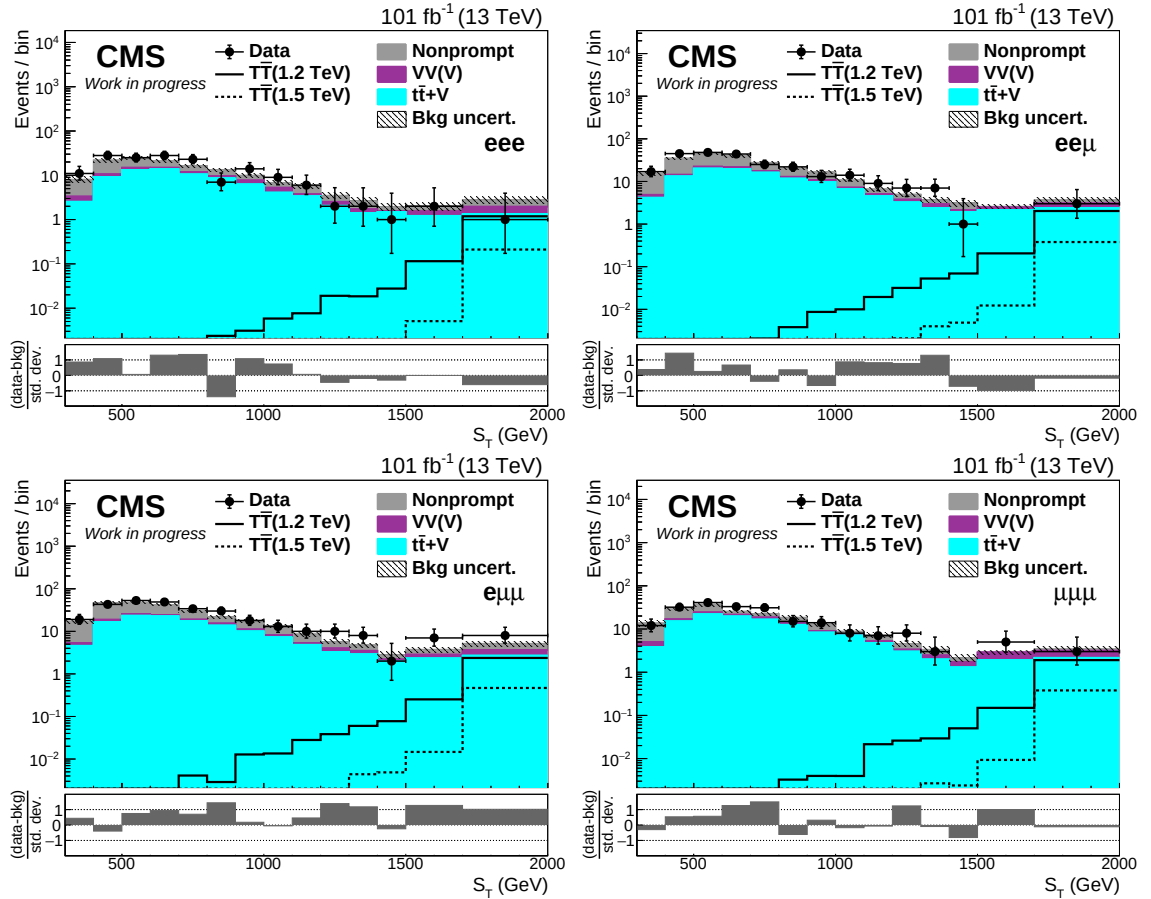


Figure 5.48: Distribution of S_T in the multi-lepton channel SR with final selections in each flavour category. Data from 2017 and 2018 is shown using black markers and $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines. Estimated background after fitting to Run 2 data (Section 5.12) is shown with filled histograms. Statistical and systematic uncertainties are shown by the hatched region.

Combining distributions of S_T in 2017 and 2018, the observed (expected) upper limits intersect with the theory cross sections at T mass of 1192 (1188) GeV for branching ratios in the singlet scenario and 1328 (1324) GeV for branching ratios in the doublet scenario as shown in Figure 5.49. At the extreme of $T \rightarrow tZ$ only, T mass is excluded below 1395 (1384) GeV in data (background). For B signals, the observed (expected) upper limits on B mass are 1180 (1165) GeV for in singlet scenario, 989 (986) GeV for in doublet scenario, and 1280 (1295) GeV in $B \rightarrow tW$ -only scenario, as shown in Figure 5.50. When combining distributions in 2017 and 2018, some systematic uncertainties are correlated across the years, as indicated in Table 5.31.

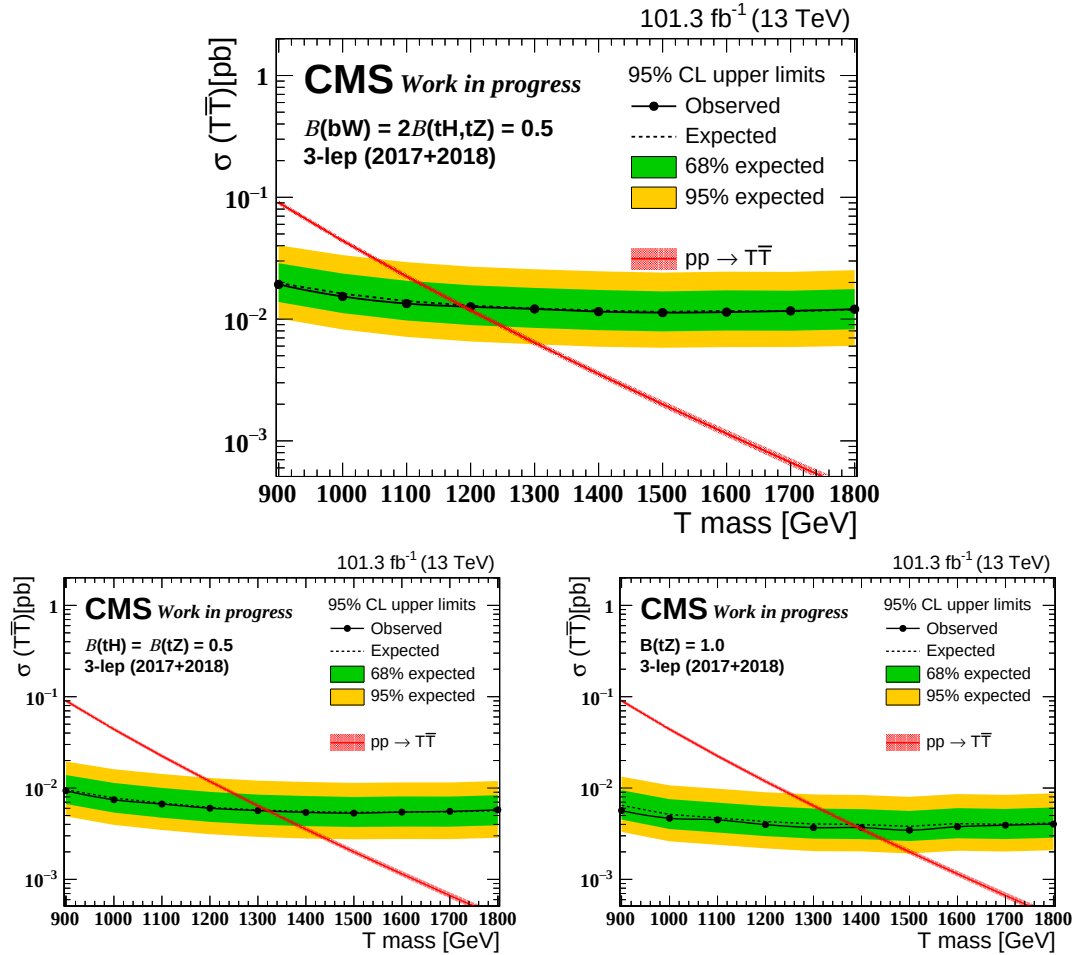


Figure 5.49: The 95% CL upper limits on $T\bar{T}$ production cross sections as a function of T masses for singlet (top), doublet (bottom left), and $T \rightarrow tZ$ only (bottom right) scenarios in 2017 and 2018 in 3+L channel.

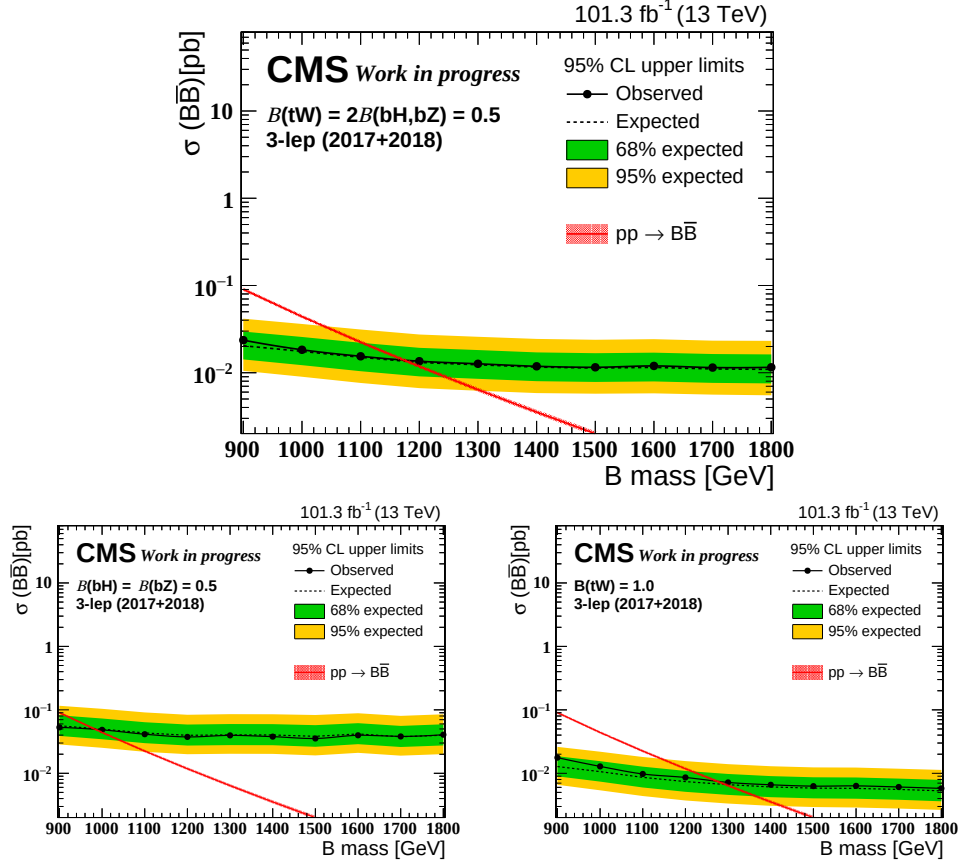


Figure 5.50: The 95% CL upper limits on $B\bar{B}$ production cross sections as a function of B masses for singlet (top), doublet (bottom left), and $B \rightarrow tW$ only (bottom right) scenarios in 2017 and 2018 in 3+L channel.

Combining distributions of S_T in 2017 and 2018 from this analysis with that in 2016 from a published analysis [21], the exclusion limits are observed (expected) at T mass of 1248 (1210) GeV for the singlet scenario and 1382 (1350) GeV for doublet scenario as shown in Figure 5.51. At the extreme of $T \rightarrow tZ$ only, T mass is excluded below 1458 (1401) GeV in data (background). For B signals, the observed (expected) upper limits on B mass are 1233 (1204) GeV for the singlet scenario, 1046 (1002) GeV for doublet scenario, and 1342 (1325) GeV for $B \rightarrow tW$ only, as shown in Figure 5.52. Correlations of systematic uncertainties between years are shown in Table 5.31. The improvements in T mass limits in this analysis from that of the 2016 analysis [21] are tabulated in Table 5.32.

Table 5.32: The exclusion limits on T mass measured in the multi-lepton channel for the singlet and doublet scenarios in the 2016 analysis [21], in 2017 and 2018, and in full Run 2.

Year	2016	2017+2018	full Run 2
Singlet	1.08 TeV	1.19 TeV	1.25 TeV
Doublet	1.22 TeV	1.33 TeV	1.38 TeV

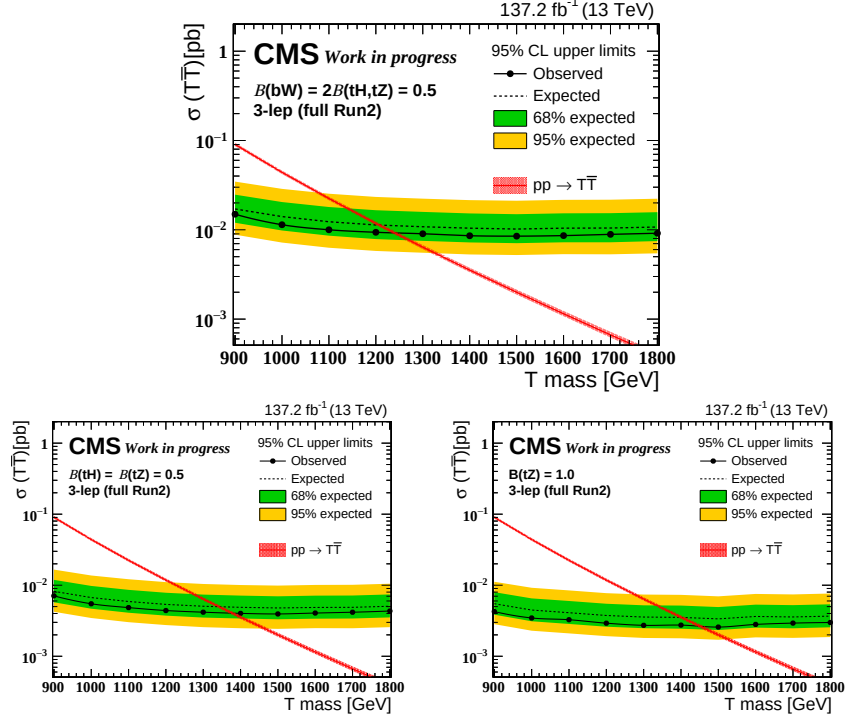


Figure 5.51: The 95% CL upper limits on $T\bar{T}$ production cross sections as a function of T masses for singlet (top), doublet (bottom left), and $T \rightarrow tZ$ only (bottom right) scenarios in the full Run 2 in 3+L channel.

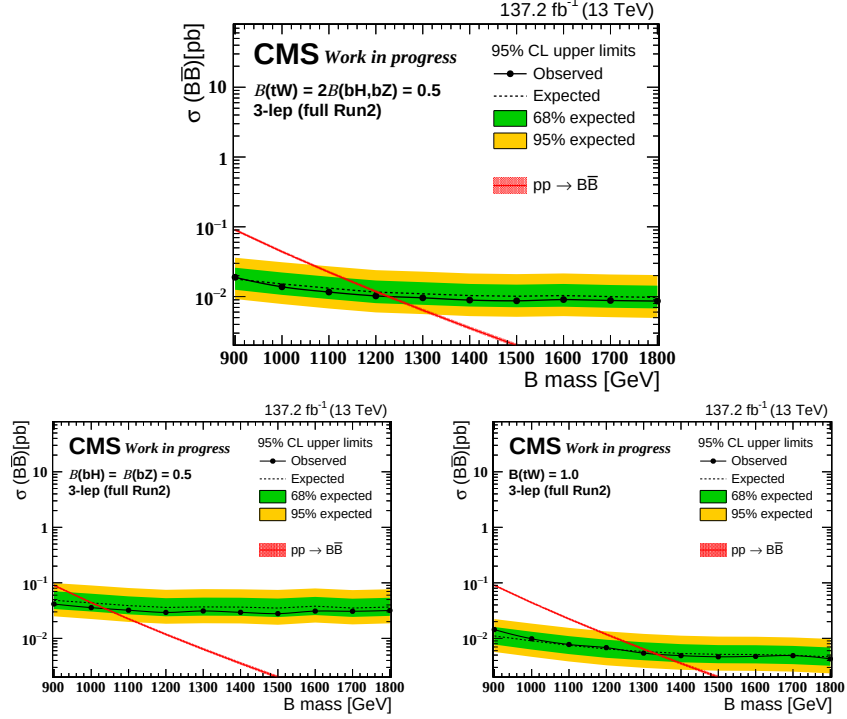


Figure 5.52: The 95% CL upper limits on $B\bar{B}$ production cross sections as a function of B masses for singlet (top), doublet (bottom left), and $B \rightarrow tW$ only (bottom right) scenarios in the full Run 2 in 3+L channel.

5.10.2 Limits in SS2L Channel

In the 2016 publication which is combined with this analysis [21], exclusion limits were obtained with the total event yields of $H_T^{\text{lep}} > 1200$ GeV in the three flavour categories. As an improvement in this analysis, shape-based likelihood fits are performed on the binned distributions of $H_T^{\text{lep}} > 400$ GeV. The bins at $H_T^{\text{lep}} < 200$ GeV are removed due to lack of statistics. In addition, the turn-ons of the H_T^{lep} distributions, in bins between 200 and 400 GeV, are found to be pulling fits off in statistical tests similar to those in Section 5.11.6. Thus, the signal sensitive region $H_T^{\text{lep}} > 400$ GeV is used.

Figure 5.53 shows the background distributions in H_T^{lep} , for 2017 and 2018, after a likelihood fit to all leptonic channels and full Run 2 data, which will be explained in Section 5.12. Data and $T\bar{T}$ signal at VLQ mass of 1.0 (1.5) TeV (before the fit) are overlaid in the same figure.

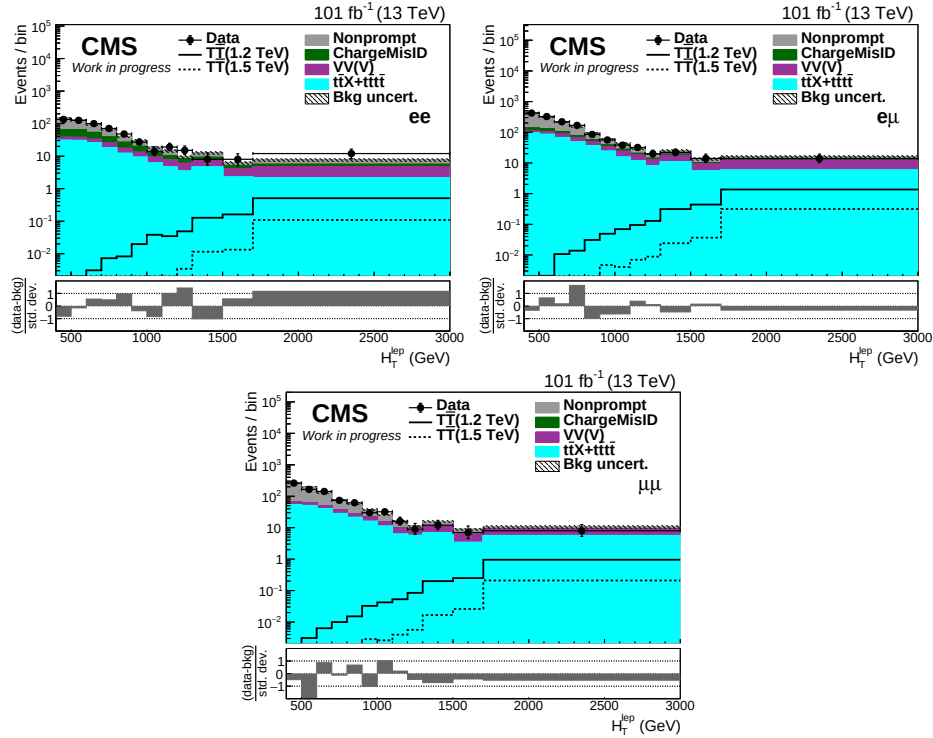


Figure 5.53: Distribution of H_T^{lep} in the SS2L channel SR in each flavour category. Data from 2017 and 2018 is shown using black markers, $T\bar{T}$ signals using solid lines, and background using filled histograms. Statistical and systematic uncertainties in the background prediction before performing the fit to data are shown by the hatched region.

In the following, when limits of production cross sections do not intersect with the theory cross sections (red line), we conclude that the mass exclusion limit is found at < 900 GeV as previous searches have excluded VLQ masses below 0.9 TeV.

Combining distributions of H_T^{lep} in 2017 and 2018, the observed (expected) exclusion limits on T mass are < 900 (< 900) GeV in singlet scenario and 952 (960) GeV in doublet scenario, as shown in Figure 5.54. For B signals, the observed (expected) limits on B mass are 976 (986) GeV in singlet scenario and < 900 (< 900) GeV in doublet scenario, as shown in Figure 5.55. Correlations between years in systematic uncertainties are listed in Table 5.31.

