

SEARCH FOR VECTOR-LIKE HEAVY QUARK PARTNERS
DECAYING TO LEPTONS AND JETS IN pp COLLISIONS
AT $\sqrt{s} = 13$ TEV BY THE CMS EXPERIMENT

by

Rizki Syarif

Sc. M., Brown University, 2015

B.Sc. Honours, University of Sydney, 2013

B.Sc., University of Sydney, 2006

A dissertation submitted in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy
in the Department of Physics at Brown University

Providence, Rhode Island

May 2019

© Copyright 2019 by Rizki Syarif

This dissertation by Rizki Syarif is accepted in its present form by
the Department of Physics as satisfying the dissertation requirement
for the degree of Doctor of Philosophy.

Date _____
_____ Meenakshi Narain, Advisor

Recommended to the Graduate Council

Date _____
_____ Ulrich Heintz, Reader

Date _____
_____ David Cutts, Reader

Approved by the Graduate Council

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

Curriculum Vitae

Education

- **Ph.D. Physics**, Brown University, October 2018.
- **M.Sc. Physics**, Brown University, 2015.
- **B.Sc. (Advanced) Honours in Physics**, First Class Honours, University of Sydney, 2013.
- **B.Sc. (Advanced) in Mathematics and Physics**, University of Sydney, 2006.

Research Experience

- Graduate research assistant, CMS Experiment, supervisor: Prof. Meenakshi Narain,
June 2014 - 2018
 - *Performance measurements of b-tagging algorithms for boosted Higgs ($b\bar{b}$) jets* This work involves studying the performance of various jet clustering algorithm, tagging algorithm, and also their pT and pile up dependence, for the purpose of tagging Higgs boson decaying to b-quark pair jets. More specifically, the jet algorithm considered were C/A jets, AK and kT with various distance parameters. The tagging algorithms considered were IVF CSV and CSV on Fatjets and on Pruned Subjets with Explicit JTA and SV clustering. Various background were also considered for this study such as inclusive QCD, b-jets, b-jets from gluon splitting, light quark jets and gluon jets. All these performances were then compared to previous results published in BTV-13-001.
 - *Performance measurements of 3D integrated circuit sensors for future detectors and detector upgrades.*

In this project, we tested the performance of the new VIP2B chip that had been developed for particle tracking purposes for future high energy physics experiments. The work involves setting up an FPGA system with the VIP2B chip and performing electrical measurements to see how well the chip responds and to see whether the chip was performing as originally designed.

– *Beyond the standard model T' vector-like quark search using 8 TeV data*

In this project, I helped to produce data ntuples and plots of various discrimination variables for the search of the Beyond the Standard model vector-like quark T' search using 8 TeV data. In addition, I worked on the T' to tH decay channel and studied various selection strategies to discriminate against top quark pair background using multivariate analysis based upon Boosted Decision Tree in TMVA framework.

– *Studies on Shower Deconstruction Algorithm as Higgs to b-quark pair tagger*

Shower Deconstruction is a jet tagger algorithm that can be used to discriminate Higgs boson decaying to b-quark pair signal from (QCD) background in the boosted regime. This method has not been studied before using CMS (simulated) data. The work involves studying the performance and integrating the Shower Deconstruction code to CMS software framework. In addition, the work also involves optimizing the input parameters of the algorithm and comparing the resulting performance with other jet discriminants such as n-subjettiness and q-jet volatility. Future work would involve studying Shower Deconstruction in combination with existing jet tagger algorithms.

– *Commissioning and validation of the algorithm for identification of high momentum heavy mass resonance decaying to double b-quarks*

A multivariate (Boosted Decision Tree) approach to tagging boosted double b-quark jet has been developed. Originally developed as boosted Higgs to b-quark pair decay jet tagger, this method is in principle applicable to a range of massive resonance decaying to a pair of boosted b-quark pair. The method involves separately optimizing track related variables and soft lepton related variables using multivariate analysis based upon Boosted Decision Tree to produce a final discriminant. My involvement in the project has been to validate this new tagging method using CMS 13 TeV data. Specifically, I am collaborating to produce the final commissioning and validation plots and also maintaining the validation (software) packages.

- *Exploring N-Jettiness and X Cone as a method to determine jet multiplicity and event shape*

During my fellowship at Fermilab as an LPC Graduate Scholar, I performed studies of N-jettiness used as an observable to determine the multiplicity of jets in an event and the number of X Cone jets to cluster, particularly when the event topology is boosted. The studies were validated using 13 TeV CMS dataset. The results were made public and published in CERN Document Server <http://cds.cern.ch/record/2275226>.

- *Maintenance of the Cosmic Rack Stand.*

During the course of my PhD research, I have worked on the maintenance of the cosmic rack (c-rack) at CERN as service work to the CMS collaboration under the tracker data acquisitions (DAQ) group. The c-rack consists of genuine CMS silicon strip modules of the tracker outer barrel (TOB) subdetector.

- *PhD thesis: Search for Vector-Like Heavy Quark Partners Decaying to Leptons and Jets in pp Collisions at $\sqrt{s} = 13$ TeV by the CMS Experiment*

- Undergraduate researcher, ATLAS experiment, supervisor: Dr. Bruce Yabsley, March 2012 - December 2012.

- Honours thesis: J/ψ pair analysis as a probe of Double Parton Scattering using ATLAS data

Awards & Grants

- Dissertation Fellowship Award, Brown University Physics Department, 2017.
- Graduate Scholar Fellowship, D.O.E / Fermilab LHC Physics Center (LPC) , 2016.
- BOOST2015 conference at University of Chicago Travel Grant, 2015.
- Graduate Teaching Assistant Award, Brown University Physics Department, 2014.

Teaching Experience

- Teaching assistant :
 - Foundations of Mechanics (PHYS 0050), Fall 2013

- Foundation of Electromagnetism and Modern Physics (PHYS 0060), Spring 2014
- Basic Physics (PHYS 0030), Fall 2014
- Introduction to Relativity and Quantum Physics (PHYS 0160), Spring 2014

Selected Publications

- *”Search for vector-like T or B quark pairs in leptonic final states in 36 fb^{-1} of proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ ”* - CMS collaboration, JHEP 08 (2018) 177. [http://dx.doi.org/10.1007/JHEP08\(2018\)177](http://dx.doi.org/10.1007/JHEP08(2018)177), <http://cds.cern.ch/record/2308309>, [arXiv:1805.04758](https://arxiv.org/abs/1805.04758)
- *”Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV ”* - CMS Collaboration, JINST 13 (2017) P05011. [doi:10.1088/1748-0221/13/05/P05011](https://doi.org/10.1088/1748-0221/13/05/P05011), [arXiv:1712.07158](https://arxiv.org/abs/1712.07158).

Talks, Posters, Summer School

- *“ J/ψ pair analysis as a probe of Double Parton Scattering using ATLAS data”* Poster. Australian Institute of Physics (AIP) Congress, Sydney, Australia, 2012
- *“Identification of jets originating from high-momentum $H \rightarrow b\bar{b}$ ”*. Presentation. ”APS April 2016: American Physical Society April Meeting 2016, 16-19 Apr 2016, Salt Lake City, UT (United States).
- *“Search for vector-like T quark pairs in single lepton and trilepton final states in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ ”*. Presentation. DPF2017: 2017 Meeting of the Division of Particles and Fields of the American Physical Society , 31 Jul-4 Aug 2017, Fermilab, Batavia, IL (United States).
- *The 2016 European School of High-Energy Physics (ESHEP2016): Skeikampen, Norway, June 15-28 2016*. Attended and completed.

Preface and Acknowledgements

I would like to thank my parents, my family, my partner, my dear friends and my (former) band mates for their endless support during my entire journey towards obtaining a PhD in physics. I would like to thank my supervisor Meenakshi Narain for her guidance and for the opportunity to work with her for the past 5 years. Thank you to the readers, David Cutts, Ulrich Heintz. Thank you also to my previous mentors, colleagues, friends at the University of Sydney, Brown University, Fermilab, CERN, all of whom have helped me pave the way to where I am today.

I have moved across 5 cities in 5 different countries in the past decade in order to pursue this PhD. It was a goal almost 8 years in the making. It has not always been smooth sailing. In fact, it has been long and often hard, but it has been absolutely gratifying. I am ever so grateful to everyone and everything that has helped made this journey possible.

And to Mother Nature, thank you for your random coincidences that have aligned in my favor up to this day when I completed this PhD, but most of all thank you for your wonderful mysteries.

Contents

List of Tables	xv
List of Figures	xx
1 Introduction	1
2 The standard model theory of particle physics	3
2.1 Fundamental particles and forces of nature	4
2.2 Quantum field theory	5
2.3 Quantum chromodynamics	7
2.4 Electroweak theory	9
3 Beyond the standard model and vector-like quarks (VLQ)	15
3.1 The hierarchy problem	16
3.2 Vector-like quarks	16
4 The Compact Muon Solenoid (CMS)	19
4.1 The Large Hadron Collider (LHC)	19
4.2 The CMS detector	22
4.2.1 Coordinate system	22
4.2.2 The pixel detector	22
4.2.3 The silicon strip tracker	24
4.2.4 The electromagnetic calorimeter (ECAL)	28
4.2.5 The hadron calorimeter (HCAL)	29

4.2.6	The muon detector	30
4.3	Trigger systems in CMS	32
4.4	Particle identification and physics objects reconstruction	33
4.4.1	Particle-flow algorithm	35
4.4.2	Track reconstruction	35
4.4.3	Muon reconstruction	35
4.4.4	Electron reconstruction	36
4.4.5	Hadrons and jets	37
4.4.6	Missing transverse energy reconstruction	38
4.4.7	Primary vertex reconstruction	38
4.4.8	Identification of b-quark jets	39
5	Search for vector-like-quarks decaying to leptons and jets final states	42
5.1	Trilepton final state	44
5.2	Data and simulated samples	45
5.3	Physics objects reconstruction	48
5.3.1	Leptons	48
5.3.2	Jets and Missing transverse energy	51
5.4	Event selection	52
5.4.1	HLT Triggers	52
5.4.2	Offline event selections	53
5.4.3	Control Region and Signal Region	53
5.5	Background Estimation	55
5.5.1	Matrix method for trilepton nonprompt background estimation	56
5.5.2	Prompt rates	57
5.5.3	Fake rates	57
5.6	Systematic uncertainties	86
5.7	Kinematics distributions after selection	87
5.7.1	Kinematics distributions in control region	88
5.7.2	Kinematics distributions in signal region	90

5.8	Final event selection and limits	93
5.9	Same-sign dilepton final state	97
5.9.1	Kinematic plots	101
5.9.2	Results	111
5.10	Combined results: trilepton, SS dilepton, single-lepton	117
6	The XCone jet clustering algorithm and the search for VLQ $T\bar{T} / B\bar{B} \rightarrow e/\mu + \text{jets}$	121
6.1	The XCone Jet Algorithm	121
6.1.1	N -Jettiness and XCone	121
6.1.2	Samples and Event Selections	123
6.1.3	Determining Event Shape	123
6.1.4	Results	128
6.1.5	Summary: The XCone jet algorithm	131
6.2	Validation using 2016 CMS data	131
6.2.1	N -jettiness in Drell-Yan events	131
6.2.2	N -jettiness in semi leptonic $t\bar{t}$	135
6.2.3	Public results for the BOOST 2017 conference	140
6.2.4	Summary: validation in data	142
6.3	Utilizing XCone jets in the search of $T\bar{T} / B\bar{B} \rightarrow e/\mu + \text{jets}$	143
6.3.1	Exploring observables based on XCone jets	144
6.3.2	Optimizing selections using XCone jets multiplicity	154
6.3.3	Summary and discussion: XCone jets in $T\bar{T} / B\bar{B} \rightarrow \text{single lepton}$	163
7	Summary and conclusions	166
A	Identification of high momentum Higgs decaying to b-quark pairs (boosted double-b tagging)	168
A.1	Introduction	168
A.2	Event samples, reconstruction and wide jet identification	169
A.3	Double-b tagger algorithm	171
A.3.1	Discriminating variables	171

A.4	Efficiency measurement in 2015 data	175
B	Shower Deconstruction algorithm as a high momentum $H \rightarrow b\bar{b}$ tagger	181
B.1	Introduction	181
B.2	Shower Deconstruction	183
B.2.1	Sanity Check (using 8 TeV samples and Anti- k_T R=1.2 fatjets)	184
B.2.2	Microjet b-tagging	185
B.3	Efficiency (using 8 TeV samples and Anti- k_T R=1.2 fatjets)	188
B.3.1	Number of microjets and b-tagged microjets	191
B.4	Efficiency (using 13 TeV samples, AK8 fatjets)	192
B.4.1	SD efficiency using selected fatjets	194
B.5	Microjet b-tag rate and fake rate	194
B.5.1	Tag rate and fake rate using k_T microjets	196
B.5.2	Comparing k_T , AK and CA microjets	200
B.6	Optimizing performance	201
B.6.1	Microjet b-tag rate	201
B.6.2	Microjet b-tag fake rate	202
B.6.3	Higgs mass window	202
B.6.4	Microjet cone size	203
B.6.5	Discussion on discrimination between different backgrounds	206
B.7	Mass distribution shaping	206
B.8	Performance comparison	207
B.9	Conclusion	213
C	The Cosmic Rack	214
C.1	Hardware identification maps and visual mapping	216
C.2	Cluster latency measurement	219
	Bibliography	221

List of Tables

2.1	Electroweak quantum numbers of the fundamental fermions (where the left-handed and right-handed have different quantum numbers) and the fundamental bosons. The electromagnetic charge (Q), the third component of the weak isospin I_3 and the weak hypercharge Y obey the Gell-Mann-Nishijima relation $Q = I_3 + \frac{Y}{2}$	6
3.1	Allowed representations of VLQ, in quantum numbers under $SU(2)_L$ and $U(1)_Y$ in comparison to SM quarks. The subscript L denotes the left hand component and the subscript R denotes the right hand component. The superscript in $H^{(c)}$ denotes that it could either be H or H^c depending on the representation chosen. The q denotes the SM quark flavors, u (d) denotes the up(down)-type SM quarks, whereas T, B, X, Y denotes the VLQs. In the Lagrangian rows, Q_L, u_R, d_R denotes the left and right handed SM quarks (as defined in Section 2.4), whereas T_R, B_R, ψ denotes the VLQs, λ denotes the Yukawa coupling involving the VLQs, and σ^a are Pauli matrices. [1, 2].	18
5.1	Signal efficiencies in the single-lepton, same-sign dilepton, and trilepton channels, split into the six possible final states of both $T\bar{T}$ and $B\bar{B}$ production, for three mass points. Efficiencies, stated in percent, are final yields calculated with respect to the expected number of events in the corresponding decay mode, before any selection. The most sensitive decay modes for each channel are noted in bold. The efficiency for bWbW events in the same-sign dilepton and trilepton channels is negligible, as is the efficiency for bZbZ events in the same-sign dilepton channel. [3]	44
5.2	Data sample definitions.	46

5.3	Background MC sample definitions, from RunIISummer16MiniAODv2 campaign. Uncertainties include contributions from energy scale variations and PDF uncertainties.	47
5.4	$T\bar{T}$ signal MC samples at various mass points from RunIISummer16MiniAODv2 campaign. These samples were generated with inclusive decays of the $T\bar{T}$ with equal branching ratios (33%) for $T \rightarrow tH, tZ,$ and bW . Uncertainties include contributions from energy scale variations and PDF uncertainties.	47
5.5	Minimum MVA values for the 88% and 95% electron MVA discriminant working points (WP).	49
5.6	HLT trigger paths used for the trilepton analysis.	53
5.7	Monte Carlo background prediction in percentages. In the table, V stands for W or Z vector boson. In addition we also show the signal ($T\bar{T}_{M1000}$) significance in each regions.	54
5.8	electron prompt rates	57
5.9	Number of events, in the control region with 2 jets and exactly 3 leptons using the optimal fakerates that corresponds the minimum average χ^2 point (shown in 'CR2' column of Table?). The uncertainties includes both statistical and systematic.	61
5.10	Number of events, in control region with 2 jets (and exactly 3 leptons) using the fakerates measured in control region with 1 jet (shown in 'CR1' column of Table?(a)). The uncertainties includes both statistical and systematic.	68
5.11	Fake rate measurement using MC truth in 1 or more leptons events in $t\bar{t}$	78
5.12	Fake rate measurement using MC truth in 3 leptons events in $t\bar{t}$	78
5.13	Fake rate measurement closure test using MC truth in 3 leptons events in $t\bar{t}$, in the four trilepton channels where two leptons are prompt (p) and one is nonprompt/fake (f) but all pass the tight ID (t).	80
5.14	(a): Lepton fake rates measured in data in control region with 2 jets (CR2), in control region with 1 jet (CR1) and their differences. The uncertainties are statistical. (b): Deviation of yields between observed and predicted number of prompt-prompt-nonprompt $t\bar{t}$ MC trilepton events where all leptons pass the tight ID, considering all sources of nonprompt leptons. (c): Fake rate measurement using the χ^2 minimization in trilepton events using $t\bar{t}$ powheg sample compared to the fake rate measurement by its MC truth. (d): Fake rate measurement using the χ^2 minimization of the matrix method in trilepton events using $t\bar{t}$ powheg sample requiring CR2 and SR selections.	84

5.15	Summary of uncertainties due to the fake rate measurement and the matrix method.	85
5.16	Summary of the systematic uncertainties on Monte Carlo samples	86
5.17	Summary of the uncertainties on the data driven non prompt background. (*) Uncertainties are one-sided in the positive direction.	86
5.18	Number of events after selection requirements for the trilepton final state, in control region. The uncertainties includes both statistical and systematic.	89
5.19	Number of events after selection requirements for the trilepton final state, in signal region. The uncertainties include both statistical and systematic.	90
5.20	Number of events with at least 1 loose lepton after the selection requirements in the signal region. These are inputs to construct the data-driven background. The uncertainties include both statistical and systematic. Note: ddbkg is the resulting data-driven background yield and the N's are counts events in data.	91
5.21	Signal efficiencies, in percentages, after selection.	92
5.22	HLT trigger paths used for the trilepton analysis.	98
5.23	Background yields for requiring two same-sign tight leptons and after restricting only 2 leptons in the event.	99
5.24	SS dilepton signal to background sensitivity measurement.	99
5.25	Systematic uncertainties, in percentages. Signal PDF and scale uncertainties are acceptance only. The values in the table are the maximum shifts from the nominal yield.	100
5.26	Summary of values for normalization uncertainties and dependencies for shape uncertainties. The symbol σ denotes one standard deviation of the uncertainty and “env” denotes an envelope of values. Background from opposite-sign dilepton events is denoted “OS”, background from nonprompt leptons is denoted “NP”, while other backgrounds modeled from simulation are denoted “MC”. For signals, theoretical uncertainties are labeled as “Shape” for shape-based searches, and “Accept.” for counting experiments. Additionally, “CR” denotes control region and “RMS” denotes root mean square. [3]	118
6.1	MC samples used in this study.	123
6.2	MC samples	135

6.3	Comparison of VLQ T mass upper limits (in GeV) using various different discriminators. The limits are based on the 95% CL upper limits on the production cross section of $T\bar{T}$	151
6.4	Baseline expected 95% CL upper limits (pb) for the production cross section of $T\bar{T}$, only considering luminosity, lepton identification, lepton isolation, lepton reconstruction, and QCD renormalization scale uncertainties.	154
6.5	Baseline expected 95% CL upper limits (pb) for the production cross section of $B\bar{B}$, only considering luminosity, lepton identification, lepton isolation, lepton reconstruction, and QCD renormalization scale uncertainties.	158
6.6	Expected 95% CL upper limits for the mass of T (GeV) with XCone jet multiplicity requirements. The '*' labels the most sensitive limits for the particular signal decay scenario.	161
6.7	Observed 95% CL upper limits for the mass of T (GeV) with XCone jet multiplicity requirements. The '*' labels the most sensitive (expected) limits for the particular signal decay scenario.	161
6.8	Expected 95% CL upper limits for the mass of B (GeV) with XCone jet multiplicity requirements. The '*' labels the most sensitive limits for the particular signal decay scenario.	163
6.9	Observed 95% CL upper limits for the mass of B (GeV) with XCone jet multiplicity requirements. The '*' labels the most sensitive (expected) limits for the particular signal decay scenario.	163
A.1	Loose double-b tag efficiency (ϵ) and Data/MC efficiency ratio (SF). Uncertainties are both statistical and systematic for the SF and data efficiency, while for the MC efficiency only the statistical uncertainty is reported. Jets with $p_T > 700$ GeV are included in the last bin.	178
A.2	Medium double-b tag efficiency (ϵ) and Data/MC efficiency ratio (SF). Uncertainties are both statistical and systematic for the SF and data efficiency, while for the MC efficiency only the statistical uncertainty is reported. Jets with $p_T > 700$ GeV are included in the last bin.	178
A.3	Tight double-b tag efficiency (ϵ) and Data/MC efficiency ratio (SF). Uncertainties are both statistical and systematic for the SF and data efficiency, while for the MC efficiency only the statistical uncertainty is reported. Jets with $p_T > 500$ GeV are included in the last bin.	180
B.1	8 TeV simulated samples used	182
B.2	13 TeV simulated samples used	182

B.3	The tag rate calculation results.	200
B.4	The fake rate and rate of IVF SV found in the microjets.	200
B.5	Configurations varied for optimization, and values looked at in this study. The values underlined are the observed to be the optimal values. For the microjet cone size the optimal value depends on the background.	206
B.6	QCD inclusive background rejection (%) for SD, τ_2/τ_1 and q-jet volatility using working points with signal efficiencies being, $\varepsilon_s = 90\%$, $\varepsilon_s = 70\%$, and $\varepsilon_s = 50\%$. The values are obtained based on the plots in Figure B.34 and Figure B.35.	210
B.7	$Z' \rightarrow t\bar{t}$ background rejection (%) for SD, τ_2/τ_1 and q-jet volatility using working points with signal efficiencies being $\varepsilon_s = 90\%$, $\varepsilon_s = 70\%$, and $\varepsilon_s = 50\%$. The values are obtained based on the plots in Figure B.34 and Figure B.35.	210
C.1	State of the c-rack on May 23rd, 2018.	218

List of Figures

2.1	Summary of the standard model particles. [4]	6
2.2	Sketch of the Higgs potential appearing in \mathcal{L}_H [5].	11
4.1	Schematic layout of CERN accelerators (top) and an artist's view of of the LHC (bottom) [6].	21
4.2	LHC cryodipole [7, 8].	21
4.3	Schematic diagram of the CMS detector [9].	23
4.4	Left: Map of the magnetic field. Right: Field lines (right) predicted for a longitudinal section of the CMS detector by a magnetic field model at a central magnetic flux density of 3.8 T. Each field line represents a magnetic flux increment of 6 Wb. [10].	23
4.5	Illustrations of coordinates used by the CMS experiment. [11].	24
4.6	Geometrical layouts of the CMS pixel detector [12].	25
4.7	Exploded view of the barrel pixel detector showing the parts of the module, where SDM is the surface mounted device, TBM is the token bit manager chip which controls several read-out chips (ROCs), and HDI is the high density interconnect. [12].	25
4.8	Schematic cross section through the top half of the CMS tracker in the r-z plane. The star at the center indicates the approximate position of the pp collision. Green dashed lines are imaginary lines that partitions the tracker into the four subsystems. Thin black lines indicate strip tracker modules that provide 2-D hits and thick blue lines indicate strip tracker modules that provide reconstruction of hit positions in 3-D, by having two back-to-back strip modules where one module is rotated by a 'stereo' angle. The red lines indicate the pixel detector modules which also provide 3-D hits. [13].	27

4.9	Left: Diagram of silicon strip tracker module housing two sensors. Right: photograph of a TEC subsystem sensor module, mounted on a carrier plate. [12].	27
4.10	Total thickness t of the CMS pixel and strip tracker material expressed in units of interaction lengths λ_l (left) and radiation lengths X_0 (right), as a function of the pseudorapidity η . [13].	28
4.11	Schematic view of the CMS electromagnetic calorimeter [14].	29
4.12	ECAL energy resolution as a function of electron energy. [12].	30
4.13	A schematic drawing of one quadrant of CMS showing the HCAL and the locations of the HB, HE, HO, and HF [15].	31
4.14	A schematic drawing of one quadrant of CMS showing the muon system [16].	32
4.15	Sketch of a transverse slice the CMS detector with illustrations of how various particles would traverse and leave energy deposits in the detector [17].	34
4.16	Muon transverse-momentum resolutions as a function of the muon transverse momentum measured using the CMS tracker. [12].	36
4.17	The resolutions of the magnitude of the jet transverse momentum vector ($\sigma(E_T)/E_T$), as a function of the magnitude (E_T) in various $ \eta $ regions, measured using the CMS calorimeters. [12].	38
4.18	Illustration of a jet originating from a b-quark, labeled as a heavy-flavored jet, with a secondary vertex (SV) from the decay of a B hadron. The charged-particle tracks (including possibly a soft lepton) are displaced with respect to the primary vertex, resulting in a large impact parameter (IP) value. [18].	40
5.1	Feynman diagrams showing pair production of $T\bar{T}$ (left) and $B\bar{B}$ (right) decaying to SM particles. [3]	43
5.2	Scale factors for electron in bins of η and p_T for MVA tight (custom) working point (top) and miniIsolation < 0.1 (bottom).	50
5.3	Jet multiplicity distribution using only Monte Carlo to simulate background, after selection (omitting jet multiplicity requirement). Only statistical + lumi + lepton Iso + lepton ID uncertainties are included in the plot.	54
5.4	χ^2 values for various electron and muon fake rate values. The intersection of the horizontal and vertical lines indicate the minimum value.	59

5.5	2D Gaussian probability distribution for the electron and muon fake rate values. The intersection of the horizontal and vertical lines indicate the most probable value.	59
5.6	Gaussian probability distribution for each fake rate after marginalization.	60
5.7	Lepton p_T distributions with the optimal fakerates that corresponds the minimum average χ^2 , in the four different trilepton categories, in the control region with 2 jets with exactly 3 leptons. (Statistical and systematic uncertainties are included in the plots)	60
5.8	Lepton η distributions with the optimal fakerates that corresponds the minimum average χ^2 , in the four different trilepton categories, in the control region with 2 jets with exactly 3 leptons. (Statistical and systematic uncertainties are included in the plots)	61
5.9	Lepton multiplicity distributions with the optimal fakerates that corresponds the minimum average χ^2 , in the four different trilepton categories, in the control region with 2 jets with exactly 3 leptons. (Statistical and systematic uncertainties are included in the plots)	62
5.10	Electron multiplicity distributions with the optimal fakerates that corresponds the minimum average χ^2 , in the four different trilepton categories, in the control region with 2 jets with exactly 3 leptons. (Statistical and systematic uncertainties are included in the plots)	63
5.11	μ multiplicity distributions with the optimal fakerates that corresponds the minimum average χ^2 , in the four different trilepton categories, in the control region with 2 jets with exactly 3 leptons. (Statistical and systematic uncertainties are included in the plots)	64
5.12	Highest p_T lepton p_T distributions with the optimal fakerates that corresponds the minimum average χ^2 , in the four different trilepton categories, in the control region with 2 jets with exactly 3 leptons. (Statistical and systematic uncertainties are included in the plots)	65
5.13	2nd highest p_T lepton p_T distributions with the optimal fakerates that corresponds the minimum average χ^2 , in the four different trilepton categories, in the control region with 2 jets with exactly 3 leptons. (Statistical and systematic uncertainties are included in the plots)	66
5.14	3rd highest p_T lepton p_T distributions with the optimal fakerates that corresponds the minimum average χ^2 , in the four different trilepton categories, in the control region with 2 jets with exactly 3 leptons. (Statistical and systematic uncertainties are included in the plots)	67
5.15	χ^2 scan for various electron and muon fake rate values as a cross check in control region with exactly 1 jet. The intersection of the horizontal and vertical lines indicate the minimum value.	68

5.16	Lepton p_T distributions, in control region with 2 jets and exactly 3 leptons, with the fakerates that corresponds the minimum average χ^2 measured in control region with 1 jet (column 'CR1' in Table 5.14(a)), of the four different trilepton categories. (Statistical and systematic uncertainties are included in the plots)	69
5.17	Lepton η distributions, in control region with 2 jets and exactly 3 leptons, with the fakerates that corresponds the minimum average χ^2 measured in control region with 1 jet (column 'CR1' in Table 5.14(a)), of the four different trilepton categories. (Statistical and systematic uncertainties are included in the plots)	70
5.18	Lepton multiplicity distributions, in control region with 2 jets and exactly 3 leptons, with the fakerates that corresponds the minimum average χ^2 measured in control region with 1 jet (column 'CR1' in Table 5.14(a)), of the four different trilepton categories. (Statistical and systematic uncertainties are included in the plots)	71
5.19	Electron multiplicity distributions, in control region with 2 jets and exactly 3 leptons, with the fakerates that corresponds the minimum average χ^2 measured in control region with 1 jet (column 'CR1' in Table 5.14(a)), of the four different trilepton categories. (Statistical and systematic uncertainties are included in the plots)	72
5.20	μ multiplicity distributions, in control region with 2 jets and exactly 3 leptons, with the fakerates that corresponds the minimum average χ^2 measured in control region with 1 jet (column 'CR1' in Table 5.14(a)), of the four different trilepton categories. (Statistical and systematic uncertainties are included in the plots)	73
5.21	Highest (in p_T) lepton p_T distributions, in control region with 2 jets and exactly 3 leptons, with the fakerates that corresponds the minimum average χ^2 measured in control region with 1 jet (column 'CR1' in Table 5.14(a)), of the four different trilepton categories. (Statistical and systematic uncertainties are included in the plots)	74
5.22	2nd highest (in p_T) lepton p_T distributions, in control region with 2 jets and exactly 3 leptons, with the fakerates that corresponds the minimum average χ^2 measured in control region with 1 jet (column 'CR1' in Table 5.14(a)), of the four different trilepton categories. (Statistical and systematic uncertainties are included in the plots)	75

5.23	3rd highest (in p_T) lepton p_T distributions, in control region with 2 jets and exactly 3 leptons, with the fakerates that corresponds the minimum average χ^2 measured in control region with 1 jet (column 'CR1' in Table 5.14(a)), of the four different tripleton categories. (Statistical and systematic uncertainties are included in the plots)	76
5.24	HT distributions from the fake rate measurement closure test using MC truth in 3 leptons events in $t\bar{t}$, in the four tripleton channels where two leptons are prompt (p) and one is nonprompt/fake (f) but all pass the tight ID (t).	79
5.25	χ^2 scan for various electron and muon fake rate values using $t\bar{t}$ MC. The intersection of the horizontal and vertical lines indicate the minimum value.	82
5.26	Lepton p_T distributions with the fake rates that corresponds the minimum χ^2 , in the four different tripleton categories, using $t\bar{t}$ powheg sample. 'Observed' events are tripleton events that pass the tight ID and 'Predicted' events are the result of using the matrix method.	82
5.27	Predicted background distributions of AK4 Jet kinematic variables in the tripletonic final state after selection requirements are applied, in the control region. Top left: AK4 Jet p_T distribution. Top right: AK4 Jet η distribution. Botton left: Number of AK4 Jets. Bottom right: Number of b-tagged AK4 Jets	88
5.28	Predicted background distributions of kinematic variables in the tripletonic final state after selection requirements are applied, in the control region. Top left: Lepton p_T distribution. Top right: Lepton η distribution. Botton left: p_T^{miss} distribution. Bottom right: S_T distribution.	89
5.29	Predicted background distributions of AK4 Jet kinematic variables in the tripletonic final state after selection requirements are applied, in the signal region. Top left: AK4 Jet p_T distribution. Top right: AK4 Jet η distribution. Botton left: Number of AK4 Jets. Bottom right: Number of b-tagged AK4 Jets	91
5.30	Predicted background distributions of kinematic variables in the tripletonic final state after selection requirements are applied, in the signal region. Top left: Lepton p_T distribution. Top right: Lepton η distribution. Botton left: p_T^{miss} distribution. Bottom right: S_T distribution.	92
5.31	Predicted background distributions of final yield in the tripletonic final state after final selection requirements are applied, used as template to generate the limits. The lower panel shows the difference between data and background divided by the total uncertainty. [3]	94