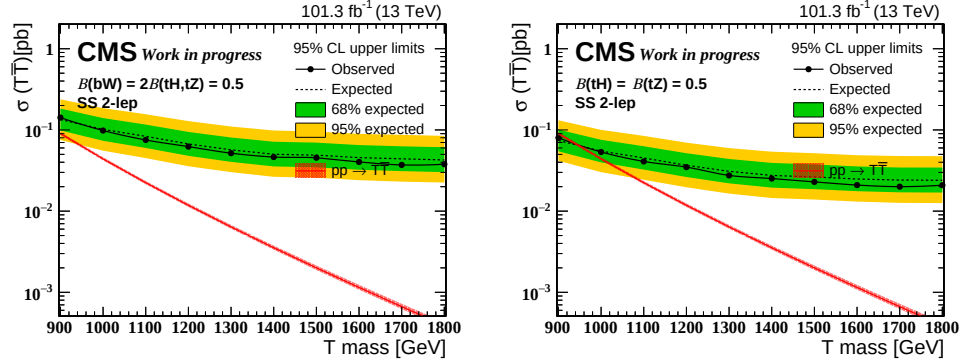


Figure 5.54: The 95% CL upper limits on $T\bar{T}$ production cross sections as a function of T masses for singlet (left) and doublet (right) scenarios in 2017 and 2018 in SS2L channel.

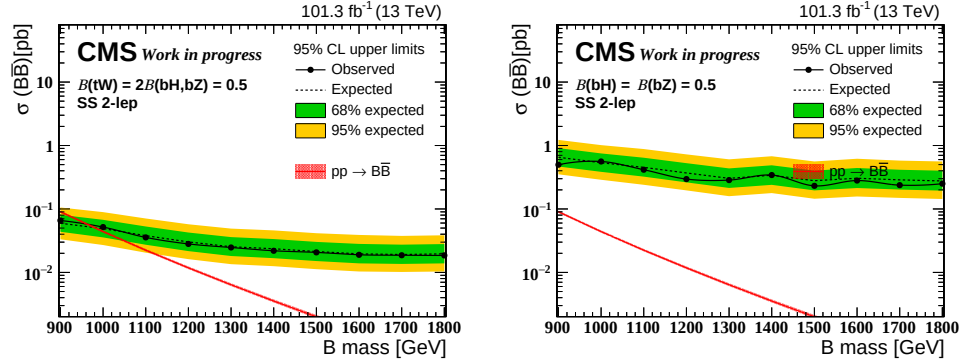


Figure 5.55: The 95% CL upper limits on $B\bar{B}$ production cross sections as a function of B masses for singlet (left) and doublet (right) scenarios in 2017 and 2018 in SS2L channel.

Combining distributions of H_T^{lep} in 2017 and 2018 in this analysis with $H_T^{\text{lep}} > 1200$ GeV event yields in 2016 [21], the observed (expected) exclusion limits on T mass are < 900 (< 900) GeV in singlet scenario and 1020 (998) GeV in doublet scenario, as shown in Figure 5.56. For B signals, the observed (expected) limits are found at B mass of 1029 (1024) GeV in singlet scenario and < 900 (< 900) GeV in doublet scenario, as shown in Figure 5.57. Correlations of systematic uncertainties between years can be found in Table 5.31. The improvements in T mass limits in this analysis from that of the 2016 analysis [21] are tabulated in Table 5.33.

Table 5.33: The exclusion limits on T mass measured in the SS2L channel for the singlet and doublet scenarios in the 2016 analysis [21], in 2017 and 2018, and in full Run 2.

Year	2016	2017+2018	full Run 2
Singlet	< 0.8 TeV	< 0.9 TeV	< 0.9 TeV
Doublet	0.91 TeV	0.95 TeV	1.02 TeV

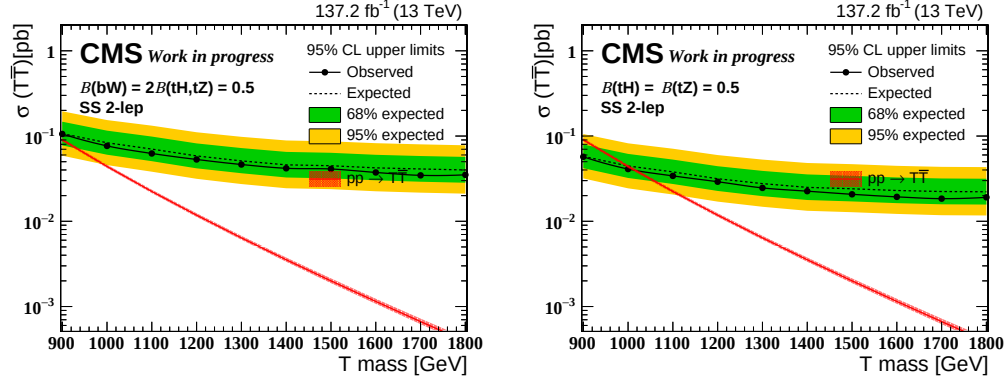


Figure 5.56: The 95% CL upper limits on $T\bar{T}$ production cross sections as a function of T masses for singlet (left) and doublet (right) scenarios in the full Run 2 in SS2L channel.

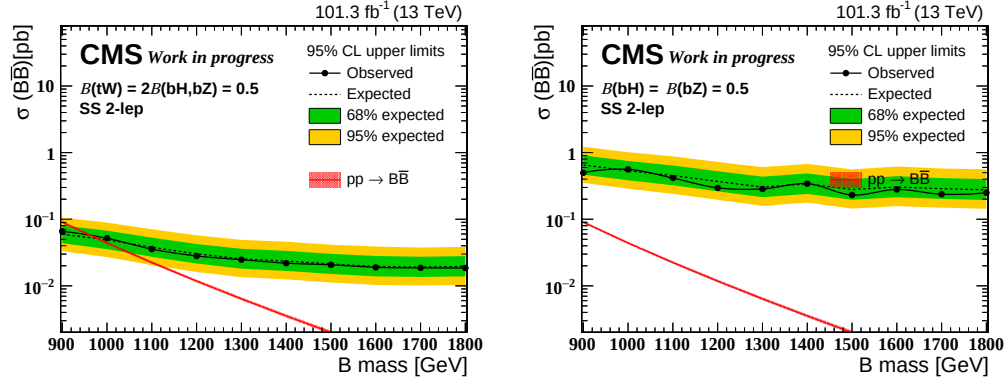


Figure 5.57: The 95% CL upper limits on $B\bar{B}$ production cross sections as a function of B masses for singlet (left) and doublet (right) scenarios in the full Run 2 in SS2L channel.

Analyzing 2017 and 2018 data, exclusion limits on VLQ masses have been extended. Respectively in the multi-lepton and SS2L channel, the updated exclusion limits are about > 100 GeV stronger in this analysis, compared to that measured in 2016 data, in most branching ratio scenarios.

5.11 Statistical Treatment and Tests

In Section 5.10, the upper limits on signal production cross sections are calculated by likelihood fits on the shape of distributions of a discriminant in SR with final selections using the Higgs Combine program [137]. To check the health of the fits, a number of tests are performed according to recommendations of the CMS Beyond-two-Generations (B2G) group.

For all the tests in this section, statistical uncertainties are treated by the Barlow-Beeston lite method with “autoMCstats” method in Higgs Combine [138, 139]. When yield is below 1.0, Poisson PDF is used instead. Systematic uncertainties in Table 5.24 are imported to Higgs Combine as nuisance parameters. Normalization-only uncertainties are configured as log-normal priors, and shape uncertainties are treated with templates morphing with Gaussian priors. Signal cross section is configured as a flat prior.

The tests displayed in this dissertation are:

1. Nuisance Pulls: compare the value and uncertainty of each nuisance parameter before and after the maximum likelihood fit.
2. Impacts: the change in fitted signal strength by varying each nuisance parameter one by one, computed by the “impacts” method in the Higgs Combine
3. Bias Test: check whether fits are biased by comparing the fitted signal strength r against zero for the null hypothesis (no signal) in Asimov pseudo-data generated using background with no artificially injected signal. The comparison is visualized by plotting the signal strength pull, which is the fitted r and divided by the uncertainty.
4. Goodness-of-fit: compare the test statistic obtained from data against that from Asimov pseudo-data generated using background-only model in a CR.

5.11.1 Statistical Tests in 3+L Channel

In multi-lepton channel, the shape of $S_T > 300$ GeV distributions in SR with final selections (Figure 5.48) are fitted using the Higgs Combine. Distributions of S_T in 2016 are directly taken from the published 2016 analysis [21, 131].

5.11.2 3+L Channel: Nuisance pulls

Values (θ) and uncertainties (σ) of nuisance parameters before and after maximum likelihood fit on distributions of S_T in the multi-lepton channel SR are compared. Denoting the scenario before the fit with a subscript I , nuisance pull is defined by $\frac{1}{\sigma_I} \left(\frac{\theta - \theta_I}{\theta_I} \right)$, where σ_I is set to 1.0 in this analysis [137]. It is equivalent to the changes in nuisance parameter's value relative to its initial value. The constraint on a nuisance pull is defined by σ/σ_I . Figure 5.58 shows the pulls and their corresponding constraints of the first 25 nuisance parameters in alphabetical order. Nuisance parameters related to fake rates are generally pulled more significantly.

5.11.3 3+L Channel: Impacts

The impact of each nuisance parameter is found for fitting distributions of S_T in the multi-lepton channel SR. Nuisance parameters of the top 30 greatest impacts are shown in Figures 5.59. Nuisance parameters such as renormalization scale, PDF, and jet energy correction have relatively high impacts overall. For statistical nuisances, bins at the tail of the distribution tend to have higher impacts, as expected from distributions of simulated signals.

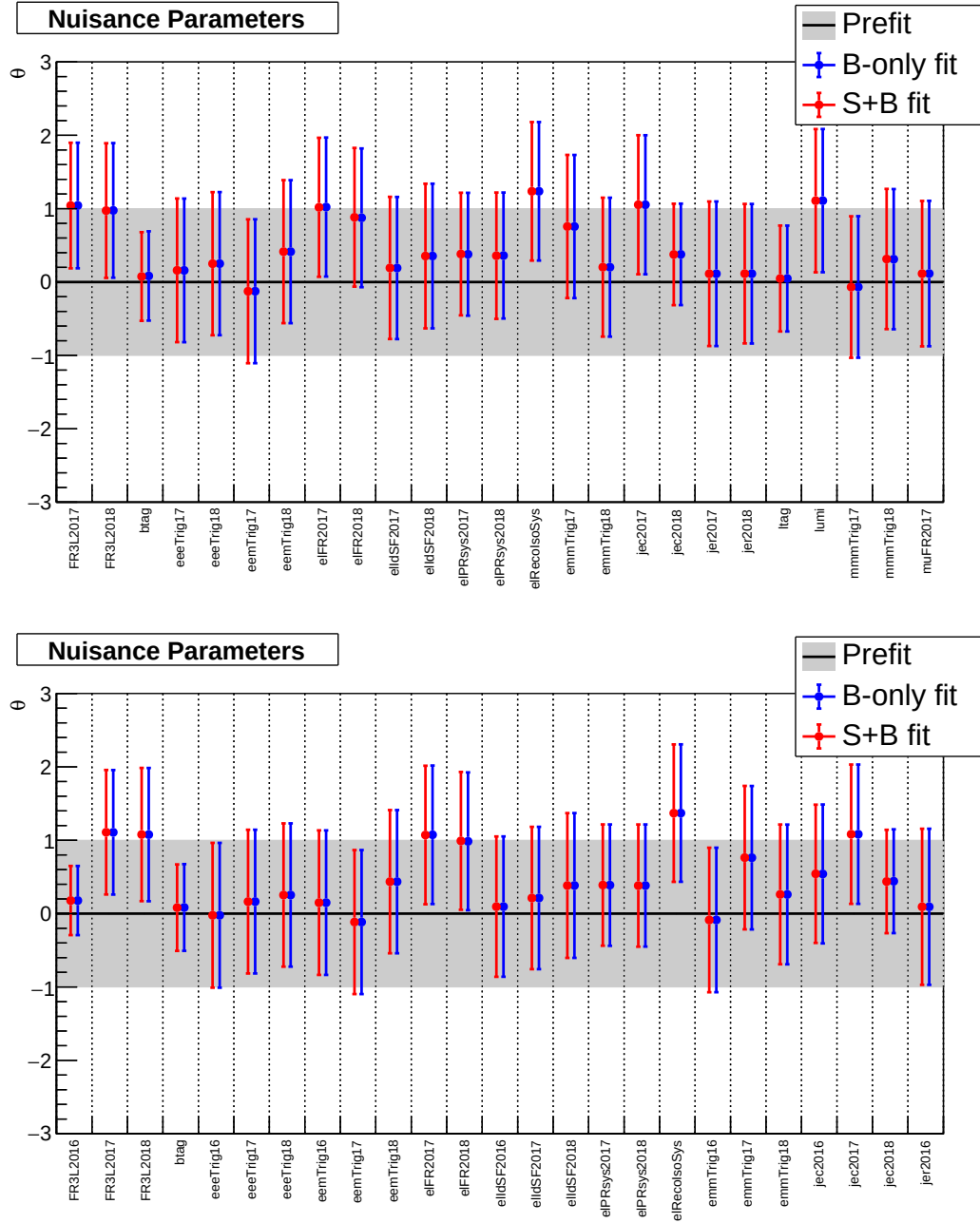


Figure 5.58: Normalized (unitless) nuisance pulls and constraints of the maximum likelihood fits in the multi-lepton channel SR in 2017 and 2018 (top) and the full Run 2 (bottom). For each nuisance parameter in x -axis, the marker shows the difference in its value before and after the fit, normalized by its value before the fit, and the error bar shows the uncertainty after the fit. The blue markers indicate results from the fit with background-only model, fixing signal strength to zero, and the red markers denote the result for a fit with background+signal model with a floating r . For prefit, the difference is always zero, shown with a solid black line, and the normalized uncertainty range of 1 to -1 is shown with shaded area.

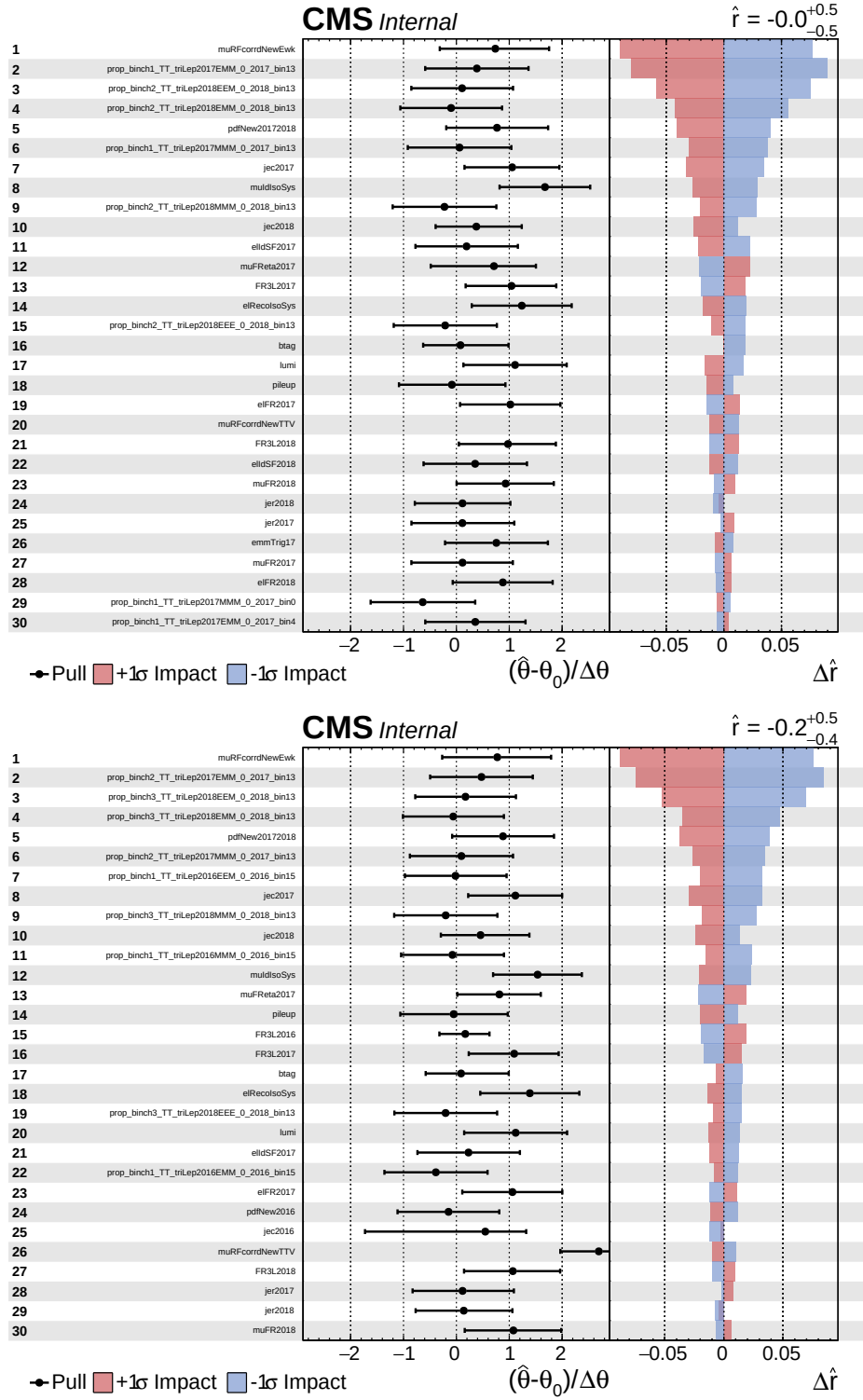


Figure 5.59: The top 30 nuisance parameter impacts from maximum likelihood fits in the multi-lepton channel SR in 2017 and 2018 (top) and the full Run 2 (bottom). The unitless fitted signal strength is shown on the top right corner of each plot.

5.11.4 3+L Channel: Bias Test

Figure 5.60 shows the distributions of signal strength pulls $\left(\frac{0-r_{\text{fitted}}}{\text{unc.}}\right)$ at T or B mass of 1000 GeV in 2017+2018 and the full Run 2. Signal strength pulls show a slight bias near -0.2 which is within accepted range.

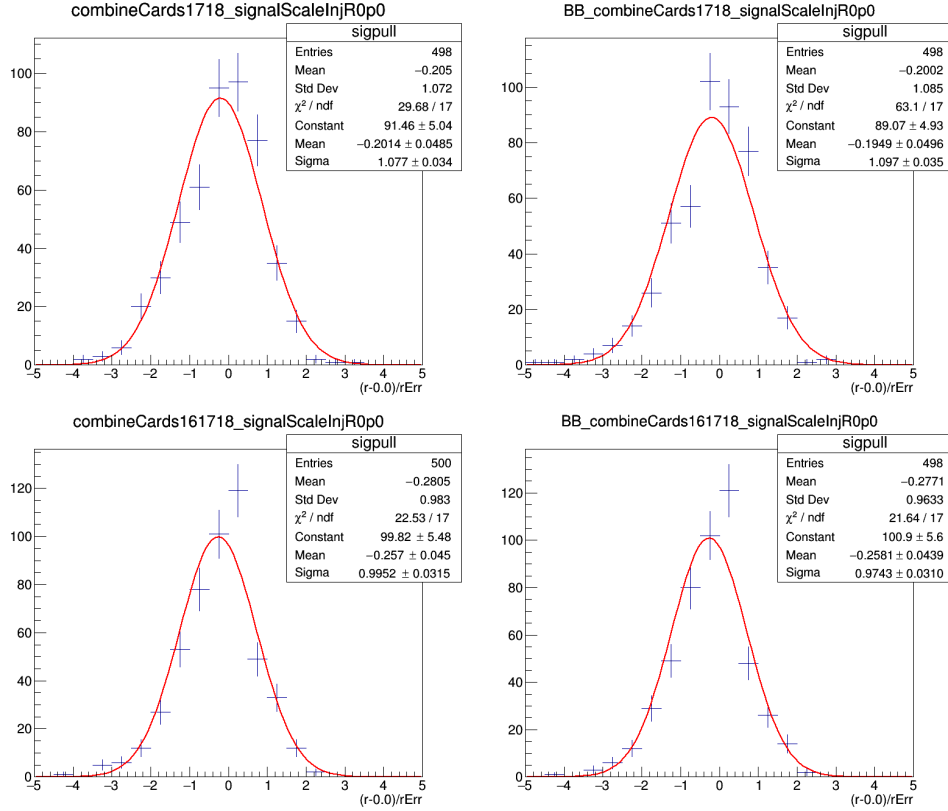


Figure 5.60: Distributions of signal strength pulls when injecting $r = 0$ for mass points 1000 GeV in $T\bar{T}$ (left) and $B\bar{B}$ (right) analysis in the multi-lepton channel in 2017+2018 combined (top) and 2016+2017+2018 combined (bottom).

5.11.5 3+L Channel: Goodness of Fit

This test compares test statistic obtained from data against that from Asimov pseudo-data. This test was performed in the multi-lepton channel CR1 when the analysis was blinded. Figures 5.62–5.63 show the distributions of S_T in CR1. The result for the maximum likelihood fit in 2017 and 2018 is shown in Figure 5.61. It is observed that data lies within the bulk of the distribution of toy data (pseudo-data), inferring a proper estimation of background. Since distributions of S_T in CR1 in 2016 are not available, goodness of fit test for the full Run 2 is skipped.

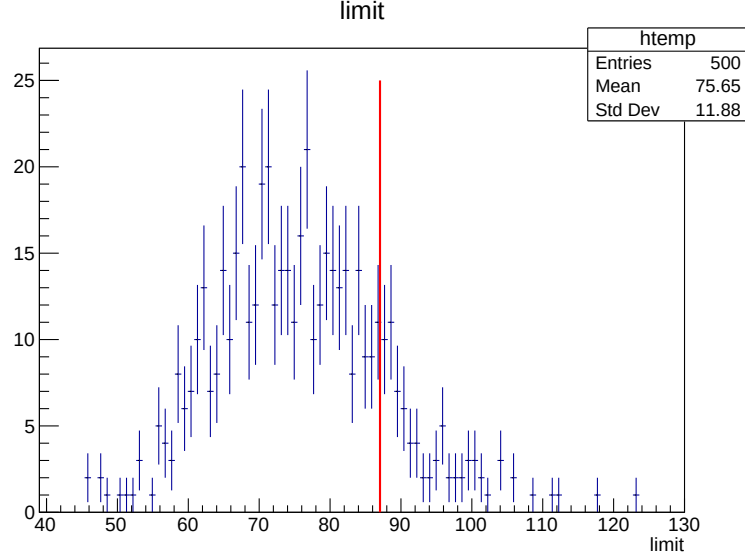


Figure 5.61: Distributions of the saturated χ^2 metric in toys (markers) generated with background-only model and data (red line) in maximum likelihood fit in the multi-lepton CR1 in 2017 and 2018.

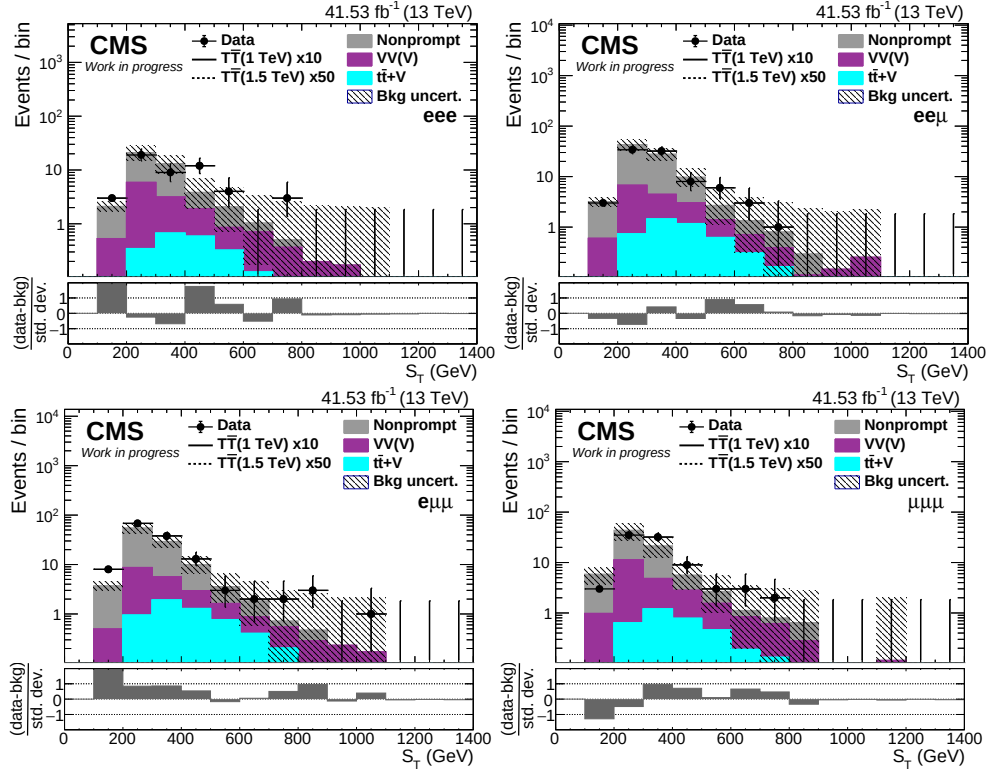


Figure 5.62: Distribution of S_T in the multi-lepton channel CR1 in each flavour category. Data from 2017 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV scaled by a factor of 10 (50) using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with fake rates measured in CR2. Statistical and systematic uncertainties are shown by the hatched region.

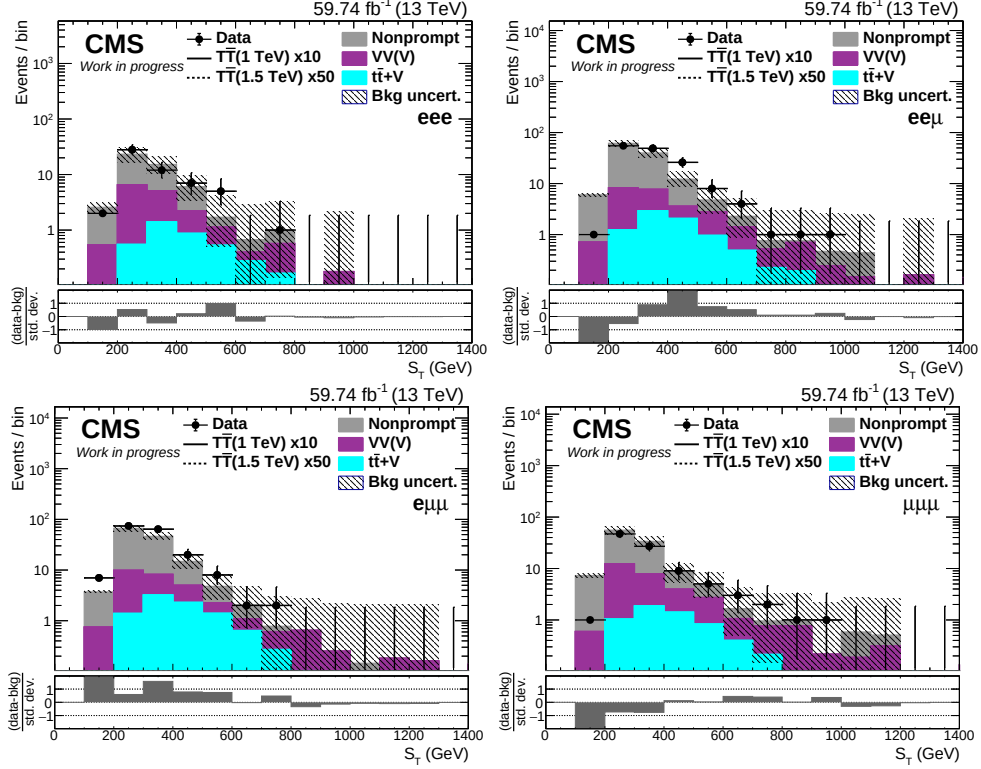


Figure 5.63: Distribution of S_T in the multi-lepton channel CR1 in each flavour category. Data from 2018 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV scaled by a factor of 10 (50) using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with fake rates measured in CR2. Statistical and systematic uncertainties are shown by the hatched region.

5.11.6 Statistical Tests in SS2L Channel

In SS2L channel, distributions of $H_T^{\text{lep}} > 400$ GeV in SR with final selections (Figure 5.53) are fitted. H_T^{lep} is the scalar sum of all AK4 jets' p_T and the leptons' p_T . The 2016 analysis used for full Run 2 result conducted a counting experiments for the SS2L channel in which event yields are fitted [21, 131]. Yields are directly taken from the 2016 analysis to combine with the shapes of H_T^{lep} in 2017 and 2018.

5.11.7 SS2L Channel: Nuisance pulls

Values and uncertainties of nuisance parameters before and after maximum likelihood fit on distributions of $H_T^{\text{lep}} > 400$ GeV in the SS2L channel SR are compared. Figure 5.64 shows the pulls and their corresponding uncertainties on the first 25 nuisance parameters in alphabetical order. Nuisance parameters related to nonprompt rates and PDF are pulled more significantly.

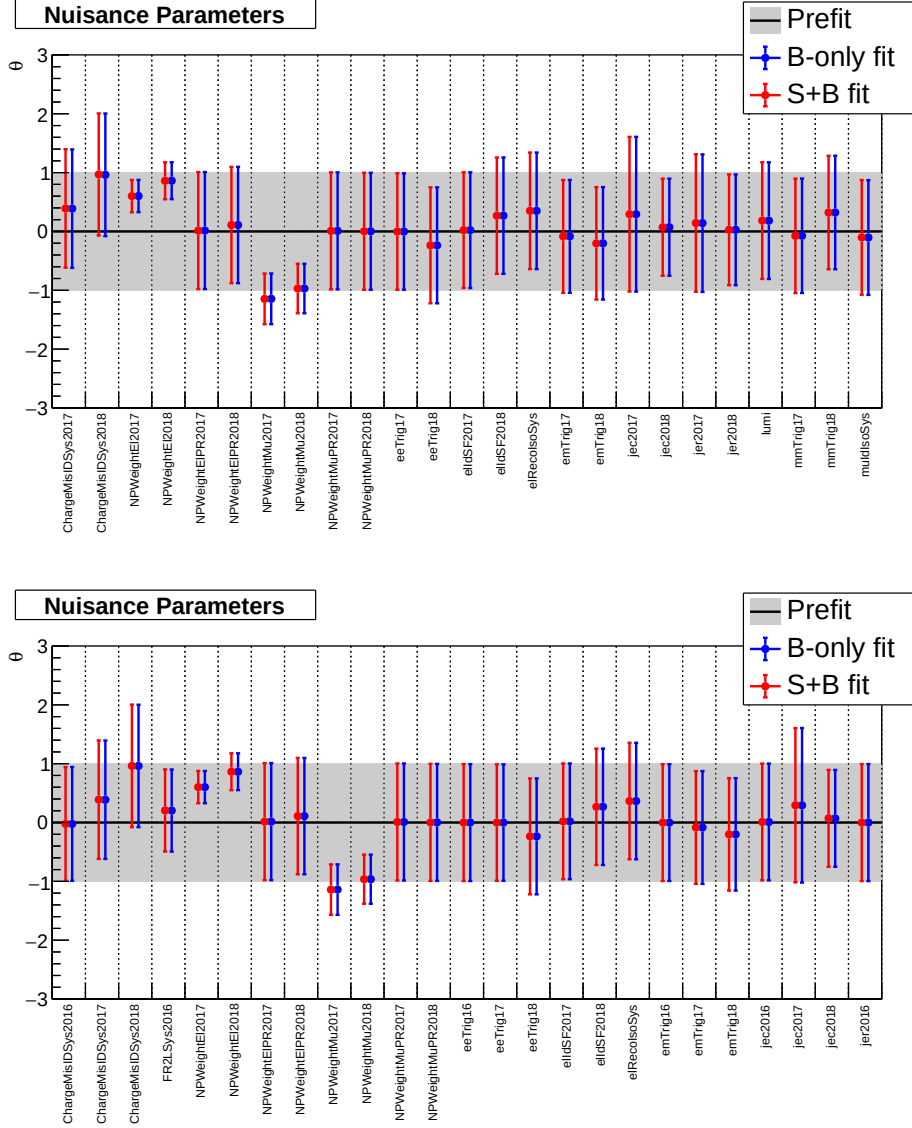


Figure 5.64: Normalized (unitless) nuisance pulls and constraints of the maximum likelihood fits in the SS2L channel SR in 2017 and 2018 (top) and the full Run 2 (bottom).

5.11.8 SS2L Channel: Impacts

The impact of each nuisance parameter is found for maximum likelihood fit on distributions $H_T^{\text{lep}} > 400$ GeV in the SS2L channel SR. Nuisance parameters of the top 30 greatest impacts are shown in Figures 5.65. Similar to the multi-lepton channel, nuisance parameters such as renormalization scale, PDF, and jet energy correction have relatively high impacts. For statistical nuisances, bins at the tail of the distribution tend to have higher impacts, as expected from distributions of signals.

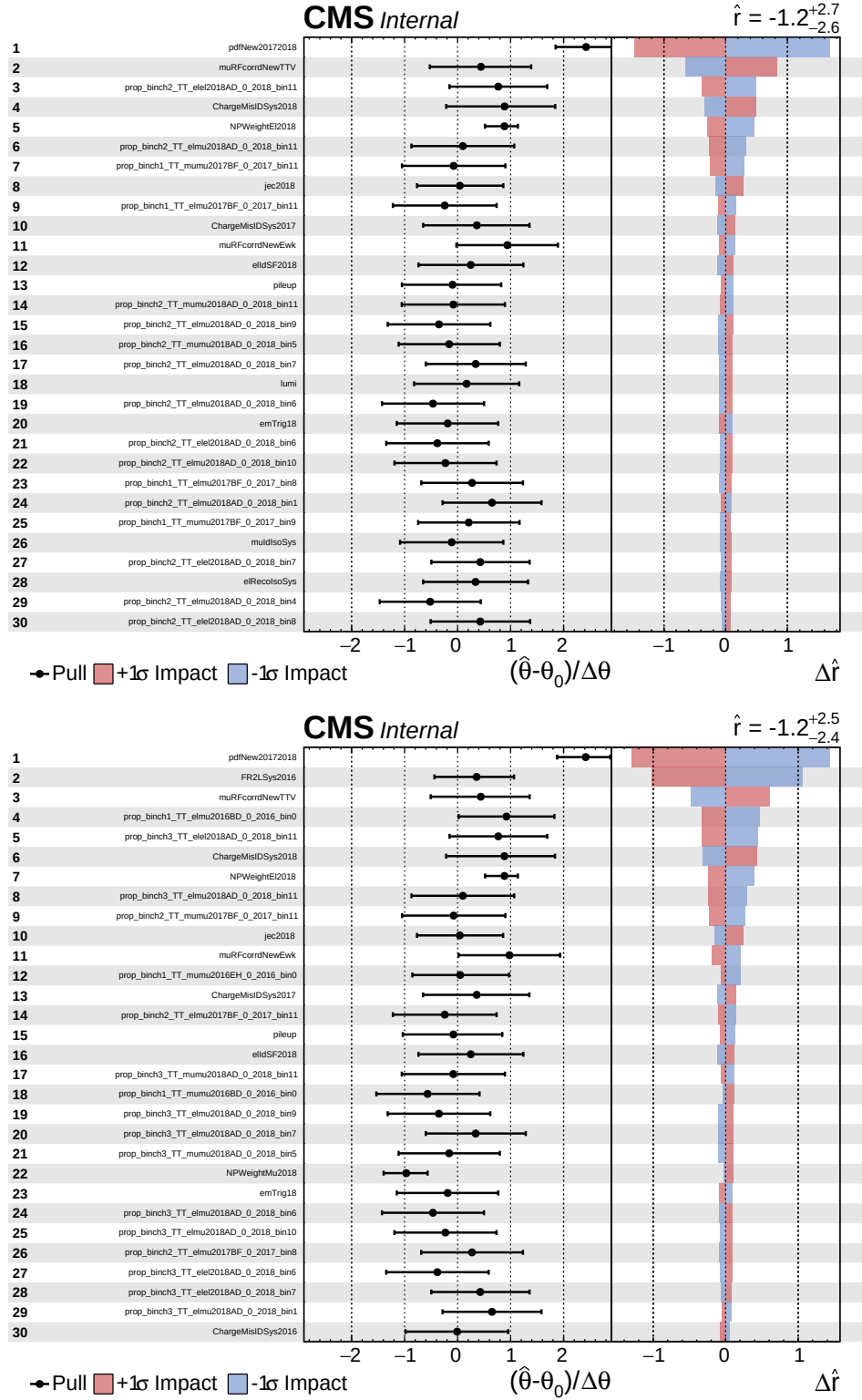


Figure 5.65: The top 30 nuisance parameter impacts from maximum likelihood fits in the SS2L channel SR in 2017 and 2018 (top) and the full Run 2 (bottom).

5.11.9 SS2L Channel: Bias Test

Figure 5.66 shows the distributions of signal strength pulls $\left(\frac{0-r_{\text{fitted}}}{\text{unc.}}\right)$ at T or B mass of 1000 GeV in 2017+2018 and the full Run 2. The signal strength pulls peaks very close to zero in maximum likelihood fit of 2017 and 2018 but show a small acceptable bias near -0.15 in the full Run 2.

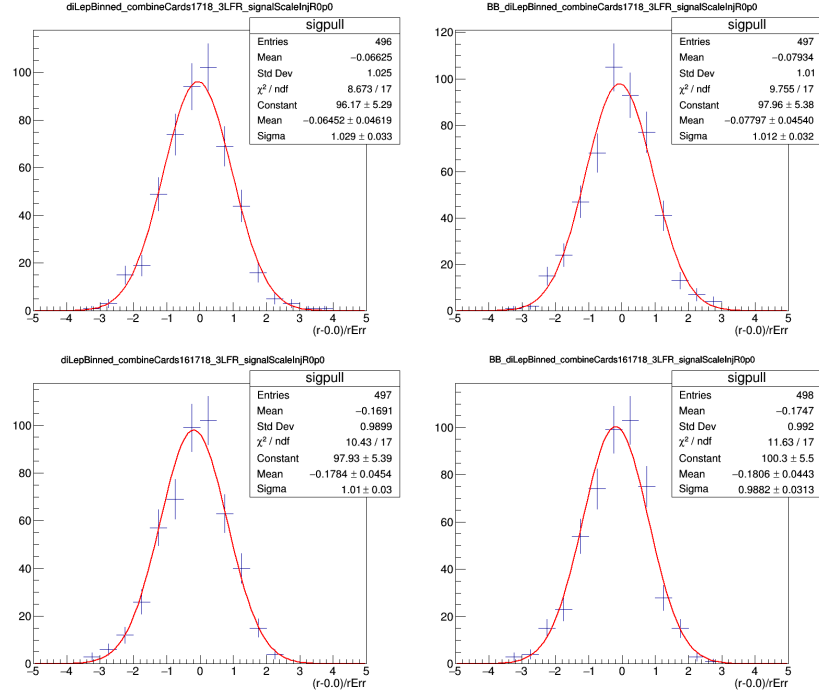


Figure 5.66: Distributions of signal strength pulls when injecting $r = 0$ for mass points 1000 GeV in $T\bar{T}$ (left) and $B\bar{B}$ (right) analysis in the SS2L channel in 2017+2018 combined (top) and 2016+2017+2018 combined (bottom).

5.11.10 SS2L Channel: Goodness of Fit

This test compares test statistic obtained from data against that from pseudo-data. To eliminate the effect of the turn-on in distributions of H_T^{lep} , this test was performed in the low H_T^{lep} CR with final selections of $H_T^{\text{lep}} > 400$ GeV, i.e., $400 < H_T^{\text{lep}} < 1000$ GeV as shown in Figures 5.68. The results for the maximum likelihood fits in 2017, 2018, and 2017+2018 are shown in Figure 5.67. Since corresponding event yields in 2016 are not available, goodness of fit test for the full Run 2 is skipped. It is observed that data lies outside or near the edge to the right of the bulk of the distribution of toy data (pseudo-data) in 2018, and thus 2017+2018, inferring underestimations of background. The fit in 2017 is good.