5.32	trilepton final state cross-section limits (with singlet branching fractions: 50% bW, 25% tH, 25% tZ), shown as a function of mass of T.	94
5.33	trilepton final state cross-section limits with the doublet branching fractions: 0% bW, 50% tH, 50% tZ (left) and with the branching fractions: 0% bW, 0% tH, 1000% tZ (right), shown as a function of mass of T.	95
5.34	Distributions of ST in the four flavor channels after the limit calculation fit. The uncertainties plotted include all uncertainties.	95
5.35	Post-fit pulls of the nuisance parameters (top) and the their correlations (bottom).	96
5.36	Distribution of leading Lepton p_T for the combined flavor channel (All) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty.101	101
5.37	Distribution of sub-leading Lepton p_T for the combined flavor channel (All) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty.	102
5.38	Distribution of leading lepton η for the combined flavor channel (All) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty.103	103
5.39	Distribution of Sub-leading lepton η for the combined flavor channel (All) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty.104	104
5.40	Distribution of leading Jet p_T for the combined flavor channel (All) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty.105	105
5.41	Distribution of sub-leading Jet p_T for the combined flavor channel (All) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty.106	106
5.42	Distribution of leading Jet η for the combined flavor channel (All) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty.107	107

5.43	Distribution of sub-leading Jet η for the combined flavor channel (All) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty.	108
5.44	Distribution of nConst for the combined flavor channel (All) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty.	109
5.45	Distribution of H_T^{lep} for the combined flavor channel (All) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty. [3]	110
5.46	Summary of background yields and observed data after the full analysis selection for the 2016B-H datasets. Also shown are the number of expected events for a 1000 GeV VLQ T . . .	112
5.47	Expected Number of Background Events vs. Analysis Cut for all channels combined, where NP is nonprompt and CMID is charge misidentified SS dilepton events.	112
5.48	Expected Number of Signal Events vs. Analysis Cut for all channels combined	112
5.49	Background rejection after applying SSDL and exactly two leptons requirements for all channels combined.	113
5.50	Signal efficiency after applying SSDL and exactly two leptons requirements for all channels combined	113
5.51	SS dilepton final state cross-section limits (top: with singlet branching fractions: 50% bW, 25% tH, 25% tZ), bottom: with doublet branching fractions: 0% bW, 50% tH, 50% tZ), shown as a function of mass of T	114
5.52	SS dilepton final state cross-section limits (top: with singlet branching fractions: 50% tW, 25% bH, 25% bZ), bottom: with doublet branching fractions: 0% tW, 50% bH, 50% bZ), shown as a function of mass of B.	115
5.53	Postfit and correlation plots. The top right postfit plot corresponds to an alternate limit calculation setting where the electron and muon fake rate nuisances are kept separate.	116
5.54	The 95% CL expected and observed upper limits on the cross section of $T\bar{T}$ (upper row) and $B\bar{B}$ (lower row) production after combining all channels for the singlet (left) and doublet (right) branching fraction scenarios. The predicted cross sections are shown by the red curve, with the uncertainty indicated by the width of the line. [3]	119

5.55	The 95% CL expected (left) and observed (right) lower limits on the T quark (upper row) and B quark (lower row) mass, expressed in GeV, after combining all channels for various branching fraction scenarios. [3]	120
6.1	N -jettiness definition.	122
6.2	XCone definitions and parameters used in this study. [19]	122
6.3	N -jettiness plots for N from 0 to 19 with labels that shows the number of quarks (q), gluons (g), partons (p).	124
6.4	$\mathcal{T}_N - \mathcal{T}_{N-1}$, ($\Delta\mathcal{T}_N$), plots for N from 2 to 19 with labels that shows the number of quarks (q), gluons (g), partons (p).	124
6.5	$\frac{\mathcal{T}_N}{\mathcal{T}_{N-1}}$ plots for N from 2 to 19 with labels that shows the number of quarks (q), gluons (g), partons (p).	124
6.6	$\frac{\mathcal{T}_N - \mathcal{T}_{N-1}}{\mathcal{T}_N}$ plots for N from 2 to 19 with labels that shows the number of quarks (q), gluons (g), partons (p).	125
6.7	AK4 multiplicities in $t\bar{t}$ and $T'T'$	126
6.8	Optimal N in $t\bar{t}$ and $T'T'$ based on a forward, (a)-(c), and reverse, (d)-(f), scanning the $\mathcal{T}_N - \mathcal{T}_{N-1}$ distribution.	126
6.9	Optimal N in $t\bar{t}$ and $T'T'$ based on forward, (a)-(c), and reverse, (d)-(f), scanning the $\frac{\mathcal{T}_N}{\mathcal{T}_{N-1}}$ distribution.	127
6.10	Optimal N in $t\bar{t}$ and $T'T'$ based on forward, (a)-(c), and reverse, (d)-(f), scanning the $\frac{\mathcal{T}_N - \mathcal{T}_{N-1}}{\mathcal{T}_N}$ distribution.	127
6.11	In some rare cases when the threshold (red line) is not broken (left), the threshold is lowered (right) and the vertical lines shows how forward (blue line) and reverse (black line) scanning give different result.	128
6.12	ROC curves comparing various methods to distinguish 6 prong $t\bar{t}$ events from 10-prong $T'T' \rightarrow tHtH$ events. The higher the T' mass the more collimated the jets are.	128
6.13	Jet p_T distribution (GeV) of AK4 and XCone jets in various samples.	129
6.14	XCone multiplicities in $t\bar{t}$ and $T'T'$	129
6.15	ROC curves comparing the performance of XCone and AK4 to distinguish 6-prong $t\bar{t}$ events from 10-prong $T'T' \rightarrow tHtH$ events. The higher the T' mass the more collimated the jets are. 130	130

6.16	$M_{\mu\mu}$ distribution.	132
6.17	Various \mathcal{T}_N plots and distributions for Drell-Yan events. The left most plot in (b) shows the mean of the distribution and the vertical bands are the standard deviation of \mathcal{T}_N distribution for a particular N	133
6.18	Various $\mathcal{T}_N - \mathcal{T}_{N-1}$ plots and distributions for Drell-Yan events. The left most plot in (b) shows the mean of the distribution and the vertical bands are the standard deviation of $\mathcal{T}_N - \mathcal{T}_{N-1}$ distribution for a particular N	134
6.19	Jet multiplicity distribution.	134
6.20	Kinematics distribution.	135
6.21	Jet mass distributions distribution with AK8 $p_T > 200$ Gev.	136
6.22	Jet mass distributions distribution with AK8 $p_T > 400$ Gev.	137
6.23	Jet mass distributions distribution with AK8 $p_T > 600$ Gev.	137
6.24	Jet mass distributions distribution with AK8 $p_T > 800$ Gev.	137
6.25	Various \mathcal{T}_N plots and distributions for $t\bar{t}$ events. The left most plot in (b) shows the mean of the distribution and the vertical bands are the standard deviation of \mathcal{T}_N distribution for a particular N	138
6.26	Various $\mathcal{T}_N - \mathcal{T}_{N-1}$ plots and distributions for $t\bar{t}$ events. The left most plot in (b) shows the mean of the distribution and the vertical bands are the standard deviation of $\mathcal{T}_N - \mathcal{T}_{N-1}$ distribution for a particular N	139
6.27	Jet multiplicity distribution.	139
6.28	Kinematics distribution.	140
6.29	Distributions of \mathcal{T}_N and $\mathcal{T}_N - \mathcal{T}_{N-1}$, ($\Delta\mathcal{T}_N$), in semileptonic standard model $t\bar{t}$ MC plotted over a range of jet multiplicity for N from 0 to 9, and where the y-axis mean value is also shown.	140
6.30	Distributions of Jet Multiplicities.	141
6.31	Kinematics distributions of AK4 and XCone jets: p_T (left) and HT (right).	141
6.32	Comparisons of top mass reconstruction using AK8, AK4 and XCone jets.	142
6.33	Comparison of XCone based observables in $t\bar{t}$ (in red) and $T\bar{T} \rightarrow bWbW$ (in black), where T has a mass of 1.0 GeV and preselection cuts are applied. S_T is labelled "AK4pHTpMETpLeptonPt", $\min[M(l, b)]$ is labelled "minMleppBjet", AK4 jets multiplicity is labelled "NJets_JetSubCalc", and XCone jets multiplicity is labelled "NXConeJets".	146

6.34	Comparison of XCone based observables in $t\bar{t}$ (in red) and $T\bar{T} \rightarrow tZtZ$ (in black), where T has a mass of 1.0 GeV and preselection cuts are applied. S_T is labelled "AK4pHTpMETpLeptonPt", $\min[M(l, b)]$ is labelled "minMleppBjet", AK4 jets multiplicity is labelled "NJets_JetSubCalc", and XCone jets multiplicity is labelled "NXConeJets".	146
6.35	Comparison of XCone based observables in $t\bar{t}$ (in red) and $T\bar{T} \rightarrow tHtH$ (in black), where T has a mass of 1.0 GeV and preselection cuts are applied. S_T is labelled "AK4pHTpMETpLeptonPt", $\min[M(l, b)]$ is labelled "minMleppBjet", AK4 jets multiplicity is labelled "NJets_JetSubCalc", and XCone jets multiplicity is labelled "NXConeJets".	147
6.36	ROC curves for observables, in samples of $t\bar{t}$ and $T\bar{T} \rightarrow bWbW$, where T has a mass of 1.0 GeV. S_T is labelled "AK4pHTpMETpLeptonPt", $\min[M(l, b)]$ is labelled "minMleppBjet", AK4 jets multiplicity is labelled "NJets_JetSubCalc", and XCone jets multiplicity is labelled "NXConeJets".	148
6.37	ROC curves for observables, in samples of $t\bar{t}$ and $T\bar{T} \rightarrow tZtZ$, where T has a mass of 1.0 GeV. S_T is labelled "AK4pHTpMETpLeptonPt", $\min[M(l, b)]$ is labelled "minMleppBjet", AK4 jets multiplicity is labelled "NJets_JetSubCalc", and XCone jets multiplicity is labelled "NXConeJets".	149
6.38	ROC curves for observables, in samples of $t\bar{t}$ and $T\bar{T} \rightarrow tHtH$, where T has a mass of 1.0 GeV. S_T is labelled "AK4pHTpMETpLeptonPt", $\min[M(l, b)]$ is labelled "minMleppBjet", AK4 jets multiplicity is labelled "NJets_JetSubCalc", and XCone jets multiplicity is labelled "NXConeJets".	150
6.39	Distributions of various observables based on pseudo b-tagged XCone jets, comparing $t\bar{t}$ (in red) and $T\bar{T} \rightarrow bWbW$ (in black), where T has a mass of 1.0 GeV, and only the preselection cuts are applied. (a) <i>MassWcand1BtagXCone</i> , (b) <i>MassWcand2BtagXCone</i> , (c) <i>Mass1XCone123</i> , (d) <i>Mass2XCone123</i>	153
6.40	Distributions of the <i>M_WcandBTagXCone_XCone123</i> observable, comparing $t\bar{t}$ (in red) and $T\bar{T} \rightarrow bWbW$ (in black), where T has a mass of (a) 1.0 GeV, (b) 1.4 GeV, (c) 1.8 GeV. Only the preselection cuts are applied.	153
6.41	ROC curves for various observables, in samples of $t\bar{t}$ and $T\bar{T} \rightarrow bWbW$, where T has a mass of 1.0 GeV.	155

6.42	ROC curves for various observables, in samples of $t\bar{t}$ and $T\bar{T} \rightarrow bWbW$, where T has a mass of 1.4 GeV.	156
6.43	ROC curves for various observables, in samples of $t\bar{t}$ and $T\bar{T} \rightarrow bWbW$, where T has a mass of 1.8 GeV.	157
6.44	(a) Shows the optimal number of X Cone jet multiplicity selection using the $\min[M(l, b)]$, S_T discriminants in search for $T\bar{T}$, and (b) shows the corresponding limits improvement in comparison to the baseline limit shown in Table 6.4	158
6.45	(a) Shows the optimal number of X Cone jet multiplicity selection using only the S_T discriminant in search for $T\bar{T}$, and (b) shows the limits improvement in comparison to the baseline limit shown in Table 6.4	159
6.46	(a) Shows the optimal number of X Cone jet multiplicity selection using the $\min[M(l, b)]$, S_T discriminants in search for $B\bar{B}$, and (b) shows the corresponding limits improvement in comparison to the baseline limit shown in Table 6.5	159
6.47	(a) Shows the optimal number of X Cone jet multiplicity selection using only the S_T discriminant in search for $B\bar{B}$, and (b) shows the limits improvement in comparison to the baseline limit shown in Table 6.4	160
6.48	The 95% CL upper limits for the mass of T (GeV), which corresponds to the optimal choice of X Cone jet multiplicity requirement as labelled ‘*’ in Tables 6.6 and 6.7	162
6.49	The 95% CL upper limits for the mass of B (GeV), which corresponds to the optimal choice of X Cone jet multiplicity requirement as labelled ‘*’ in Tables 6.8 and 6.9	164
A.1	Distributions of 2D IP significance for the most displaced track raising the SV invariant mass above the bottom quark threshold, number of secondary vertices associated to the AK8 jet, the vertex energy ratio for SV_0 , and the z variable. Comparison between $H \rightarrow b\bar{b}$ jets from simulated samples of KK-Graviton decaying to HH and QCD jets containing zero, one or two b quarks are used. AK8 jets are selected with $p_T > 300\text{GeV}$ and pruned jet mass $70 < m < 200\text{GeV}$. The distributions are normalized to unit area. [20]	174

A.2	Distributions of 2D IP significance for the most displaced track raising the SV invariant mass above the b quark threshold, number of secondary vertices associated to the AK8 jet, the vertex energy ratio for SV_0 , and the z variable. Data and simulated events are shown for the double-muon tagged jets selection. Simulated events are normalized to the yield observed in data, the overflow is in last bin. The bottom panel in each figure shows the ratio of the number of events observed in data to that of the MC prediction. [20]	176
A.3	Double-b tagger discriminant distribution in data and simulated samples for the double-muon tagged jets selection. Simulated events are normalized to the yield observed in data. The loose, medium and tight operating points are also reported. The bottom panel shows the ratio of the number of events observed in data to that of the MC prediction. [20]	176
A.4	Comparison of the JP discriminant distribution for the data and the sum of the fitted templates for all selected jets (left) and those jets passing the loose double-b tagger requirement (right) with p_T between 500 and 600 GeV. The shaded area represents the statistical and systematic (refer to the text for details) uncertainties on MC templates. Double-muon tagged AK8 jets are used for this measurement. The overflow is included in the last bin. [20]	177
A.5	Data/MC efficiency ratio (SF) for loose, medium and tight double-b tagger requirement obtained with single and double-muon tagged selections. Central values of scale factors are artificially shifted along the x-axis for better visibility. [20]	179
B.1	p_T (left) and mass (right) distributions of signal Radion(M800) $\rightarrow HH$ jets, and backgrounds $t\bar{t}$ jets and QCD jets	185
B.2	p_T (left) and mass (right) distributions of signal Radion(M800) $\rightarrow HH$ jets, and backgrounds $t\bar{t}$ jets and QCD jets, which have valid SD discriminator χ value, ie. $\chi > 0$	185
B.3	Log χ distributions of signal Radion(M800) $\rightarrow HH$ jets, and backgrounds $t\bar{t}$ jets and QCD jets	186
B.4	Number of microjets (left) and number of b-tagged microjets (right) in signal and background fatjets.	187
B.5	Distributions of $\Delta R(\text{Microjet, MC truth b-quark})$ (left) and $\Delta R(\text{Microjet, IVF SV})$ (right) of signal sample.	187
B.6	Log χ (left) and p_T (right) distributions of signal Radion (M800) $\rightarrow HH$ jets, and backgrounds $Z' \rightarrow t\bar{t}$ jets	188

B.7	(Top left) p_T distribution of unmatched fatjet and generator Higgs particle. (Top right) p_T distribution of matched fatjet and generator Higgs particle with $\Delta R(\text{fatjet, gen Higgs}) < 0.05$. (Bottom left) p_T distribution of matched fatjets and generator Higgs particles with $\Delta R(\text{fatjet, gen Higgs}) < 0.05$ and accepted by SD. (Bottom right) p_T distribution of matched fatjets and generator Higgs particles with $\Delta R(\text{fatjet, gen Higgs}) < 0.5$ and rejected by SD.	189
B.8	Log χ performance curve of signal against $Z' \rightarrow t\bar{t}$ background with three different ΔR 's: $\Delta R < 0.05$ (tight), $\Delta R < 0.2$ (medium), $\Delta R < 1.2$ (loose). The performance of cutting at $\log \chi = -10.8$ is plotted on each line.	190
B.9	(Left) Distribution of the number of microjets in signal and background fatjets. (Middle) Distribution of the number of microjets in signal and background fatjets that have a valid SD interpretation or $\chi > 0$. (Right) Distribution of the number of microjets in signal and background fatjets that have an invalid SD interpretation or $\chi = 0$	191
B.10	(Left) Distribution of the number of (IVF) b-tagged microjets in signal and background fatjets. (Middle) Distribution of the number of (IVF) b-tagged microjets in signal and background fatjets that have a valid SD interpretation or $\chi > 0$. (Right) Distribution of the number of (IVF) b-tagged microjets in signal and background fatjets that have an invalid SD interpretation or $\chi = 0$	191
B.11	(Left) 2D histogram of number of microjets versus $\log \chi$ distribution of signal sample. (Right) 2D histogram of number of b-tagged microjets versus $\log \chi$ distribution of signal sample. Signal fatjets are matched with generator level Higgs with $\Delta R < 0.05$ and with $p_T > 300\text{GeV}$ selection.	192
B.12	(Left) Signal p_T distribution of matched fatjets with generator Higgs, with $\Delta R < 0.05$. (Middle) p_T distribution of those that are accepted by SD. (Right) p_T distribution of those that that are rejected by SD.	193
B.13	(Left) $Z' \rightarrow t\bar{t}$ background p_T distribution of matched fatjets with generator top quarks, with $\Delta R < 0.05$. (Middle) p_T distribution of those that are accepted by SD. (Right) p_T distribution of those that that are rejected by SD.	193
B.14	Signal p_T distribution before matching with generator Higgs, after matching, and after matching and applying SD algorithm (using $\Delta m_H = 20\text{ GeV}$).	194

B.15	p_T distribution of matched fatjets (ΔR (fatjet, gen) < 0.05), have minimum p_T of 300 GeV and those that have at least 2 microjets and at least 2 b-tagged microjets.	195
B.16	Performance curve or ROC of $\log \chi$ distribution of signal against $Z' \rightarrow t\bar{t}$ background with various Higgs mass windows as input parameter to SD, where the denominator is the number of entries in Figure B.15.	195
B.17	(Left) B-quark matched microjets p_T distribution of signal sample. (right) B-quark matched microjet p_T distribution of signal sample that are IVF b-tagged.	197
B.18	(Left) B-hadron matched microjets p_T distribution of signal sample. (right) B-hadron matched microjet p_T distribution of signal sample that are IVF b-tagged.	197
B.19	(Left) p_T distribution of matched $Z' \rightarrow t\bar{t}$ microjets. (Middle) p_T distribution of matched $Z' \rightarrow t\bar{t}$ microjets that are IVF b-tagged. (Right) fake rate p_T dependance obtained by graphing the ratio of middle plot againts the left plot.	198
B.20	(Left) p_T distribution of matched QCD300-470 microjets. (Middle) p_T distribution of matched QCD300-470 microjets that are IVF b-tagged. (Right) fake rate p_T dependance obtained by graphing the ratio of middle plot againts the left plot.	198
B.21	(Left) p_T distribution of matched QCD470-600 microjets. (Middle) p_T distribution of matched QCD470-600 microjets that are IVF b-tagged. (Right) fake rate p_T dependance obtained by graphing the ratio of middle plot against the left plot.	198
B.22	Number of IVF SV among u/d/s/g flavoured fatjets in $Z' \rightarrow t\bar{t}$ sample (left), QCD300-470 sample (middle), and QCD470-600 sample (right)	199
B.23	Number of microjets among u/d/s/g flavoured fatjets in $Z' \rightarrow t\bar{t}$ sample (left), QCD300-470 sample (middle), and QCD470-600 sample (right)	199
B.24	(Left) p_T distribution QCD470-600 microjets clustered from u/d/s/g flavoured fatjets, using various jet algorithms. (Right) p_T distribution QCD470-600 b-tagged microjets clustered from u/d/s/g flavoured fatjets, using various jet algorithms.	200
B.25	(Left) Number of microjets from fatjets in QCD470-600 sample. (Center) Number of microjets from u/d/s/g fatjets in QCD470-600 sample. (Right) Number of b-tagged microjets from u/d/s/g flavored fatjets in QCD470-600 sample.	201
B.26	Signal efficiency versus background rejection curve comparing different tag rate input parameter for lower p_T (left) and higher p_T (right) fatjets.	202

B.27 Signal efficiency versus background rejection curve comparing different b-tag fakerate input parameter for lower p_T (left) and higher p_T (right) fatjets.	203
B.28 Signal efficiency versus background rejection curve comparing different higgs mass window input parameters of 20 GeV and 30 GeV for lower p_T (left) and higher p_T (right) fatjets.	204
B.29 Signal efficiency versus background rejection curve comparing higgs mass window input parameters of 30 GeV and 50 GeV for lower p_T (left) and higher p_T (right) fatjets.	204
B.30 Signal efficiency versus background rejection curve comparing microjet cone size $R=0.10$ with $R=0.15$ for lower p_T (left) and higher p_T (right) fatjets.	205
B.31 Signal efficiency versus background rejection curve comparing microjet cone size $R=0.20$ with $R=0.15$ for lower p_T (left) and higher p_T (right) fatjets.	205
B.32 Mass distribution of signal (bottom) and QCD (top left) and $Z' \rightarrow t\bar{t}$ (top right) background before and after $SD, \tau_2/\tau_1$, q-vol selections.	207
B.33 p_T distribution of signal and backgrounds after selection of $p_T > 300$ GeV, $95 < m_{pruned} < 155$, $\chi > 0$	208
B.34 Signal efficiency versus background rejection (1-efficiency) of SD in comparison to τ_2/τ_1 and q-jet volatility, in p_T bins.	209
B.35 Signal efficiency versus background rejection (1-efficiency) of SD in comparison to τ_2/τ_1 and q-jet volatility, inclusive in p_T ranging from 300 GeV to 10000 GeV	210
B.36 Signal efficiency versus p_T for a given QCD (inclusive) background rejection based on $\log(chi)$, τ_2/τ_1 and q-vol selections	211
B.37 Signal efficiency versus p_T for a given $Z' \rightarrow t\bar{t}$ background rejection based on $\log(chi)$, τ_2/τ_1 and q-vol selections	212
B.38 Signal efficiency versus p_T for a given QCD inclusive (left) and $Z' \rightarrow t\bar{t}$ (right) background rejection based on a $\log(\chi)$ selection	212
C.1 Sketches of the c-rack at design stage. [21]	215
C.2 Visualisation of the c-rack modules in x,y,z coordinates as registered by the CMS software, showing the layer number and the module number (layer/module number).	216
C.3 State of the c-rack on May 23rd, 2018.	217

C.4	Visualisation of the c-rack modules in x,y,z coordinates, showing CCU address and channel (Address/channel)	219
C.5	Cluster latency measurement of 50 to 80 bx units of 300 events per time step.	220
C.6	Cluster latency measurement where the x-axis has been inverted in order to fit using the CR-RC pulse shape function.	220

Chapter 1

Introduction

Many of us have wondered, at some point in our lives, what the world is made out of and how do they come together. This body of work is a humble attempt to contribute to the vast amount of research work in trying to answer these questions and further understand what the universe is composed of, in the context of particle physics. *The standard model* (SM) theory of particle physics is the best and most robust theory that we know of that could provide the partial answers. Specifically, this dissertation is on the search for particles called *vector-like quarks* (VLQ) which are hypothetical particles, predicted to exist by theories *beyond the standard model* (BSM). The data sample used in this research was collected by the CMS experiment in 2016 at CERN's Large Hadron Collider in Geneva, Switzerland.

The last remaining particle that is predicted by the SM, the Higgs boson, was discovered in 2012 [22, 23]. The discovery closed one chapter of searches but opened a whole new chapter of searches for new physics beyond the SM. Vector-like quarks are particles predicted by a class of BSM theories. They are predicted to couple with 3rd generation SM quarks, SM vector-bosons, and the Higgs boson. Previously, VLQ searches has been performed by CMS [24, 25] and ATLAS [26] experiments using center-of-mass energy $\sqrt{s} = 8$ TeV proton-proton (pp) collision data. This dissertation continues the search effort using $\sqrt{s} = 13$ TeV pp collisions and focuses the search on pair produced top/bottom-like VLQs decaying to final states with multiple leptons (electrons/muons) and jets.

The thesis begins with an overview of the relevant particle physics theories (Chapter 2 and 3), followed by a description of the experimental apparatus, the CMS detector, in Chapter 4. Chapter 5 discusses the strategy, analysis and results of the search for vector-like heavy quark partners decaying to final states with

leptons and jets. Next, Chapter 6 explores the use of a new jet clustering algorithm XCone in the context of the search with single lepton final states, and lastly Chapter 7 summarizes the findings.

Chapter 2

The standard model theory of particle physics

The current understanding of physics tells us that our universe is composed of a finite set of unimaginably small indivisible elementary particles. They make up all of known physical matter, and their interactions with one another are governed by distinct forces of nature. Until now, physicists have experimentally identified four such forces: the electromagnetic force, the weak nuclear force, the strong nuclear force, and the gravitational force. The standard model (SM) theory of particle physics is an elaborate theoretical and mathematical construct that attempts to explain the nature of the dynamics and interactions of all these known particles, excluding the gravitational force.

The SM theory is a quantum gauge field theory where the gauge invariance is based on the symmetry product group $SU(3)_C \times SU(2)_L \times U(1)_Y$. Quantum field theory combines the framework of field theory, quantum mechanics, and special relativity into one consistent framework. Despite its limitations, the SM has provided the means to perform physics calculations that have led to stunningly accurate and precise agreements between theoretical predictions and experimental observations. It has stood the test of time for over 50 years, and it is commonly considered to be the best theory that we have for describing the particle nature of our universe.

In this chapter we will give an overview of SM particles and describe the theoretical frameworks upon which SM is built, and that following that we will describe the explicit mathematical formulation of SM.

This chapter are mainly drawn from the following references: [27, 28, 29, 30, 31, 32].

2.1 Fundamental particles and forces of nature

SM particles can be classified into two types: *fermions* and *bosons*. Fermions are particles whose spin momentum take half-integer values, and they obey the Fermi-Dirac statistics due the the Pauli exclusion principle. By contrast, bosons are those with integer values, and they obey the Bose-Einstein statistics. As fundamental particles SM fermions are spin-half particles, and all matter particles are fermions wheras all force particles which mediate the interactions between matter are bosons.

Fundamental fermions can be further classified into particles which do not interact at all with the strong force, called *leptons*, and those which interact with the strong force, called *quarks*. The strong force is what binds the quarks to make composite particles or bound states called *hadrons*, such as protons and neutrons. However, both leptons and quarks interact with the electromagnetic and the weak force. There are three *generations* of both leptons and quarks, and there are two leptons and two quarks for each generation. The generations differ mainly by their masses where the higher generation particles are generally the heavier siblings of the corresponding lower generation particles, except for neutrinos ¹. Another difference is that the first generation particles are stable and do not decay, whereas the higher generation particles have short lifetimes and decay to lighter particles ². Apart from the mass and lifetime differences, all of the other fundamental properties are identical across the generations. The first generation leptons are electron (e) and electron-neutrino (ν_e), the second generation are muon (μ) and muon-neutrino (ν_μ), and the third generation are tau (τ) and tau-neutrino (ν_τ). The electron, muon, and tau leptons each have a negative electric charge, whereas all the neutrino leptons are electrically neutral. For the quarks, the first generation are up (u) and down (d), second generation are charm (c) and strange (s), third generation are top (t) and bottom (b) quarks. The u , c , and t quarks each have a positive fractional electric charge of $+2/3$, and they are often referred to as the "up" type quarks, whereas the d , s , and b quarks each have a negative fractional electric charge of $-1/3$, and they are often referred to as the "down" type quarks. In terms of helicity, there are only left-handed neutrinos (and right-handed antineutrinos), as we have not observed any right-handed neutrinos in nature [33]. By contrast, all other fermions have been observed as right-handed and left-handed.

¹Neutrinos have been measured to be at least about nine orders of magnitude lighter compared to all other particles, however the absolute value of the masses are currently still not known.

²Again, neutrinos are an exception to this decay behaviour. Neutrinos from one generation transform into another generation when propagating over large distances

The fundamental gauge bosons which mediate the three forces of nature are the following: the photon (γ) mediates the electromagnetic force interaction, the W^\pm and Z bosons mediate the weak force interaction, and the gluon (g) mediates the strong force. Both the photon and the gluon are massless bosons, whereas the W^\pm and Z are massive bosons. These latter two bosons are massive due to the Higgs field whose quantum particle manifestation is the Higgs boson, the most recent and the last of the predicted SM particles to be discovered by the CMS and ATLAS experiments at CERN in 2012 [34, 35].

The strong force is described by Quantum Chromodynamics (QCD) which introduces the concept of *color* charges, and the electromagnetic force and the weak force are two manifestations of one force whose symmetry has been broken where the symmetry breaking is possible due to the presence of the Higgs field. This unified theory of electromagnetic force and the weak force is called the electroweak (EW) theory. The theory is a chiral gauge theory where left-handed fermions are treated differently than the right-handed fermions. The EW theory introduces the *weak isospin* quantum number I and the *hypercharge* quantum number Y . The SM is the combination of QCD and EW theories, as these theories assemble all known particles.

All matter particles and force mediator particles of the SM are summarized in Figure 2.1, which also show their current best mass measurements [36]. Table 2.1 shows their electroweak quantum numbers, and in addition, to each of these particles mentioned, there exist a corresponding anti-particle, which differs only by its opposite electric charge.

2.2 Quantum field theory

The SM is built under the theoretical framework of quantum field theory (QFT) which is a framework combining field theory, quantum mechanics, and special relativity. In this theory, fields are quantized, and a particle is a quanta of the field which is constructed as an operator on the quantum mechanical Hilbert space of a particle state. The field acts as creation and annihilation operators for particles (and antiparticles). Moreover, the fields, hence the particles, must obey an equation of motion which is Lorentz invariant.

Typically, a QFT model is constructed by first formulating a Lagrangian, more precisely a Lagrangian density, which consists of the fields that represent the types of particles one would like to include in the model, eg. leptons, quarks, and bosons. The Lagrangian would generally be categorized into kinetics terms, which are bilinear field terms), and interaction terms, which consist of three or more fields. Using perturbation

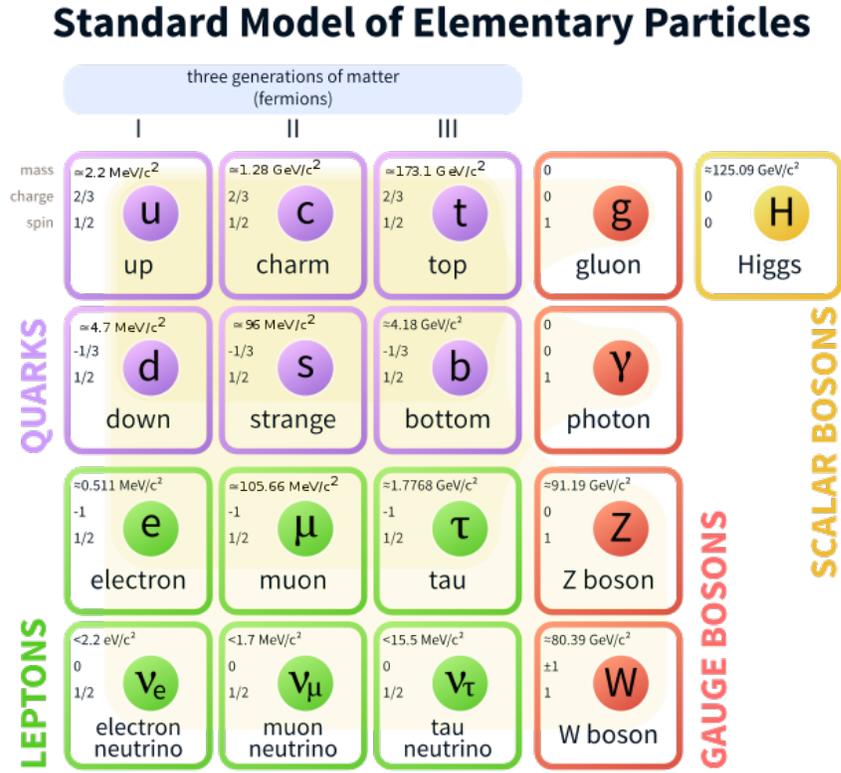


Figure 2.1: Summary of the standard model particles. [4]

Table 2.1: Electroweak quantum numbers of the fundamental fermions (where the left-handed and right-handed have different quantum numbers) and the fundamental bosons. The electromagnetic charge (Q), the third component of the weak isospin I_3 and the weak hypercharge Y obey the Gell-Mann-Nishijima relation $Q = I_3 + \frac{Y}{2}$.

Particles	EM charge Q (Q)	Weak Isospin (I_3) (I_3)	Weak Hypercharge (Y) (Y)
Left handed fermions			
ν_e, ν_μ, ν_τ	0	$+\frac{1}{2}$	-1
e, μ, τ	-1	$-\frac{1}{2}$	-1
u, c, t	$+\frac{2}{3}$	$+\frac{1}{2}$	$+\frac{1}{3}$
d, s, b	$-\frac{1}{3}$	$-\frac{1}{2}$	$+\frac{1}{3}$
Right-handed fermions			
e, μ, τ	-1	0	-2
u, c, t	$+\frac{2}{3}$	0	$+\frac{4}{3}$
d, s, b	$-\frac{1}{3}$	0	$-\frac{2}{3}$
Fundamental bosons			
W	± 1	± 1	0
Z	0	0	0
γ	0	0	0
H	0	$-\frac{1}{2}$	+1

theory and path integral methods, the Lagrangian allows the derivation of all the possible interactions that can happen between the fields and provides the means for calculating the matrix element of the interaction processes. In the end, it provides the means to calculate the scattering amplitudes, decay rates, hence cross sections of the various interactions, which then can be compared to experimental observations. Pictorial representation of the perturbation expansions, also known as *Feynman diagrams*, along with the specific rules of how the diagrams come together, are commonly used to perform the calculations. In this context, one speaks of a *tree level* diagrams to refer to the leading order terms of the expansion, whereas higher order terms are represented by diagrams containing one or more *loops*. It is worth noting that often calculations in QFT would involve infinity values rendering the calculation useless. However, in many instances these infinities are unphysical and can be removed once physicality constraints are enforced. The method for handling these infinities are called *renormalization*, which involves redefining and rescaling the parameters of the model in such way that the infinities are side stepped.

There are only three classes of fields present in the formulation of the Lagrangian of the standard model. Spin-0 particles are described by scalar fields $\phi(x)$, spin-1 particles by vector fields $V_\mu(x)$, and spin-1/2 fermions by spinor fields $\psi(x)$. Furthermore, various symmetries are imposed on a model. The Lagrangian formulation allows for having space-time symmetry in terms of Lorentz invariance and also internal symmetries, hence the choice of formulating the theory in terms of the Lagrangian³. Furthermore, gauge symmetry is central to the formulation the standard model Lagrangian. As we will discuss in the next sections, the model for the electroweak force is constructed to be gauge invariant based on the $SU(2)_L \times U(1)_Y$ symmetry group, and the model for the QCD is constructed using the $SU(3)_C$ symmetry group.

In the rest of the chapter we will describe the mathematical formulation the strong force and the electroweak force under the framework of QFT.

2.3 Quantum chromodynamics

The strong interactions is described by QCD, which is a non-Abelian gauge theory based on the $SU(3)_C$ Lie group symmetry in the triplet representation (where the subscript C stands for color), ie. the quark fields are color charge triplet. In this representation, the $SU(3)_C$ color group has eight generators,

$$T^a = \frac{1}{2}\lambda^a, \quad a = 1, \dots, 8 \quad (2.1)$$

³In contrast, the Hamiltonian does not preserve Lorentz symmetry, it instead preserves conservation of energy.

where λ_a are the 3×3 Gell-Mann matrices [37]. They obey the commutation relations,

$$[T^a, T^b] = if^{abc}T^c, \quad (2.2)$$

where f^{abc} is the structure constant of the SU(3) group. Consequently, there are eight gluons as the gauge bosons and as quanta of the vector fields G_μ^a . The QCD Lagrangian [28] describing the six flavors (f) quark fields $\Psi^{(f)}$ interacting via the strong force can be written as,

$$\mathcal{L}_{\text{QCD}} = \sum_f \left(\bar{\Psi}^{(f)} (i\gamma^\mu D_\mu - m_f) \Psi^{(f)} - \frac{1}{4} G_{\mu\nu}^a G^{a,\mu\nu} \right), \quad (2.3)$$

where m_f is the mass of the quark of flavor f and D_μ is the covariant derivative, defined as,

$$D_\mu = \partial_\mu - ig_s \frac{\lambda_a}{2} G_\mu^a, \quad (2.4)$$

and $G_{\mu\nu}^a$ are the field strengths, defined as,

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial G_\mu^a + g_s f_{abc} G_\mu^b G_\nu^c, \quad (2.5)$$

with g_s being the strong coupling constant, often also expressed in terms of $\alpha_s = \frac{g_s^2}{4\pi}$.

A feature of QCD, being a non-Abelian gauge theory, is that the strength of the strong coupling constant becomes asymptotically smaller at higher energies, a property also known as *asymptotic freedom*. This serves as an explanation why color charged quarks and gluons are never observed in isolation and why the strong force is short-ranged. This feature can be seen from the QCD beta function or the renormalization group equation [29], where the leading term can be written as,

$$\alpha_s(\mu) = \frac{2\pi}{11 - \frac{2n_f}{3}} \frac{1}{\ln \frac{\mu}{\Lambda_{\text{QCD}}}} \quad (2.6)$$

where μ is the some arbitrary (energy) scale, n_f is the number of flavors, Λ_{QCD} is the scale of the *Landau pole*⁴ of QCD, and this equation is valid for $\mu > \Lambda_{\text{QCD}}$. As there are 6 quark flavors, the denominator of the first factor is always positive and therefore $\alpha_s(\mu)$ decreases with increasing μ . As a consequence, perturbation methods only works well at very large energy scales, and becomes less useful and ultimately invalid at lower energies. Due to this property and the many self-interaction terms of QCD, perturbative calculations in this theory are often very cumbersome, impractical or impossible.

⁴This is the energy scale where the coupling strength becomes infinite.

2.4 Electroweak theory

At high energies the electromagnetic force and the weak are unified as one force, the electroweak force. The EW theory, also known as the Glashow-Weinberg-Salam model [38, 39, 40] based on the authors who developed the model independently, is constructed based on a gauge symmetry that spontaneously breaks from $SU(2)_L \times U(1)_Y \rightarrow U(1)_{em}$ due to the presence of the Higgs field. The high-energy $U(1)_Y$ symmetry is the hypercharge and it is different from the low-energy $U(1)_{em}$ which is the symmetry for electromagnetism [29].

The EW lagrangian can be broken down into four parts: The pure gauge field terms \mathcal{L}_G , the fermion fields and fermion-gauge interaction terms \mathcal{L}_F , the higgs terms \mathcal{L}_H , and the fermion-scalar (Yukawa) interactions \mathcal{L}_y ,

$$\mathcal{L}_{EW} = \mathcal{L}_G + \mathcal{L}_F + \mathcal{L}_H + \mathcal{L}_y. \quad (2.7)$$

The pure gauge field terms

The gauge symmetry $SU(2)_L \times U(1)_Y$ has four generators, the weak isospin operators I_i , where $i = 1, 2, 3$, associated with the non-Abelian $SU(2)$ part and the weak hypercharge generator Y associated with the Abelian $U(1)$ part. These generators satisfy,

$$[I_i, I_j] = i\epsilon_{ijk}I_k \quad (2.8)$$

$$[I_i, Y] = 0,$$

where ϵ_{ijk} is the Levi-Civita symbol. The gauge vector field associated to I_i is denoted by W_μ^i whereas the gauge field associated with Y is denoted by B_μ and the pure gauge field terms can be written as,

$$\mathcal{L}_G = -\frac{1}{4}W_{\mu\nu}^i W^{\mu\nu,i} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}, \quad (2.9)$$

where $W_{\mu\nu}^i$ and $B_{\mu\nu}$ are gauge fields strengths,

$$W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i + g\epsilon_{ijk}W_\mu^j W_\nu^k, \quad (2.10)$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu,$$

and where g is the $SU(2)$ gauge coupling constant.

The fermion fields and fermion-gauge interaction terms

The lagrangian containing the fermion fields and the fermion-gauge interaction terms can be written as,

$$\mathcal{L}_F = \sum_j \bar{\psi}_L^j i\gamma^\mu D_\mu^L \psi_L^j + \sum_{j,\sigma} i\gamma^\mu D_\mu^R \psi_{R\sigma}^j. \quad (2.11)$$

where there are left and right versions of the covariant derivative are

$$\begin{aligned} D_\mu^L &= \partial_\mu - ig \frac{1}{2} \sigma_i W_\mu^i + ig' \frac{Y}{2} B_\mu, \\ D_\mu^R &= \partial_\mu + ig' \frac{Y}{2} B_\mu, \end{aligned} \quad (2.12)$$

where g' is the $U(1)$ gauge coupling constant associated and σ_i are the Pauli matrices.

The L in $SU(2)_L$ indicates that only the left-handed components of the Dirac fermion fields, ψ_L , are doublets under the $SU(2)_L$ transformation, whereas the right-handed components, ψ_R , are singlets. Consequently, the fermions are grouped as follows,

$$\psi_L^j = \begin{pmatrix} \psi_{L,f+}^j \\ \psi_{L,f-}^j \end{pmatrix}, \quad \psi_{R,f}^j, \quad (2.13)$$

where $j = 1, 2, 3$ denotes the fundamental fermion generations. For the quarks, 'f+' denotes the "up" flavors and "f-" denotes the down flavor quarks. For the leptons, the "f+" denotes the neutrino particles and the "f-" denotes the non-neutrino leptons. Only the quarks and non-neutrino leptons have right-handed components.

The Higgs terms and spontaneous symmetry breaking

The scalar potential whose vacuum expectation value (VEV) breaks the gauge symmetry from $SU(2)_L \times U(1)_Y \rightarrow U(1)_{em}$ is the complex doublet Higgs field, H . It is a complex scalar field which has the weak hypercharge quantum number $Y = 1$ and the weak isospin quantum number $I = 1/2$, hence it is a weak isospin doublet,

$$H(x) = \begin{pmatrix} \phi^+(x) \\ \phi^0(x) \end{pmatrix}, \quad (2.14)$$

where $\phi^{+,0} = \text{Re}(\phi^{+,0}) + i\text{Im}(\phi^{+,0})$. The lagrangian containing the Higgs terms and potential is

$$\mathcal{L}_H = (D_\mu H)^\dagger (D^\mu H) - V(H), \quad (2.15)$$

where D_μ the covariant derivative is written as

$$D_\mu = \partial_\mu - ig \frac{\sigma_i}{2} W_\mu^i + ig' \frac{1}{2} B_\mu, \quad (2.16)$$

and the potential containing the Higgs self-interaction terms is written as

$$V(H) = -\mu^2 H^\dagger H + \frac{\lambda}{4} (H^\dagger H)^2, \quad \text{with } \mu^2, \lambda > 0, \quad (2.17)$$

where μ is the constant for the mass term and λ is constant for the quartic self-interaction term. Due to setting $\mu^2 > 0$, the mass term has a 'wrong' sign in comparison to the Lagrangian for a free scalar field. Consequently, the true minimum of this potential is non-zero and they occur at infinitely many degenerate points, at $H^\dagger H = \frac{2\mu^2}{\lambda}$ or at any equivalent values a $U(1)$ symmetry transformation. This results in having a VEV of

$$\langle 0|H|0 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad \text{where } v = \frac{2\mu}{\sqrt{\lambda}}. \quad (2.18)$$

Due to the nature of the potential (see Figure 2.2) and the resulting ground state, the original $SU(2)_L \times U(1)_Y$ gauge symmetry is now broken by the vacuum configuration and thus symmetry is spontaneously broken. However, the vacuum configuration is still symmetric under a gauge $U(1)$ symmetry, which is identified as the unbroken $U(1)_{em}$ gauge symmetry of electromagnetism, generated by the charge Q . Evidently, we can

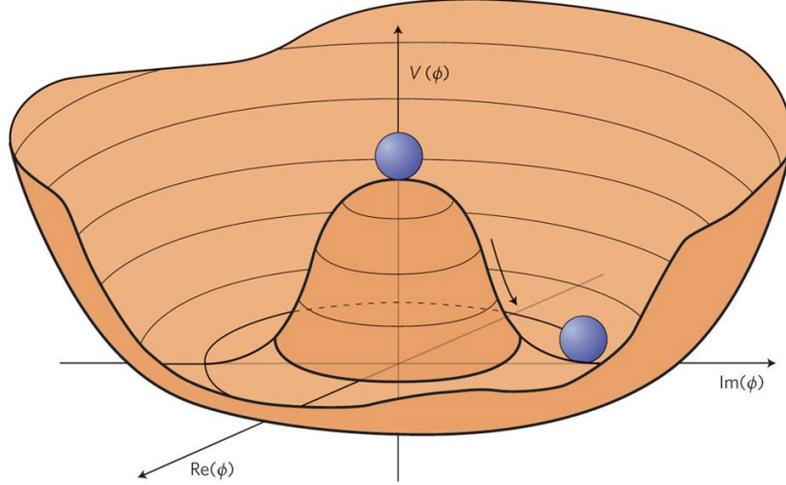


Figure 2.2: Sketch of the Higgs potential appearing in \mathcal{L}_H [5].

rewrite the Higgs field, replacing ϕ^+ and $\phi^0(x)$, in terms of v and real scalar fields $h(x)$, and parameterization fields θ_a ($a = 1, 2, 3$),

$$H(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \theta_1 + i\theta_2 \\ v + h(x) + i\theta_3 \end{pmatrix}. \quad (2.19)$$

If we Taylor expand around the vacuum configuration we can write

$$H(x) = \frac{\exp(i\frac{\theta_a \sigma^a}{v})}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \quad (2.20)$$

we observe that the exponential factor outside serves as a gauge transformation for a $SU(2)_L$ doublet. With the choice of the *unitary gauge*, $\phi \rightarrow \exp(i\frac{\theta_a \sigma_a}{v})\phi$ [31] the field parameters θ_a vanish and we can write

$$H(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \quad (2.21)$$

Substituting Eq. 2.21 to firstly the Higgs potential in Eq 2.4,

$$\begin{aligned} V &= \mu^2 h^2 + \frac{\mu}{v} h^3 + \frac{\mu}{4v^2} h^4 \\ &= \frac{M_H^2}{2} h^2 + \frac{M_H^2}{2v} h^3 + \frac{M_H^2}{8v^2} h^4, \end{aligned} \quad (2.22)$$

where it is now evident that $h(x)$ describes the physical Higgs boson with mass $M_H = \sqrt{2}\mu$. Next, the substitution on rest of Eq 2.4 yields,

$$\frac{1}{2} \left(\frac{g}{2}v\right)^2 (W^{(1)}W^{(1)} + W^{(2)}W^{(2)}) + \frac{1}{2} \left(\frac{v}{2}\right)^2 \begin{pmatrix} W_\mu^{(3)} & B_\mu \end{pmatrix} \begin{pmatrix} g^2 & g'g \\ g'g & g'^2 \end{pmatrix} \begin{pmatrix} W^{(3),\mu} \\ B^\mu \end{pmatrix}. \quad (2.23)$$

We introduce the fields W_μ^\pm , Z_μ and A_μ as the vector fields associated with the physical W^\pm , Z gauge bosons of the weak interactions and photon (γ) of the electromagnetic interactions respectively, and they are defined as mixtures of W_μ^i and B_μ as follows,

$$W_\mu^\pm \equiv \frac{1}{\sqrt{2}} \left(W_\mu^{(1)} \mp iW_\mu^{(2)} \right), \quad (2.24)$$

$$\begin{pmatrix} Z_\mu & A_\mu \end{pmatrix} \equiv \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} W_\mu^{(3)} \\ B_\mu \end{pmatrix}, \quad (2.25)$$

where we have also defined $\theta_W \equiv \arctan(\frac{g'}{g})$ to be the *weak mixing angle*. Using Eq 2.24 and Eq. 2.25, Eq. 2.22 is now diagonalized and can be written as,

$$M_W^2 W_\mu^+ W^{\mu,-} + \frac{1}{2} \begin{pmatrix} A_\mu & Z_\mu \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & M_Z^2 \end{pmatrix} \begin{pmatrix} A^\mu \\ Z^\mu \end{pmatrix} \quad (2.26)$$

The gauge bosons W^\pm and Z are now massive, with masses M_W and M_Z ,

$$M_W = \frac{1}{2}g'v \quad M_Z = \frac{1}{2}\sqrt{g'^2 + g^2}v. \quad (2.27)$$

The vector bosons acquire mass due to the VEV of the Higgs potential which spontaneously brakes the original $SU(2)_L \times U(1)_Y$ gauge symmetry to $U(1)_{em}$. This is commonly referred to as the Higgs mechanism however to be more fair to the many other authors who discovered it independently, we will call it the

Englert-Brout-Higgs-Guralnik-Hagen-Kibble mechanism⁵ [41, 42, 43]. In addition, the A_μ vector field is identified as the familiar photon field which is massless as required.

The fermion-Higgs (Yukawa) interactions terms

To generate the masses of the charged fermions, the interaction terms between the fermion field and the Higgs field are introduced. The interactions between Dirac fields and scalar fields are commonly known as the Yukawa interactions [44]. The Lagrangian can be written as follows,

$$\mathcal{L}_y = \sum_j \left(-y_l^j \bar{L}_L^j H l_R^j - y_d^j \bar{Q}_L^j H d_R^j - y_u^j \bar{Q}_L^j H^c u_R^j + h.c. \right) \quad , \quad (2.28)$$

where $j=1,2,3$ indicate the fermion generations, $y_{l,d,u}^j$ are the individual yukawa couplings and we define the following left-handed lepton and quark fields (consistent with Eq 2.13),

$$L_L^j = \begin{pmatrix} \nu_L^j \\ l_L^j \end{pmatrix}, \quad \nu_R^j, \quad l_R^j \quad Q_L^j = \begin{pmatrix} u_L^j \\ d_L^j \end{pmatrix}, \quad u_R^j, \quad d_R^j \quad (2.29)$$

where ν are the neutrino leptons, l are the non-neutrino leptons, u are the up-type quarks d are the down-type quarks. Whereas, the right-handed fermions fields are $SU(2)$ singlets. (Note that the last two terms in Eq. 2.28 is $SU(3) \times SU(2) \times U(1)$ invariant [29]. In the unitary gauge, (substituting Eq. 2.20), Eq. 2.28 becomes,

$$\mathcal{L}_y = \sum_f -y_f \frac{v}{\sqrt{2}} \bar{\psi}_f \psi_f - y_f \frac{1}{\sqrt{2}} \bar{\psi}_f \psi_f h \quad , \quad (2.30)$$

where the charged fermions $f = l, u, d$ are now massive with masses

$$m_f = y_f \frac{v}{\sqrt{2}} \quad , \quad (2.31)$$

which are proportional to the coupling constants of the interaction between the massive fermions and the physical Higgs field $h(x)$. Though it has been experimentally verified that neutrinos are massive particles through the observation of neutrino oscillations [45, 46, 47], the standard model does not require it to be massive. The definite mass values of neutrinos are still not known, the nature of its field and the mechanism by which it acquires mass is not yet well established⁶ [49].

⁵There were many people around about the same time who discovered the same idea in slightly different contexts, including Anderson, Brout, Englert, Ginzburg, Guralnik (who was a Brown University Professor), Hagan, Kibble, Landau, other than Higgs [29].

⁶Neutrino mass can be constructed without altering basic structure of the standard model [48].

Currents

We conclude this chapter by describing briefly the Noether currents in the electroweak theory after symmetry breaking. Expanding Eq. 2.11 we can write the Noether currents for the electromagnetic current J_{em}^μ , weak neutral current (NC) J_{NC}^μ , weak charged current (CC) J_{CC}^μ with their corresponding vector fields,

$$\begin{aligned}
 \mathcal{L}_{FG} &= J_{em}^\mu A_\mu + J_{NC}^\mu Z_\mu + J_{CC}^\mu (W_\mu^+ + W_\mu^-) \\
 J_{em}^\mu &= \frac{g'g}{\sqrt{g'^2 + g^2}} \sum_{f=l,q} Q_f \bar{\phi}_f \gamma^\mu \psi_f \quad , \\
 J_{NC}^\mu &= \frac{g}{2 \cos(\theta_W)} \sum_{f=l,q} \bar{\phi}_f (v_f \gamma^\mu - a_f \gamma^\mu \gamma_5) \psi_f \quad , \\
 J_{CC}^\mu &= \frac{g}{\sqrt{2}} \left(\sum_{i=1,2,3} \bar{\nu}^i \gamma^\mu \frac{1-\gamma_5}{2} e^i + \sum_{i,j=1,2,3} \bar{u}^i \gamma^\mu \frac{1-\gamma_5}{2} V_{ij} d^j \right) \quad ,
 \end{aligned} \tag{2.32}$$

where, $e^i = e, \mu, \tau$, $\nu^i = \nu_e, \nu_\mu, \nu_\tau$, and v_f, a_f are the neutral current coupling constants,

$$\begin{aligned}
 v_f &= I_3^f - 2Q_f \sin(\theta_W)^2 \\
 a_f &= I_3^f \quad ,
 \end{aligned} \tag{2.33}$$

where Q_f is the electromagnetic charge, I_3^f is the third component of the weak isospin and V_{ij} are the elements of the CKM matrix [50, 51].

We have described the standard model in terms of the collection of particles it describes and also in terms of the mathematical formulation. In the next chapter we will discuss several problems with the SM, their possible solutions and beyond the SM theories which inspires the measurements performed in this thesis.

Chapter 3

Beyond the standard model and vector-like quarks (VLQ)

Despite its tremendous successes, the standard model (SM) has some unresolved mysteries, problems, and limitations. To begin with, the SM has not incorporated the force of gravity and therefore it has nothing to say about the quantum nature of gravity. Furthermore, the neutrinos have been observed to be massive and yet the SM does not require them to be massive, let alone predict the values of their masses. In addition, the SM relies on about 20 seemingly arbitrary parameters that need to be manually set by experimental observation. Moreover, QFT, the theoretical framework upon which the SM is built, incorrectly predicts the observed limits of the cosmological vacuum energy density, and the cosmological constant, to a factor of $\approx 10^{120}$ disagreement [52]. Also, dark matter, which has been determined to be the most abundant matter occupying the vast majority of the universe, is not explained by the SM. One particular issue with the SM that pertains to this thesis is the *hierarchy problem*, namely the large discrepancy between the scale of the electroweak force, which is in the ≈ 100 GeV scale, and the scale of gravitational force, which is in the 10^{19} GeV scale. In any case, all of the mentioned problems guarantee that the SM is not the full picture of our universe, and this situation begs for extension or modification of the SM.

The formulation of the beyond-the-standard-model (BSM) theories often involves introducing new fields, or interaction terms, or incorporating new symmetries, or adding extra spatial dimensions, or incorporating a combination of all or some of the things mentioned. Even though different BSM models may be motivated

by different problems, some of the models introduce a common type or class of particles. This thesis is particularly interested in a class of particles called *vector-like quarks* (VLQ). Incidentally there are many BSM theories that predict the existence of this class of particles and these theories are often motivated by the hierarchy problem.

In this chapter, firstly we will briefly discuss the hierarchy problem and following that we will describe VLQs and how they are produced in the LHC. The section on the hierarchy problems is mainly drawn from Ref. [53], and the VLQ section is mainly drawn from Ref. [2, 54].

3.1 The hierarchy problem

One way to pose the hierarchy problem is as follows. The physical Higgs mass correction can be written as,

$$m_{h,\text{physical}}^2 = m_{h,\text{bare}}^2 + \delta m_h^2 \quad , \quad (3.1)$$

where

$$\delta m_h^2 = \frac{\Lambda}{32\pi^2} \left(6\lambda + \frac{1}{4}(g^2 + g'^2) - y_t \right) \quad , \quad (3.2)$$

where Λ is some higher energy cutoff scale, λ is the self-interaction coupling, g and g' are the gauge fields interaction couplings, and y_t is the yukawa top quark interaction coupling. It is apparent that if Λ is much larger than the electroweak scale which is of order $\approx 100\text{GeV}$, then we see that δm_H will be much larger than the physical mass m_H . In more precise terms, the Higgs mass is quadratically sensitive to this cut off scale. If there are no new physics between the electroweak scale and the scale of the gravitational force, which is at the Plank mass scale m_{Plank} , then $\Lambda \approx m_{\text{Plank}}$, which is 10^{17} magnitude greater than the relatively light Higgs mass. This situation is known as the hierarchy problem.

3.2 Vector-like quarks

Vector-like quarks are color-charged fermions whose left-hand and right-hand components transform in the same way under the SM gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$ [54, 55]. In other words, they transform as triplets under the color gauge group, and both chiralities have the same electroweak quantum numbers.

These exotic quarks are predicted by several BSM theories such as the *Composite Higgs models* [56, 57, 58, 59, 60, 61], *Extra Dimensions models* [62, 63], *Gauged Flavour Group models* [64, 65], *Little Higgs models* [66], and *Supersymmetric non-minimal extensions of the SM* [67, 68, 69]. In some of these BSM

models (the composite higgs, little higgs and the minimal supersymmetry models), the existence of VLQs in the TeV energy scale as heavy quark partners of the SM model quarks would ameliorate the hierarchy problem.

The term "vector-like" can be illustrated as follows. The SM charge current, J^μ , can be written as,

$$\mathcal{L}_W = \frac{g}{\sqrt{2}} (J^{\mu,+} W^+ + J^{\mu,-} W^-). \quad (3.3)$$

The SM quarks are chiral, such that

$$\begin{aligned} J^{\mu,+} &= J_L^{\mu,+} + J_R^{\mu,+}, \text{ where} \\ J_L^{\mu,+} &= \bar{u} \gamma^\mu (1 - \gamma^5) d \\ J_R^{\mu,+} &= 0. \end{aligned} \quad (3.4)$$

Therefore, SM charge current is known to be a *vector minus axial vector* current, or " $V - A$ ". In contrast, the VLQs have the following current,

$$J_{\text{VLQ}}^{\mu,+} = J_{\text{VLQ},L}^{\mu,+} + J_{\text{VLQ},R}^{\mu,+} = \bar{U}_L \gamma^\mu D_L + \bar{U}_R \gamma^\mu D_R = \bar{U} \gamma^\mu D, \quad (3.5)$$

which is a fully vector current, hence the term "vector-like".

VLQs do not generate their masses through the Yukawa couplings to a Higgs doublet, and they are not excluded by experimental data (up to some mass and branching ratio assumptions). VLQs predominantly interact and mix with the third generation of the SM quarks and so they are often referred to as "top partners" [70, 71]. These exotic quarks can be electroweak gauge group singlets, doublets, or triplets, as shown in Table 3.1, which summarizes all the allowed representations of VLQs with various quantum numbers and the Yukawa and mass terms in comparison to the SM quarks.

This thesis focuses on the search for pair produced vector-like T quark of charge $2e/3$ and vector-like B quark of charge $-e/3$.

Table 3.1: Allowed representations of VLQ, in quantum numbers under $SU(2)_L$ and $U(1)_Y$ in comparison to SM quarks. The subscript L denotes the left hand component and the subscript R denotes the right hand component. The superscript in $H^{(c)}$ denotes that it could either be H or H^c depending on the representation chosen. The q denotes the SM quark flavors, u (d) denotes the up(down)-type SM quarks, whereas T , B , X , Y denotes the VLQs. In the Lagrangian rows, Q_L , u_R , d_R denotes the left and right handed SM quarks (as defined in Section 2.4), whereas T_R , B_R , ψ denotes the VLQs, λ denotes the Yukawa coupling involving the VLQs, and σ^a are Pauli matrices. [1, 2].

	SM quarks	VLQ singlets	VLQ doublets	VLQ triplets
	$\begin{pmatrix} u \\ d \end{pmatrix}$ $\begin{pmatrix} c \\ s \end{pmatrix}$ $\begin{pmatrix} t \\ b \end{pmatrix}$	(T) (B)	$\begin{pmatrix} X \\ T \end{pmatrix}$ $\begin{pmatrix} T \\ B \end{pmatrix}$ $\begin{pmatrix} B \\ Y \end{pmatrix}$	$\begin{pmatrix} X \\ T \\ B \end{pmatrix}$ $\begin{pmatrix} T \\ B \\ Y \end{pmatrix}$
$SU(2)_L$	$q_L = 2$ $q_R = 1$	1	2	3
$U(1)_Y$	$q_L = \frac{1}{3}$ $u_R = \frac{4}{3}$ $d_R = -\frac{2}{3}$	$\frac{2}{3}$ $-\frac{1}{3}$	$\frac{7}{6}$ $\frac{1}{6}$ $-\frac{5}{6}$	$\frac{2}{3}$ $-\frac{1}{3}$
\mathcal{L}_Y	$-y_u^i \bar{Q}_L^i H^c u_R^i$ $-y_d^i \bar{Q}_L^i V_{ij} H d_R^j$	$-\lambda_u^i \bar{Q}_L^i H^c T_R$ $-\lambda_d^i \bar{Q}_L^i H B_R$	$-\lambda_u^i \psi_L^i H^{(c)} u_R^i$ $-\lambda_d^i \psi_L^i H^{(c)} d_R^i$	$-\lambda^i \bar{Q}_L^i \sigma^a H^{(c)} \psi_R^a$
\mathcal{L}_m	not allowed	$-M \bar{\psi} \psi$	$-M \bar{\psi} \psi$	$-M \bar{\psi} \psi$

Chapter 4

The Compact Muon Solenoid (CMS)

The Compact Muon Solenoid (CMS) is a particle physics detector stationed at one of the collision points of the Large Hadron Collider (LHC) at European Organization for Nuclear Research / Conseil Européen pour la Recherche Nucléaire (CERN) around Geneva area, Switzerland.

In this chapter I will summarize the LHC and the CMS experiment, its sub-detector and components, and the process of reconstructing events and physics particles that will be relevant for the physics analysis in this thesis.