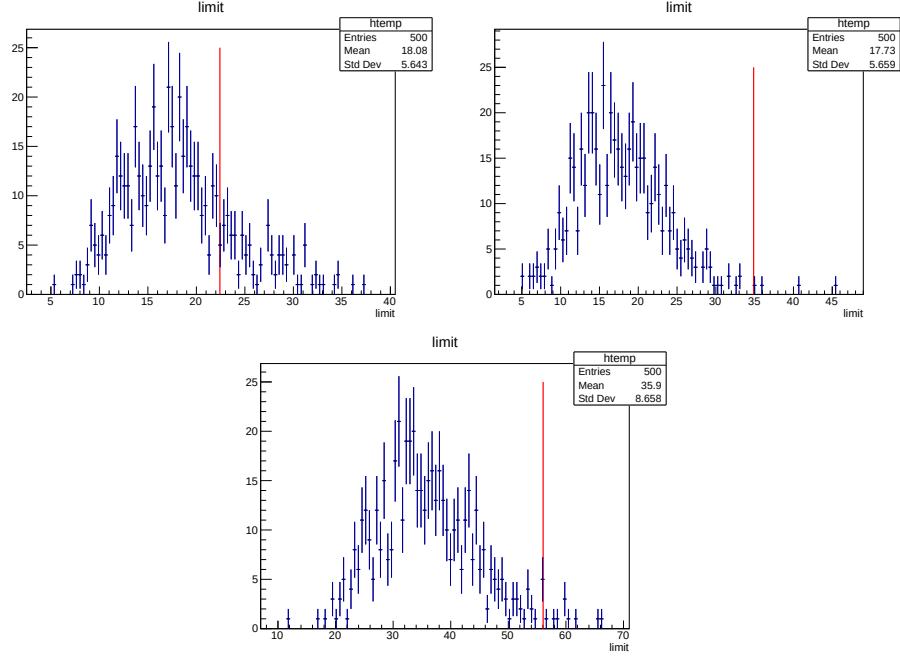


Figure 5.67: Distributions of the saturated χ^2 metric in toys (markers) generated with background-only model and data (red line) in maximum likelihood fit in the SS2L channel control region of $400 < H_T^{\text{lep}} < 1000$ GeV in 2017, 2018, and 2017+2018.

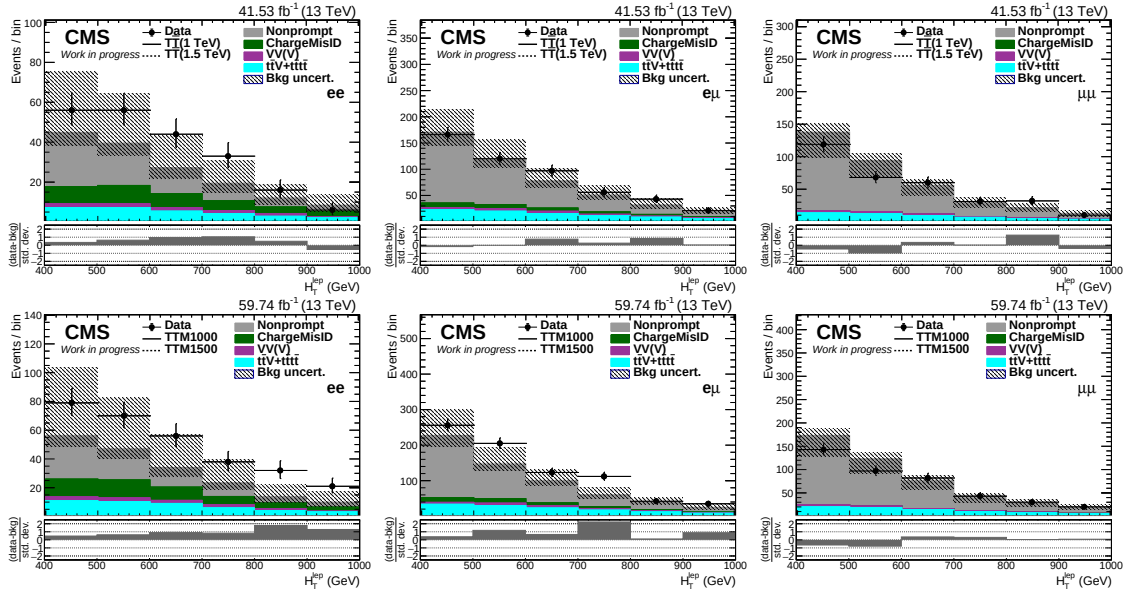


Figure 5.68: Distribution of $400 < H_T^{\text{lep}} < 1000$ GeV in the SS2L channel low H_T^{lep} CR in each flavour category. Data from 2017 (top) and 2018 (bottom) is shown using black markers, $TT\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Nonprompt background is estimated with fake rates measured in CR2. Statistical and systematic uncertainties are shown by the hatched region.

Investigating into the disagreement in 2018, maximum likelihood fits are done separately for each flavour categories and the corresponding goodness of fits are shown in Figure 5.69. These plots show that goodness of fits agree in ee and $\mu\mu$ categories and the background underestimation is dominantly due to the $e\mu$ category in 2018. Comparing data to the backgrounds before and after maximum likelihood fit (pre- and post-fit) in 2018 as shown in Figure 5.70, which are produced by running “FitDiagnostics” with option “-plots” in Combine package, the third bin in the distribution of H_T^{lep} in $e\mu$ category in 2018 is not well-fitted. This could be the cause of the data excess observed in goodness of fit.

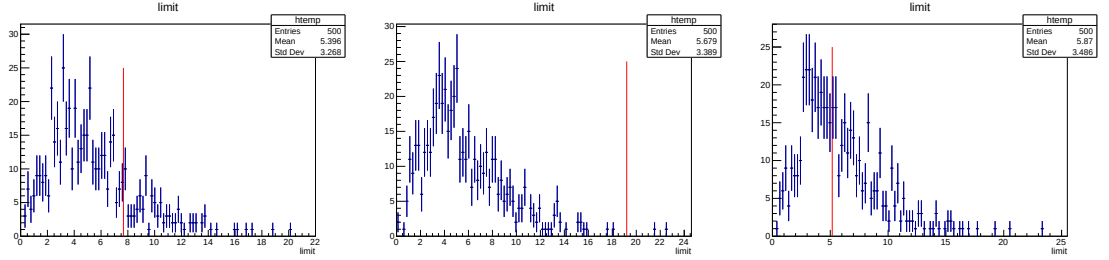


Figure 5.69: Distributions of the saturated χ^2 metric in toys (markers) generated with background-only model and data (red line) in maximum likelihood fit in the SS2L channel control region of $400 < H_T^{\text{lep}} < 1000$ GeV in 2018 in ee (left), $e\mu$ (mid), and $\mu\mu$ (right) categories independently.

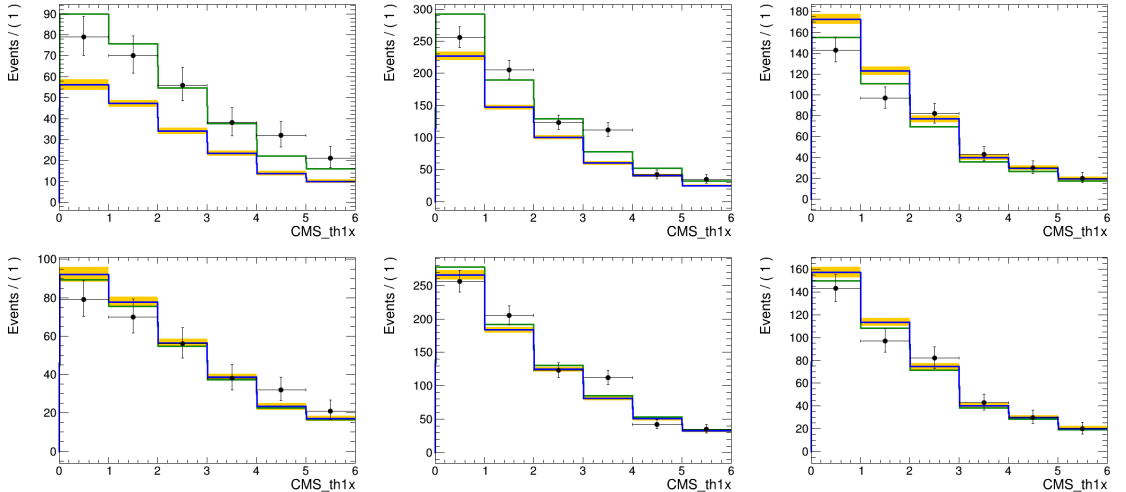


Figure 5.70: Distributions of H_T^{lep} with pre-fit (top) and post-fit (bottom) backgrounds in the SS2L channel control region of $400 < H_T^{\text{lep}} < 1000$ GeV in ee (left), $e\mu$ (middle), and $\mu\mu$ (right) categories. Data from 2018 is shown using black markers, background using blue solid lines, and upward error using green solid lines. Statistical uncertainties are shown by the yellow filled region.

The third bin in the distribution corresponds to $700 < H_T^{\text{lep}} < 800$ GeV. To check if this bin is the source of the observed data excess in goodness of fit, a separate maximum likelihood fit is performed in $400 < H_T^{\text{lep}} < 700$ GeV and the result of the goodness of fit test is shown in Figure 5.71. Excluding the $700 < H_T^{\text{lep}} < 800$ GeV bin in which data is above fitted background, the goodness of fit test found data within the bulk of toys.

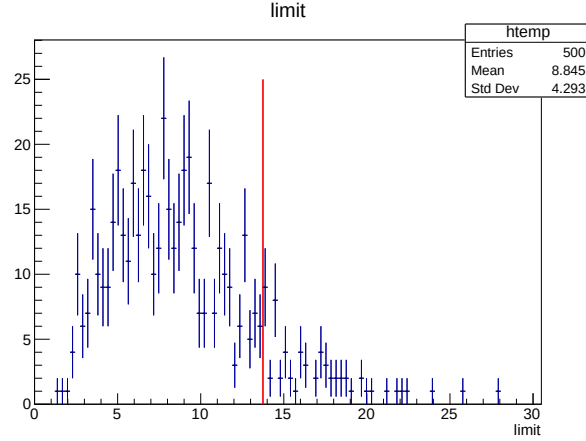


Figure 5.71: Distributions of the saturated χ^2 metric in toys (markers) generated with background-only model and data (red line) in maximum likelihood fit in the SS2L channel control region of $400 < H_T^{\text{lep}} < 700$ GeV in 2018.

As an alternative validation, the goodness of fit is tested independently in the SS2L channel nJet = 3 CR in 2018, with the full H_T^{lep} distributions as shown in Figure 5.73. The results for all flavour categories and the $e\mu$ category alone are shown in Figure 5.72. Data excess is not observed in nJet = 3 CR.

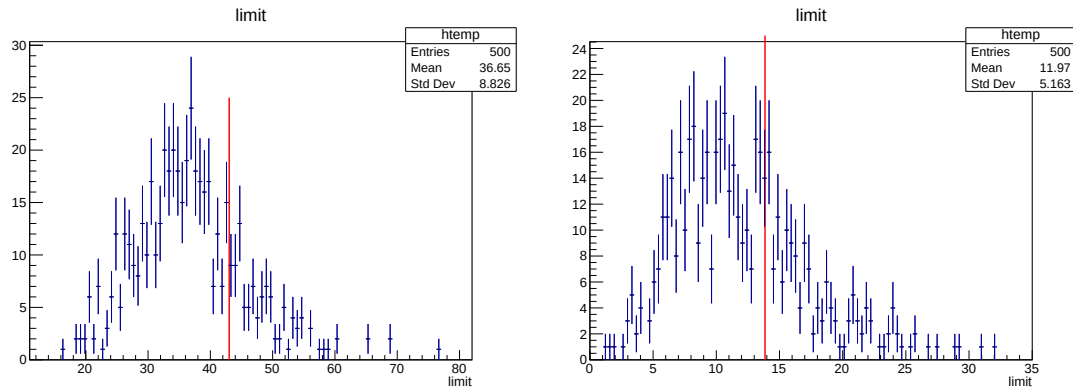


Figure 5.72: Distributions of the saturated χ^2 metric in toys (black markers) generated with background-only model and data (red line) in analysis of 2018 in combined (left) and $e\mu$ (right) categories in nJet = 3 control region.

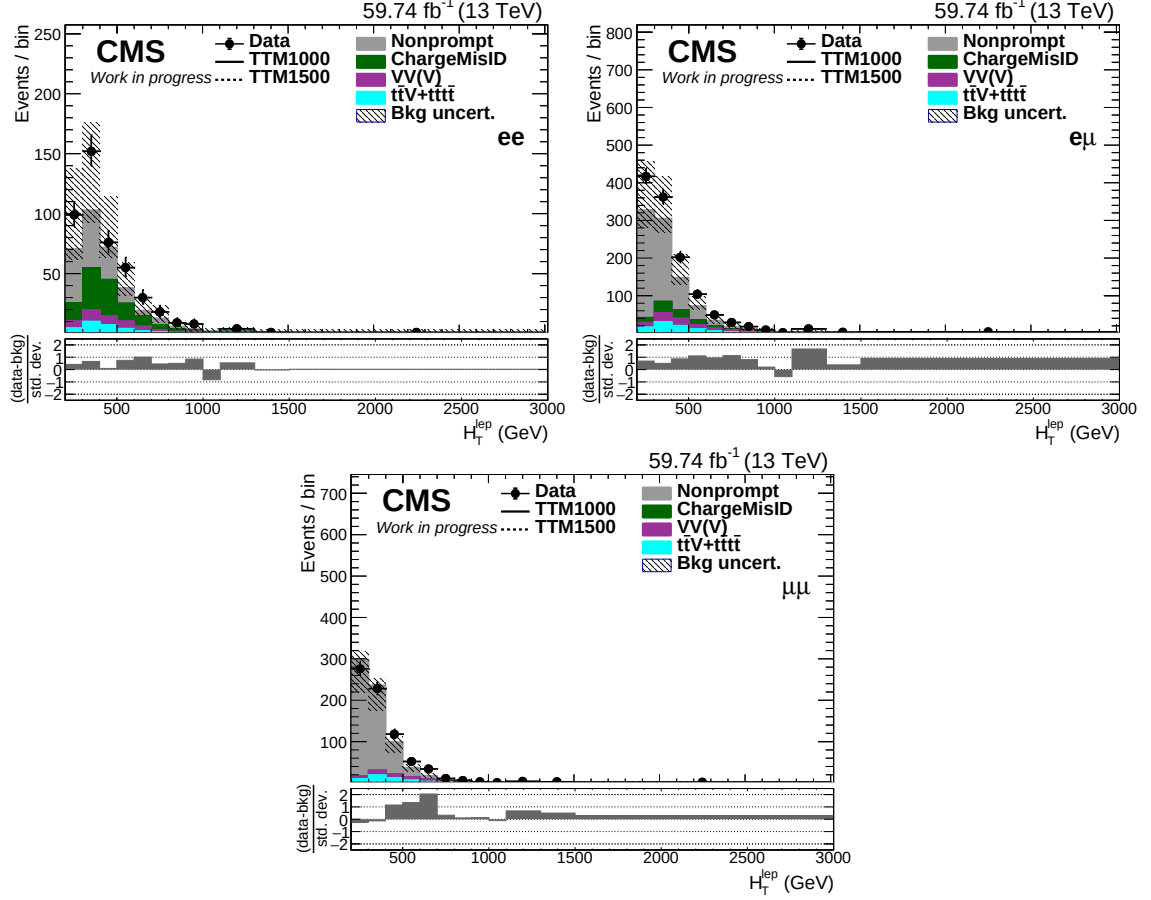


Figure 5.73: Distribution of H_T^{lep} in the SS2L channel $n\text{Jet} = 3$ CR in each flavour category. Data from 2018 is shown using black markers, $T\bar{T}$ with VLQ mass of 1.0 (1.5) TeV using solid (dashed) lines, and background using filled histograms. Statistical and systematic uncertainties are shown by the hatched region.

It is then interesting to study the correlation between the data excess and AK4 jet multiplicity. Focusing on $700 < H_T^{\text{lep}} < 1000$ GeV where the data excess appear, the goodness of fit is tested on distributions of $700 < H_T^{\text{lep}} < 1000$ GeV in the SS2L channel in 2018 requiring different jet multiplicities, from exactly 4 up to exactly 11 jets. Results are shown in Figure 5.74. Data excess is observed only in $n\text{Jet} = 6$ and $n\text{Jet} = 11$. For $n\text{Jet} = 11$, since it is the maximum jet multiplicity in most samples, the “excess” is likely caused by lack of statistics. The $n\text{Jet} = 6$ scenario is more interesting and might be worth investigating into in other analyses.

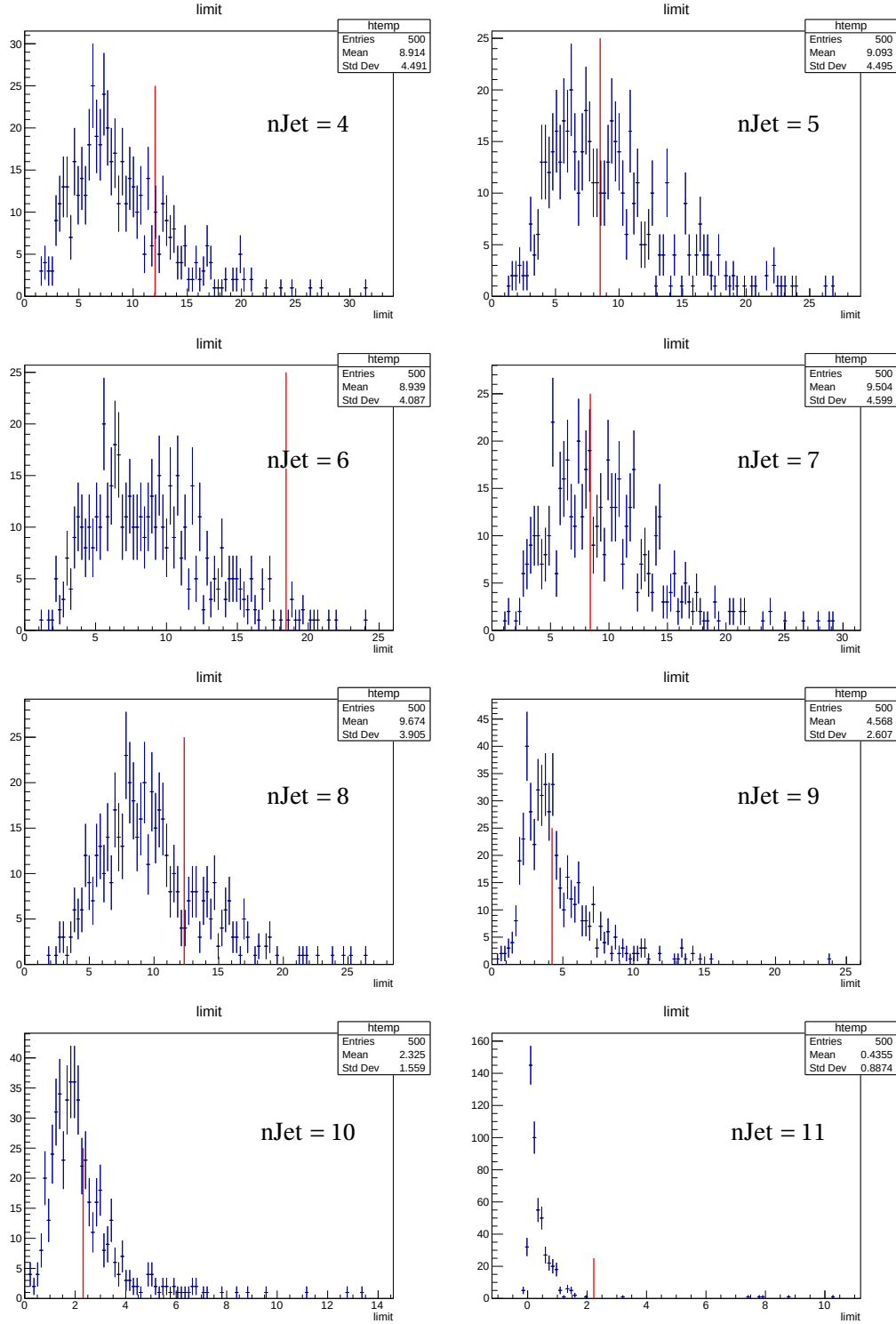


Figure 5.74: Distributions of the saturated χ^2 metric in toys (markers) generated with background-only model and data (red line) in maximum likelihood fit in the SS2L channel control region of $700 < H_T^{\text{lep}} < 1000$ GeV in 2018 with $n\text{Jet} = 4$ to 11 , from left to right and top to bottom.

5.12 Combined Result: 1 + 2 + 3 Leptons in full Run 2

To cover all possible final states with leptons, a search for $T\bar{T}$ and $B\bar{B}$ pairs in final states with exactly one lepton is combined with the SS2L and multi-lepton channels. Final state with opposite-sign dilepton are less sensitive to VLQ searches because of the dominating SM background of Drell-Yan processes, and thus is not considered.

Analyzed by our collaborator Prof. Julie Hogan from Bethel University, the single-lepton channel, in which VLQ pairs decay to two third generation quarks and two bosons with only one W decaying leptonically, is most sensitive to the three $T\bar{T}$ decay modes and $B \rightarrow tW$ decays. The hadronic decays of the two quarks and the remaining boson produce at least three energetic jets with a cone radius of $\Delta R = 0.8$ (AK8 jets). VLQ candidates are reconstructed from the lepton, p_T^{miss} and one AK8 jet or from two AK8 jets. A multilayer perceptron network (MLP) is trained to separate signal from background and is used to divide events into twelve categories in the SR. Instead of a single kinematic variable, the prediction score of a deep neural network (DNN) on VLQ reconstruction is used as the discriminating variable.

The 95% CL upper limits on the production cross sections of $T\bar{T}$ and $B\bar{B}$ for masses of 0.9–1.8 TeV are calculated using Higgs Combine with asymptotic approximation. Combining the three types of leptonic final states, 54 (48) templates are fed into the binned maximum likelihood fits of the $T\bar{T}$ ($B\bar{B}$) analysis. Using the Barlow-Beeston lite method, statistical uncertainties of MC samples enter the likelihood fit as Poisson-distributed nuisance parameters. Systematic uncertainties in Table 5.34 are included as nuisance parameters in the form of log-normal priors (normalization-only) or Gaussian priors (shape) while signal cross section is assigned a flat prior.

5.12.1 Limits

In the full Run 2 $T\bar{T}$ analysis, the observed (expected) upper limits intersect the theory cross sections at T mass of 1.49 (1.41) TeV for branching ratios in singlet scenario and 1.50 (1.50) TeV for branching ratios in doublet scenario, as shown in Figure 5.75. In the $B\bar{B}$ analysis, the observed(expected) upper limits intersect the theory cross sections at B mass of 1.47 (1.38) TeV for branching ratios in singlet

Table 5.34: Summary of systematic uncertainty sources for the various analysis channels, grouped according to channel. The second column shows uncertainties for the single-lepton channel, evaluated with 2016–2018 data combined. The third and fourth columns show uncertainties for the other channels, evaluated using 2017 or 2018 data. Ranges indicate representative values across different lepton flavour categories. Uncertainties that affect different sources are indicated as “MC” for all simulation (including signal), “OS” for opposite-sign dilepton events used to derive CMID background, and “NP” for nonprompt lepton background.

Source	1 ℓ	Uncertainty		Samples
		2–3 ℓ 2017	2–3 ℓ 2018	
Luminosity	1.6%	2.3%	2.5%	MC
μ ID + Iso	4.3%	2.5–7.5%	2.5–7.5%	MC
e Reco + Iso	3.1%	1.4–4.2%	1.8–5.4%	MC
e ID		$\pm\sigma(p_T, \eta)$		MC
Pileup		$\sigma_{inel.} \pm 4.6\%$		MC
Trigger prefiring		$\pm\sigma$		MC
Jet energy scale		$\pm\sigma(p_T, \eta)$		MC
Jet energy res.		$\pm\sigma(\eta)$		MC
1, 3 ℓ b tag: b/c		$\pm\sigma(p_T)$		MC
1, 3 ℓ b tag: light		$\pm\sigma(p_T)$		MC
Ren./Fact. scale		env($\times 2, \times 0.5$)		MC
PDF		$\pm\sigma$		MC
<i>Single-lepton channel</i>				
e trigger	$\pm\sigma(p_T, \eta)$	-	-	MC
μ trigger	$\pm\sigma(p_T, \eta)$	-	-	MC
H_T W+jets corr.	env(upper, lower fits)	-	-	W+jets
H_T $t\bar{t}$ corr.	$\pm\sigma$	-	-	$t\bar{t}$
DeepAK8: light	$-\sigma$	-	-	MC
DeepAK8: light (higgs)	$\pm\sigma$	-	-	MC
DeepAK8: heavy eff.	$\pm 25\%/\text{jet/flavour}$	-	-	MC
DeepAK8: heavy mis.	$\pm 25\%/\text{jet/flavour}$	-	-	MC
<i>SS dilepton channel</i>				
ee trigger	-	$\pm\sigma(p_T, \eta)$		MC
$e\mu$ trigger	-	$\pm\sigma(p_T, \eta)$		MC
$\mu\mu$ trigger	-	$\pm\sigma(p_T, \eta)$		MC
Charge misid.	-	34.5%	60.0%	OS
2 ℓ e/μ prompt rate	-	$\pm\sigma(p_T, \eta)$		NP
2 ℓ e NP rate	-	NP rate + 0.07	NP rate + 0.08	NP
2 ℓ μ NP rate	-	NP rate - 0.02		NP
<i>Multi-lepton channel</i>				
3 ℓ e prompt rate	-	1.0–6.6%	0.9–7.3%	NP
3 ℓ μ prompt rate	-	1.0–6.6%	1.2–5.5%	NP
3 ℓ ℓ NP rate	-	3.9–23.0%	1.4–24.6%	NP
3 ℓ μ NP rate in $\eta(\mu)$	-	0–38.7%	1.0–1.4%	NP
3 ℓ e/μ NP rate	-	$\pm\sqrt{\sigma_{\text{gaus}}^2 + \sigma_{\Delta(f_{\text{SR}}, f_{\text{CR2}})}}}$		NP

scenario and 1.11 (1.09) TeV for branching ratios in doublet scenario, as shown in Figure 5.76. The numerical values of observed and expected mass limits at the two benchmarks and 100% branching ratio of the three individual decay modes respectively are documented in Table 5.35.

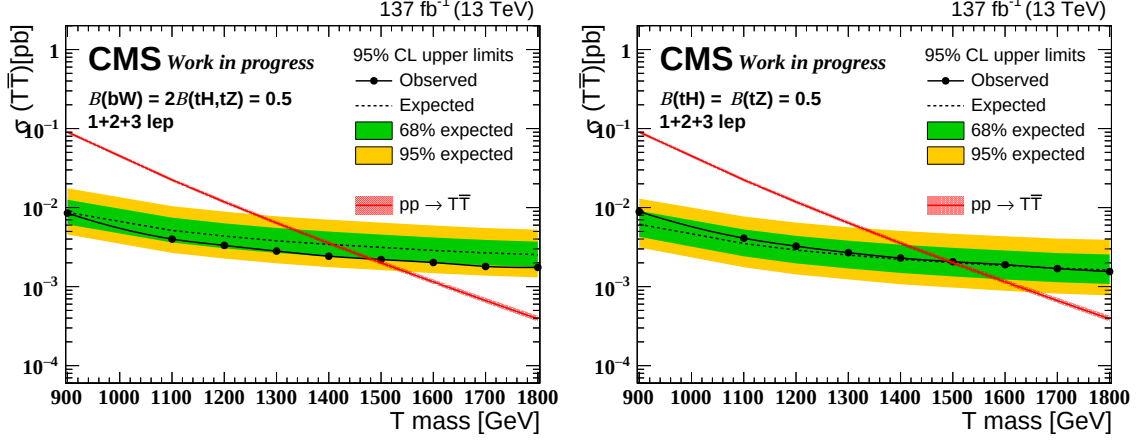


Figure 5.75: Expected 95% CL upper limits on $T\bar{T}$ production for the singlet (left) and doublet (right) benchmarks.

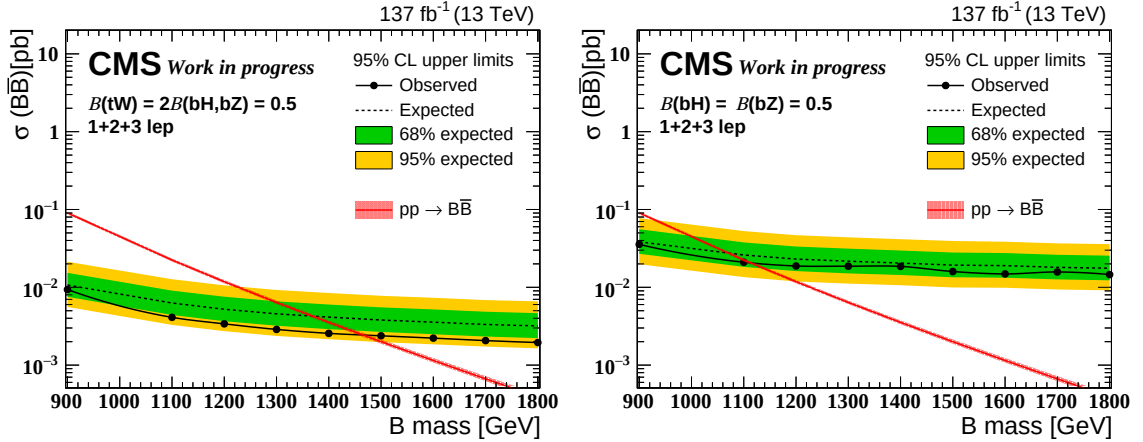


Figure 5.76: Expected 95% CL upper limits on $B\bar{B}$ production for the singlet (left) and doublet (right) benchmarks.

For a model-independent search of VLQ, all possible combinations of branching ratios are scanned, from 0 to 1.0 at a step of 0.2 for each of the three decay modes. The scan for $T\bar{T}$ analysis, shown in Figure 5.77, excludes T masses below 1.48–1.54 TeV, depending on the branching fractions. The $B\bar{B}$ analysis excludes B masses below 1.11–1.56 TeV, as shown in Figure 5.78. The maximum limits on T and B mass are found at scenarios of $\mathcal{B}(bW) = 100\%$ and $\mathcal{B}(tW) = 100\%$ respectively. The exclusion limit on T mass is the strongest observed in CERN, considering all publications from the CMS and ATLAS collaborations.

Table 5.35: Expected and observed lower limits on the T and B quark masses.

$\mathcal{B}(bW)$	$\mathcal{B}(tZ)$	$\mathcal{B}(tH)$	Expected Limit (TeV)	Observed limit (TeV)
1.0	0	0	1.44	1.54
0	1.0	0	1.47	1.48
0	0	1.0	1.55	1.50
0.50	0.25	0.25	1.41	1.49
0	0.50	0.50	1.50	1.50

$\mathcal{B}(tW)$	$\mathcal{B}(bZ)$	$\mathcal{B}(bH)$	Expected Limit (TeV)	Observed Limit (TeV)
1.0	0	0	1.51	1.56
0	1.0	0	1.08	1.12
0	0	1.0	1.12	1.11
0.50	0.25	0.25	1.38	1.47
0	0.50	0.50	1.09	1.12

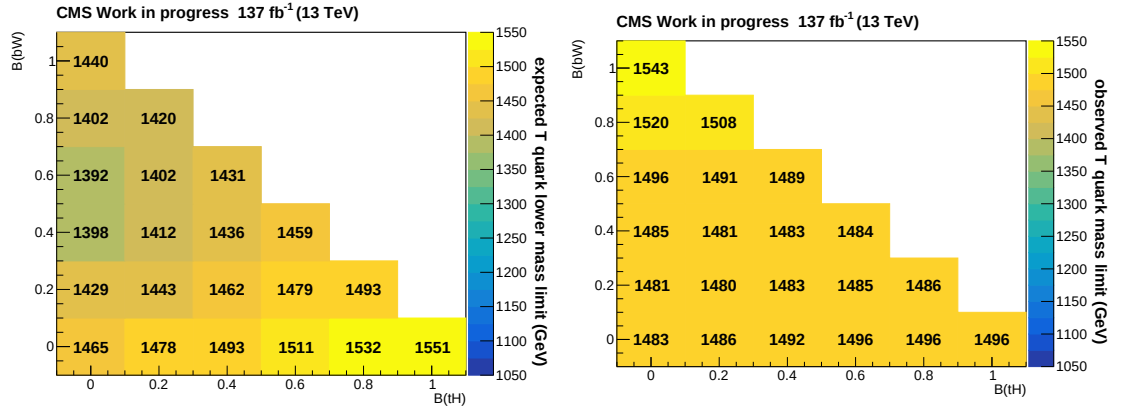


Figure 5.77: The 95% CL expected (left) and observed (right) lower limits on the T quark masses for many possible branching fraction combinations.

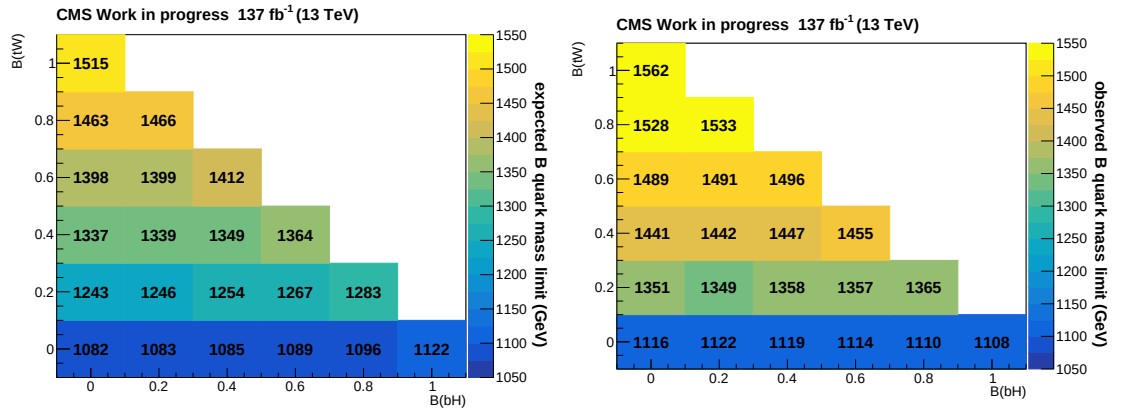


Figure 5.78: The 95% CL expected (left) and observed (right) lower limits on the B quark masses for many possible branching fraction combinations.

Chapter 6

Summary

In 2016–2018, the LHC produced approximately 150 fb^{-1} of proton-proton collisions at centre-of-mass energy of 13 TeV at the CMS detector. At this high energy frontier, it is possible to probe for hypothetical new heavy particles. This dissertation presented a search for pairs of non-chiral vector-like quarks of flavour T and B in the LHC in final states with leptons. I am responsible for analysing data in 2017–2018 for $T\bar{T}$ and $B\bar{B}$ signals in same-sign dilepton and multi-lepton final states. Previously published same-sign dilepton and multi-lepton searches on 2016 data [21] and a single-lepton search in full Run 2 data by Dr. Hogan from Bethel University are combined with my work to find stronger exclusion limits on VLQ masses. The combined search is the most complete one performed to date in the CMS collaboration, covering all search-able leptonic final states and all existing 13 TeV data.

In the single-lepton channel, signal and control regions are defined by predictions from a multilayer perceptron network trained to discriminate signal from background. In same-sign dilepton and multi-lepton channel, due to the rare SM background, signal can be separated from background more easily with event signatures, primarily the jets. Signal and control regions are defined based on jet multiplicity. The discriminants H_T^{lep} and S_T , which include the sum of jets' p_T , are sensitive to signal at high value as jets from the heavy VLQs are more boosted than SM background.

In both same-sign dilepton and multi-lepton channel, there is a significant contribution to background from nonprompt leptons. Instead of studying each of the different sources of nonprompt leptons, they are estimated as one entity by a data-driven method. According to the matrix method, the total nonprompt background yield can be predicted by weighing events with loose lepton(s) using prompt

and fake rates. In same-sign dilepton channel, there is also background due to misidentification of leptons' charge, which is again predicted from data. Opposite-sign dilepton events in data are weighted by the CMID rates.

Incorporating all three channels, a combined binned maximum likelihood fit calculate the 95% CL upper limits on the production cross sections for VLQ masses of 0.9–1.8 TeV. Scanning over possible combinations of branching fractions, ***T* mass is excluded below 1.48–1.54 TeV and *B* mass 1.11–1.56 TeV.**

In retrospect, before Run 2, *T* and *B* quarks masses were excluded for masses below 0.72–0.95 TeV and 0.58–0.81 TeV respectively at 95% C.L., for different combinations of branching fractions, in pair production searches by the CMS and ATLAS Collaborations at $\sqrt{s} = 7$ TeV [14, 15, 16], 8 TeV [17, 18, 19, 20]. At the beginning of Run 2, an equivalent search for pair production of *T* and *B* quarks in leptonic final states based on data collected in 2016 excluded their masses below 1.14–1.3 TeV and 0.91–1.24 TeV [21], which explains why this analysis searches for VLQ with masses at or above 0.9 TeV. Considering all searches for VLQ pair by both CMS and ATLAS on partial or full Run 2 data at $\sqrt{s} = 13$ TeV, *T* and *B* quarks masses are excluded below 1.31–1.42 TeV and 1.22–1.58 TeV [21, 22, 23, 24, 25, 26].

Table 6.1: The exclusion limits on VLQ masses in pair production searches reported by the CMS and ATLAS collaborations over the years.

Year	2010–2012	the 2016 Search[21]	2016–before Ours	Our Results
$\sqrt{s}$	7–8 TeV	13 TeV		
VLQ	exclude masses below (TeV)			
T	0.72–0.95	1.14–1.30	1.31–1.42	1.48–1.54
B	0.58–0.81	0.91–1.24	1.22–1.58	1.11–1.56

Comparing the result of this analysis with other published results, as shown in Table 6.1, this combined analysis has extended the exclusion limits on *B* quark mass, from 0.91–1.24 TeV measured in 2016 data to 1.11–1.56 TeV with the full Run 2 data. **More excitingly, the updated exclusion limits of *T* quark mass at 1.48–1.54 TeV and that of *B* quark mass with significant decays to tW are the strongest to date,** in both CMS and ATLAS Collaborations.

Using similar strategies, with slight fine-tuning, it is possible to search for other flavours of VLQs or other new physics signals which are capable of producing the same final states, such as $X_{5/3}\bar{X}_{5/3}$ and

$t\bar{t}t\bar{t}$. Extension of search to $X_{5/3}\bar{X}_{5/3} \rightarrow tW\bar{t}W$ signals is more trivial. The same event selections and background estimation can be applied and the preliminary expected limits found are comparable to previous $X_{5/3}$ analyses. Unfortunately, such an analysis might not set stronger limits than existing searches. It is recommended to optimize the final selections for $X_{5/3}\bar{X}_{5/3}$ signals, which is not completed at the time of this writing.

In addition, $t\bar{t}t\bar{t}$ production processes can also be studied. It is analogous to a new physics signal with only one mass point. However, applying the same event selections and background estimation, the preliminary expected limits found are a lot weaker than previous searches. This is because $t\bar{t}t\bar{t}$ processes have fairly different event signatures compared to VLQs. Event selections, as well as signal and control regions, need to be modified. At the very least, jet multiplicity cut should be raised. $t\bar{t}t\bar{t}$ decays with three leptons should produce four b-tagged jets and one or two jets from gauge bosons.

In prospect, it is also possible to improve the analysis with machine learning. The immediate thought is to replace the discriminant with a machine learning prediction score, like the single-lepton channel. For example, using a deep neural network, input variables are automatically tested and compared between different combinations to find the one with the strongest discriminating power. The technical difficulty is to incorporate the data-driven backgrounds. The architecture that is currently available to me uses MC simulations for all background. One can either use MC simulations to mimic data-driven background, such as $t\bar{t}$ which can be scaled to look like nonprompt background, or modify the package to accept data-driven background, which is not trivial. Nevertheless, I believe that this is worth investing into.

Chapter A

Photon and Muon Identification/Isolation Performances

Section 4.4 briefly described how analysis objects are reconstructed from detector signals by the Particle Flow (PF) algorithm. These reconstructed candidates include the real prompt objects from the primary interaction and reducible detection backgrounds such as mis-reconstruction, pileup, and decays within jets [85]. For the rejection of detection backgrounds, identification and isolation algorithms are designed to select reconstructed prompt objects using detector information [140]. Object identification (ID) and isolation (ISO) are also applied in my analysis (Section 5.3).

A.1 Photons

For photons, the most significant source of background is the di-photon decays of light neutral mesons π^0 or η from jet fragmentation (or hadronization) [141]. In the energy regime of most analysis, the neutral mesons are boosted. Thus, di-photon may be detected in the same ECAL crystal, and thus mis-reconstructed as a single-photon.