4.1 The Large Hadron Collider (LHC)

The LHC is a circular particle accelerator and collider with a circumference of 26.7 km and is located inside a tunnel that varies approximately between 45 and 170 meters beneath the surface of Geneva, Switzerland. It is a two-ring superconductor accelerator designed mainly for proton-proton (pp) collisions and to achieve center-of-mass collision energy of 14 TeV with luminosities $\geq 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ [8] and with 25 ns bunch spacings.

The instantaneous luminosity, L , delivered by the LHC can be described by the following formula,

$$L = \frac{N_p^2 \cdot n_b \cdot f_{rev}}{4\pi \cdot \sigma_x \cdot \sigma_y} F(\theta_c, \sigma_x, \sigma_y), \quad (4.1)$$

$$F = \left(1 + \left(\frac{\sigma_z}{\sigma_x} \right)^2 \left(\frac{\theta_c}{2} \right) \right)^{-1/2}, \quad (4.2)$$

where N_p is the number of protons in a bunch, n_b is the number of bunches in a beam, f_{rev} the frequency of the beam revolution, σ_x and σ_y are transverse beam width characterizing the beam optics, F accounts for luminosity reduction due to the beam crossing angle, θ_c , the hourglass effect leading to a varying transverse bunch size in the collision point because of the several cm longitudinal bunch extension and other effects, and σ_z is the RMS bunch length [8, 72]. Given some time integrated luminosity, $\int L dt$, the number of observed events from the collisions is calculated using the following formula,

$$N_{obs} = \xi \cdot \sigma \cdot \int L dt, \quad (4.3)$$

where N_{obs} is the number of observed events, ξ is the efficiency due to the measurement and experimental apparatus, and σ is the cross section of the physics process as determined by nature.

The proton source comes from a tank of hydrogen gas. The electrons of the the hydrogen atoms are removed by applying an electric field, obtaining the protons to be injected to the LHC. There are several pre-injectors prior to delivering the protons to the LHC. First, these protons are accelerated by a linear accelerator called Linac2 to the energy of 50 MeV before being injected to the Proton Synchrotron Booster (PSB), which further accelerates them to 1.4 GeV. Then, the protons are injected to the Proton Synchrotron (PS), which accelerates them increasing their energy to 25 GeV. Following that, the protons are injected into the Super Proton Synchrotron (SPS) where the protons achieve energy up to 450 GeV before the beams are finally transferred to the two rings of the LHC, where the two beams will travel in clockwise and anti-clockwise direction [73]. Figure 4.1 shows a diagram of the LHC underneath Geneva area and the CERN accelerator complex, consisting of Linac2, PSB, PS, and SPS.

The LHC uses dipole superconducting magnets to guide the beams along the roughly circular trajectory of the ring and it uses quadrupole superconducting magnets to focus the beams at collision points. The superconducting magnets are operated at a temperature of 1.9 K in superfluid helium [8]. Figure 4.2 shows a computer-generated diagram of an LHC dipole, showing the beam pipes and the superconducting magnets.

The LHC was built to enable investigations into the physics of the Standard Model in the TeV energy range, the search of the Brout-Englert-Higgs-Guralnik-Hagen-Kibble Boson (or more popularly known simply as the Higgs boson) and also explorations of new physics beyond the Standard Model.

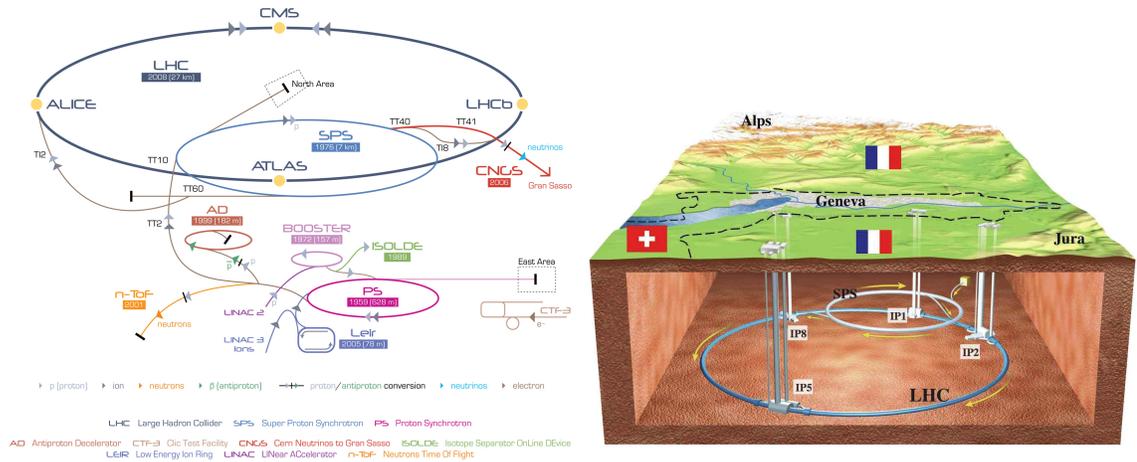


Figure 4.1: Schematic layout of CERN accelerators (top) and an artist's view of the LHC (bottom) [6].

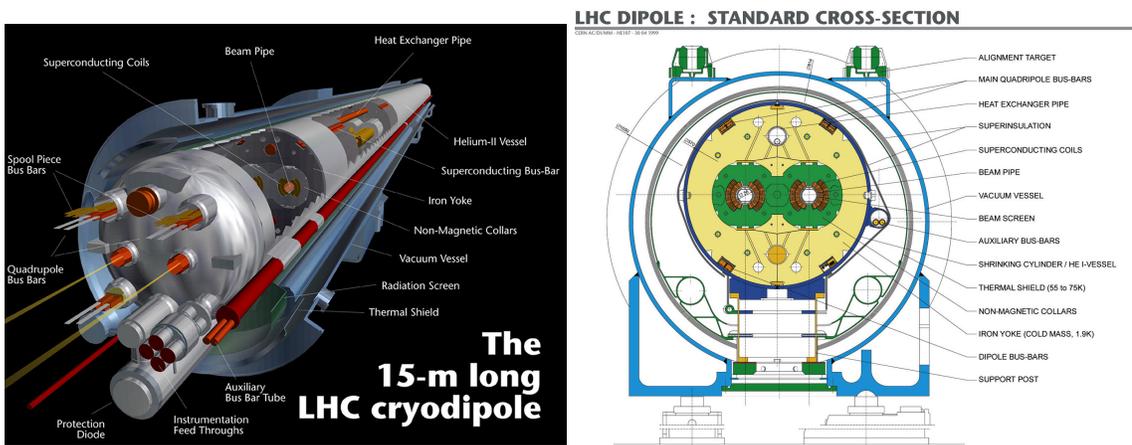


Figure 4.2: LHC cryodipole [7, 8].

There are four main detectors located in the four collision points of the LHC namely, CMS [12], ATLAS [74], ALICE [75] and LHCb [76]. CMS and ATLAS are general purpose detectors whereas ALICE and LHCb are mainly dedicated to heavy-ion and flavour physics studies, respectively.

4.2 The CMS detector

The CMS detector has a cylindrical geometry and is composed of several layers of sub-detector units. A central feature of the detector is a superconducting solenoid of 6 m internal radius, which provides 3.8 T of magnetic field. In total, the detector weighs about 14000 tons, has a radius of 15 m and a length of 28.7 m. The innermost part of the detector is the tracking system consisting of the pixel detector and the silicon strip tracker. As we move outwards, the tracker is surrounded by the electromagnetic calorimeter (ECAL), and then the hadron calorimeter (HCAL). All of these are encapsulated within the volume of the solenoid. Finally, the outermost part is the muon chamber, which is integrated with the iron return yoke of the magnet system. Figure 4.3 shows a schematic diagram of the CMS detector and Figure 4.4 shows the magnetic field produced by the CMS magnet system. The coordinate system used and more details of each sub-detectors are explained in the following subchapters.

4.2.1 Coordinate system

CMS uses a coordinate system where the origin is centered at the nominal pp collision point, the y -axis points vertically upwards, the x -axis points radially inwards towards the LHC center, and the z -axis points along the anti-clockwise direction of the LHC seen from above. The azimuthal angle ϕ is measured from the x -axis and it is transverse to plane of the beam. The radial coordinate in this plane is denoted by r . The polar angle, θ , is measured from the z -axis. However, in practice CMS uses pseudorapidity $\eta = -\ln[\tan(\theta/2)]$ to represent polar angles. This quantity is useful due to its property of being Lorentz invariant in the boost direction, ie. the z direction. Figure 4.5 shows some illustrations of the coordinates used by CMS.

4.2.2 The pixel detector

The pixel detector is a silicon based detector designed to precisely track the trajectories of charged particles in r - ϕ and z space and it is especially responsible for enabling precise measurements of a track's impact parameter, which is important for secondary vertex reconstruction, and which in turn is crucial identifying

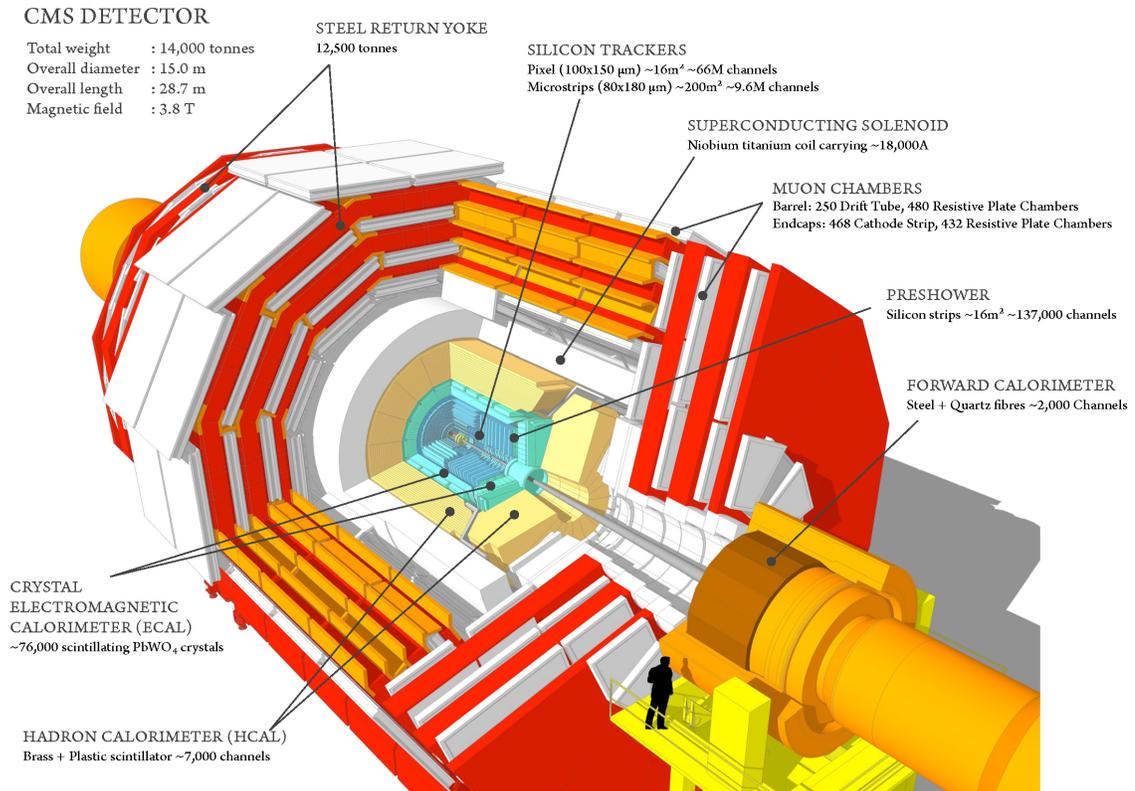


Figure 4.3: Schematic diagram of the CMS detector [9].

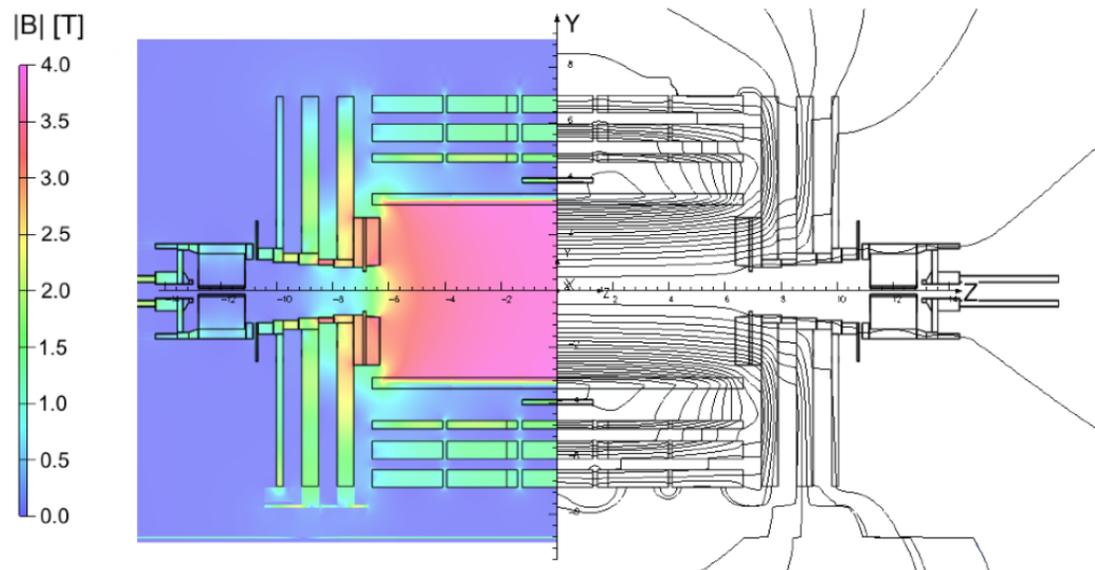


Figure 4.4: Left: Map of the magnetic field. Right: Field lines (right) predicted for a longitudinal section of the CMS detector by a magnetic field model at a central magnetic flux density of 3.8 T. Each field line represents a magnetic flux increment of 6 Wb. [10].

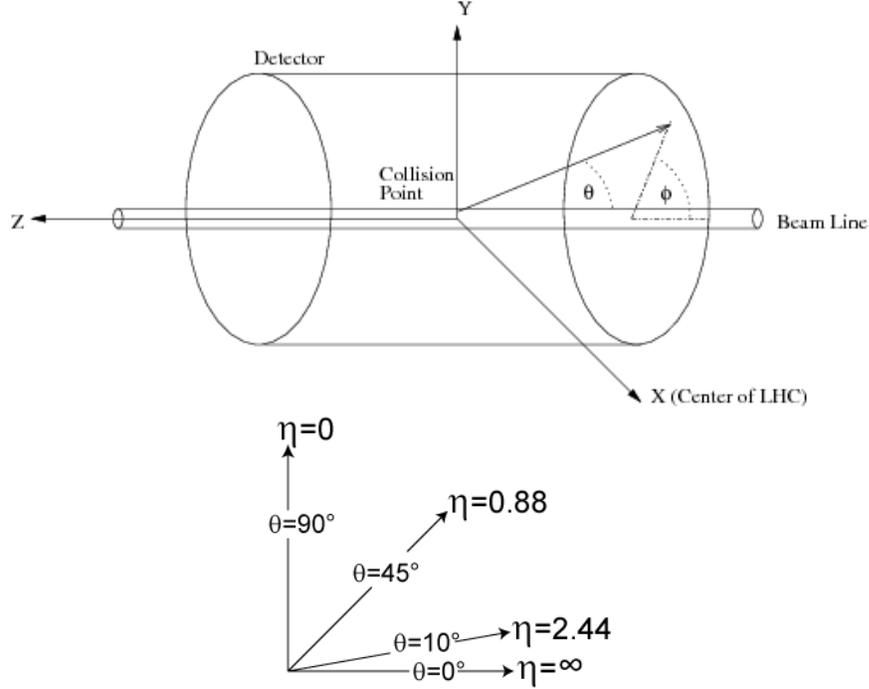


Figure 4.5: Illustrations of coordinates used by the CMS experiment. [11].

b-quark jets. The pixel detector measures three-dimensional (3D) coordinates of hits from charged particles interacting with the silicon pixel sensors as they pass through the detector. The detector has hit spatial resolutions of about $10 \mu\text{m}$ in the transverse coordinate and $20\text{-}40 \mu\text{m}$ in the longitudinal coordinate [13].

This innermost CMS sub-detector is composed of 1440 pixel detector modules consisting of 66 million pixels, and it adopts the $n\text{-on-}n$ silicon sensor concept [12]. A pixel cell is $100 \times 150 \mu\text{m}^2$ in size and the modules are arranged in three 53-cm-long barrel layers (BPix) at $r = 4.4, 7.3,$ and 10.2 cm and also arranged in endcap layer disks (FPix), which consist of 2 disks at each end at $z = \pm 34.5$ and $z = \pm 46.5 \text{ cm}$, extending from ≈ 6 to 15 cm in radius. This arrangement provides three tracking points for a particle's trajectory for most of the whole pseudorapidity range of $|\eta| < 2.5$. Figure 4.6 shows sketches of the CMS pixel detector layout and Figure 4.7 shows the parts of the BPix detector module.

4.2.3 The silicon strip tracker

The CMS strip tracker subdetector was designed to enable the determination of the trajectories of charged-particles with good momentum resolution and reconstruction efficiency. It is composed of 15148 strip detector modules, which has a total of 9.3 million strips. The elements of the strip tracker are single-sided $p^+\text{-in-}n$

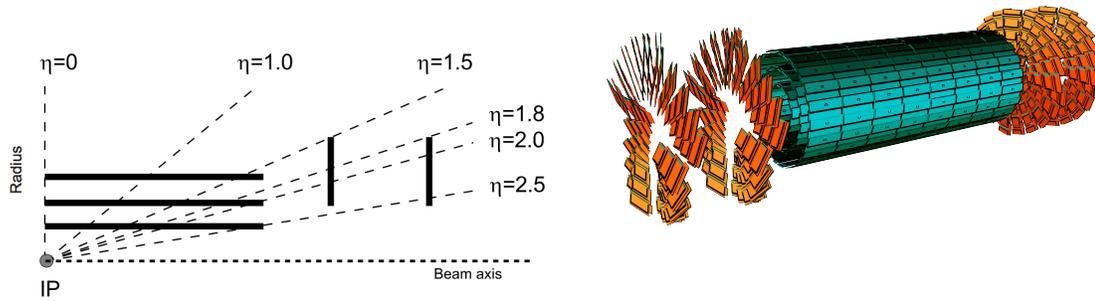


Figure 4.6: Geometrical layouts of the CMS pixel detector [12].

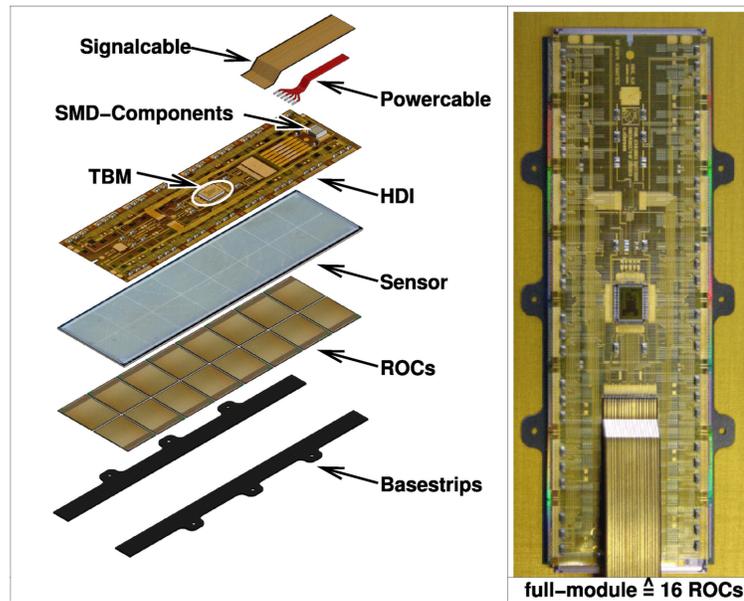


Figure 4.7: Exploded view of the barrel pixel detector showing the parts of the module, where SDM is the surface mounted device, TBM is the token bit manager chip which controls several read-out chips (ROCs), and HDI is the high density interconnect. [12].

silicon sensors. The modules are arranged in four subsystems: the tracker inner barrel (TIB), outer barrel (TOB), inner disk (TID), and end caps (TEC). The modules typically have dimensions of $6 \times 12 \text{ cm}^2$ in the inner barrel and $10 \times 9 \text{ cm}^2$ in the outer barrel.

The TIB consists of four cylindrical layers at $r=255.0, 339.0, 418.5,$ and 498.0 mm and they extend along $z = \pm 700 \text{ mm}$. The strip pitch (ie. the distance between neighboring strips) on average varies between $80\text{-}120 \text{ }\mu\text{m}$. The TID modules are arranged in three disks at the ends of the TIB between $z = \pm 800$ and $z = \pm 900 \text{ m}$ and the strip mean pitch varies between $100\text{-}141 \text{ }\mu\text{m}$. TIB and TID together provide position measurements with resolution of about $13\text{-}38 \text{ }\mu\text{m}$ in the $r\phi$ space [13].

The TOB consists of 6 cylindrical barrel layers at $r = 608, 692, 780, 868, 965, 1080 \text{ mm}$ and extends along $z = \pm 118 \text{ cm}$. The strip mean pitch varies between $122\text{-}183 \text{ }\mu\text{m}$. Its position measurement resolution in $r\phi$ is approximately $18\text{-}47 \text{ }\mu\text{m}$ [13].

The TEC consists of nine disks, each with 7 concentric rings that hold the modules. Disks 1-3, counting from smaller $|z|$, have modules in all 1-7 rings, counting from smaller r . Disks 4-6 only has modules in rings 2-7, disk 7-8 has modules only in rings 3-7, and disk 9 has modules only in rings 4-7. The TEC extends from $r = 220$ to 1135 mm and along $z = \pm 2800 \text{ mm}$. The strip pitch varies between $97\text{-}184 \text{ }\mu\text{m}$ on average. In terms of its position measurement resolution, the TEC has a similar resolution as the TOB [13].

Two different silicon sensor thicknesses have been used. The inner four layers of the barrel, the inner disks, and the inner four of the end cap disks use sensors of $320 \text{ }\mu\text{m}$ thick. Otherwise, for all modules at positions $r > 60 \text{ cm}$, the sensor thickness is $500 \text{ }\mu\text{m}$.

The strips are oriented along the beam axis on the barrels, and along the radial direction on the end caps. The first two layers of the TIB, TID, TOB, and ring 1,2 and 5 of the TEC are double-sided modules where two strip sensors are mounted back-to-back with a stereo angle of 100 mrad . These double-sided modules provide a z coordinate measurement for the barrel modules and r coordinate measurement for the end caps module, enabling a precise 3D measurement of a hit position with a resolution of $230 \text{ }\mu\text{m}$ ($530 \text{ }\mu\text{m}$) in TIB (TOB) and a resolution that varies with pitch in TID and TEC.

With the four subsystems, the strip tracker ensures at least ≈ 9 hits on a particle's trajectory where at least ≈ 4 of them are 3D hit position measurements by the stereo modules, in the full range of $|\eta| < 2.4$,

Figure 4.8 shows a schematic cross section of the top half of the CMS tracker, which illustrates how the silicon tracker modules are arranged. Figure 4.9 shows a diagram and photograph of a CMS silicon strip module.

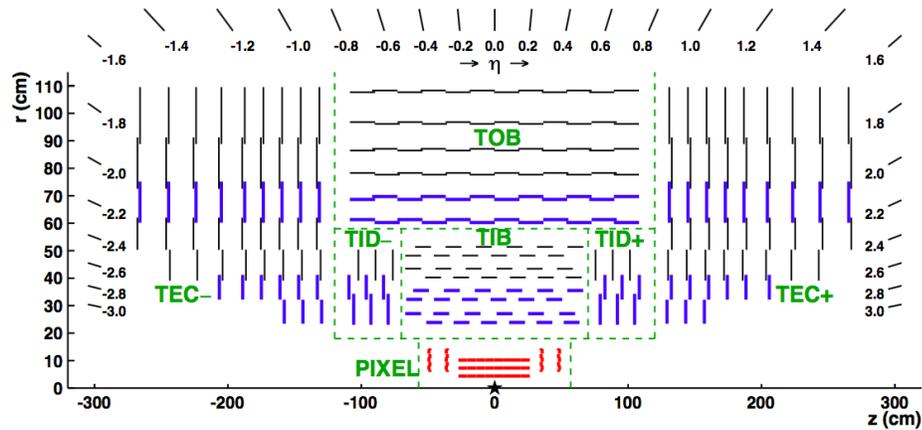


Figure 4.8: Schematic cross section through the top half of the CMS tracker in the r - z plane. The star at the center indicates the approximate position of the pp collision. Green dashed lines are imaginary lines that partition the tracker into the four subsystems. Thin black lines indicate strip tracker modules that provide 2-D hits and thick blue lines indicate strip tracker modules that provide reconstruction of hit positions in 3-D, by having two back-to-back strip modules where one module is rotated by a 'stereo' angle. The red lines indicate the pixel detector modules which also provide 3-D hits. [13].

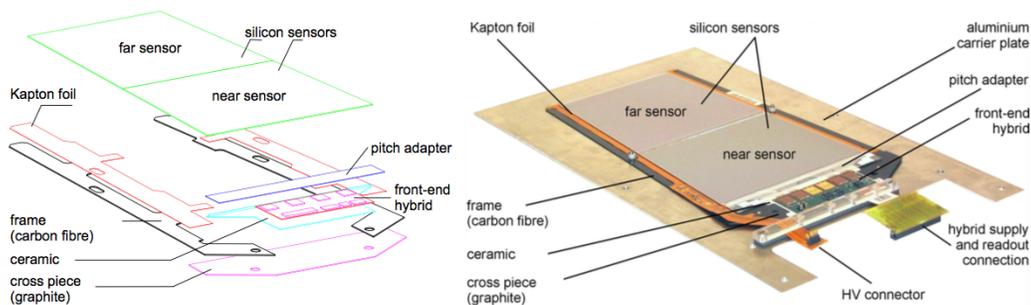


Figure 4.9: Left: Diagram of silicon strip tracker module housing two sensors. Right: photograph of a TEC subsystem sensor module, mounted on a carrier plate. [12].

Together with the pixel detector, and other pertaining services such as cables, support, cooling, the CMS tracker represents a substantial amount of material. The thickness is at most 0.5 interaction lengths (λ_I) or 1.8 radiation lengths (X_0). Figure 4.10 shows the layers and thickness in λ_I and X_0 as a function of η .

At about $|\eta| = 1.5$, there is a probability of 85% that an electron will emit a bremsstrahlung photon or a photon will convert to an electron-positron pair. For a hadron at $|\eta| = 1.5$, there is a probability of 20% that it will experience a nuclear interaction before it reaches the ECAL.

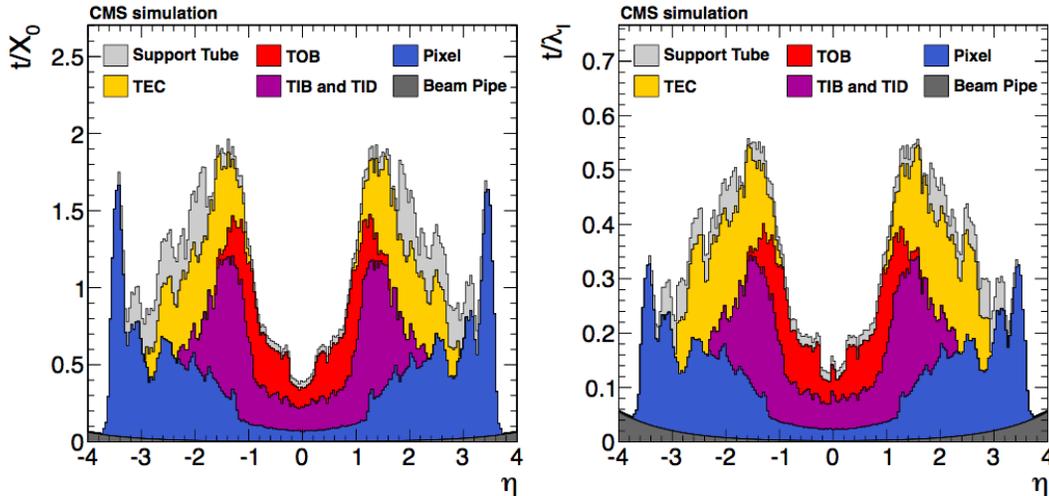


Figure 4.10: Total thickness t of the CMS pixel and strip tracker material expressed in units of interaction lengths λ_I (left) and radiation lengths X_0 (right), as a function of the pseudorapidity η . [13].

4.2.4 The electromagnetic calorimeter (ECAL)

The electromagnetic calorimeter is a homogeneous, hermetic, crystal calorimeter sub-detector designed to measure the energy of particles, particularly electrons and photons. It is composed of 61200 lead tungstate (PbWO_4) crystal blocks, of approximately $2.2 \times 2.2 \times 23 \text{ cm}^3$, grouped in 36 super-modules in the barrel (EB) region and 7324 crystal blocks, of approximately $3 \times 3 \times 22 \text{ cm}^3$, in 4 half circle units (Dee) at each end cap (EE) region surrounding the silicon strip tracker. The barrel covers a pseudorapidity range of $|\eta| < 1.48$ and the end cap covers $1.48 < |\eta| < 3.0$. Following the ECAL endcaps, a pre-shower detector based on lead absorber and silicon strips sensors is installed, covering $1.65 < |\eta| < 2.6$, mainly for detecting and rejecting neutral pions. Figure 4.11 shows a schematic diagram of the ECAL.

The PbWO_4 crystals have a high density of 8.28 g/cm^3 , a short radiation length of $X_0 = 0.85 \text{ cm}$, and a

Molière radius of $R_M = 2.19$ cm. These features allow the calorimeter to be compact yet high in granularity. The homogeneous medium allows for a better energy resolution by minimizing sample fluctuations [77]. The energy resolution, for electrons between 20 to 250 GeV having a central impact on a 3×3 array crystal, typically has the form $\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.128 \text{ GeV}}{E}\right)^2 + (0.3\%)^2$, where the first, second, and third terms are the stochastic, the noise, and the constant term respectively, and E is the energy measured in GeV [14, 77].

With the chosen sizes, the crystals can contain more than 98% of the energy of electrons and photons up to 1 TeV and cause up to two thirds of hadrons traversing through them to shower [78].

The crystal blocks scintillate as electromagnetic particles traverse through them. The scintillating light is captured by silicon avalanche photodiodes (APDs) in the barrel region and by vacuum phototriodes (VPTs) in the endcap region [12].

The ECAL has a good energy resolution and an angle acceptance that matches the tracker's coverage of $|\eta| < 2.5$. Figure 4.12 shows the energy resolution, for incident electrons, as measured in a beam test.

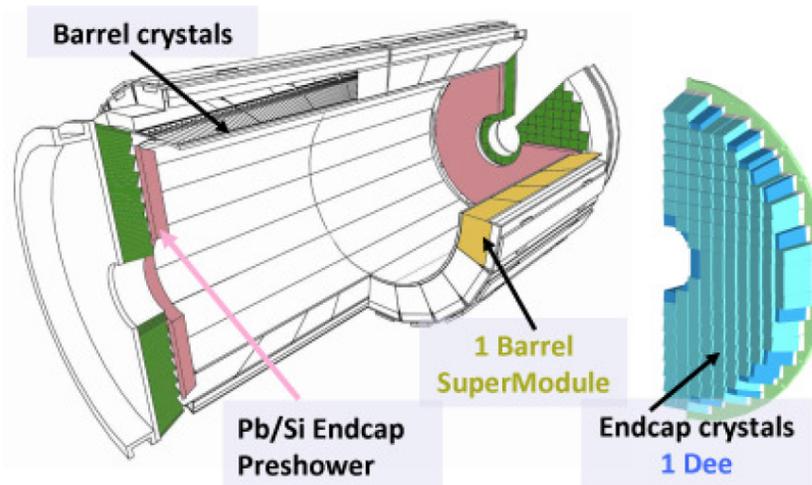


Figure 4.11: Schematic view of the CMS electromagnetic calorimeter [14].