To identify a real single-photon, three discriminating variables are considered: isolation sums, H/E , and $\sigma_{i\eta i\eta}$. An isolation sum is the sum of p_T of all reconstructed candidates, other than the photon itself, inside a cone of radius $\Delta R = 0.3$ around the photon. The isolation sums are calculated for charged hadron I_{ch} , neutral hadron I_n , and photon I_γ candidates separately. They are corrected for pileup by subtracting contributions in an effective area. Setting an upper threshold on isolation sums

can reject collimated photons which have larger energy spreads.

H/E is the ratio of energy deposited in HCAL to that in ECAL within a cone of radius $\Delta R = 0.15$ around the most energetic crystal (seed) of an ECAL energy cluster (supercluster). It is useful for discriminating between prompt photons and photons from jet, as the latter but not the former will carry significant hadronic energy, i.e., a large H/E ratio.

$\sigma_{i\eta i\eta}$ describes the spread of the electromagnetic shower along η -direction. Considering the 5×5 ECAL readouts around a supercluster seed, the mean pseudorapidity position and the energy deposition of the 5×5 crystals are $\bar{\eta}_{5 \times 5}$ and $E_{5 \times 5}$ while those of the i -th crystal are η_i and E_i . Defined in Equation A.1, $\sigma_{i\eta i\eta}$ is the log-energy-weighted RMS of η position of ECAL crystals. For a single-photon, the distribution of $\sigma_{i\eta i\eta}$ is narrow and centres at a lower value compared to collimated di-photon. Therefore, cutting on $\sigma_{i\eta i\eta}$ can effectively discriminate single-photon from the reducible detection background.

$$\sigma_{i\eta i\eta} = \sqrt{\frac{\sum_i^{5 \times 5} w_i (\eta_i - \bar{\eta}_{5 \times 5})^2}{\sum_i^{5 \times 5} w_i}} \quad \text{where } w_i = \max\left(0, 4.7 + \ln\left(\frac{E_i}{E_{5 \times 5}}\right)\right) \quad (\text{A.1})$$

There are two ID algorithms available for photons in the CMS experiment: cut-based ID and multivariate-analysis (MVA) ID. The former has working points tight, medium, and loose for average passing rates of 90%, 80%, and 70% respectively. The latter has two working points WP90 and WP80 for efficiencies of 90% and 80% respectively.

A.2 Muons

Muon candidates considered by the identification algorithm are tracker muons, which are defined in Section 4.4 as inner tracks matched to at least one muon chamber segment. Detection background arises from in-flight decays or energetic charged hadrons that penetrates the entire HCAL (punch-through) and triggered the innermost layer of muon chambers [77, 84].

Identification algorithms use variables that characterize the muon track and track fitting for discrimination. Applying cuts on the number of pixel detector hits and number of inner tracker hits,

background from decays in flight is suppressed. Prompt muons are selected by setting an upper limit on the distance of a muon track from PV. Setting a minimum on number of hits in muon chambers and a maximum on χ^2 of track fitting can reduce background due to punch-throughs.

There are a number of cut-based IDs available for muons in the CMS experiment, optimized for different stringencies. There are three commonly used working points: tight, medium, and loose.

In my analysis and many BSM searches, mini-Isolation is used in addition to cut-based ID. The definition of mini-Isolation and its tight and loose working points are described in Section 5.3.1.

A.3 Tag-and-Probe Method

The ID and ISO efficiencies of leptons and photons are measured by the tag-and-probe (TnP) method [77, 112, 142]. It is a common technique used to estimate the passing rate of a selection.

In $Z \rightarrow \ell\bar{\ell}$ events, Z decays to two leptons of the same flavour. The “tag” and “probe” refer to the two opposite-sign (OS) same-flavour leptons respectively. A tag lepton is identified with stringent selections, such as passing ID at tight working point and/or ISO requirements, in addition to some basic requirements on triggers, p_T and η acceptance. The collection of tags therefore has a high purity of real electrons or muons.

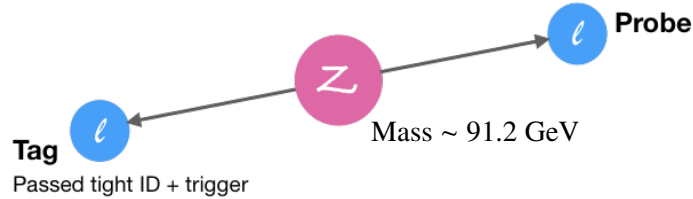


Figure A.1: A diagram illustrating the tag and probe leptons from a Z mother.

For each tag lepton, all same-flavour lepton candidates are looped through to find the matching probe lepton. The tag and probe leptons must have OS electric charges and an invariant dilepton mass $m_{\ell\bar{\ell}}$ within a range around the Z boson mass. For muons, $60 < m_{\mu\bar{\mu}} < 120$ GeV is required. The crux of the TnP method is that the collection of probes is unbiased, since they are selected with requirements on dilepton mass and charge sign with very loose p_T and η cuts. However, the loose selection also means that not all selected tag-probe pairs are real dilepton pairs from Z decays. To

find the actual yields of $Z \rightarrow \ell\bar{\ell}$ events, the distributions of $m_{\ell\bar{\ell}}$ are fitted with a Gaussian function that peaks at Z mass plus a smooth background. To find the efficiencies of passing a ID/ISO, TnP events are categorized by whether the probe lepton passes or fails the ID/ISO requirements, and the two categories are fitted simultaneously. The ID/ISO efficiencies is then estimated by ratios of the Gaussian signal yields of passing probes to that of passing and failing probes $\left(\frac{\text{pass}}{\text{pass}+\text{fail}}\right)$.

The same measurement method is applicable to photon with some corrections [141, 142, 143]. In $Z \rightarrow e\bar{e}$ events, electrons pretend photons when cuts on tracks and electric charge are ignored. The shape of the electromagnetic shower is corrected for the differences between electrons and photons. The tag and probe lepton pairs are determined with $50 < m_{\gamma\bar{\gamma}} < 130$ GeV. Similar to leptons, selection efficiencies on photons can be measured by $\left(\frac{\text{pass}}{\text{pass}+\text{fail}}\right)$ in probes.

To account for the discrepancies in ID/ISO efficiencies between data and MC simulations, the scale factors which are ratio of efficiencies in data to that in MC are calculated.

A.4 Photon Loose ID

Selection criteria for photon cut-based ID at loose working point are defined in Table A.1. Efficiencies and scale factors of the photon loose ID are calculated for 2016–2018 data with ultra-legacy (UL) reprocessing in which calibrations are improved [70].

Table A.1: Requirements of cut-based photon ID at loose working point.

Variable	ECAL Barrel (EB)	ECAL Endcap (EE)
H/E	< 0.04596	< 0.0590
$\sigma_{i\eta i\eta}$	< 0.0106	< 0.0272
I_{ch}	< 1.694 GeV	< 2.089 GeV
I_n	$< (24.032 + 0.01512p_T + 2.26 \times 10^{-5}p_T^2)$ GeV	$< (19.722 + 0.0117p_T + 2.3 \times 10^{-5}p_T^2)$ GeV
I_γ	$< 2.876 + 0.004017p_T$	$< 4.162 + 0.0037p_T$

ID efficiencies are measured in data stored with the single-electron trigger “HLT_Ele32_WPTight_Gsf_L1DoubleEG_v” and MC simulations of Drell-Yan events with di-electron final state. Tags are selected with $p_T > 35$ GeV, $|\eta| < 2.17$, and cut-based tight ID. Probes are selected with electric charge and invariant dilepton mass. Passing and failing probes in MC sample are fitted with a Crystal Ball $\times$ Gaussian signal function and an exponential background function with power law tail

(MC-altSig). Data samples are fitted three individual times with a Gaussian signal function plus a exponential background function with turnover at low mass (nominal), a MC-fitted signal function plus nominal background function (altSig), and a Gaussian signal function plus a purely exponential background function (altBkg).

Fits in a p_T - η bin in 2017 UL is shown in Figure A.2 as an example. Fits in some bins requires manually tuning fit parameters and modifying fit functions. For example, at low p_T , an additional Gaussian function is added to better fit the distributions of failing probes, as shown in Figure A.3.

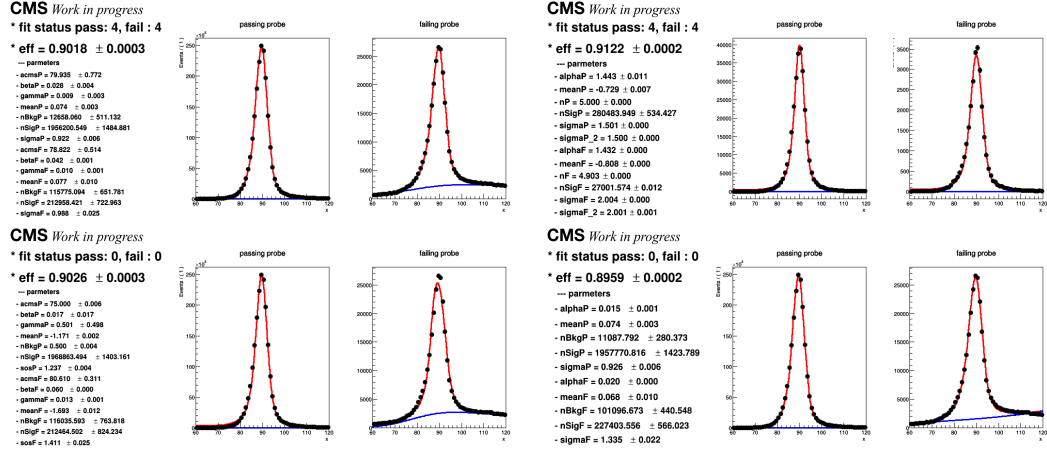


Figure A.2: Fits to distributions of dilepton masses for passing and failing probe in bin $35 < p_T < 50$ GeV and $-1.44 < \eta < -0.8$. 2017 UL data is fitted by nominal (top left), altSig (bottom left), and altBkg (bottom right) fit functions respectively. DY MC fitted by MC-altSig fit function (top right).

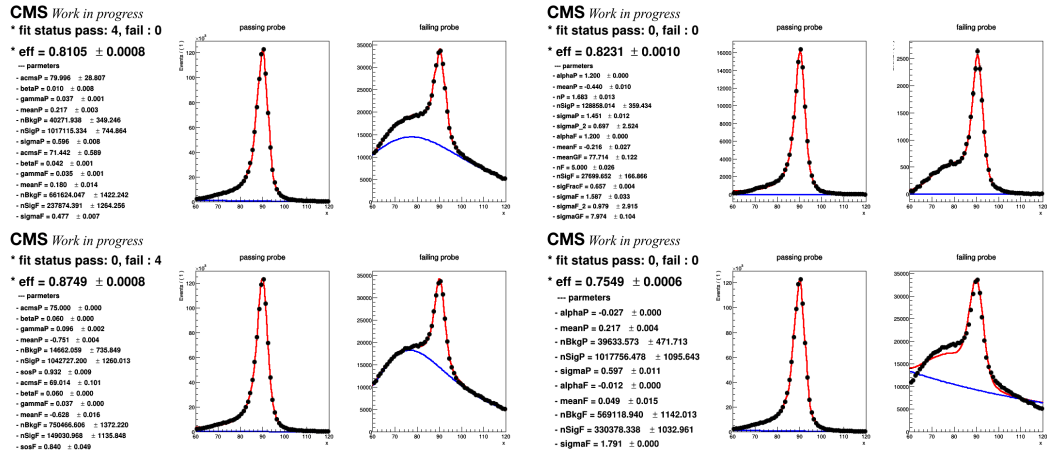


Figure A.3: Fits to distributions of dilepton masses for passing and failing probe in bin $20 < p_T < 35$ GeV and $0.0 < \eta < 0.8$. 2017 UL data is fitted by nominal (top left), altSig (bottom left), and altBkg (bottom right) fit functions respectively. DY MC fitted by MC-altSig fit function (top right).

Scale factors for photon Loose ID were measured for 2016, 2017, and 2018 UL respectively, and the results are shown in Figures A.4. Uncertainties of the scale factors are calculated by the quadrature

sum of statistical errors, differences between altBkg and nominal fits, differences between altSig and nominal fits, differences between MC-altSig fits and counting in MC, and differences between counting in MC with tag selection of $p_T > 37$ GeV and $p_T > 35$ GeV. Large uncertainty in bin $20 < p_T < 35$ GeV arises from systematic uncertainties.

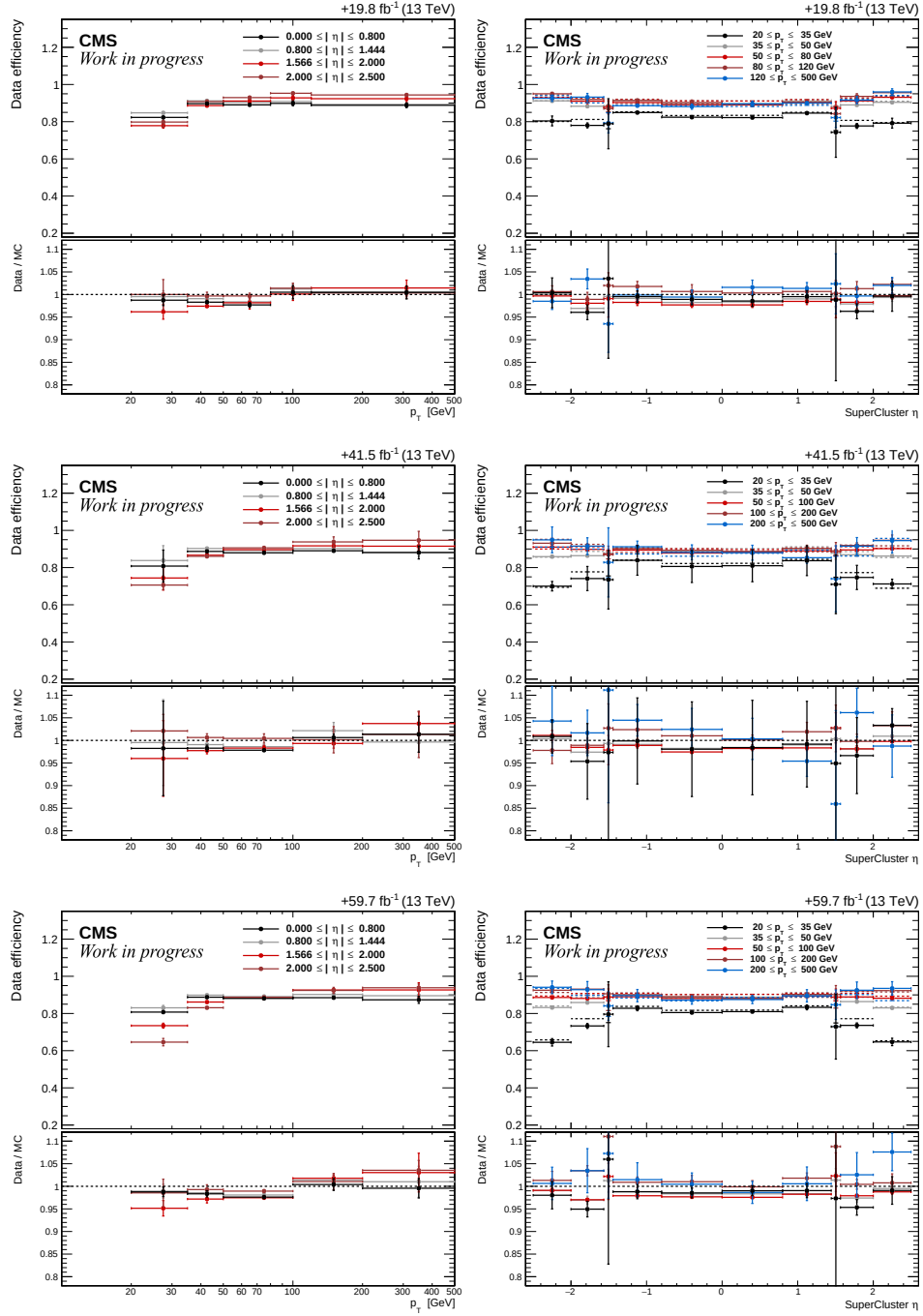


Figure A.4: The measured data efficiencies and scale factors as a function of p_T (left) and η (right) in 2016 UL (top), 2017 UL (middle), and 2018 UL (bottom).

A.5 Muons in Different Event Topologies

I initiated this project together with the CMS Muon POG, hoping to optimize muon ID/ISO efficiencies for analyses. Conventionally, SFs are measured by TnP method in Drell-Yan (DY) events. However, the SM background in many analyses is more $t\bar{t}$ -like. Drell-Yan and $t\bar{t}$ events have different topologies, which can be parametrized by variables such as jet multiplicity, muon multiplicity, and lepton-jet ΔR . In this project, the ID/ISO efficiencies as a function of p_T and $|\eta|$ are computed, in $t\bar{t}$ and DY MC events respectively. Noticing the differences between the two distributions, efficiencies in DY are computed as a function of different topology parameters to find which can be used to reproduce the efficiencies observed in $t\bar{t}$ sample. If such a topology dependence is observed, it can be included in future measurement of efficiencies which are applied to $t\bar{t}$ -like processes.

During the proposal of this project, Muon POG was developing a new TnP analysis framework and I worked with the validations and developments of the involved software packages as part of this project. The new TnP analysis tool `spark_tnp` is designed to measure muon ID/ISO efficiencies by fitting the distributions of tag-probe dilepton masses. Previously, these distributions are produced by looping over the large number of events one by one with ROOT [144, 145]. In `spark_tnp`, events are stored in a **pandas** DataFrame [146] which is similar to an advanced version of spreadsheet. Data handling becomes a lot easier and data processing is a lot faster using Apache Spark environment [147]. `MuonAnalysis-MuonAnalyzer` is designed specifically to interface between `spark_tnp` and the CMS data storage format. CMS data are stored in Analysis Object Data (AOD) format, or the slimmed Mini-AOD [148]. `MuonAnalysis-MuonAnalyzer` reads the CMS AOD or Mini-AOD files and output flat ntuples in which information is stored as lists of values. Flat ntuples can be easily converted to a DataFrame which can then be stored in Parquet files [149]. These new programs are used in my TnP study

To compare $t\bar{t}$ against DY events, “MC Truth” efficiencies are measured using information from the MC generators. As a starting point, efficiencies of muon cut-based tight ID and mini-Isolation < 0.1 which are used in my VLQ search (Section 5.3) are studied.

A.5.1 Framework Validation and Development

For this project, the information of the generated (GEN) particles must be kept. The ntuple producer `MuonAnalysis-MuonAnalyzer` is updated with new collections that stores the properties of relevant GEN particles and the calculated mini-Isolation values of each muon. The analyzer `spark_tnp` is also updated to process GEN particle information and evaluate mini-Isolation working points.

To cross-check the new programs (`MuonAnalysis-MuonAnalyzer+spark_tnp`) with the preceding equivalent (`TagProbeFitTreeProducer+TnPUtils`), the same type of ID/ISO efficiencies are measured by each software package with the same set of samples and event selections. Figures A.5 and A.6 show the efficiencies that a cut-based tight ID muon passes mini-Isolation < 0.1 measured in 2017 data and DY MC sample respectively. In 2017 data, the differences in measured efficiencies between the programs are generally small, ± 0.003 to 0.005 , except for two low- p_T bins. In $20 < p_T < 25$ and $0.9 < |\eta| < 1.2$, despite a slightly large deviation of 0.013 , measured mini-ISO efficiencies agree between the two programs within each other's error. Similarly, in DY MC events, the measured mini-ISO efficiencies agree between the two packages. Thus, the new framework is put in use.

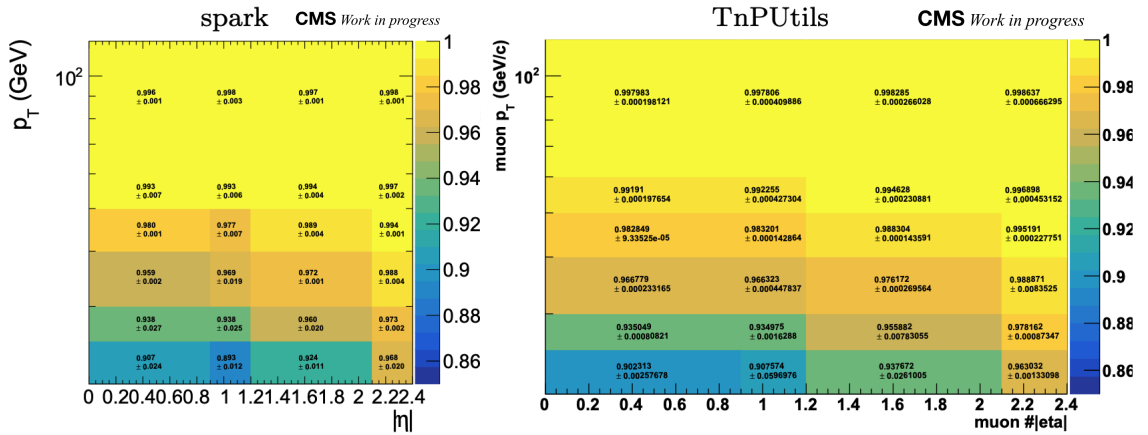


Figure A.5: Measured efficiencies of cut-based tight muons passing mini-Isolation < 0.1 as a function of p_T and $|\eta|$ in 2017 data by `spark_tnp` (left) and `TnPUtils` (right).

Though the new framework serves its purpose, I still tried to understand why deviations exist. The fits on passing and failing probes in each p_T - $|\eta|$ bin are examined and compared between the

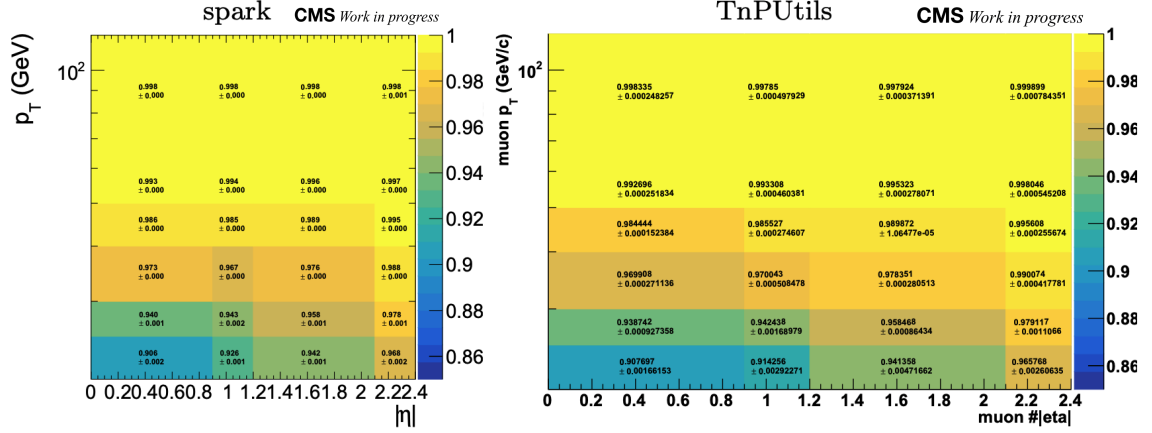


Figure A.6: Measured efficiencies of cut-based tight muons passing mini-Isolation < 0.1 as a function of p_T and η in DY MC events for 2017 by `spark_tnp` (left) and `TnPUtils` (right).

frameworks, and there is no noticeable difference. I studied both framework closely and realized that efficiencies are computed differently, which might be the source of deviations. In `TnPUtils`, the $m_{\mu\bar{\mu}}$ distributions of passing and failing probes are fitted simultaneously for the efficiency in each bin. In `spark_tnp`, distributions of passing and failing probes are fitted independently and the efficiency in each bin is calculated by $\frac{\text{pass}}{\text{pass}+\text{fail}}$. It is hard to tell which method is more accurate, but the latter one is more consistent with the TnP analysis tools for electrons and photons by EGamma POG.

Additionally, I made a replication of `MuonAnalysis-MuonAnalyzer` which is modified to process $t\bar{t}$ MC samples for my topology study. For $t\bar{t}$ events, since there is no real tag and probe leptons, the $m_{\mu\bar{\mu}}$ requirement is removed. Instead, each lepton is checked for a top quark mother and for being a prompt lepton in its GEN information. In $t\bar{t}$ MC simulations, it is possible to find two prompt leptons from top quarks, which are stored in the output ntuple. There is no need for a $t\bar{t}$ analyzer, as the ID/ISO efficiencies in $t\bar{t}$ events are obtained by cut-and-count. As once described in Section 5.6.2, cut-and-count means applying a cut and counting the number of events passing and failing the cut.

A.5.2 MC Truth Efficiencies

It took repeated discussion with the CMS Muon POG to correctly define the MC truth efficiencies in both DY and $t\bar{t}$ events, before studying their typology-dependences. In DY events, GEN di-muon

pairs are selected if they have the same Z mother. In $t\bar{t}$ events, a t quark decays to bW and muon may arise from the semi-leptonic decay of W boson. Thus, GEN muons are stored if they have a W mother and a t ‘grandmother’. Each $t\bar{t}$ MC event contains zero to two prompt GEN muons.

Each selected GEN muon is matched to a reconstructed (reco) muon geometrically. ID/ISO requirements are applied on reco muons. In DY events, tag muons are selected by $p_T > 29$ GeV and $|\eta| < 2.4$ and probe muons $p_T > 10$ GeV and $|\eta| < 2.4$, in addition to di-muon requirement of OS charges and $70 < m_{\mu\mu} < 110$ GeV. In $t\bar{t}$ events, one of the two muon must pass tag’s cuts and the other muon passes probe’s. In each event, both combinations of tag-probe or probe-tag are considered when the p_T and η cuts are fulfilled. The efficiencies are determined by counting the number of matched reco probes passing and failing the ID/ISO.

MC truth efficiencies of a GEN muon passing the muon cut-based tight ID were measured for DY and $t\bar{t}$ processes respectively, using MC samples generated with 2018 LHC conditions. Results are shown in Figures A.7–A.8. The plots of measured MC truth efficiencies against p_T are very different between the two MC samples.

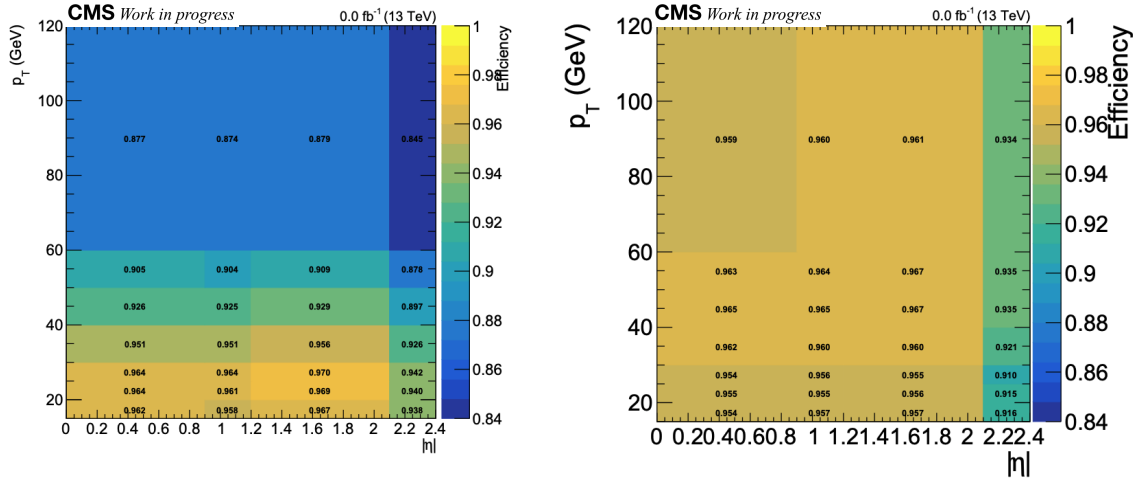


Figure A.7: Measured MC truth efficiencies of GEN muons passing cut-based tight ID as a function of p_T and $|\eta|$ in 2018 MC samples of $t\bar{t}$ (left) and Drell-Yan (right) processes.

Eight topology parameters are studied: z -coordinate of the PV (dZ), the minimum separation between lepton and jet, number of b -tagged jets, number of muons, number of AK4 jets, relative p_T between the lepton-jet pair with smallest ΔR , relative isolation of the probe muon, and scalar sum of jets’ p_T (H_T). Recall that the aim is to find a topology parameter which affects the efficiency

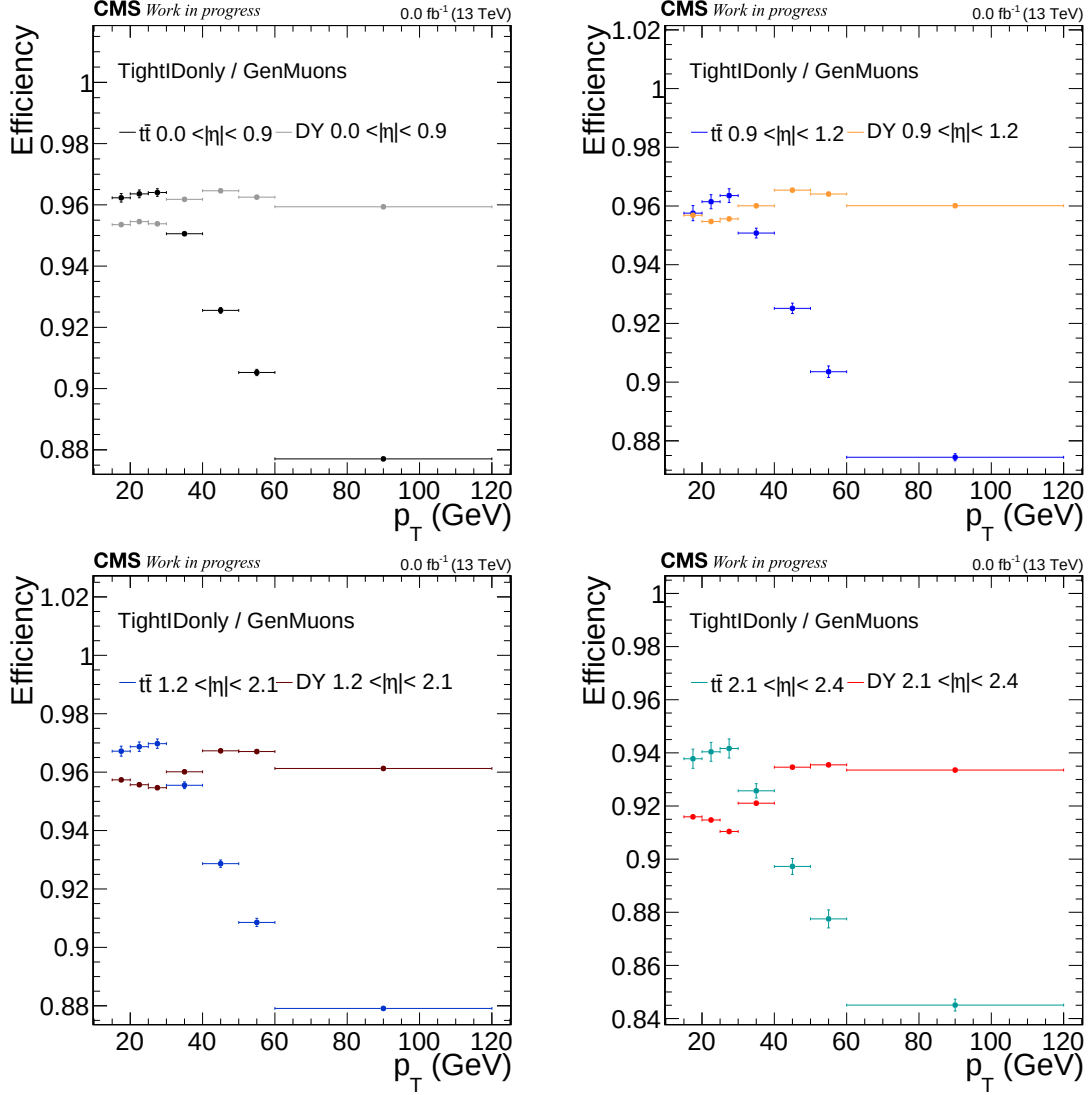


Figure A.8: Measured MC truth efficiencies of GEN muons passing cut-based tight ID as a function of p_T in $|\eta| < 0.9$ (top left), $0.9 < |\eta| < 1.2$ (top right), $1.2 < |\eta| < 2.1$ (bottom left), and $1.2 < |\eta| < 2.4$ (bottom right) in 2018 for $t\bar{t}$ and Drell-Yan processes.

measurement in DY events and produces MC truth efficiencies that falls with p_T as observed in $t\bar{t}$ events. Among all the parameters studied, only dZ modifies the MC truth efficiencies sufficiently. In Figure A.9, the MC truth efficiencies measured in DY sample is plotted against p_T for three ranges of dZ . At higher dZ , especially in $dZ > 0.2$, efficiencies falls with p_T and the shape looks similar to that observed in $t\bar{t}$ events. But the measured efficiencies are a lot lower in $dZ > 0.2$ than in $t\bar{t}$ events.

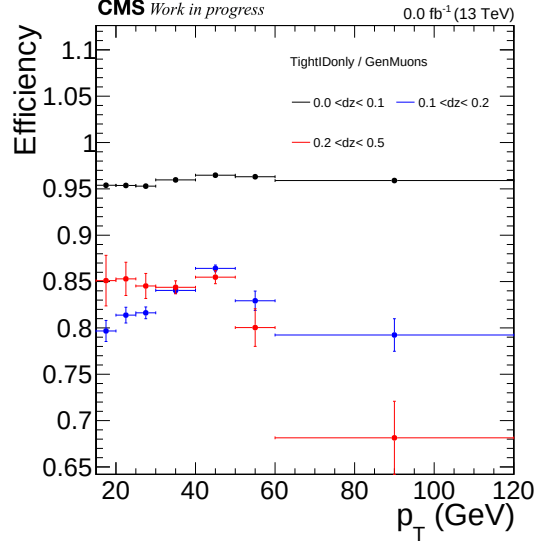


Figure A.9: Measured MC truth efficiencies of GEN muons passing cut-based tight ID as a function of p_T in 2018 DY MC events for different dZ ranges.

So far, this project has observed that ID efficiencies can be very different between $t\bar{t}$ and DY events. One can conclude that there is a large room for improvements in measuring ID/ISO efficiencies for analyses with $t\bar{t}$ -like backgrounds. My study finds that the efficiency- p_T distribution changes with the z position of the primary vertex. I propose at the end of this preliminary investigation that, besides studying more topology parameters, a different or finer binning of dZ can be tested.

Bibliography

- [1] RD Archer-Hind. “Plato Timaeus 51 B”. In: *The Journal of Philology* 24.48 (1895), p. 49.
- [2] Charles T. Sebens. “What’s Everything Made Of?” In: *American Scientist* 108.1 (2020), p. 42.
- [3] David Griffiths. “Introduction to Elementary Particles”. In: WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2004.
- [4] J Ballam, GA Loew, and RB Neal. *STATUS OF DESIGN, CONSTRUCTION, AND RE-SEARCH PROGFUMS AT SLAC*. 1965.
- [5] J Ballam. *The Initial Experimental Program of SLAC*. 1966.
- [6] J. -E. Augustin et al. “Discovery of a Narrow Resonance in e^+e^- Annihilation”. In: *Phys. Rev. Lett.* 33 (23 Dec. 1974), pp. 1406–1408. DOI: 10.1103/PhysRevLett.33.1406. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.33.1406>.
- [7] J. J. Aubert et al. “Experimental Observation of a Heavy Particle J ”. In: *Phys. Rev. Lett.* 33 (23 Dec. 1974), pp. 1404–1406. DOI: 10.1103/PhysRevLett.33.1404. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.33.1404>.
- [8] G Arnison et al. “Experimental observation of lepton pairs of invariant mass around 95 GeV/c² at the CERN SPS collider”. In: *Physics Letters B* 126.5 (1983), pp. 398–410.
- [9] Geoffrey Arnison et al. “Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s}=540$ GeV”. In: *Physics letters B* 122.1 (1983), pp. 103–116.

- [10] Georges Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214.
- [11] Serguei Chatrchyan et al. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235.
- [12] Serguei Chatrchyan et al. “Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV”. In: *JHEP* 06 (2013), 081. 117 p. DOI: 10.1007/JHEP06(2013)081. arXiv: 1303.4571.
- [13] Paul Langacker. *The standard model and beyond*. Taylor & Francis, 2017.
- [14] Serguei Chatrchyan et al. “Search for a vector-like quark with charge $2/3$ in $t + Z$ events from pp collisions at $\sqrt{s} = 7$ TeV”. In: *Phys. Rev. Lett.* 107 (2011), p. 271802. DOI: 10.1103/PhysRevLett.107.271802. arXiv: 1109.4985 [hep-ex].
- [15] Serguei Chatrchyan et al. “Search for pair produced fourth-generation up-type quarks in pp collisions at $\sqrt{s} = 7$ TeV with a lepton in the final state”. In: *Phys. Lett. B* 718 (2012), p. 307. DOI: 10.1016/j.physletb.2012.10.038. arXiv: 1209.0471 [hep-ex].
- [16] Georges Aad et al. “Search for pair production of a new quark that decays to a Z boson and a bottom quark with the ATLAS detector”. In: *Phys. Rev. Lett.* 109 (2012), p. 071801. DOI: 10.1103/PhysRevLett.109.071801. arXiv: 1204.1265 [hep-ex].
- [17] V. Khachatryan et al. “Search for vector-like charge $2/3$ T quarks in proton-proton collisions at $\sqrt{s} = 8$ TeV”. In: *Phys. Rev. D* 92 (2016), p. 012003. DOI: 10.1103/PhysRevD.92.012003. arXiv: 1509.04177 [hep-ex].
- [18] Serguei Chatrchyan et al. “Inclusive search for a vector-like T quark with charge $\frac{2}{3}$ in pp collisions at $\sqrt{s} = 8$ TeV”. In: *Phys. Lett. B* 729 (2014), p. 149. DOI: 10.1016/j.physletb.2014.01.006. arXiv: 1311.7667 [hep-ex].
- [19] Georges Aad et al. “Search for pair production of a new heavy quark that decays into a W boson and a light quark in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”. In:

- Phys. Rev. D* 92 (2015), p. 112007. DOI: 10.1103/PhysRevD.92.112007. arXiv: 1509.04261 [hep-ex].
- [20] Georges Aad et al. “Search for production of vector-like quark pairs and of four top quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”. In: *JHEP* 08 (2015), p. 105. DOI: 10.1007/JHEP08(2015)105. arXiv: 1505.04306 [hep-ex].
 - [21] Albert M Sirunyan et al. “Search for vector-like T and B quark pairs in final states with leptons at $\sqrt{s} = 13$ TeV”. In: *JHEP* 08 (2018), p. 177. DOI: 10.1007/JHEP08(2018)177. arXiv: 1805.04758 [hep-ex].
 - [22] Albert M Sirunyan et al. “Search for pair production of vector-like quarks in the $bW\bar{b}W$ channel from proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Phys. Lett. B* 779 (2018), p. 82. DOI: 10.1016/j.physletb.2018.01.077. arXiv: 1710.01539 [hep-ex].
 - [23] Albert M Sirunyan et al. “Search for vector-like quarks in events with two oppositely charged leptons and jets in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 79.4 (2019), p. 364. DOI: 10.1140/epjc/s10052-019-6855-8. arXiv: 1812.09768 [hep-ex].
 - [24] Albert M Sirunyan et al. “Search for pair production of vectorlike quarks in the fully hadronic final state”. In: *Phys. Rev. D* 100.7 (2019), p. 072001. DOI: 10.1103/PhysRevD.100.072001. arXiv: 1906.11903 [hep-ex].
 - [25] Albert M Sirunyan et al. “A search for bottom-type, vector-like quark pair production in a fully hadronic final state in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Phys. Rev. D* 102 (2020), p. 112004. DOI: 10.1103/PhysRevD.102.112004. arXiv: 2008.09835 [hep-ex].
 - [26] Morad Aaboud et al. “Combination of the searches for pair-produced vector-like partners of the third-generation quarks at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Phys. Rev. Lett.* 121.21 (2018), p. 211801. DOI: 10.1103/PhysRevLett.121.211801. arXiv: 1808.02343 [hep-ex].
 - [27] Mark Srednicki. *Quantum field theory*. Cambridge University Press, 2007.
 - [28] Abdelhak Djouadi. “The anatomy of electroweak symmetry breaking: Tome I: The Higgs boson in the Standard Model”. In: *Physics reports* 457.1 (2008), pp. 1–216.