4.2.5 The hadron calorimeter (HCAL)

The HCAL is a sampling calorimeter built with the main purpose of measuring the energy of hadrons. The absorber material is brass and the active material is plastic scintillator. It surrounds the ECAL and it radially extends from $r = 1.77$ m to $r = 2.95$ m. It has four subsystems: the barrel (HB), endcap (HE),

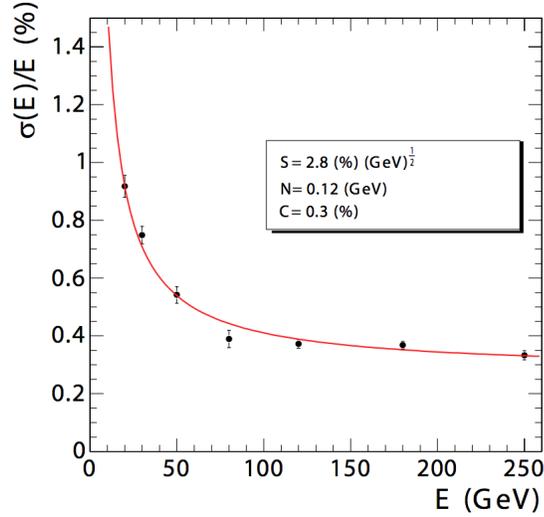


Figure 4.12: ECAL energy resolution as a function of electron energy. [12].

outer (HO) and forward (HF) calorimeters. HB and HE are installed within the volume of superconducting solenoid, the HF is placed at about $z=11.2$ m and HO is placed out of the solenoid because of limited space between ECAL and the solenoid. In HB, there are 36 brass/scintillator wedges with 17 longitudinal layers of 5 cm of brass and 3.7 mm of scintillator. In HE, there are 19 longitudinal layers of 8 cm of brass and 3.7 mm of scintillator. HO consists of scintillator tiles with 1 or 2 longitudinal layers of 10 mm scintillator and covers an $|\eta| < 1.3$ range. HF is made out of cylindrical steel structure absorber / quartz fiber and covers a $3.0 < |\eta| < 5.0$ range. The HB and HE cover $|\eta| < 1.3$ and $1.3 < |\eta| < 3.0$, respectively. Figure 4.13 shows the schematic diagram of the HCAL.

Most particles will start showering in the ECAL and therefore the response and resolution of the CMS calorimeter depends on both the HCAL and the ECAL. The hadronic energy resolution of the HB and the EB combined is measured to be $\left(\frac{\sigma}{E}\right)^2 = \left(\frac{0.847 \pm 0.016}{\sqrt{E}}\right)^2 + (0.074 \pm 0.008)^2$ [79, 15], where the first and second terms are the stochastic and the noise terms respectively. The endcaps have similar energy resolutions to the barrel.

4.2.6 The muon detector

The muon detector is designed to detect muons and to reconstruct their momentum and charge over the whole kinematic range of the LHC. It is integrated with the return iron yoke of the magnet system and it consists of three different gas-based detectors: the drift tube chamber (DT), which uses 85%/15% of Ar/CO₂

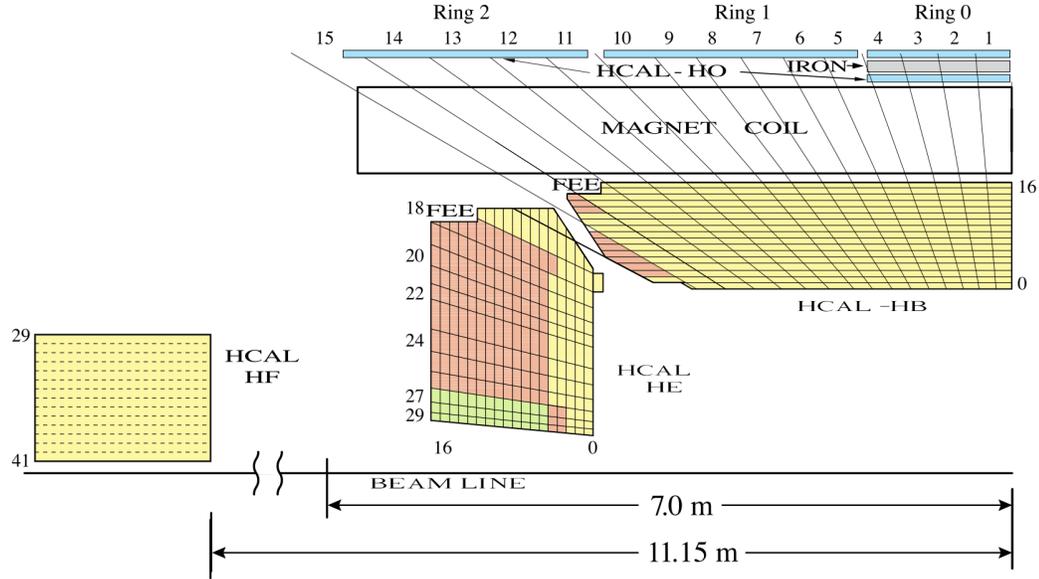


Figure 4.13: A schematic drawing of one quadrant of CMS showing the HCAL and the locations of the HB, HE, HO, and HF [15].

gas mixture, the cathode strip chamber (CSC), which uses 50%/40%/10% $\text{CO}_2/\text{Ar}/\text{CF}_4$ gas mixture, and the resistive plate chamber (RPC), which uses 95.2%/4.5%/0.3% of $\text{C}_2\text{H}_2\text{F}_4/i\text{-C}_4\text{H}_{10}/\text{SF}_6$ gas mixture.

The DTs are installed in the barrel region surrounding the solenoid and cover the pseudorapidity range of $|\eta| < 1.2$. The CSC are installed in the end cap regions and is capable of detecting muons in the $0.9 < |\eta| < 2.4$ range. These muon detectors have position measurement resolutions of about 75-150 μm [10]. In addition, both the DT and CSC subsystems can trigger on the muon transverse momentum with resolutions of about 15% in the barrel and 25% in the end cap [12], independent of the other CMS subdetectors. The RPCs are detectors solely dedicated for triggering. The RPCs have sharp transverse momentum thresholds, faster responses but coarser position measurements than the DTs and CSCs. The RPCs are installed both in the barrel region and the end cap region, and covers the pseudorapidity range of $|\eta| < 1.6$. Figure 4.14 shows a diagram of the arrangements of each of the muon detectors with respect to other sub-units of the CMS detector.

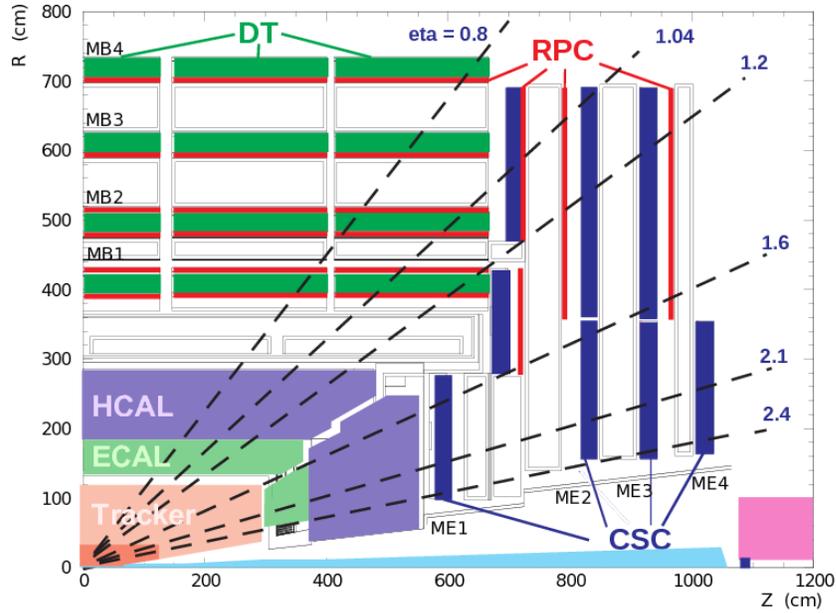


Figure 4.14: A schematic drawing of one quadrant of CMS showing the muon system [16].

4.3 Trigger systems in CMS

The LHC provides pp bunch collisions every 25 ns, which amounts to 40 million bunch collisions per second. Only a very small fraction of those collisions will produce the interesting physics signatures that we call events. CMS employs a two-step triggering system to reduce the collision rate down to, first, just below 100 kHz, done by the hardware-based Level-1 (L1) trigger which, further down to 1 kHz, accomplished by the software-based high level trigger (HLT). Only the output events from the HLT trigger will be recorded and stored.

The L1 trigger system considers information only from the calorimeter system and the muon system. This system consists of custom-designed, mostly programmable electronics that decides whether or not to keep an event in about 3.8 μ s. On the other hand, the HLT system considers the full CMS detector information. The software-based HLT is implemented in a computer processing farm separate from the detector. Between 2015 and 2016 the CMS trigger system was upgraded and improved in order to accommodate the harsh beam conditions because of the increased beam luminosity during the 2016 data taking period [80], on which data this thesis is based.

4.4 Particle identification and physics objects reconstruction

The CMS was designed such that different particles would leave distinct signatures in the detector. These signatures consists of combinations of *hits* registered in the silicon trackers and energy deposits in the calorimeters. In the process of identifying these particles, “physics objects” are reconstructed using the hits and the energy deposits. There are five categories of stable particles that CMS can directly identify: muons, electrons, photons, charged hadrons and neutral hadrons.

A muon will traverse the CMS detector and be detected as a series of hits or a *track* in the silicon detectors and/or the muon chambers. It will deposit little or no energy in the calorimeters.

When traversing the thick materials of the CMS tracker, an electron will often emit bremsstrahlung photons, and photons will convert into electrons-positron pairs, before arriving at the ECAL. This creates an electromagnetic shower detected in the ECAL as a group of clusters. For this reason, electrons and photons will both have a similar signature. However, an electron will leave an associated track in the the silicon detectors whereas photons will not. In addition, most of the electromagnetic energy of the electron or photon will mainly be collected by the ECAL.

A charged hadron will create a track and either start to hadronize in the silicon detector or in the calorimeters and it will generally deposit energy clusters both in the ECAL and HCAL. In contrast, a neutral hadron only deposits clusters of energy in the ECAL and HCAL with no registered hits in the tracker. Quarks and gluons created in the collisions will manifest themselves as *hadronic jets* detected as collection of tracks and energy clusters.

A neutrino will traverse through the detector without any interaction but it can be indirectly detected or inferred by the magnitude of the missing transverse momentum vector, or also known as missing transverse energy, $|\vec{p}_T^{\text{miss}}|$. All other neutral particles that are not part of the currently known SM particle zoo will also pass through CMS undetected, just as a neutrino.

The electromagnetic charge of a particle is determined by measuring the direction of curvature of the track, and the momentum of a charged particle is determined by measuring the degree of curvature of the tracks due to the magnetic field.

Figure 4.15 shows a sketch of a transverse slice the CMS detector showing the trajectories and energy deposits of various particles.

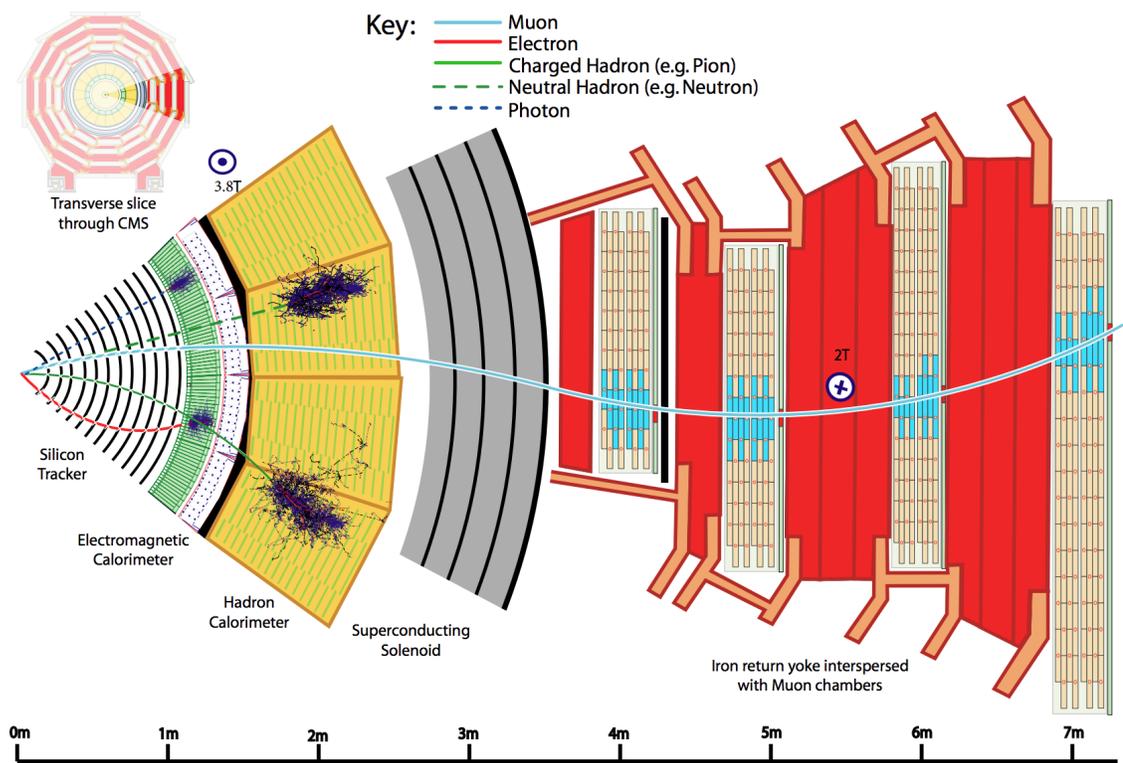


Figure 4.15: Sketch of a transverse slice the CMS detector with illustrations of how various particles would traverse and leave energy deposits in the detector [17].

The following sub-sections describe briefly the software algorithm CMS uses to perform particle reconstruction, and the various methods for reconstructing specific physics objects, including tracks, muons, electrons, photons, jets, missing transverse energy, primary vertex, and b-jet identification.

4.4.1 Particle-flow algorithm

For particle reconstruction and identification, the CMS experiment employs the software called the *particle-flow* (PF) reconstruction algorithm [17]. The main feature of the algorithm is that it utilizes (global) information from all CMS sub-detectors in identifying what physics objects the hits and clusters represent, as opposed to using only (local) information from a particular of sub-detector. The PF algorithm optimizes particle identification and measurements of their kinematic properties. The resulting PF objects are used to create higher level physics objects such as jets and missing transverse energy.

4.4.2 Track reconstruction

Track reconstruction uses the hits to obtain measurements of the momentum and position parameters of the charged particles that created the hits. There are three stages to the reconstruction [13]. The first is seed generation using a few hits in the pixel tracker. Second, extrapolating a trajectory by pattern recognition using an algorithm based on a Kalman-filter technique [81] to gather hits from all tracker layers, a process known as track-finding. The third is the track-fitting process that consists of the final fit, also based on a Kalman filter, to determine the best possible estimate of the charged particle properties: origin, transverse momentum, and direction.

4.4.3 Muon reconstruction

For muon reconstruction, tracks are first reconstructed in two independent ways. One is using only the pixel and silicon strip trackers (*tracker track*), and the other is using only the muon chamber (*stand-alone muon track*). Using these tracks, the following muon objects that can be reconstructed: *global muon* and *tracker muon*. A global muon is reconstructed (outside-in) by taking a stand-alone muon track and matching it with a tracker track, using the Kalman-filter technique [81]. A tracker muon is reconstructed (inside-out) using tracker tracks and extrapolating them to the muon chamber where at least one muon segment is matched to the extrapolated track [82]. A PF muon is reconstructed by using information from both global and tracker

muon properties and applying further selection criteria [17].

Figure 4.16 shows the expected the CMS transverse-momentum resolutions of muons reconstructed only using the inner tracker, the muon chamber (or muon system), and using both (full system).

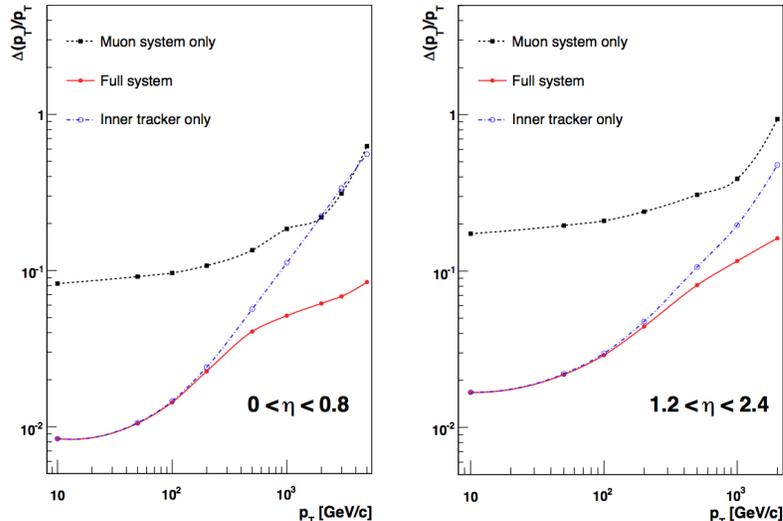


Figure 4.16: Muon transverse-momentum resolutions as a function of the muon transverse momentum measured using the CMS tracker. [12].

4.4.4 Electron reconstruction

Electron reconstruction uses information both from the inner tracker and the ECAL. The ECAL reconstructs the energy deposits into clusters and these are further grouped into a *supercluster* in a small window in η and an extended window in ϕ around the electron direction, taking into account the azimuthal bending of the electron due to the magnetic field [17]. However, some electrons will be missed by this small window and so a tracker-based electron reconstruction is also used. Using a dedicated modeling of the electron energy loss to the bremsstrahlung radiation in the tracker material, electron trajectories are fitted with a Gaussian sum filter (GSF) [83]. Complete electron reconstruction combines variables that measure the track quality and association between the track and the supercluster properties, and the likelihood of the electron being produced in a photon conversion in the detector. As mentioned earlier in the main section, due to having a similar physics signature in detector, electron and (isolated) photon reconstruction follow similar steps except that photon candidates are seeded from ECAL clusters that do not have an associated GSF track.

4.4.5 Hadrons and jets

After muons and electrons (and isolated photons) have been identified among all the identified PF candidates, what remain are particles from jet fragmentation and hadronization. These particles can be detected as charged and neutral hadrons, non-isolated photons, and though rarely, as low-momentum-muons. Within the tracker acceptance, $|\eta| < 2.5$, ECAL clusters that are not associated with any tracks are identified as non-isolated photons and HCAL clusters that are not associated with any tracks are identified as neutral hadrons. These particles are treated as input to a jet clustering algorithm to create a PF jet object. The analysis in this thesis uses the anti-kT algorithm [84, 85] with distance parameter $R=0.4$ (AK4). To avoid clustering particles that originate from secondary proton-proton collisions (*pileup*), charge hadrons associated with pileup are removed before clustering, and the resulting jets are called PF charge-hadron-subtracted (CHS) jets.

Due to the non-uniform and non-linear response of the CMS calorimeters, measured jet energies are different from the true energy of the particle jet and therefore jet energies are corrected. A multiplicative factorized approach is applied for the correction [86]. Given the four-momentum of the raw jet, p_μ^{raw} , the corrected four-momentum of the jet, p_μ^{corr} , can be obtained by applying the following equation:

$$p_\mu^{\text{corr}} = p_\mu^{\text{raw}} \cdot C_{\text{offset}}(p_T^{\text{raw}}) \cdot C_{\text{MC}}(p_T', \eta) \cdot C_{\text{rel}}(\eta) \cdot C_{\text{abs}}(p_T'') \quad (4.4)$$

where C_{offset} is the offset correction, which is based on a measure of soft radiation, electronic noise, and pile-up, C_{MC} is the Monte Carlo simulation (MC) calibration factor that is based on simulation to correct the reconstruction jet to be equal to the the energy of the generated MC particle jets, C_{rel} and C_{abs} are the relative and absolute energy scale, respectively, and both are residual corrections that are derived using data driven method. The p_T' is the transverse momentum after applying the offset correction, and p_T'' is the transverse momentum after all previous corrections.

In addition to energy scale corrections, the jet p_T resolution is also corrected for resolution broadening as a result of extra radiation activity in a realistic collision event. It is removed by comparing to ideal case of a two-body process, both in data and in MC [86].

Figure 4.17 shows the expected resolutions of the magnitude of the jet transverse momentum vector, as a function of the magnitude in various $|\eta|$ regions.

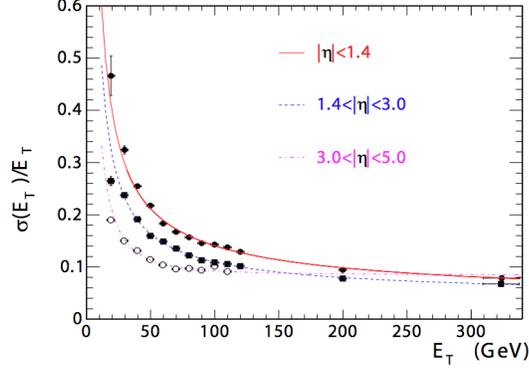


Figure 4.17: The resolutions of the magnitude of the jet transverse momentum vector ($\sigma(E_T)/E_T$), as a function of the magnitude (E_T) in various $|\eta|$ regions, measured using the CMS calorimeters. [12].

4.4.6 Missing transverse energy reconstruction

Missing transverse energy is the magnitude of the missing transverse momentum vector, \vec{p}_T^{miss} , which is a result of all the particles undetected by the detector. This vector is defined as the negative vector sum of the p_T of all final-state particles detected, in this case all PF particles. Before jet-energy-correction, it is defined as follows,

$$\vec{p}_{T,\text{PF}}^{\text{miss}} = - \sum_{i=1}^{N_{\text{particles}}} \vec{p}_{T,i} \quad (4.5)$$

and after jet-energy-correction, it is defined as,

$$\vec{p}_{T,\text{PF}}^{\text{miss}}(\text{raw}) = - \sum_{i=1}^{N_{\text{particles}}} \vec{p}_{T,i} - \sum_{j=1}^{\text{PF}_{\text{jets}}} (\vec{p}_{T,j}^{\text{corr}} - \vec{p}_{T,j}) \quad (4.6)$$

where $\vec{p}_{T,j}^{\text{corr}}$ is the corrected p_T of the PF jets and $\vec{p}_{T,j}$ is the uncorrected or raw p_T of the PF jets.

4.4.7 Primary vertex reconstruction

Using the reconstructed tracks, primary vertex (PV) reconstruction involves measuring the intersecting origin of the tracks both from the primary pp collision and from the background or secondary collisions (pileup) because of the multiple pp collisions at any given bunch crossing. The vertices originating from background pp collisions are referred to as pileup vertices.

There are three steps to all vertex reconstruction. The first is selecting the tracks, second is clustering the tracks intersect at the same interaction point, and third, using the associated tracks, position of each vertex is determined by fitting [13].

The vertex with the largest value of the summed physics-object transverse momentum squared, $\sum p_T^2$,

is taken to be the primary pp interaction vertex where the physics objects are jets with the tracks assigned to the vertex as inputs, and the associated missing transverse momentum.

4.4.8 Identification of b-quark jets

In the context of CMS, b-jets are defined as jets containing B hadron and identification of jets originating from b-quarks relies on the fact that hadrons containing a b-quark have distinctive properties compared to other hadrons that contain only lighter quarks. These properties allows b-jets to be distinguished from jets originating from lighter quarks. In particular, B hadrons have a relatively long lifetime, in the order of 1.5 ps, which results in these hadrons traveling a few mm to one cm away, depending on their momentum, from the PV before decaying. As a result, the decay products of the B hadron will have displaced tracks, with respect to the PV, from which a secondary vertex (SV) can be reconstructed. In addition, B hadrons have a relatively large mass, which causes the tracks of the decay products to have a larger momentum transverse to the jet axis in comparison to the other constituents of the jet. Moreover, there is a 20% chance that a low momentum (soft) muon will be present in the decay chain of a B hadron and the properties of this muon can also be used to help identify a b-jet.

Using the properties of a B hadron, variables related to the tracks and secondary vertices within a jet are treated as inputs to a supervised machine learning algorithm that outputs a probability measure of the likelihood the jet originated from a b-jet. In particular, the b-jet identification algorithm used in this thesis is called CSVv2 [18]. The training samples used are simulated events where the true origins of the jets are known. Several of the track variables that serves as inputs to the algorithm are the multiplicity of tracks and the closest distance between the PV and the track, which is also known as the impact parameter (IP). In addition, the input SV related variables include the multiplicity of SVs, the invariant mass calculated from the tracks that make up the SV, and the distance between the SV and PV, also known as the flight distance. For the IP and flight distance variables, the values used in the algorithm are the ratio of the values with their uncertainties and these are also known as the 'significance'. Figure 4.18 illustrates the displaced tracks, and the SV as a result of e jet originating from a b quark.

There are several b-tagging qualities that can be defined, based on the desired identification efficiency and the misidentification efficiency. For the chosen quality level (or working point) of b-tagging used in this thesis, the efficiency for tagging true b hadron jets in a $t\bar{t}$ simulation is $\approx 65\%$, averaged over jets with $p_T > 30$ GeV, and the misidentification efficiency is $\approx 1\%$ for b-tagging light-quark jets, such as jets originating

from u , d , s quarks, or gluon jets, measured in multi-jet events. Such a good level of b-tagging efficiency

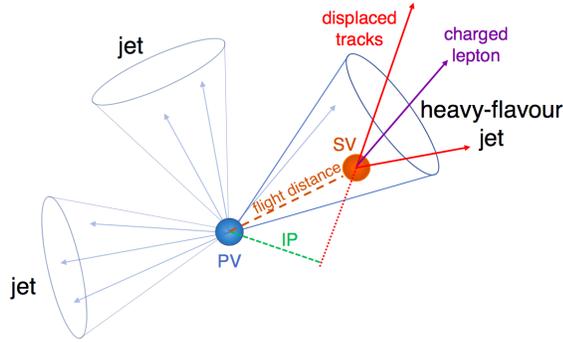


Figure 4.18: Illustration of a jet originating from a b-quark, labeled as a heavy-flavored jet, with a secondary vertex (SV) from the decay of a B hadron. The charged-particle tracks (including possibly a soft lepton) are displaced with respect to the primary vertex, resulting in a large impact parameter (IP) value. [18].

in CMS is achieved because the tracking system provides an impact parameter (IP) resolution of about 15 (30) μm at a p_T of 100 (5) GeV/c, as described in Sections 4.2.2 and 4.2.3. As a comparison, the typical IP values for tracks from b-hadron decays are in the order of few 100 μm [87].

Correction scale factor measurement for b-quark identification: LifeTime method.

The b-quark identification efficiencies in data and in simulation will not be identical, and correction scale factors are needed to be applied on the simulated samples in order to have the best agreement between data and simulation. The method used for obtaining the correction factors applied in this thesis is called the LifeTime method [18], and it uses a discriminant called the *Jet Probability* (JP). The tagging efficiency is the ratio of the number of b -jets obtained from a fit after and before applying the algorithm (with a certain working point).

$$\varepsilon_b = C_b \frac{N_b^{\text{tagged}}}{N_b}. \quad (4.7)$$

The factor C_b is a correction factor, which takes into account the fraction of jets for which the JP discriminant can be computed. It is defined as

$$C_b = \frac{n_{b,\text{MC}}^{\text{tag}}}{N_{b,\text{MC}}^{\text{tag}}} \frac{N_{b,\text{MC}}}{n_{b,\text{MC}}}, \quad (4.8)$$

with $N_{b,\text{MC}}$ the number of b -jets with JP information, $n_{b,\text{MC}}$ the number of all selected b -jets, $N_{b,\text{MC}}^{\text{tag}}$ the number of b -jets with JP information passing the algorithm working point for which the efficiency is being measured and $n_{b,\text{MC}}^{\text{tag}}$ the number of b -jets passing the tagging requirement for which the data-to-simulation scale factor is being measured, and MC refers to Monte Carlo simulation.

Other methods for identifying b-quark jets

Furthermore, despite not being utilized by the analyses in this thesis, methods for identifying high momentum jets with large distance parameters, originating from a pair of b quarks were developed and studied during the course of the PhD research. These methods were initially intended to identify jets originating from high momentum Higgs boson decaying to a pair of $b\bar{b}$, although in principle the algorithms could be used for any heavy resonances decaying to a pair of $b\bar{b}$. These studies are included in [Appendix A](#) and [Appendix B](#).

Chapter 5

Search for vector-like-quarks

decaying to leptons and jets final states

The discovery of a Higgs boson [23, 22] with a light mass points to new physics at the TeV scale. Loop corrections to the scalar H mass diverge quadratically with the calculation cut-off scale, mainly due to loops involving top quarks, W bosons, and Higgs bosons. If the standard model (SM) applies to energies past the electroweak scale, then new heavy particles are required to cancel out these contributions and stabilize the mass of the scalar boson. Little Higgs models[88, 89] predict a heavy top quark partner, T , which would cancel top quark loop contributions to the Higgs boson mass. This particle is predicted to have a mass near the TeV scale. We search here for a “vector-like” top partner which is an electroweak singlet and has vector couplings to W and Z bosons. Precision electroweak measurements from electron-positron collisions [90] place strong constraints on fourth generation quarks in the SM, but vector-like quarks are not subject to many of these constraints.

We assume that in proton-proton collisions the T quark is produced along with its antiquark, \bar{T} , through the strong interaction. Its production cross section is calculated at next-to-next-to-leading-order using perturbative QCD. The cross sections for pair production in 13TeV collisions have been calculated for T

quark masses from 800 to 1800 GeV [91].

The T quark, which has electric charge of $2/3$, can decay into three different final states: bW , tZ , or tH (Fig. 5.1), (whereas the B quark can decay into: tW , tZ or bH). For the branching fraction of 50% for $T \rightarrow bW$, and 25% for each of $T \rightarrow tZ$, tH [92], these are labelled the 'singlet' branching fractions. For the branching fraction of 50% for each of $T \rightarrow tZ$, tH [92], these are labelled the 'doublet' branching fractions (and similarly for B quark decays, the branching fraction of 50% for $B \rightarrow tW$, and 25% for each of $B \rightarrow bZ$, bH [92] are labelled 'singlet' whereas the branching fraction of 50% for each of $B \rightarrow bZ$, bH are labelled 'doublet'). The search using data taken at $\sqrt{s} = 8\text{TeV}$ set lower limits for the T quark mass between 687 and 782 GeV for all possible branching fractions into these three final states, and a lower limit of $m(T) = 696$ GeV for a T quark decaying with the nominal branching fractions [25].

In this thesis, we focus mainly on the states that produce three or more leptons (electron/muon), or trilepton, and the goal is to search for the T quark in 13 TeV collision data using the tools developed in the 8 TeV analysis [93] [94], particularly the data-driven background estimation using the matrix method [95]. In addition, we also perform measurements on the states that produce exactly two leptons with the same electrical charge, or same-sign (SS) dileptons. The SS dilepton analysis presented in this thesis uses very similar search strategies as used by the $X5/3$ pair decaying to SS dilepton search, documented and published as a CMS physics analysis summary [96]. At the end of this chapter, we will present the results combining the trilepton, SS dilepton, and the final states with exactly one lepton (single-lepton) which has been published as a CMS paper [3].