- [29] Andrew Purcell. *Go on a particle quest at the first CERN webfest. Le premier webfest du CERN se lance à la conquête des particules*. Aug. 2012. URL: <https://cds.cern.ch/record/1473657>.
- [30] Particle Data Group et al. “Review of particle physics”. In: *Progress of Theoretical and Experimental Physics* 2020.8 (2020), p. 083C01.
- [31] Maxim Perelstein, Michael E. Peskin, and Aaron Pierce. “Top quarks and electroweak symmetry breaking in little Higgs models”. In: *Phys. Rev. D* 69 (2004), p. 075002. DOI: 10.1103/PhysRevD.69.075002. eprint: hep-ph/0310039.
- [32] Oleksii Matsedonskyi, Giuliano Panico, and Andrea Wulzer. “Light top partners for a light composite Higgs”. In: *JHEP* 01 (2013), p. 164. DOI: 10.1007/JHEP01(2013)164. arXiv: 1204.6333 [hep-ph].
- [33] Roberto Contino, Leandro Da Rold, and Alex Pomarol. “Light custodians in natural composite Higgs models”. In: *Phys. Rev. D* 75 (2007), p. 055014. DOI: 10.1103/PhysRevD.75.055014. arXiv: 0612048[hep-ph] [hep-ph].
- [34] Roberto Contino et al. “Warped/composite phenomenology simplified”. In: *JHEP* 05 (2007), p. 074. DOI: 10.1088/1126-6708/2007/05/074. arXiv: 0612180[hep-ph] [hep-ph].
- [35] Yasuhiro Okada and Luca Panizzi. “LHC signatures of vector-like quarks”. In: *arXiv preprint arXiv:1207.5607* (2012).
- [36] Aldo Deandrea. “Vector-like Multiplets, Mixings and the LHC”. In: *Journal of Physics: Conference Series*. Vol. 623. 1. IOP Publishing. 2015, p. 012008.
- [37] Giacomo Cacciapaglia et al. “Next-to-leading-order predictions for single vector-like quark production at the LHC”. In: *Physics Letters B* 793 (2019), pp. 206–211.
- [38] Juan Antonio Aguilar-Saavedra et al. “Handbook of vectorlike quarks: Mixing and single production”. In: *Physical Review D* 88.9 (2013), p. 094010.
- [39] Chien-Yi Chen, Sally Dawson, and Elisabetta Furlan. “Vectorlike fermions and Higgs effective field theory revisited”. In: *Physical Review D* 96.1 (2017), p. 015006.
- [40] CMS collaboration et al. “Search for heavy vector-like quarks decaying to same-sign dileptons”. In: *CMS Physics Analysis Summary CMS-PAS-B2G-16-019* (2017).

- [41] Benedetta Belfatto, Revaz Beradze, and Zurab Berezhiani. “The CKM unitarity problem: A trace of new physics at the TeV scale?” In: *The European Physical Journal C* 80.2 (2020), pp. 1–12.
- [42] *Facts and figures about the LHC*. URL: <https://home.cern/resources/faqs/facts-and-figures-about-lhc>.
- [43] JT Boyd. “LHC Run-2 and future prospects”. In: *arXiv preprint arXiv:2001.04370* (2020).
- [44] Andreas Hoecker. “Physics at the LHC Run-2 and Beyond”. In: *CERN Yellow Reports: School Proceedings* 5 (2017), pp. 153–153.
- [45] *Linear accelerator 2*. Sept. 2012. URL: <https://cds.cern.ch/record/1997427>.
- [46] Julie Haffner. *The CERN accelerator complex. Complexe des accélérateurs du CERN*. General Photo. Oct. 2013. URL: <http://cds.cern.ch/record/1621894>.
- [47] *The History of CERN*. URL: <https://timeline.web.cern.ch/timeline-header/89>.
- [48] L Rossi. “The LHC Main Dipoles and Quadrupoles toward Series Production”. In: *IEEE Trans. Appl. Supercond.* 13 (Dec. 2002), 1221–1228. 9 p. URL: <https://cds.cern.ch/record/599389>.
- [49] Lyndon Evans. “The large hadron collider”. In: *New Journal of Physics* 9.9 (2007), p. 335.
- [50] *Pulling together: Superconducting electromagnets*. Aug. 2012. URL: <https://cds.cern.ch/record/1997395>.
- [51] Serge Dailler. “Cross section of LHC dipole. Dipole LHC: coupe transversale.” AC Collection. Legacy of AC. Pictures from 1992 to 2002. Apr. 1999. URL: <https://cds.cern.ch/record/842530>.
- [52] Jean-Luc Caron. “LHC quadrupole cross section.” AC Collection. Legacy of AC. Pictures from 1992 to 2002. May 1998. URL: <https://cds.cern.ch/record/841485>.
- [53] *The High-Luminosity LHC (HL-LHC)*. 2018-2022. URL: <https://voisins.cern/en/high-luminosity-lhc-hl-lhc>.
- [54] “LHC Guide”. Mar. 2017. URL: <http://cds.cern.ch/record/2255762>.

- [55] CMS Collaboration. “Public CMS Luminosity Information”. Feb. 2022. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#Multi_year_plots.
- [56] ATLAS Experiment. “LuminosityPublicResultsRun2”. Nov. 2021. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- [57] Tai Sakuma. *Cutaway diagrams of CMS detector*. May 2019. URL: <https://cds.cern.ch/record/2665537>.
- [58] *ATLAS: The largest volume particle detector ever built*. Feb. 2012. URL: <https://cds.cern.ch/record/1997264>.
- [59] Cms Collaboration et al. *Technical proposal for the upgrade of the CMS detector through 2020*. Tech. rep. 2011.
- [60] B Levesy et al. “CMS solenoid assembly”. In: *IEEE transactions on applied superconductivity* 16.2 (2006), pp. 517–520.
- [61] Roman Adolphi et al. “The CMS experiment at the CERN LHC”. In: *Jinst* 803 (2008), S08004.
- [62] Cheuk-Yin Wong. *Introduction to High-energy Heavy-ion Collisions*. World Scientific, 1994.
- [63] *CMS coordinate system*. URL: https://tikz.net/axis3d_cms/.
- [64] Martina Ressegotti. *Overview of the CMS detector performance at LHC Run 2*. Tech. rep. Geneva: CERN, Nov. 2018. DOI: 10.3390/universe5010018. URL: <http://cds.cern.ch/record/2649417>.
- [65] Julia Ann Gray. “The CMS Phase-1 pixel detector”. In: *Journal of Instrumentation* 8.12 (2013), p. C12047.
- [66] CMS Collaboration. “CMS Tracker Detector Performance Results”. Nov. 2021. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK>.
- [67] Erik Manuel Butz. *Operation and Performance of the CMS outer tracker*. Tech. rep. 2017.
- [68] A Benaglia. “The CMS ECAL performance with examples”. In: *Journal of Instrumentation* 9.02 (Feb. 2014), pp. C02008–C02008. DOI: 10.1088/1748-0221/9/02/c02008. URL: <https://doi.org/10.1088/1748-0221/9/02/c02008>.

- [69] Nazar Bartosik. “SISSA: Performance of the CMS Electromagnetic Calorimeter in LHC Run2”. In: *PoS* (2019), p. 126.
- [70] Francesca Cavallari and Chiara Rovelli. “Calibration and Performance of the CMS Electromagnetic Calorimeter in LHC Run2”. In: *EPJ Web of Conferences*. Vol. 245. EDP Sciences. 2020, p. 02027.
- [71] Federico De Guio. *First results from the CMS SiPM-based hadronic endcap calorimeter*. Tech. rep. Geneva: CERN, Aug. 2018. DOI: 10.1088/1742-6596/1162/1/012009. URL: <http://cds.cern.ch/record/2636475>.
- [72] Seth I Cooper, CMS collaboration, et al. “Phase i upgrade of the cms hadron calorimeter”. In: *Nuclear and particle physics proceedings* 273 (2016), pp. 1002–1007.
- [73] Sunanda Banerjee. *Simulation Geometry for the CMS HCAL*. Tech. rep. CERN-CMS-NOTE-2005-016, 2005.
- [74] J Freeman. *The CMS central hadron calorimeter: Update*. Tech. rep. Fermi National Accelerator Lab.(FNAL), Batavia, IL (United States), 1998.
- [75] G Baiatian et al. *Design, performance, and calibration of CMS hadron endcap calorimeters*. Tech. rep. CERN-CMS-NOTE-2008-010, 2008.
- [76] CMS collaboration et al. “Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data”. In: *Journal of Instrumentation* 5.03 (2010), T03012.
- [77] Albert M Sirunyan, CMS Collaboration, et al. *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}=13$ TeV*. 2018.
- [78] J. G. Layter. *The CMS muon project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: <https://cds.cern.ch/record/343814>.
- [79] CMS collaboration et al. “Calibration of the CMS drift tube chambers and measurement of the drift velocity with cosmic rays”. In: *Journal of Instrumentation* 5.03 (2010), T03016.
- [80] Daniel Francois Teyssier. *CMS Drift Tubes performance*. Tech. rep. Geneva: CERN, Dec. 2015. URL: <https://cds.cern.ch/record/2118383>.
- [81] CMS collaboration et al. “Alignment of the CMS muon system with cosmic-ray and beam-halo muons”. In: *Journal of Instrumentation* 5.03 (2010), T03020.

- [82] Priyanka Kumari et al. *Improved RPC for CMS muon system upgrade for HL-LHC. Improved-RPC for the CMS muon system upgrade for the HL-LHC*. Tech. rep. Geneva: CERN, Apr. 2020. DOI: 10.1088/1748-0221/15/11/C11012. arXiv: 2005.11396. URL: <https://cds.cern.ch/record/2792678>.
- [83] Silvio Donato. “CMS trigger performance”. In: *EPJ Web of Conferences*. Vol. 182. EDP Sciences. 2018.
- [84] CMS collaboration et al. “Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s}=13$ TeV”. In: *arXiv preprint arXiv:2006.10165* (2020).
- [85] A.M. Sirunyan et al. “Particle-flow reconstruction and global event description with the CMS detector.” In: *JINST* 12 (June 2017). Replaced with the published version. Added the journal reference and DOI. All the figures and tables can be found at <http://cms-results.web.cern.ch/cms-results/public-results/publications/PRF-14-001> (CMS Public Pages), P10003. 82 p. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965. URL: <https://cds.cern.ch/record/2270046>.
- [86] Seema Sharma. “Understanding the performance of CMS calorimeter”. In: *Pramana* 69.6 (2007), pp. 1069–1074.
- [87] W Erdmann. “Vertex reconstruction at the CMS experiment”. In: *Journal of Physics: Conference Series* 110.9 (May 2008), p. 092009. DOI: 10.1088/1742-6596/110/9/092009. URL: <https://doi.org/10.1088/1742-6596/110/9/092009>.
- [88] *Search for heavy vector-like quarks decaying to same-sign dileptons*. Tech. rep. CMS-PAS-B2G-16-019. Geneva: CERN, 2017. URL: <https://cds.cern.ch/record/2256747>.
- [89] Aram Avetisyan et al. *Search for fermionic top partners with charge $5e/3$ decaying to same-sign dileptons using 2016 data with CMS*. Tech. rep. AN-16-242. DRAFT. Geneva: CERN, Mar. 2017. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2016/242.
- [90] *CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s}=13$ TeV*. Tech. rep. CMS-PAS-LUM-17-004. Geneva: CERN, 2018. URL: <https://cds.cern.ch/record/2621960>.

- [91] *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV*. Tech. rep. CMS-PAS-LUM-18-002. Geneva: CERN, 2019. URL: <https://cds.cern.ch/record/2676164>.
- [92] *Summary table of samples produced for the 1 Billion campaign, with 25ns bunch-crossing*. URL: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/SummaryTable1G25ns>.
- [93] Richard D. Ball et al. “Parton distributions from high-precision collider data”. In: *Eur. Phys. J. C* 77 (2017), 663. 95 p. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph].
- [94] Jon Butterworth et al. “PDF4LHC recommendations for LHC Run II”. In: *J. Phys. G: Nucl. Part. Phys.* 43 (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001.
- [95] Paolo Nason. “A New method for combining NLO QCD with shower Monte Carlo algorithms”. In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040.
- [96] Stefano Frixione, Paolo Nason, and Carlo Oleari. “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”. In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070.
- [97] Simone Alioli et al. “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”. In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043.
- [98] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. “A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction”. In: *JHEP* 09 (2007), p. 126. DOI: 10.1088/1126-6708/2007/09/126. arXiv: 0707.3088 [hep-ph].
- [99] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].
- [100] Rikkert Frederix and Stefano Frixione. “Merging meets matching in MC@NLO”. In: *JHEP* 12 (2012), p. 061. DOI: 10.1007/JHEP12(2012)061. eprint: 1209.6215.
- [101] Pierre Artoisenet et al. “Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations”. In: *JHEP* 03 (2013), p. 015. DOI: 10.1007/JHEP03(2013)015. arXiv: 1212.3460 [hep-ph].

- [102] Michal Czakon and Alexander Mitov. “Top++: A Program for the calculation of the top-pair cross-section at hadron colliders”. In: *Comput. Phys. Commun.* 185 (2014), p. 2930. DOI: 10.1016/j.cpc.2014.06.021. arXiv: 1112.5675 [hep-ph].
- [103] Torbjörn Sjöstrand et al. “An Introduction to PYTHIA 8.2”. In: *Comput. Phys. Commun.* 191 (2015), p. 159. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].
- [104] S. Agostinelli et al. “GEANT4—a simulation toolkit”. In: *Nucl. Instrum. Meth. A* 506 (2003), p. 250. DOI: 10.1016/S0168-9002(03)01368-8.
- [105] *B2G group MC page*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/B2GMonteCarlo#Cross_sections.
- [106] Igor Azhgirey et al. “Beam-induced Background Simulations for the CMS Experiment at the LHC”. In: *27th Russian Particle Accelerator Conference (RuPAC’21), Alushta, Russia, 27 September-01 October 2021*. JACOW Publishing, Geneva, Switzerland. 2021, pp. 225–227.
- [107] CMS collaboration et al. “Pileup mitigation at CMS in 13 TeV data”. In: *arXiv preprint arXiv:2003.00503* (2020).
- [108] *Pileup Reweighting Utilities*. 2013. URL: <https://twiki.cern.ch/twiki/bin/view/CMS/PileupMCReweightingUtilities>.
- [109] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The catchment area of jets”. In: *JHEP* 04 (2008), p. 005. DOI: 10.1088/1126-6708/2008/04/005. arXiv: 0802.1188 [hep-ph].
- [110] *Egamma ID Recipes Run 2*. URL: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/EgammaIDRecipesRun2>.
- [111] Pedro J Fernández Manteca. “Muon Identification and Isolation efficiencies on Run II data with the CMS experiment”. In: *Sixth Annual Conference on Large Hadron Collider Physics. 4-9 June 2018. Bologna*. 2018, p. 68.
- [112] CMS collaboration et al. “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”. In: *arXiv preprint arXiv:2012.06888* (2020).
- [113] *Electron Scale Factors: MiniIsolation*. URL: https://indico.cern.ch/event/842068/contributions/3636508/attachments/1943612/3223940/ElSF_review.pdf.

- [114] *Reference muon id, isolation and trigger efficiencies for 2017 data*. URL: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonReferenceEffs2017>.
- [115] *Reference muon id, isolation and trigger efficiencies for 2018 data*. URL: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonReferenceEffs2018>.
- [116] *Muon Scale Factors: MiniIsolation*. URL: https://indico.cern.ch/event/843364/contributions/3540886/attachments/1897270/3130555/MuonSF_reviewfits.pdf.
- [117] *Jet Identification*. URL: <https://twiki.cern.ch/twiki/bin/view/CMS/JetID>.
- [118] *Jet Energy Corrections: Official Software Tools for applying JEC Corrections and Uncertainties*. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookJetEnergyCorrections>.
- [119] *Introduction to Jet Energy Corrections at CMS*. URL: <https://twiki.cern.ch/twiki/bin/view/CMS/IntroToJEC>.
- [120] Wolf Behrenhoff. *Measurement of differential $t\bar{t}$ cross sections in the dilepton decay channels with the CMS detector at 7 TeV centre-of-mass energy*. 2012.
- [121] *Recommended Jet Energy Corrections and Uncertainties For Data and MC*. URL: <https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC>.
- [122] *Jet Resolution*. URL: <https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution>.
- [123] Emil Bols et al. “Jet Flavour Classification Using DeepJet”. In: *JINST* 15.12 (2020), P12012. DOI: 10.1088/1748-0221/15/12/P12012. arXiv: 2008.10519 [hep-ex].
- [124] CMS Collaboration. *Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13 TeV with Phase 1 CMS detector*. CMS Detector Performance Note CMS-DP-2018-058. CMS, 2018. URL: <http://cds.cern.ch/record/2646773>.
- [125] *Performance of the CMS.DeepJet b tagging algorithm using 41.9 /fb of data from proton-proton collisions at 13TeV with Phase 1 CMS detector*. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/BTV13TeV2017DeepJet>.
- [126] *Recommendation for Using b-tag Objects in Physics Analyses*. URL: <https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation>.

- [127] *Methods to apply b-tagging efficiency scale factors: Jet-by-jet updating of the b-tagging status*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMMethods#2a_Jet_by_jet_updating_of_the_b.
- [128] L. Bauerdick et al. *Search for SUSY with multileptons & 2LSS in 13 TeV data (Run2 legacy)*. Tech. rep. AN-18-280. DRAFT. Geneva: CERN, July 2019. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2018/280.
- [129] *Treatment of the HEM15/16 region in 2018 data*). URL: <https://hypernews.cern.ch/HyperNews/CMS/get/JetMET/2000.html>.
- [130] CMS Collaboration et al. *Computing the contamination from fakes in leptonic final states*. Tech. rep. CMS Internal Note AN-10-261, CERN, 2010.
- [131] Anthony Barker et al. *Search for pair production of a vector-like T quark in the multi-lepton final states*. Tech. rep. AN-16-496. DRAFT. Geneva: CERN, Mar. 2018. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2016/496.
- [132] Gabriel Gallardo. *Measuring electron charge misidentification rate for SUSY analysis at ATLAS by likelihood method with $Z \rightarrow ee$ events*. Tech. rep. The University of Hong Kong, 2016.
- [133] *Reference measurements and calibrations for Run-II*. URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonReferenceSelectionAndCalibrationsRun2#Selection_and_SFs.
- [134] *Reweighting recipe to emulate Level 1 ECAL prefiring*. URL: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/L1ECALPrefiringWeightRecipe>.
- [135] *Luminosity Physics Object Group*. URL: <https://twiki.cern.ch/twiki/bin/view/CMS/TWikiLUM>.
- [136] *Egamma Run II Recommendations*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/EgammaRunIIRecommendations#E_gamma_RECO.
- [137] *Online User Manual to the Higgs Combine Tool*. URL: <https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/>.
- [138] Roger Barlow and Christine Beeston. “Fitting using finite Monte Carlo samples”. In: *Computer Physics Communications* 77.2 (1993), pp. 219–228.

- [139] JS Conway. *Nuisance parameters in likelihoods for multisource spectra, in the proceedings of PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, January 17–20, CERN, Geneva, Switzerland (2011)*. Tech. rep. CERN-2011-006, 2011.
- [140] Andreas Hinzmann. “Tools for physics analysis in CMS”. In: *Journal of Physics: Conference Series*. Vol. 331. 3. IOP Publishing. 2011, p. 032042.
- [141] CMS Collaboration et al. “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”. In: *Journal of Instrumentation* (2021).
- [142] Jonas Rembser, CMS Collaboration, et al. “CMS electron and photon performance at 13 TeV”. In: *Journal of Physics: Conference Series*. Vol. 1162. 1. IOP Publishing. 2019, p. 012008.
- [143] Morad Aaboud et al. “Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016”. In: *The European Physical Journal C* 79.3 (2019), pp. 1–41.
- [144] Ilka Antcheva et al. “ROOT—A C++ framework for petabyte data storage, statistical analysis and visualization”. In: *Computer Physics Communications* 182.6 (2011), pp. 1384–1385.
- [145] Rene Brun and Fons Rademakers. “ROOT—An object oriented data analysis framework”. In: *Nuclear instruments and methods in physics research section A: accelerators, spectrometers, detectors and associated equipment* 389.1-2 (1997), pp. 81–86.
- [146] Wes McKinney et al. “pandas: a foundational Python library for data analysis and statistics”. In: *Python for high performance and scientific computing* 14.9 (2011), pp. 1–9.
- [147] Matei Zaharia et al. “Apache spark: a unified engine for big data processing”. In: *Communications of the ACM* 59.11 (2016), pp. 56–65.
- [148] Giovanni Petrucciani et al. “Mini-AOD: A new analysis data format for CMS”. In: *Journal of Physics: Conference Series*. Vol. 664. 7. IOP Publishing. 2015, p. 072052.
- [149] Deepak Vohra. “Apache parquet”. In: *Practical Hadoop Ecosystem*. Springer, 2016, pp. 325–335.