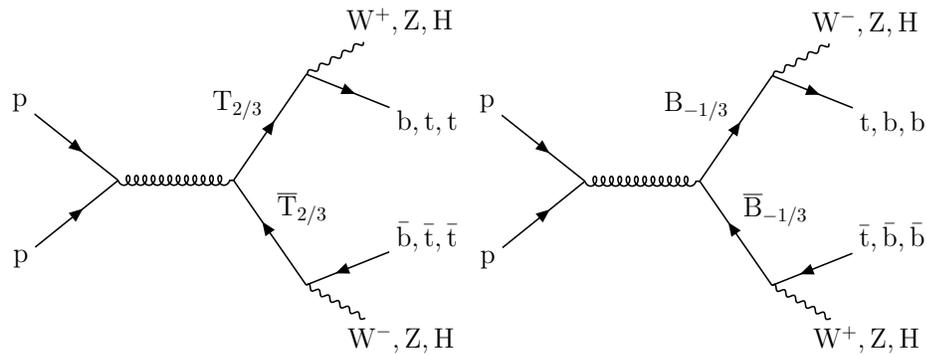


Figure 5.1: Feynman diagrams showing pair production of $T\bar{T}$ (left) and $B\bar{B}$ (right) decaying to SM particles. [3]

The strongest overall sensitivity to $T\bar{T}$ and $B\bar{B}$ production is achieved by combining the three leptonic channels, since each channel is sensitive to different VLQ decay modes. Table 5.1 shows the selection

efficiency for all three channels in each $T\bar{T}$ or $B\bar{B}$ decay mode, with respect to the total number of expected events for a given decay mode (eg., tHtH). The most sensitive decay modes for each channel are noted in bold. Comparing efficiencies across $T\bar{T}$ decay modes, the single-lepton channel has the highest efficiency for decay modes with at least one $T \rightarrow bW$ decay, the SS dilepton channel is sensitive to $B \rightarrow tW$ decays, and the trilepton channel has high efficiency for decay modes with at least one $T \rightarrow tZ$ decay.

Table 5.1: Signal efficiencies in the single-lepton, same-sign dilepton, and trilepton channels, split into the six possible final states of both $T\bar{T}$ and $B\bar{B}$ production, for three mass points. Efficiencies, stated in percent, are final yields calculated with respect to the expected number of events in the corresponding decay mode, before any selection. The most sensitive decay modes for each channel are noted in bold. The efficiency for bWbW events in the same-sign dilepton and trilepton channels is negligible, as is the efficiency for bZbZ events in the same-sign dilepton channel. [3]

$T\bar{T}$ (1.0TeV)				$B\bar{B}$ (1.0TeV)			
Decay mode	1ℓ	SS 2ℓ	$\geq 3\ell$	Decay mode	1ℓ	SS 2ℓ	$\geq 3\ell$
tHtH	9.1	1.1	0.74	bHbH	2.9	0.16	0.08
tHtZ	8.4	0.78	1.50	bHbZ	1.8	0.05	0.22
tHbW	11.0	0.61	0.29	bHtW	11.2	0.61	0.31
tZtZ	7.4	0.45	1.92	bZbZ	1.0	0.02	0.25
tZbW	9.2	0.34	0.88	bZtW	9.2	0.23	0.89
bWbW	10.8	0.02	—	tWtW	12.3	2.5	1.28
$T\bar{T}$ (1.2TeV)				$B\bar{B}$ (1.2TeV)			
Decay mode	1ℓ	SS 2ℓ	$\geq 3\ell$	Decay mode	1ℓ	SS 2ℓ	$\geq 3\ell$
tHtH	10.9	1.4	0.81	bHbH	3.2	0.19	0.08
tHtZ	10.1	0.93	1.48	bHbZ	2.0	0.08	0.19
tHbW	12.4	0.71	0.31	bHtW	12.6	0.73	0.29
tZtZ	8.8	0.53	1.98	bZbZ	1.0	0.03	0.20
tZbW	10.4	0.27	0.87	bZtW	10.4	0.28	0.87
bWbW	11.4	0.04	—	tWtW	14.1	2.8	1.33
$T\bar{T}$ (1.4TeV)				$B\bar{B}$ (1.4TeV)			
Decay mode	1ℓ	SS 2ℓ	$\geq 3\ell$	Decay mode	1ℓ	SS 2ℓ	$\geq 3\ell$
tHtH	11.7	1.5	0.81	bHbH	3.2	0.19	0.07
tHtZ	10.8	0.95	1.47	bHbZ	2.0	0.07	0.18
tHbW	13.3	0.49	0.30	bHtW	13.4	0.75	0.29
tZtZ	9.3	0.29	1.87	bZbZ	1.0	0.02	0.20
tZbW	10.9	0.75	0.85	bZtW	11.0	0.29	0.81
bWbW	11.8	0.03	—	tWtW	15.4	3.05	1.36

5.1 Trilepton final state

The trilepton channel is most sensitive to decays where one of the T quark pair decays into tZ ($T\bar{T} \rightarrow tZ + bW/tZ/tH$).

We require events to have three or more leptons passing the loose lepton ID. For signal and background simulated processes we require the events to pass the tight lepton ID requirement whereas for data we also

collect loose leptons to construct the data-driven background.

In Monte Carlo (MC) simulation we require events to have three or more leptons passing the tight ID, whereas in data we require the leptons to pass the loose ID for purpose of estimating the background events containing nonprompt leptons.

The events are differentiated into four categories based on the flavor of the three highest p_T leptons. The categories are called eee , $ee\mu$, $e\mu\mu$, and $\mu\mu\mu$.

The following sections (5.2-5.8) describe the elements involved in the trilepton search, beginning from samples, physics objects reconstruction, event selection, background estimation, systematics uncertainties and lastly the results.

5.2 Data and simulated samples

We analyze data taken during 2016 when the LHC collided protons at $\sqrt{s} = 13$ TeV, with bunch spacing of 25ns. These data samples are listed in Table 5.2. Background processes are studied using Monte Carlo (MC) simulation samples listed in Table 5.3. MC was generated with 25ns bunch spacing as part of the RunI-ISummer16MiniAODv2 campaign. Detector simulation for all MC samples is performed with GEANT4 [97]. Both data and MC were processed using CMSSW version 8_0_25.

Data events are considered for selection if they were recorded during the run B to H of 2016 runs included in the certified dataset defined by the following JSON file:

Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt.

We compare event numbers to remove duplicates of events appearing in more than one primary dataset (e.g. DoubleMu and MuonEG).

The $T\bar{T}$ signal samples (Table 5.4) are produced with Madgraph5 [98] and Pythia8 [99, 100], using equivalent branching ratios of 33% for $T \rightarrow tH$, tZ , and bW . For analysis these branching fractions are rescaled to different branching ratio scenarios. In all figures and tables we present the 'singlet' branching ratio scenario unless specified otherwise.

All samples generated with MC@NLO contain events with negative weights, introduced to cancel out double counting of next-to-leading-order correction factors during parton showering [101]. These weights are applied event-by-event in histograms and are also summed to find the effective number of events in each sample.

To simplify both limit setting and plot visualization, background samples are grouped into the following categories: diboson processes (VV), triboson processes (VVV), $t\bar{t}$ vector-boson production (TTV), and data-driven background (DDBKG). The VV group contains WZ and ZZ samples. The VVV group contain WWW, WWZ, WZZ, ZZZ processes. The TTV processes contain the rare SM processes $t\bar{t} W$ and $t\bar{t} Z$.

Table 5.2: Data sample definitions.

Primary Dataset	Reconstruction Group
DoubleEG	Run2016B-03Feb2017_ver2-v2
DoubleEG	Run2016C-03Feb2017-v1
DoubleEG	Run2016D-03Feb2017-v1
DoubleEG	Run2016E-03Feb2017-v1
DoubleEG	Run2016F-03Feb2017-v1
DoubleEG	Run2016G-03Feb2017-v1
DoubleEG	Run2016H-03Feb2017_ver2-v1
DoubleEG	Run2016H-03Feb2017_ver3-v1
DoubleMuon	Run2016B-03Feb2017_ver2-v2
DoubleMuon	Run2016C-03Feb2017-v1
DoubleMuon	Run2016D-03Feb2017-v1
DoubleMuon	Run2016E-03Feb2017-v1
DoubleMuon	Run2016F-03Feb2017-v1
DoubleMuon	Run2016G-03Feb2017-v1
DoubleMuon	Run2016H-03Feb2017_ver2-v1
DoubleMuon	Run2016H-03Feb2017_ver3-v1
MuonEG	Run2016B-03Feb2017_ver2-v2
MuonEG	Run2016C-03Feb2017-v1
MuonEG	Run2016D-03Feb2017-v1
MuonEG	Run2016E-03Feb2017-v1
MuonEG	Run2016F-03Feb2017-v1
MuonEG	Run2016G-03Feb2017-v1
MuonEG	Run2016H-03Feb2017_ver2-v1
MuonEG	Run2016H-03Feb2017_ver3-v1
Total int. lumi	35.9 fb ⁻¹

Table 5.3: Background MC sample definitions, from RunIISummer16MiniAODv2 campaign. Uncertainties include contributions from energy scale variations and PDF uncertainties.

MC for rare SM processes	Generators	Cross Section [pb]
WZTo3LNu	powheg-pythia8	4.43 [102]
ZZTo4L	powheg-pythia8	1.26 [102]
WWW_4F	amcatnlo-pythia8	0.21 [102]
WWZ	amcatnlo-pythia8	0.17 [102]
WZZ	amcatnlo-pythia8	0.056 [102]
ZZZ	amcatnlo-pythia8	0.014 [102]
TTWJetsToLNu	amcatnloFXFX-madspin-pythia8	0.204 [102]
TTZToLLNuNu_M-10	amcatnlo-pythia8	0.253 [102]
MC for checks/tests	Generators	Cross Section [pb] * k-factor
WW	pythia8	118.7 [103]
DYJetsToLL_M-50_HT-100to200	madgraphMLM-pythia8	147.4*1.23 [102]
DYJetsToLL_M-50_HT-200to400	madgraphMLM-pythia8	40.99*1.23 [102]
DYJetsToLL_M-50_HT-400to600	madgraphMLM-pythia8	5.678*1.23 [102]
DYJetsToLL_M-50_HT-600to800	madgraphMLM-pythia8	1.367*1.23 [102]
DYJetsToLL_M-50_HT-800to1200	madgraphMLM-pythia8	0.6304*1.23 [102]
DYJetsToLL_M-50_HT-1200to2500	madgraphMLM-pythia8	0.1514*1.23 [102]
DYJetsToLL_M-50_HT-2500toInf	madgraphMLM-pythia8	0.003565*1.23 [102]
WJetsToLNu	amcatnloFXFX-pythia8	61526.7 [102]
TT	powheg-pythia8	832 [102]

Table 5.4: $T\bar{T}$ signal MC samples at various mass points from RunIISummer16MiniAODv2 campaign. These samples were generated with inclusive decays of the $T\bar{T}$ with equal branching ratios (33%) for $T \rightarrow tH$, tZ , and bW . Uncertainties include contributions from energy scale variations and PDF uncertainties.

Signal	Generators	Cross Section [fb] [91]
TprimeTprime_M-800	madgraph-pythia8	196
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

5.3 Physics objects reconstruction

Event selection is performed by reading events from the miniAOD [104] file format with the LJMet software framework [105].

Dedicated event filters remove events that are affected by: known noise patterns in the HCAL, accelerator-induced particles traveling along the beam direction at large radius (up to 5m), anomalously high energy deposits in certain ECAL “superclusters” [83], ECAL cell triggers that are not performing optimally, and muon candidates with large track uncertainties matched to misreconstructed tracks or charged hadrons.

The primary vertex (PV) is reconstructed from all tracks in the event that are compatible with the LHC beam spot in x,y plane. The one with the highest $\sum p_T^2$ is selected as the primary interaction vertex.

To correct for effects of different pileup distributions in data and MC, distributions of the mean number of interactions per bunch crossing are created. In MC we apply no selection (to create the distribution of N true interactions), and in data we use the pileupCalc.py tool to filter events according to the JSON file. The minimum bias cross section is set to 69.2 mb and varied by $\pm 4.6\%$ to calculate uncertainty due to pileup reweighting.

We use the particle flow algorithm [17] to reconstruct the final state objects such as electrons, muons, jets, and missing transverse energy (p_T^{miss}).

5.3.1 Leptons

Electrons are reconstructed [83] taking into account track quality, association between the track and electromagnetic shower, shower shape, and the likelihood of the electron being produced in a photon conversion in the detector. A multivariate (MVA) discriminant is used to identify well-reconstructed electrons at two quality levels: a “tight” level with $\approx 88\%$ efficiency ($\approx 4\%$ misidentification efficiency) and a “loose” level with $\approx 95\%$ efficiency ($\approx 5\%$ misidentification efficiency).

In addition to electron identification selections, we require electrons (and muons) to be isolated using the mini-isolation variable, which is defined as the p_T sum of PF particles within a p_T -dependent cone around the lepton, corrected for the effects of pileup using the effective area of the cone [106] and divided by the lepton p_T . The radius of the isolation cone in $\eta - \phi$ space, \mathcal{R} , is determined by:

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

Using a p_T -dependent cone size allows for greater efficiency at high energies when jets and leptons are

more likely to overlap. For tight electrons mini-isolation must be less than 0.1, and for loose electrons mini-isolation must be less than 0.4.

The details of the electron selection criteria are summarized in Table 5.5

Table 5.5: Minimum MVA values for the 88% and 95% electron MVA discriminant working points (WP).

Identification	Tight	Loose
Electron $ \eta $	Min. MVA, 88% WP	Min. MVA, 95% WP
0.0 – 0.8	0.674	-0.041
0.8 – 1.479	0.744	0.383
1.479 – 2.4	0.170	-0.515
Mini-Isolation	< 0.1	< 0.4

Efficiencies of the electron identification were measured using the “tag-and-probe” method [82] and a scale factor is defined as the ratio of data efficiency to MC efficiency in bins of p_T and η . The scale factors are applied to MC as weights. Scale factors for the tight MVA ID and mini-isolation working points have been measured and are shown in Figure 5.2. We include a 1% systematic uncertainty for isolation scale factors and 2% for the ID. In addition, we apply the (official CMS) electron reconstruction correction scale factor ¹.

Muons are selected using the “tight” and “loose” muon identification criteria². Tight muons have mini-isolation < 0.1 and pass the following cuts:

- Reconstructed as “global muon” with particle-flow muon identification criteria
- A goodness of fit of $\chi^2/ndof < 10$
- Number of muon system hits > 0
- Number of chambers with matched segments > 1
- The minimum radial distance of the track with respect to the primary vertex $d_{XY} < 2$ mm
- The minimum longitudinal distance of the track with respect to the primary vertex $d_Z < 5$ mm
- Number of pixel hits > 0
- Number of tracker layers with hits > 5

¹This is provided by the CMS electron and photon physics object group (Egamma POG) [107]

²This is provided by the CMS Muon Physics Object Group [108].

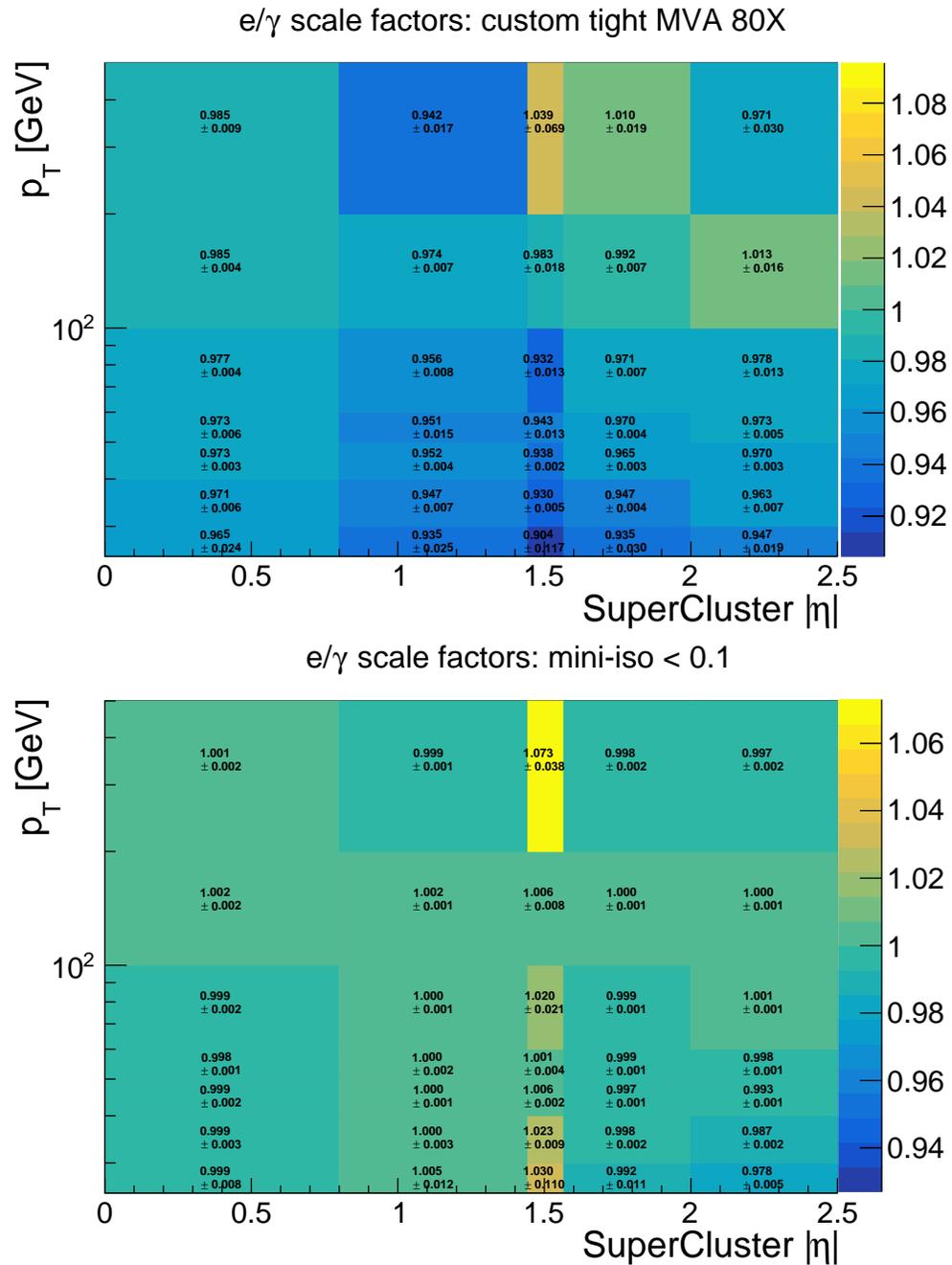


Figure 5.2: Scale factors for electron in bins of η and p_T for MVA tight (custom) working point (top) and minisolation < 0.1 (bottom).

Loose muons have mini-isolation < 0.4 , and are reconstructed as a global or tracker muon with the particle-flow muon identification criteria. Efficiencies and scale factors of the muon identification requirements and mini-isolation working points that are used in this analysis are those calculated by authors for the 13 TeV search for the exotic quark $X_{5/3}$ [96] [109]. The scale factors corrections for both the identification criteria and the isolation were derived using the tag-and-probe technique by the Muon POG [110] and muon tracks reconstruction scale factors provided by the Tracking POG released for full 2016 data are applied [111]. We estimate an uncertainty of 2% per muon for the identification criteria and 1% per muon for the mini-isolation scale factor and weight MC events by these scale factors to correct for discrepancies with the data.

5.3.2 Jets and Missing transverse energy

Jets are reconstructed using the anti-kT algorithm [112] with a distance parameter of 0.4 (AK4). Pileup contributions are corrected by removing the PF objects identified as charged hadrons from the jets. All AK4 jets with $p_T > 30$ GeV that lie within the tracker acceptance of $|\eta| < 2.4$ are considered in this search. In addition the “loose” particle flow jet identification requirements [113] are applied:

- Neutral hadron fraction < 0.99
- Neutral electromagnetic fraction < 0.99
- Number of constituents > 1
- Charged hadron fraction > 0
- Charged multiplicity > 0
- Charged electromagnetic fraction < 0.99

The selection criteria are applied to reject events containing noise and mismeasured jets. Leptons that pass loose requirements are removed from jets that have an angular separation of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.4$ with the leptons (where ϕ is azimuthal angle in radians), before jet energy corrections are applied. This is done by matching PF particles in the lepton and jet collections and subtracting the four-momentum of a matched lepton candidate from the jet four-momentum. Loose leptons, as well as tight leptons, are removed from jets because these leptons are used to estimate nonprompt lepton backgrounds.

Jets are tagged as b quark jets using a multivariate discriminant, specifically the combined secondary

vertex (CSVv2) algorithm [114], which uses information about secondary vertices within the jet. For simulated $t\bar{t}$ events, our requirement on this discriminant has an efficiency for tagging true b quark jets of $\approx 65\%$, averaged over jets with $p_T > 30\text{GeV}$. The efficiency for falsely tagging light-quark or gluon jets, measured in multijet event data, is $\approx 1\%$. Efficiency differences in data and simulation are corrected by applying scale factors, which are functions of jet p_T [114]. Uncertainties due to jet energy scale (JEC), jet energy resolution (JER), b/light (mis-)tagging scale factors uncertainties are evaluated by raising and lowering the respective uncertainty by one standard deviation.

The missing transverse momentum vector \vec{p}_T^{miss} is defined as the projection onto the plane perpendicular to the beam axis of the negative vector sum of the momenta of all reconstructed PF objects in an event. Its magnitude is referred to as p_T^{miss} . The energy scale corrections applied to jets are propagated to p_T^{miss} . We define H_T as the scalar p_T sum of all reconstructed jets in the event that have $p_T > 30\text{ GeV}$ and $|\eta| < 2.4$. In addition, we define the S_T as the scalar sum of p_T^{miss} , the p_T of leptons, and the H_T in the event.

5.4 Event selection

5.4.1 HLT Triggers

This analysis uses the following dilepton HLT triggers (Table 5.6):

- HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v,
- HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL(.DZ)_v,
- HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL(.DZ)_v,
- HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL(.DZ)_v.

Specifically for Run 2016 H dataset, we use the triggers with extra "DZ" filtering, as labelled with .DZ and highlighted in blue in Table 5.6. The DZ filter is an extra requirement on the difference of the z-coordinates at the points of closest approach with the beamline. The reason for this is because the non-DZ version is prescaled for Run 2016 H era. Using the DZ filtering ensures that the trigger objects originate from the same the primary vertex [115].

Trigger efficiency correction scale factor of $1.0 \pm 0.03(\text{sys.})$ is applied in MC based on the tag and probe calculation performed by CMS for the search particles of supersymmetry [116] decaying to trilepton final states [117], where the correction scale factors for trilepton events passing the dilepton triggers in Table 5.6

were found to very close to unity. In addition, they also have found that the DZ leg of the trigger does not significantly effect the trigger efficiency in MC, and so, following their conclusion, we use the non DZ version of the triggers in MC throughout.

Table 5.6: HLT trigger paths used for the trilepton analysis.

Event category	Trigger path
eee	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*
$e\mu\mu$	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*, HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL(<u>DZ</u>)_v*, HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL(<u>DZ</u>)_v*
$e\mu\mu$	HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL(<u>DZ</u>)_v, HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL(<u>DZ</u>)_v*, HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL(<u>DZ</u>)_v*,
$\mu\mu\mu$	HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL(<u>DZ</u>)_v

5.4.2 Offline event selections

The leptons are mostly coming from W/Z bosons and we require them to have $p_T > 30$ GeV and be centrally produced within $|\eta| < 2.4$, based on the efficiencies of the dilepton triggers listed in Table 5.6. The tight and loose lepton ID definitions for both electrons and muons are as described in Section 5.3.1.

We also require missing transverse energy, p_T^{miss} , to be greater than 20 GeV to account for the neutrinos from W decays.

Our most sensitive signal decay will produce at least 3 jets and so we use AK4 jets and we require the event to have at least 3 of them, each with $p_T > 30$ and $|\eta| < 2.4$. We also require there be at least 1 b-tagged AK4 jet in the events since each vector-like T quark pair decay will produce at least two b-quarks. For b-tagging, we require the b-tagged jets to pass the medium CSVIVFv2 working point (0.8484).

In addition, we require that if there are same-flavored opposite-sign (OS) lepton pairs among the leptons in the event any pair must produce invariant mass $M_{l(OS)} > 20$ GeV. This is to reject background leptons originating from low mass resonances.

5.4.3 Control Region and Signal Region

We define a control region for the purpose of calculating lepton fake rates (described in Section 5.5.3) and also to check for data and background prediction agreement.

We choose our control region based on the fact that our signal events will produce at least 3 jets. Based

on the jet multiplicity distribution plotted in Figure 5.3 and the corresponding background fraction values in Table 5.7, we observe that events with exactly 2 jets contains effectively little or no signal and has background composition that is more similar to the signal region (in comparison to events with exactly 1 jet). Therefore we take the events having exactly 2 jets in addition to all other selections, described in Section 5.4.2, to be our control region (labelled 'CR2'). For cross checking purposes, we also define the region with exactly 1 jet (labelled as 'CR1').

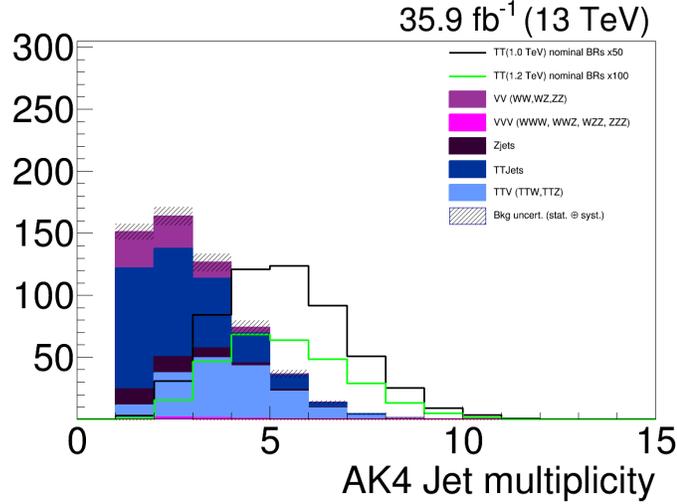


Figure 5.3: Jet multiplicity distribution using only Monte Carlo to simulate background, after selection (omitting jet multiplicity requirement). Only statistical + lumi + lepton Iso + lepton ID uncertainties are included in the plot.

Table 5.7: Monte Carlo background prediction in percentages. In the table, V stands for W or Z vector boson. In addition we also show the signal (TT_{M1000}) significance in each regions.

Background (MC)	NJet=1 (CR1)	NJet=2 (CR2)	NJet \geq 3 (SR)
VV	19.35	15.99	7.52
VVV	0.33	1.09	1.04
V +jets	8.58	7.84	5.09
$t\bar{t}$ +jets	64.43	53.32	37.56
$t\bar{t}$ + V	7.29	21.76	48.80
ewk (VV, VVV, V +jets)	28.26	24.91	13.66
top ($t\bar{t}$ +jets, $t\bar{t}$ + V)	71.72	75.08	86.37
$\frac{TT_{M1000}}{\sqrt{bkg}}$	-	0.05	0.64

Summary of (common) offline selection criteria:

- Lepton $p_T > 30$ GeV
- Lepton $|\eta| < 2.4$
- AK4 jets $p_T > 30$ GeV
- AK4 jets $|\eta| < 2.4$
- $p_T^{\text{miss}} > 20$ GeV
- at least 1 b-tagged AK4 jets
- if there are same-flavored OS lepton pairs: $\min(M_{l(O_S)}) > 20\text{GeV}$
- 3 or more leptons

Addition selection for control region on top of the common criteria:

- exactly 2 AK4 jets.
- specifically for lepton fake rate measurements: exactly 3 leptons.

Addition selection for signal region on top of the common criteria:

- 3 or more AK4 jets

5.5 Background Estimation

We consider two categories of background:

- Prompt background: The background in this category are events originating from Standard Model processes that have three or more leptons in the final state. These include diboson (WZ and ZZ) processes which we label VV, triboson (WWW, WZZ, WWZ and ZZZ) processes, which we label VVV and lastly top quark pair plus boson ($t\bar{t} + W$, $t\bar{t} + Z$) processes which we label TTV. We use Monte Carlo samples to predict the amount of prompt background we expect to observe.
- Nonprompt background: The background in this category are events containing nonprompt leptons passing the tight lepton ID criteria and jets faking a lepton, such as trilepton events coming from top quark pair with jets, or Z with jets processes. We use the events containing "loose" electrons in data to predict the contamination of these non prompt fake leptons. The data-driven nonprompt background estimation method used here is identical to the one used by previous analyses [93] [94] [118]. We briefly described the method in the following section. (For more details see reference [95]).

5.5.1 Matrix method for trilepton nonprompt background estimation

Given that we know the true number of prompt (p) and nonprompt (f) leptons, let the matrix \mathbf{M} relate the number of trilepton combinations consisting of prompt and nonprompt leptons with the number of trilepton combinations passing the tight (N_t) and loose (not tight) ID's (N_l) as described by Eq 5.2. The matrix \mathbf{M} consists of functions of p and f , the rate for prompt leptons to pass the tight lepton ID selection and the rate for a nonprompt lepton to pass the tight lepton ID criteria, respectively.

$$\begin{pmatrix} N_{ttt} \\ N_{ttl} \\ N_{lll} \end{pmatrix} = \mathbf{M} \begin{pmatrix} N_{ppp} \\ N_{ppf} \\ N_{pff} \\ N_{fff} \end{pmatrix} \quad (5.2)$$

As an example, for the eee or $\mu\mu\mu$ case, the matrix \mathbf{M} can be written as follows:

$$M = \begin{pmatrix} p^3 & p^2 f & p f^2 & f^3 \\ 3p^2(1-p) & p^2(1-f) + 2p(1-p)f & 2pf(1-f) + (1-p)f^2 & 3f^2(1-f) \\ 3p(1-p)^2 & 2p(1-p)(1-f) + (1-p)^2 f & p(1-f)^2 + 2(1-p)f(1-f) & 3f(1-f)^2 \\ (1-p)^3 & (1-p)^2(1-f) & (1-p)(1-f)^2 & (1-f)^3 \end{pmatrix} \quad (5.3)$$

By taking the inverse of \mathbf{M} , as shown in Eq 5.4, we could obtain the number of trilepton events originating from nonprompt background, which we label *fakes*, given the number of trilepton combinations passing the tight and loose (not tight) ID's, as described in Eqs 5.5- 5.7.

$$\begin{pmatrix} N_{ppp} \\ N_{ppf} \\ N_{pff} \\ N_{fff} \end{pmatrix} = \mathbf{M}^{-1} \begin{pmatrix} N_{ttt} \\ N_{ttl} \\ N_{lll} \end{pmatrix} \quad (5.4)$$

To find out how many trilepton events which pass the tight ID selection are actually nonprompt, using Eq 5.2 we can write the following:

$$N_{ttt} = \text{real} + \text{fakes} \quad (5.5)$$

$$N_{ttt} = \mathbf{M}[1, 1] \cdot N_{ppp} + (\mathbf{M}[1, 2] \cdot N_{ppf} + \mathbf{M}[1, 3] \cdot N_{pff} + \mathbf{M}[1, 4] \cdot N_{fff}), \quad (5.6)$$

where the terms in square brackets refer to the row and column of the matrix, respectively. Therefore we find the fakes or nonprompt background estimate by the expression:

$$fakes = (\mathbf{M}[1, 2] \cdot N_{ppf} + \mathbf{M}[1, 3] \cdot N_{pff} + \mathbf{M}[1, 4] \cdot N_{fff}) \quad (5.7)$$

where N_{ppf} , N_{pff} , and N_{fff} can be found using equation (5.4).

5.5.2 Prompt rates

We use the lepton prompt rates that were measured by the 13 TeVsearch for vector-like-quark $X_{5/3}$ [96]. The prompt rates were measured by using the tag and probe method in Drell-Yan events and are defined to be the number of probes passing the tight ID divided by the total number of events that have passed the loose ID. The numbers we use for electron prompt rates are summarized in Table 5.8. For muons, the prompt rate was found to be 0.943 ± 0.001 .

Table 5.8: electron prompt rates

$p_T(GeV)$	prompt rates
30 – 40	0.904
40 – 50	0.928
50 – 60	0.934
60 – 70	0.942
70 – 80	0.947
80 – 90	0.953
90 – 100	0.955
100 – 125	0.948
125 – 150	0.951
150 – 200	0.946
200 – 300	0.935
300 – 400	0.920
400 – 500	0.902
> 500	0.800

5.5.3 Fake rates

The lepton fake rates used in this analysis are obtained from a measurement in the control region, described Section 5.4.3, in events that have exactly three leptons.

The lepton fake rates were obtained by fitting the predicted background (data driven background distribution added with the MC background) to data. A χ^2 statistic for the fit is computed by

calculating the weighted sum of the differences squared between the number of observed events and the number of expected / predicted events in each bin. Specifically, we use the bins (i) in the leptons p_T distribution to calculate χ^2 as described in Eq 5.8, where the inputs (p, f) are the prompt rates and fake rates, N_{data} is the number events in data, N_{NP} is the number of events resulted from the nonprompt data-driven background using matrix method and N_{MC} is the sum of the yields of $VV + VVV + t\bar{t}t + V$ MC events.

$$\chi^2(p, f) = \sum_i \frac{(N_{data}^i - (N_{NP}^i(p, f) + N_{MC}^i))^2}{N_{NP}^i(p, f) + N_{MC}^i} \quad (5.8)$$

The explicit formula for $N_{NP}^i(p, f)$ can be read indirectly from Equation 5.7, where the M's are in terms of the known prompt rates, p , and fake rates, f , (see Equation 5.3), and $N_{p/f}$ are in terms of the known $N_{t/l}$ in data (see Equation 5.4). In essence, we infer what the fake rates would be given the prompt rates, the number of $t\bar{t}t, ttl, tll, ll$ events observed in data, and given the prompt backgrounds modelled by the MC samples (where we only consider $t\bar{t}t$ events in MC).

Effectively, N_{NP} is the value that varies while fitting the background to N_{data} . We calculate χ^2 values while varying the fake rates, f , (of both electron and muon) from 0.01 to 0.5, keeping the prompt rates, p , constant at the values shown in Table 5.8. This effectively gives us χ^2 values in 2D. We perform the χ^2 calculation in each of the four triplepton categories and take the sum. Figure 5.4 shows the result of the 2D χ^2 values and Figure 5.7 shows the lepton p_T distributions used to measure the fake rate, where they are plotted with the fake rates that correspond to the minimum χ^2 point. Table 5.9 shows the yields corresponding to the plots.

To compute the uncertainty, we convert the χ^2 values into a normalized 2D Gaussian probability distribution, $\chi^2 \rightarrow P(\mu_{FR}, e_{FR}) \propto \exp(-\chi^2/2)$, where FR stands for fake rate or misidentification rate. Then we marginalize, ie. take the sum of the distribution in one direction in order to obtain a 1D gaussian for one parameter, which we use to find the 1 sigma uncertainty of that parameter. See Figure 5.5 for the 2D probability distribution and Figure 5.6 for the results regarding the uncertainties.

We find that the minimum of the χ^2 corresponds to the lepton fake rates of 0.20 ± 0.02 and 0.14 ± 0.01 , for electron and muon fake rates respectively. The χ^2 minimum value is 41.2 (or $\chi^2 /$

ndf = 1.58)³.

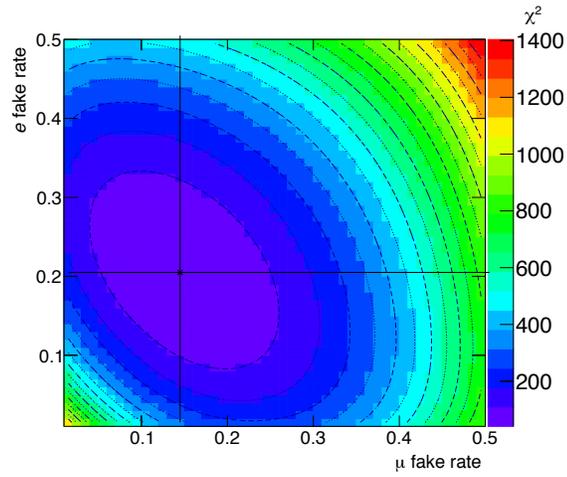


Figure 5.4: χ^2 values for various electron and muon fake rate values. The intersection of the horizontal and vertical lines indicate the minimum value.

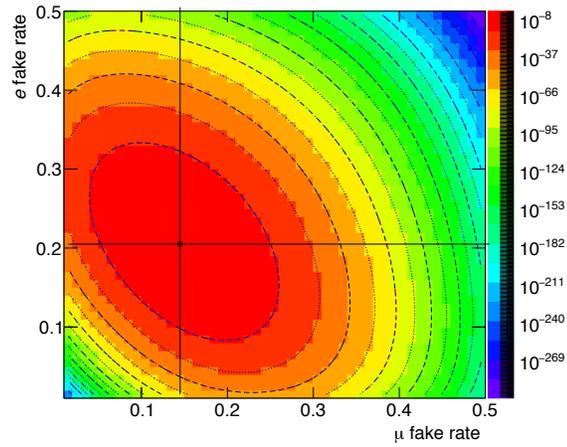


Figure 5.5: 2D Gaussian probability distribution for the electron and muon fake rate values. The intersection of the horizontal and vertical lines indicate the most probable value.

³We rebinned the lepton p_T to have 7 bins for this calculation resulting to 26 degrees of freedom.

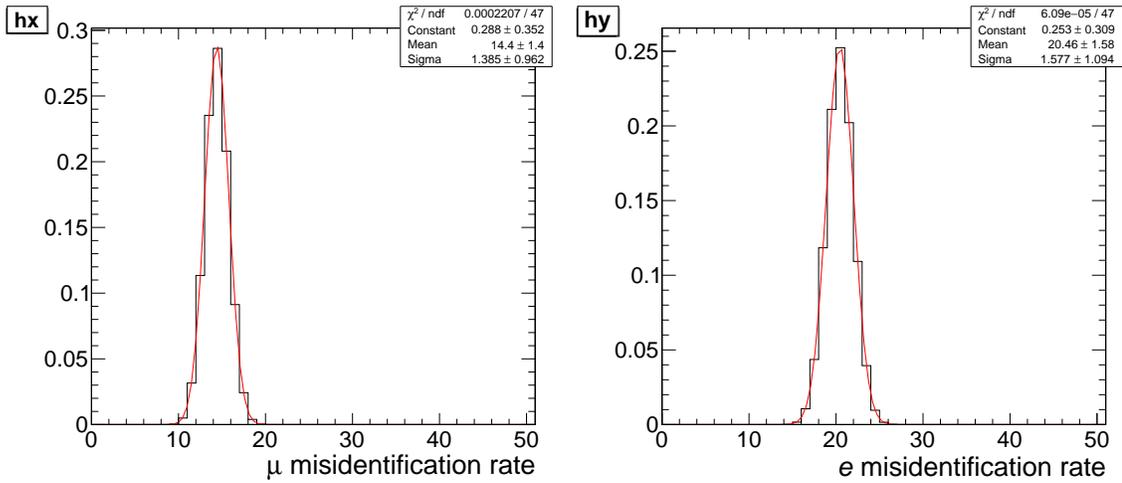


Figure 5.6: Gaussian probability distribution for each fake rate after marginalization.

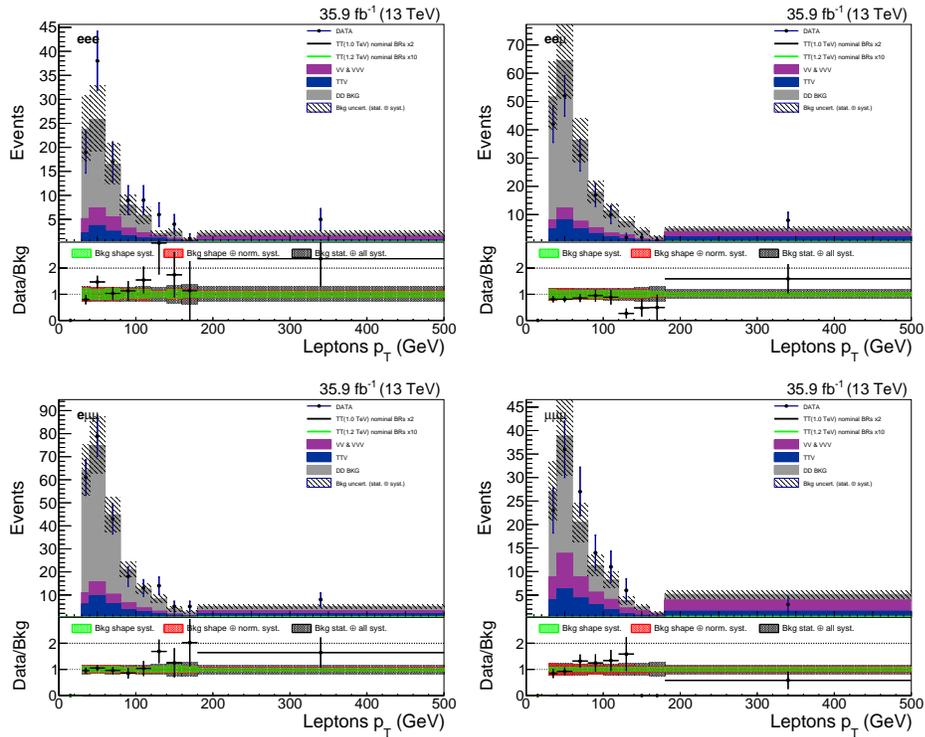


Figure 5.7: Lepton p_T distributions with the optimal fakerates that corresponds the minimum average χ^2 , in the four different trilepton categories, in the control region with 2 jets with exactly 3 leptons. (Statistical and systematic uncertainties are included in the plots)

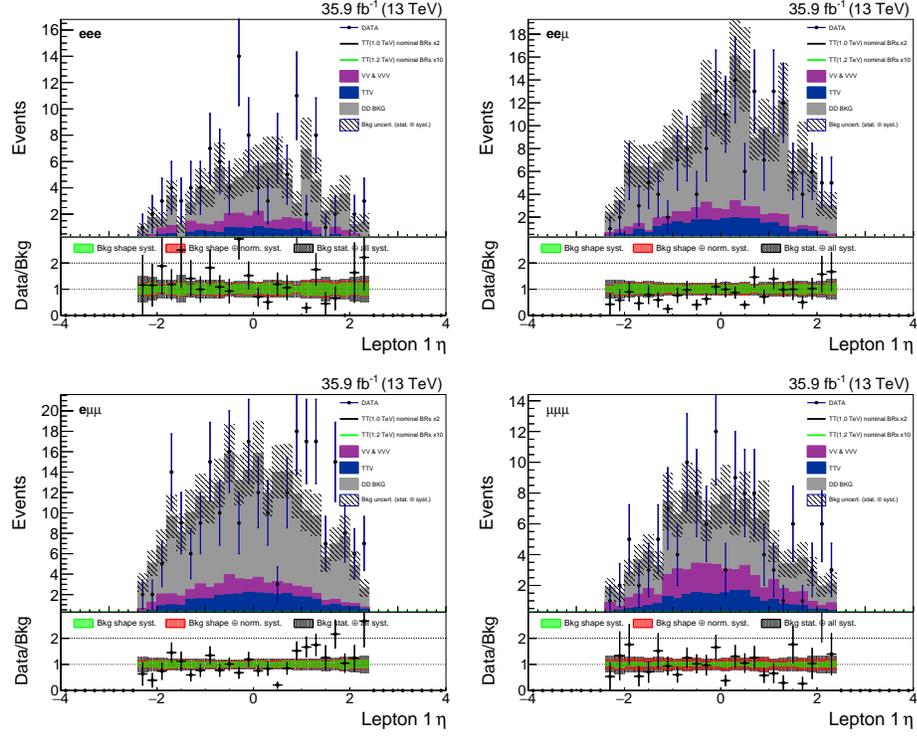


Figure 5.8: Lepton η distributions with the optimal fakerates that corresponds the minimum average χ^2 , in the four different trilepton categories, in the control region with 2 jets with exactly 3 leptons. (Statistical and systematic uncertainties are included in the plots)

Table 5.9: Number of events, in the control region with 2 jets and exactly 3 leptons using the optimal fakerates that corresponds the minimum average χ^2 point (shown in 'CR2' column of Table?). The uncertainties includes both statistical and systematic.

Sample	eee	$ee\mu$	$e\mu\mu$	$\mu\mu\mu$	Total
VV+VVV	4.77 ± 0.71	5.85 ± 0.83	7.63 ± 0.95	8.9 ± 1.3	27.2 ± 4.0
TTV	4.53 ± 0.69	9.4 ± 1.3	11.1 ± 1.5	7.6 ± 1.2	32.6 ± 5.6
ddbkg	19.76 ± 11.0	51.5 ± 23.0	60.3 ± 17.0	23.0 ± 9.6	154.4 ± 51.0
totBkg	29.06 ± 11.0	66.7 ± 23.0	79.0 ± 17.0	39.5 ± 9.8	214.1 ± 51.0
data	36	55	82	40	213
$\frac{data}{Bkg}$	1.24 ± 0.51	0.82 ± 0.31	1.04 ± 0.26	1.01 ± 0.3	0.99 ± 0.25

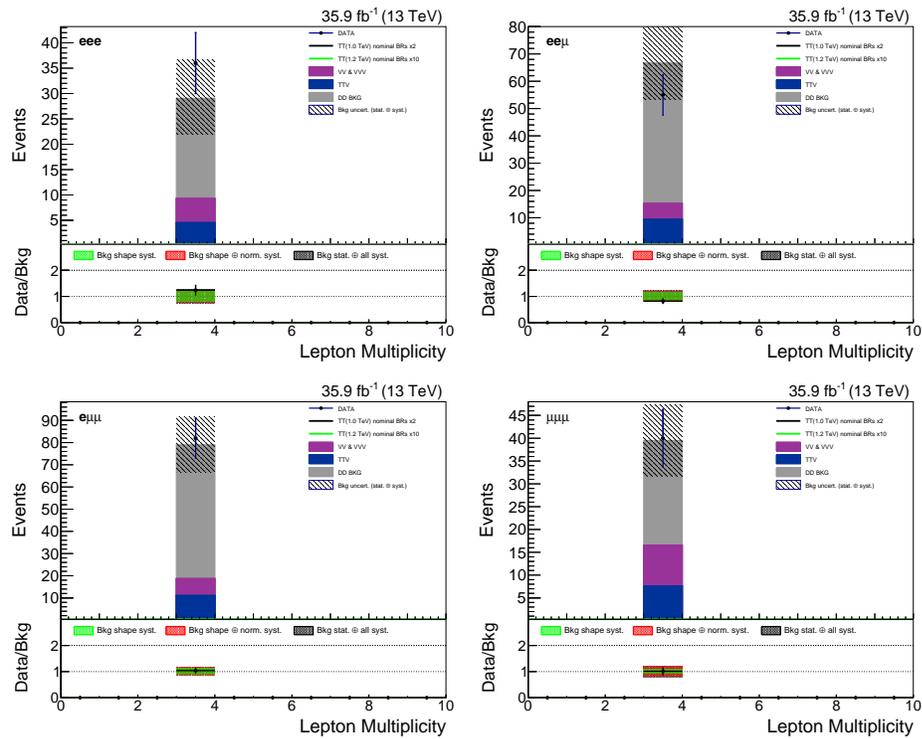


Figure 5.9: Lepton multiplicity distributions with the optimal fakerates that corresponds the minimum average χ^2 , in the four different trilepton categories, in the control region with 2 jets with exactly 3 leptons. (Statistical and systematic uncertainties are included in the plots)

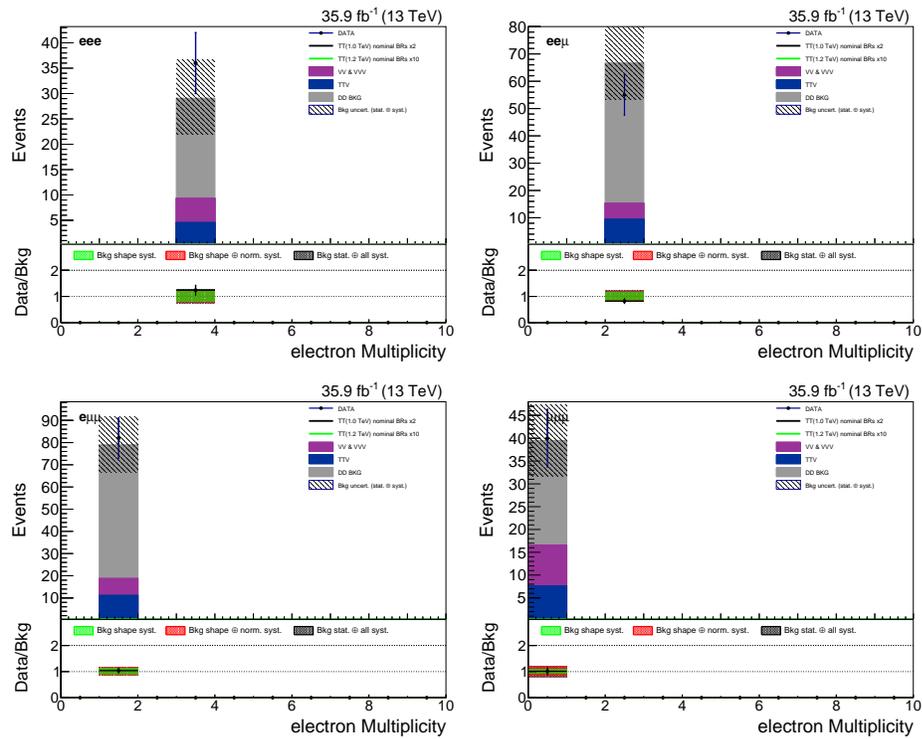


Figure 5.10: Electron multiplicity distributions with the optimal fakerates that corresponds the minimum average χ^2 , in the four different trilepton categories, in the control region with 2 jets with exactly 3 leptons. (Statistical and systematic uncertainties are included in the plots)

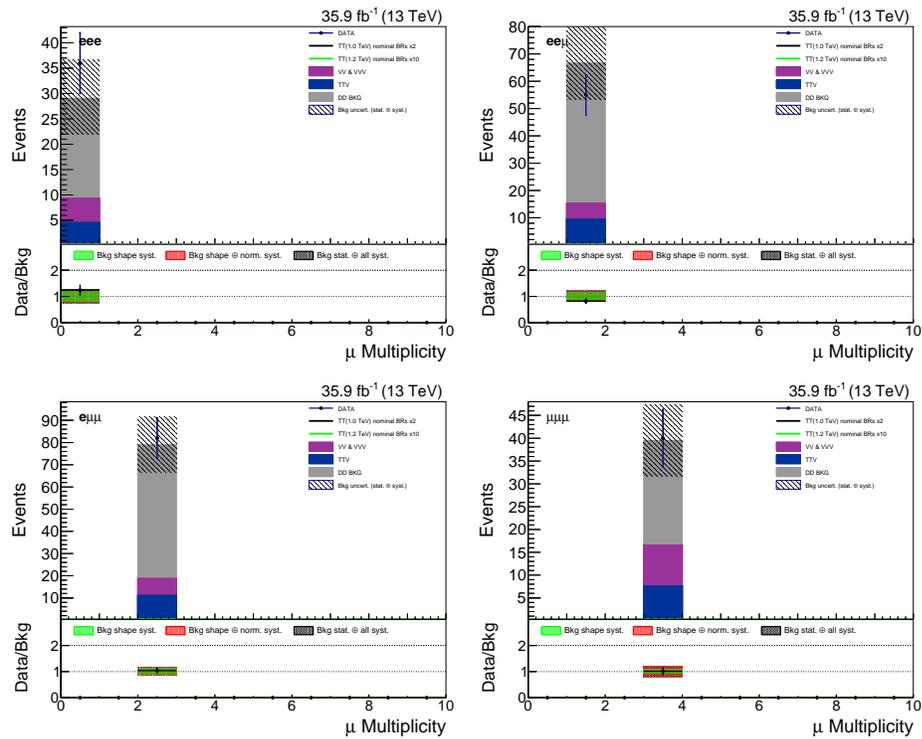


Figure 5.11: μ multiplicity distributions with the optimal fakerates that corresponds the minimum average χ^2 , in the four different trilepton categories, in the control region with 2 jets with exactly 3 leptons. (Statistical and systematic uncertainties are included in the plots)

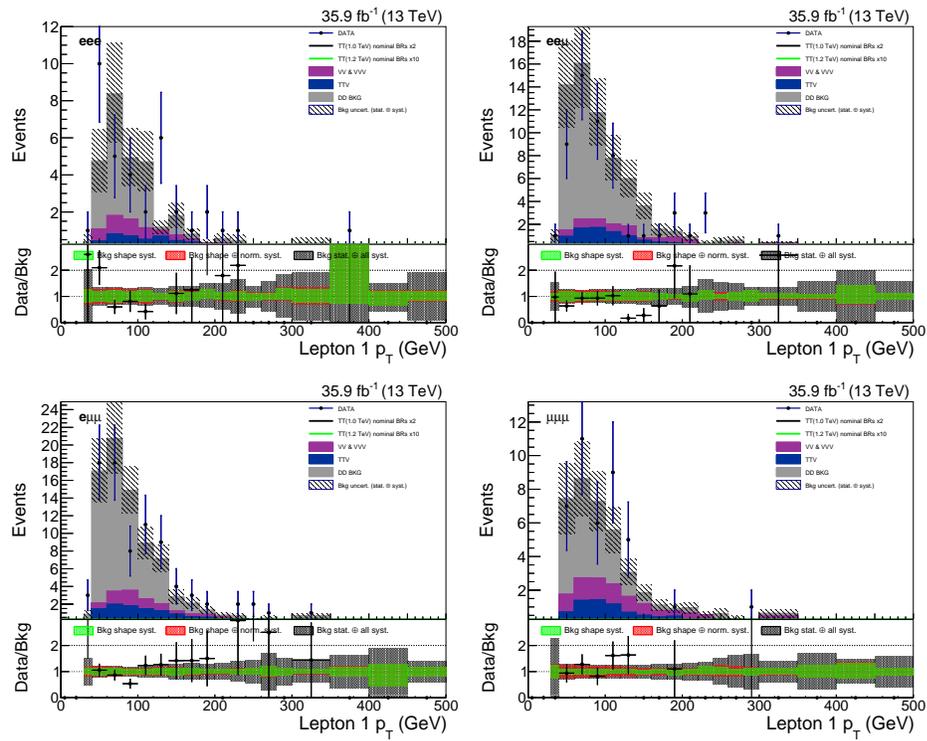


Figure 5.12: Highest p_T lepton p_T distributions with the optimal fakerates that corresponds the minimum average χ^2 , in the four different trilepton categories, in the control region with 2 jets with exactly 3 leptons. (Statistical and systematic uncertainties are included in the plots)

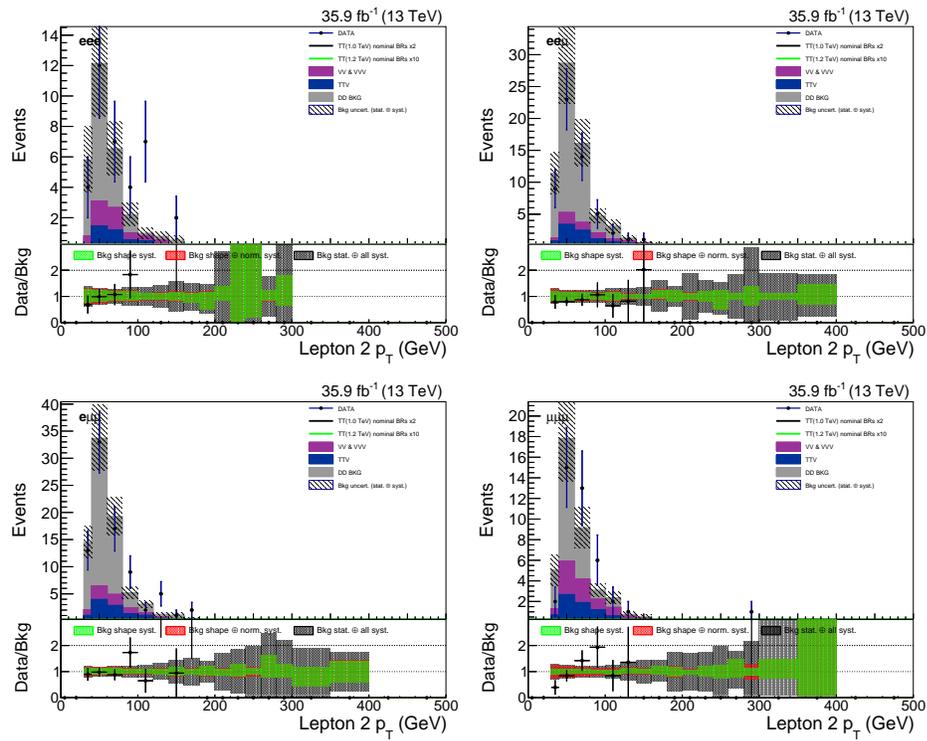


Figure 5.13: 2nd highest p_T lepton p_T distributions with the optimal fakerates that corresponds the minimum average χ^2 , in the four different trilepton categories, in the control region with 2 jets with exactly 3 leptons. (Statistical and systematic uncertainties are included in the plots)

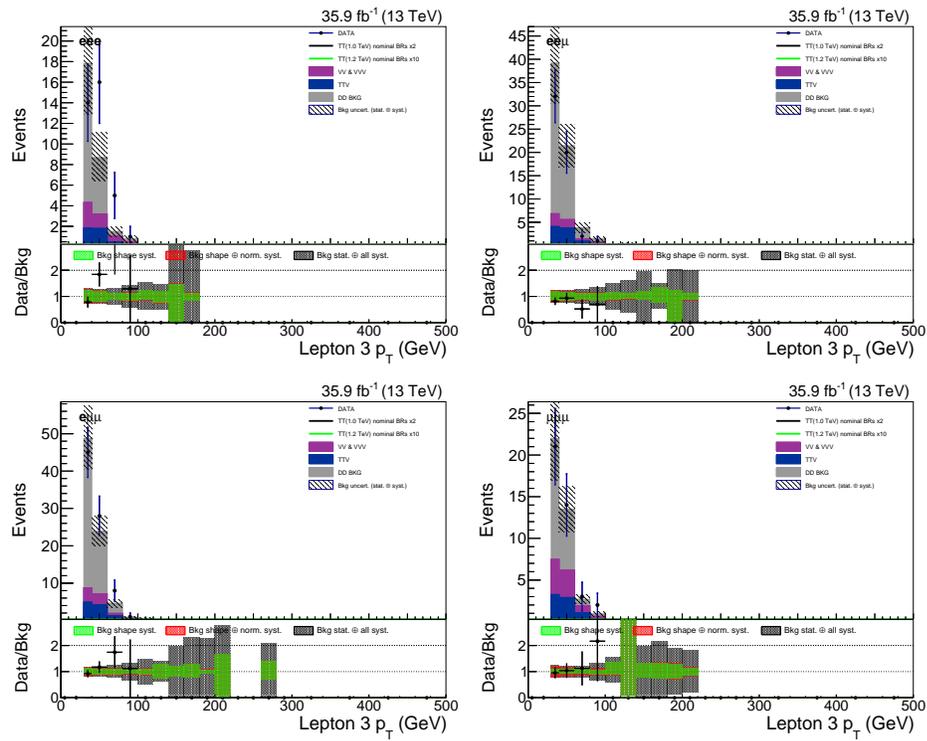


Figure 5.14: 3rd highest p_T lepton p_T distributions with the optimal fakerates that corresponds the minimum average χ^2 , in the four different trilepton categories, in the control region with 2 jets with exactly 3 leptons. (Statistical and systematic uncertainties are included in the plots)

Cross Check

We perform a cross check for our fake rate measurement by conducting the same measurement on a separate control region. This control region is defined to have exactly 1 jet, instead of 2 jets. In this control region, we find that the minimum of the χ^2 corresponds to fake rates of 0.23 ± 0.02 and 0.15 ± 0.02 , for electron and muon fake rates respectively, and the χ^2 minimum value is 48.3 ($\chi^2 / \text{ndf} = 1.26$). Figure 5.16 shows the lepton p_T distributions in the control region with 2 jets, using fake rates that correspond to the minimum χ^2 measure in the control region with 1 jet. Table 5.10 shows the yields. Figure 5.15 shows the 2D distribution of the χ^2 .

This check serves as a sanity check prior to unblinding.

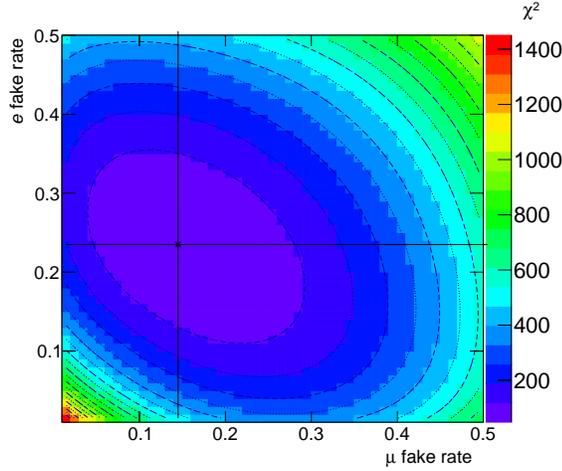


Figure 5.15: χ^2 scan for various electron and muon fake rate values as a cross check in control region with exactly 1 jet. The intersection of the horizontal and vertical lines indicate the minimum value.

Table 5.10: Number of events, in control region with 2 jets (and exactly 3 leptons) using the fakerates measured in control region with 1 jet (shown in 'CR1' column of Table?(a)). The uncertainties includes both statistical and systematic.

Sample	eee	$ee\mu$	$e\mu\mu$	$\mu\mu\mu$	Total
VV+VVV	4.77 ± 0.73	5.85 ± 0.84	7.63 ± 0.98	8.9 ± 1.3	27.2 ± 4.1
$\bar{t}t+V$	4.53 ± 0.71	9.4 ± 1.3	11.1 ± 1.5	7.6 ± 1.2	32.6 ± 5.6
dDbkg	23.7 ± 13.0	59.8 ± 27.0	68.5 ± 20.0	24.9 ± 11.0	176.8 ± 58.0
totBkg	33.0 ± 13.0	75.04 ± 27.0	87.2 ± 20.0	41.5 ± 11.0	236.6 ± 58.0
data	36	55	82	40	213
$\frac{\text{data}}{\text{Bkg}}$	1.09 ± 0.46	0.73 ± 0.28	0.94 ± 0.24	0.97 ± 0.29	0.90 ± 0.23

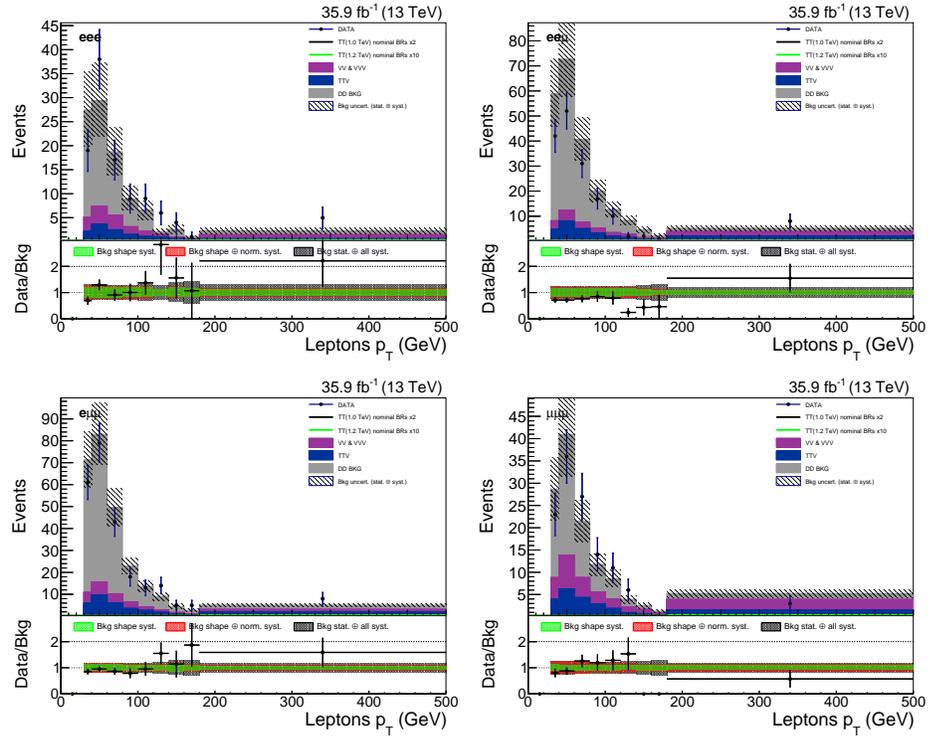


Figure 5.16: Lepton p_T distributions, in control region with 2 jets and exactly 3 leptons, with the fakerates that corresponds the minimum average χ^2 measured in control region with 1 jet (column 'CR1' in Table 5.14(a)), of the four different triplepton categories. (Statistical and systematic uncertainties are included in the plots)

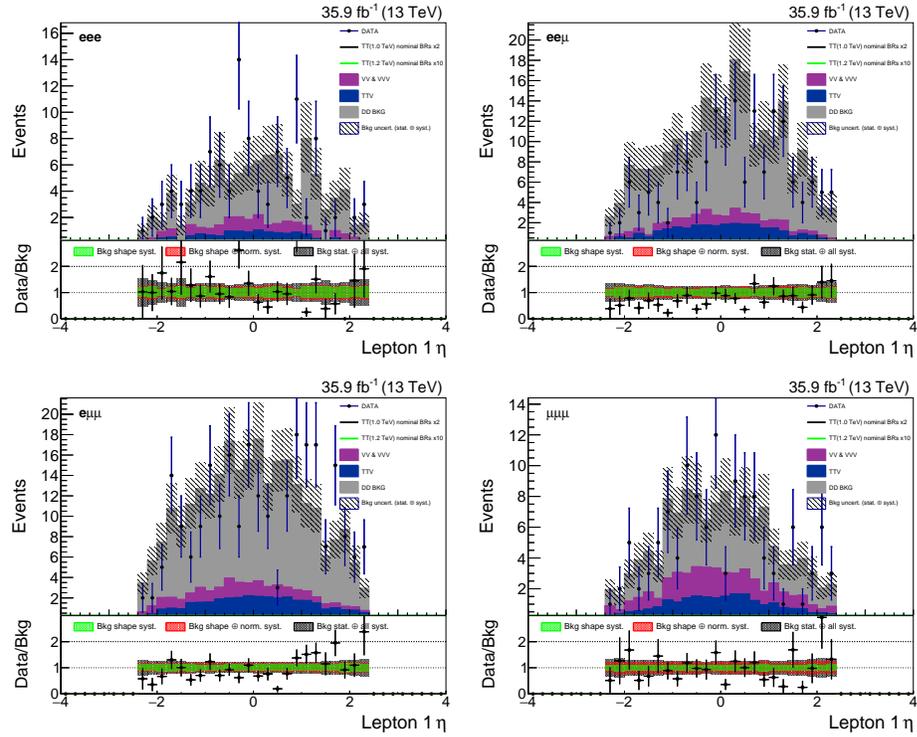


Figure 5.17: Lepton η distributions, in control region with 2 jets and exactly 3 leptons, with the fakerates that corresponds the minimum average χ^2 measured in control region with 1 jet (column 'CR1' in Table 5.14(a)), of the four different triplepton categories. (Statistical and systematic uncertainties are included in the plots)

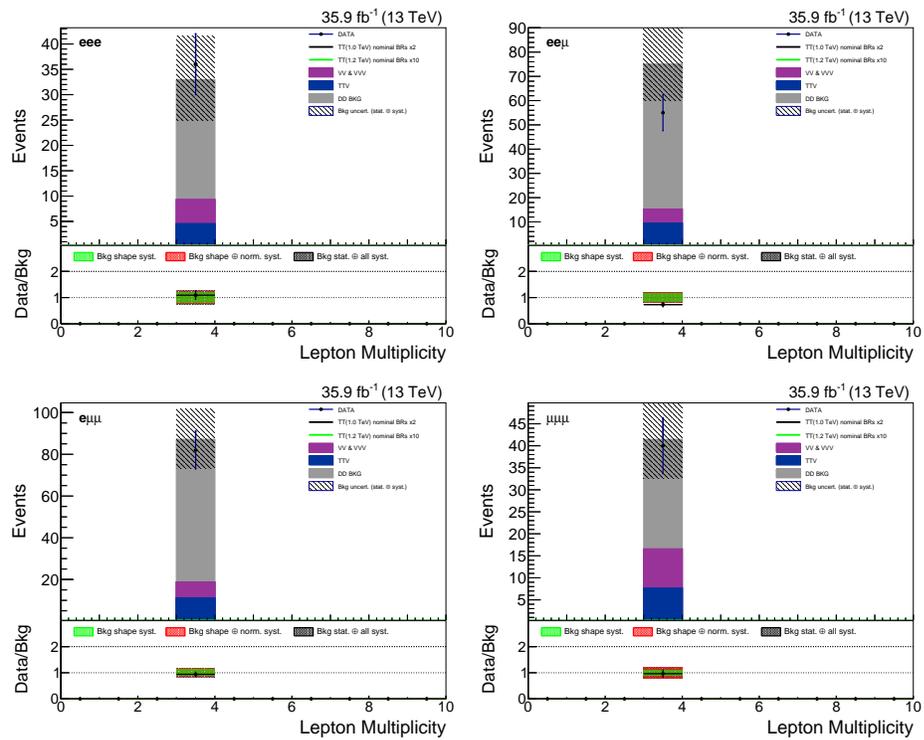


Figure 5.18: Lepton multiplicity distributions, in control region with 2 jets and exactly 3 leptons, with the fakerates that corresponds the minimum average χ^2 measured in control region with 1 jet (column 'CR1' in Table 5.14(a)), of the four different trilepton categories. (Statistical and systematic uncertainties are included in the plots)

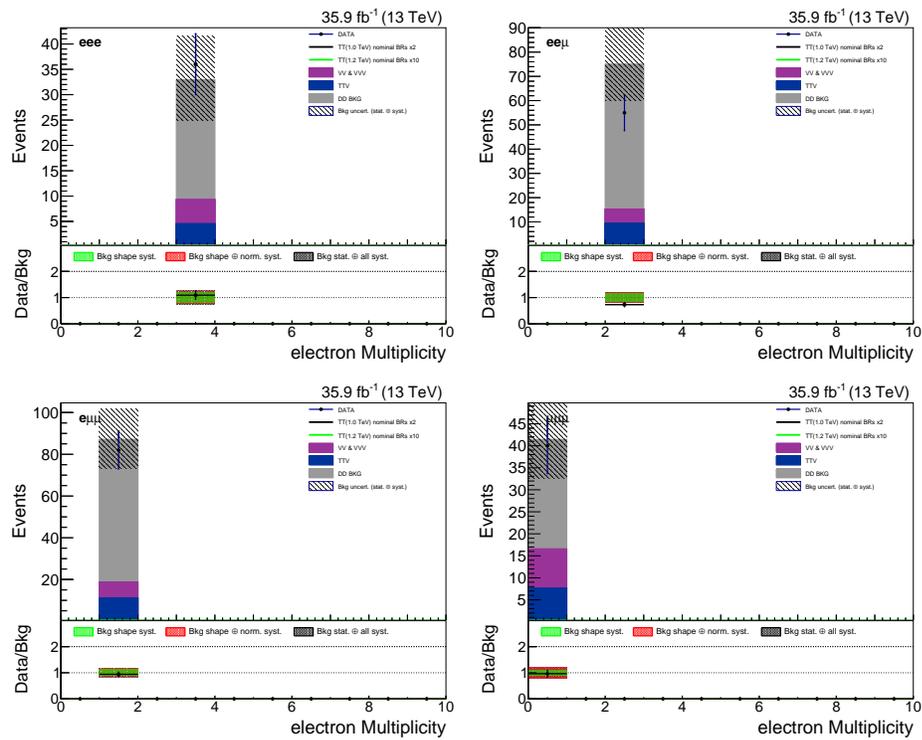


Figure 5.19: Electron multiplicity distributions, in control region with 2 jets and exactly 3 leptons, with the fakerates that corresponds the minimum average χ^2 measured in control region with 1 jet (column 'CR1' in Table 5.14(a)), of the four different trilepton categories. (Statistical and systematic uncertainties are included in the plots)

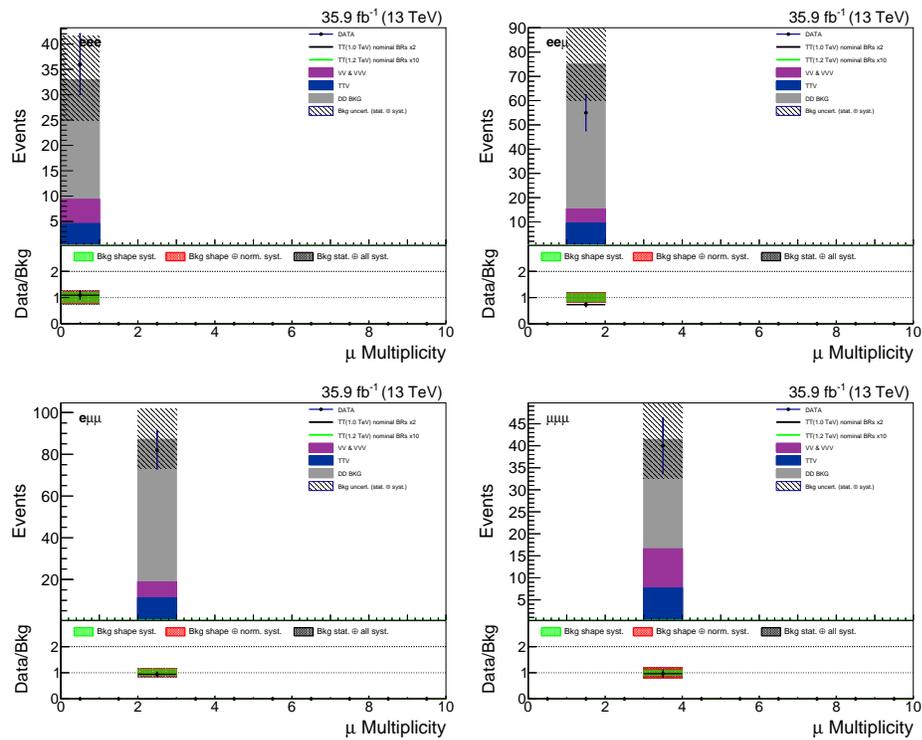


Figure 5.20: μ multiplicity distributions, in control region with 2 jets and exactly 3 leptons, with the fakerates that corresponds the minimum average χ^2 measured in control region with 1 jet (column 'CR1' in Table 5.14(a)), of the four different trilepton categories. (Statistical and systematic uncertainties are included in the plots)

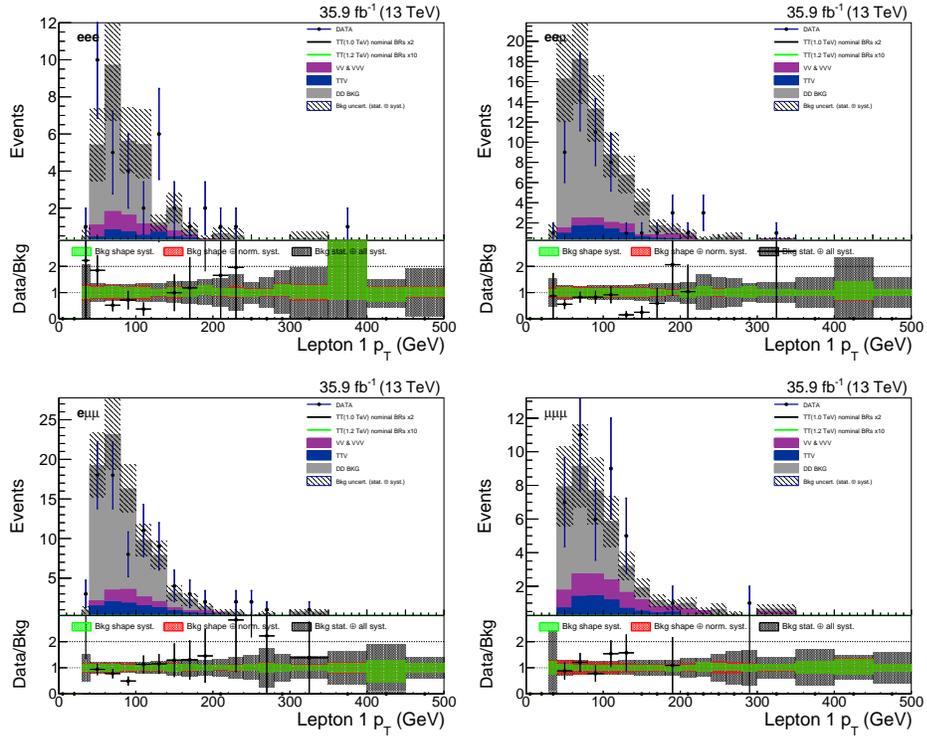


Figure 5.21: Highest (in p_T) lepton p_T distributions, in control region with 2 jets and exactly 3 leptons, with the fakerates that corresponds the minimum average χ^2 measured in control region with 1 jet (column 'CR1' in Table 5.14(a)), of the four different trilepton categories. (Statistical and systematic uncertainties are included in the plots)

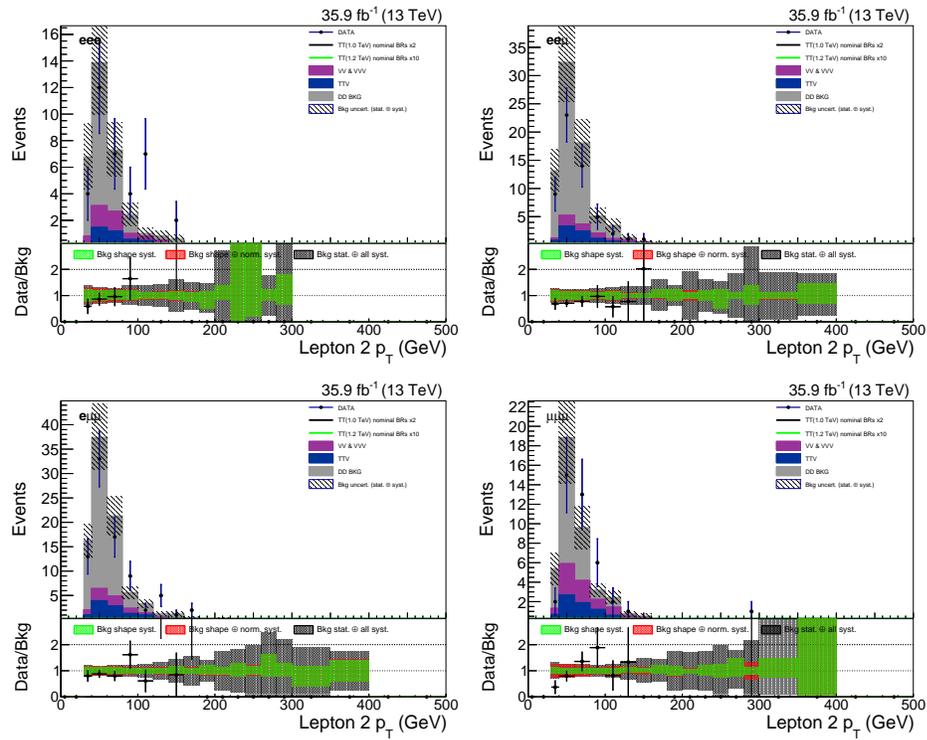


Figure 5.22: 2nd highest (in p_T) lepton p_T distributions, in control region with 2 jets and exactly 3 leptons, with the fakerates that corresponds the minimum average χ^2 measured in control region with 1 jet (column 'CR1' in Table 5.14(a)), of the four different triplepton categories. (Statistical and systematic uncertainties are included in the plots)

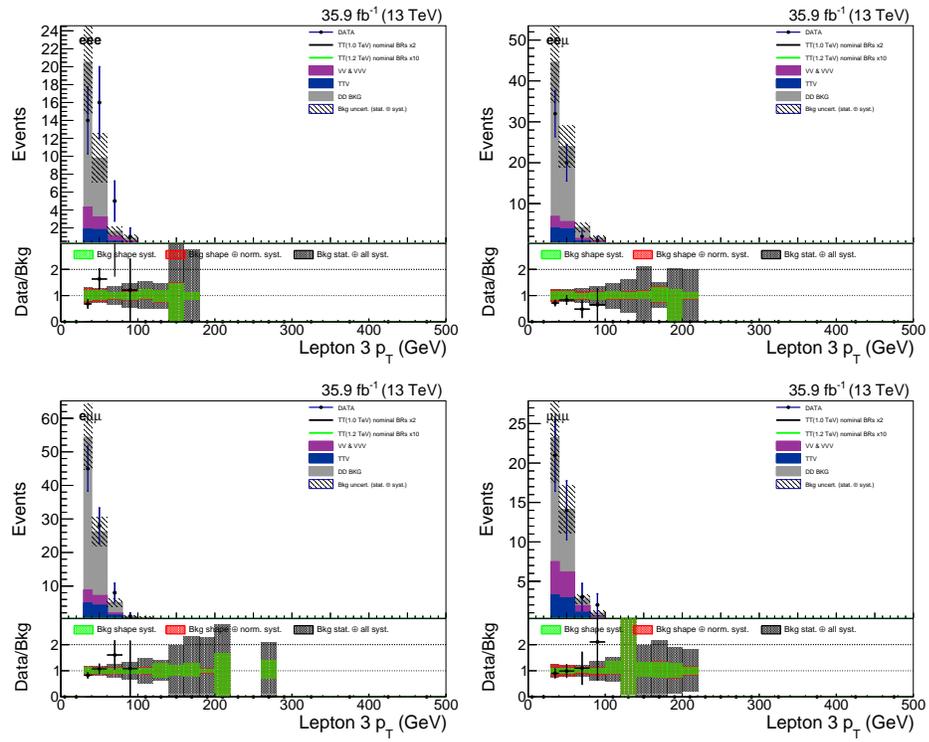


Figure 5.23: 3rd highest (in p_T) lepton p_T distributions, in control region with 2 jets and exactly 3 leptons, with the fakerates that corresponds the minimum average χ^2 measured in control region with 1 jet (column 'CR1' in Table 5.14(a)), of the four different trilepton categories. (Statistical and systematic uncertainties are included in the plots)

Closure test: Fake rate studies in MC

We perform a closure test using the MC $t\bar{t}$ powheg sample and its truth information. We require that the leptons have $p_T > 30$ GeV, $|\eta| < 2.4$, there at least be 1 jet with $p_T > 30$ GeV and $|\eta| < 2.4$, and a missing transverse energy greater than 20 GeV.

Every reconstructed (reco) lepton is matched to a generator (gen) lepton by its η, ϕ coordinates and we require $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.01$. If a match is found, the (mother) particle(s) from which the matched gen lepton decayed are identified recursively back to the proton mother. If a reco lepton is matched to a gen lepton that decayed directly from a W then we label it as a prompt lepton. Otherwise, we label it as a nonprompt lepton. A nonprompt lepton can be matched or unmatched to a gen lepton. A matched nonprompt lepton is further categorized by the origin of the fake lepton:

- 'FromB' if one of the gen mother particles is a bottom hadron.
- 'FromC' if none of the gen mother particles is a bottom hadron, and instead one of the gen mother particles is a charmed hadron.
- 'FromL' if none of the gen mother particles is a B hadron nor a C hadron, and instead one of the gen mother particles is a hadron with only light quarks.
- 'FromElse' if it doesn't satisfy any of the previous criteria.

The fake rates are determined by measuring the ratio between the number of nonprompt leptons that pass the tight ID and the number of all nonprompt leptons (that pass the loose ID). We measure this value, for electron and muons separately, for all events (with at least one lepton) and the results are shown in Table 5.11. For a later comparison, we do the same measure for exactly 3 leptons case, shown in Table 5.12.

To do the closure test we use the fake rate measured above to predict the number of tripleton events in $t\bar{t}$ that pass tight ID requirements. We consider tripleton events where 2 of the leptons are prompt and pass the tight ID and the other one is nonprompt. If the third fake lepton passes the tight ID, we label those events as 'observed' events. If the nonprompt lepton passes only the

Table 5.11: Fake rate measurement using MC truth in 1 or more leptons events in $t\bar{t}$

Source	electron fake rate	nonprompts	percentage
FromL	0.27 ± 0.04	198	0.88
FromC	0.18 ± 0.02	633	2.83
FromB	0.15	12828	57.26
FromElse	0.69 ± 0.16	48	0.21
NotMatched	0.40	8697	38.82
All sources	0.25	22404	100
source	muon fake rate	nonprompts	percentage
FromL	0.77 ± 0.32	13	0.07
FromC	0.20 ± 0.02	748	4.04
FromB	0.14	15712	84.93
FromElse	0.57 ± 0.36	7	0.04
NotMatched	0.02	2020	10.92
All sources	0.13	18500	100

Table 5.12: Fake rate measurement using MC truth in 3 leptons events in $t\bar{t}$

source	electron fake rate	nonprompts	percentage
FromL	0.44 ± 0.11	50	0.38
FromC	0.30 ± 0.07	77	0.59
FromB	0.14	10870	83.31
FromElse	0.66 ± 0.16	83	0.64
NotMatched	0.31 ± 0.01	1968	15.08
All sources	0.17	13048	100
source	muon fake rate	nonprompts	percentage
FromL	0.57 ± 0.18	28	0.19
FromC	0.25 ± 0.06	87	0.60
FromB	0.14	13192	90.84
FromElse	0.93 ± 0.21	42	0.29
NotMatched	0.01	1174	8.08
All sources	0.13	14523	100

loose ID, we weight⁴ these events with the fake rates measured in Table 5.11 and label them as 'predicted' events. In addition we require that events pass the common offline selection criteria as used in the analysis (described in Section 5.4.2 and listed in Section 5.4.3). We tabulate the results and compare them in Table 5.13 and Fig. 5.24. We observe that the predicted yields are different from the observed yields by at most 26%, when considering all sources.

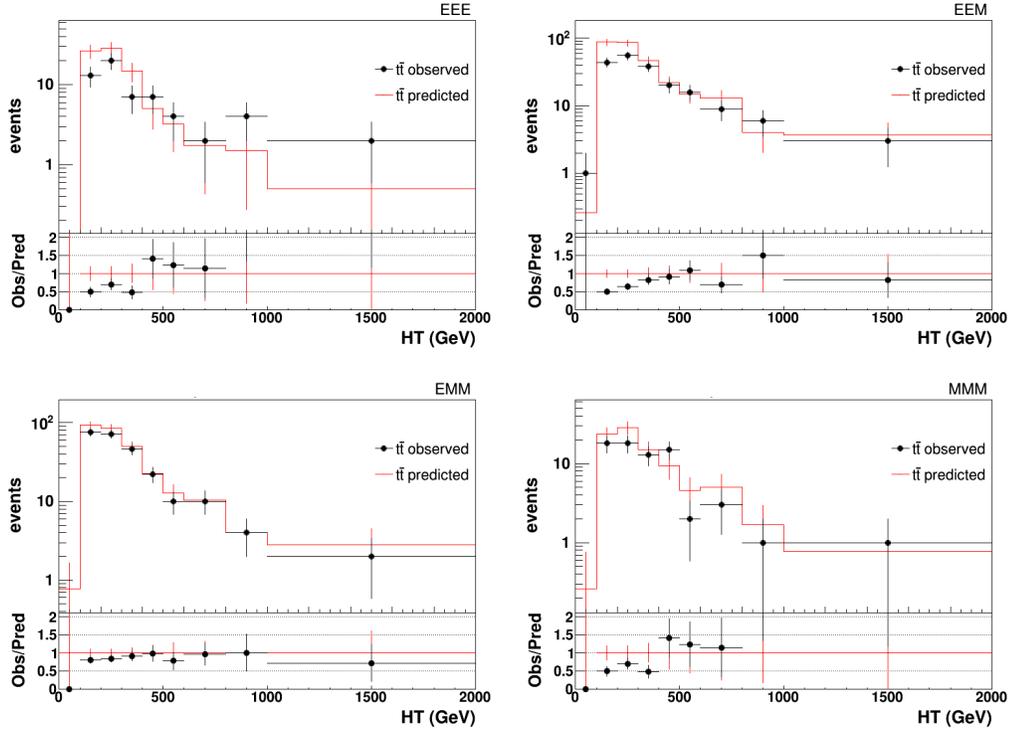


Figure 5.24: HT distributions from the fake rate measurement closure test using MC truth in 3 leptons events in $t\bar{t}$, in the four trilepton channels where two leptons are prompt (p) and one is nonprompt/fake (f) but all pass the tight ID (t).

⁴Here, the weight is the fake rate (f) since $f = \frac{N_{np(t)}}{N_{np}}$ where $N_{np(t)}$ is the total number of nonprompt leptons passing tight ID, N_{np} is total the number of nonprompt leptons (passing the inclusive loose ID). Hence, multiplying N_{np} by f gives us a prediction of $N_{np(t)}$

Table 5.13: Fake rate measurement closure test using MC truth in 3 leptons events in $t\bar{t}$, in the four trilepton channels where two leptons are prompt (p) and one is nonprompt/fake (f) but all pass the tight ID (t).

eee ppf ttt	observed	prediction
NotMatched	20 ± 4.47	14.8 ± 2.4
FromB	37 ± 6.08	42.6 ± 2.5
FromC	-	0.36 ± 0.25
FromL	2 ± 1.41	0.54 ± 0.38
FromElse	-	1.38 ± 0.98
All sources	59 ± 7.7	81.75 ± 4.52
$ee\mu$ ppf(e) ttt	observed	prediction
NotMatched	53 ± 7.28	36 ± 3
FromB	75 ± 8.66	169.5 ± 6.51
FromC	2 ± 1.41	1 ± 0.5
FromL	4 ± 2	1.5 ± 0.61
FromElse	5 ± 2.24	1.25 ± 0.56
All sources	139 ± 11.79	209.25 ± 7.23
$ee\mu$ ppf(μ) ttt	observed	prediction
NotMatched	1 ± 1	5.07 ± 0.81
FromB	48 ± 6.93	57.07 ± 2.72
FromC	0 ± 0	0.39 ± 0.23
FromL	2 ± 1.41	0.13 ± 0.13
FromElse	2 ± 1.41	0.26 ± 0.18
All sources	53 ± 7.28	67.86 ± 2.97
$e\mu\mu$ ppf(e) ttt	observed	prediction
NotMatched	53 ± 7.28	26.5 ± 2.57
FromB	46 ± 6.78	101 ± 5.02
FromC	2 ± 1.41	0.5 ± 0.35
FromL	2 ± 1.41	0.25 ± 0.25
FromElse	10 ± 3.16	1.75 ± 0.66
All sources	113 ± 10.63	130 ± 5.70
$e\mu\mu$ ppf(μ) ttt	observed	prediction
NotMatched	-	11.7 ± 1.23
FromB	118 ± 10.86	140.27 ± 4.27
FromC	1	1.17 ± 0.39
FromL	1	0.13 ± 0.13
FromElse	1	-
All sources	121 ± 11	153.27 ± 4.46
$\mu\mu\mu$ ppf ttt	observed	prediction
NotMatched	-	1.34 ± 0.16
FromB	68 ± 8.25	85.96 ± 3.47
FromC	1 ± 1	1 ± 0.44
FromL	-	-
FromElse	2 ± 1.41	-
All sources	71 ± 8.43	89.18 ± 3.40

Furthermore, to validate our fake rate measurement method performed on data in Section 5.5.3, we perform the same method using the $t\bar{t}$ sample in events with 3 leptons, where the leptons have $p_T > 30$ GeV, $|\eta| < 2.4$, and there at least be 1 jet with $p_T > 30$ GeV and $|\eta| < 2.4$, and a missing transverse energy greater than 20 GeV. We scan lepton fake rate values from 0 to 1.0 for both electron and muon and determine the fake rates that produce the best agreement, using the χ^2 test statistic, between the observed tripleton $t\bar{t}$ events (that pass tight ID requirements) and the predicted $t\bar{t}$ events using the matrix method. See Figures 5.25- 5.26 for the resulting χ^2 and the lepton p_T distributions used to calculate the χ^2 's. The fake rates corresponding to the minimum $\chi^2 = 120.1$ ($\chi^2/ndf = 3.2$) are found to be 0.16 for electron and 0.14 for muon, both with negligible statistical uncertainty. The results are tabulated in Table 5.14(c) and they are comparable to the values we have for the fake rates derived from $t\bar{t}$ MC truth for tripleton events shown in the "All Sources" entries of Table 5.12. They are also comparable to the values measured in data in the control region.

In addition, we repeat this measurement while requiring the same selection as our control region (CR2) described in Section 5.4.3, and once more requiring signal region (SR) selection. For CR2 selection, the fake rates corresponding to the minimum $\chi^2 = 72.04$ ($\chi^2/ndf = 1.9$) are found to be 0.13 for electron and 0.11 for muon and for SR selection, the fake rates corresponding to the minimum $\chi^2 = 102.32$ ($\chi^2/ndf = 2.7$) are 0.17 for electron and 0.11 for muon (see Table 5.14(d)), where fake rates stated without uncertainties indicate that the statistical uncertainties are negligible.

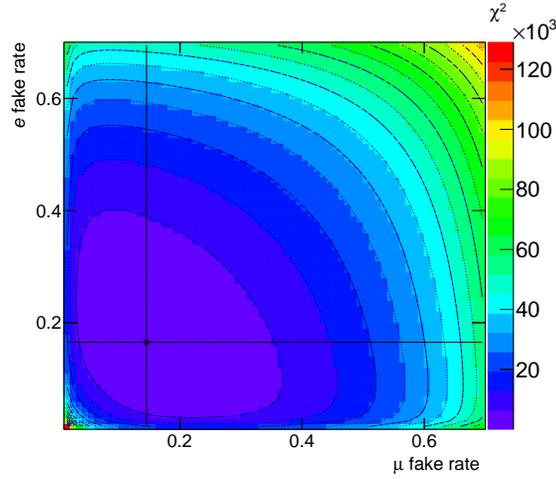


Figure 5.25: χ^2 scan for various electron and muon fake rate values using $t\bar{t}$ MC. The intersection of the horizontal and vertical lines indicate the minimum value.

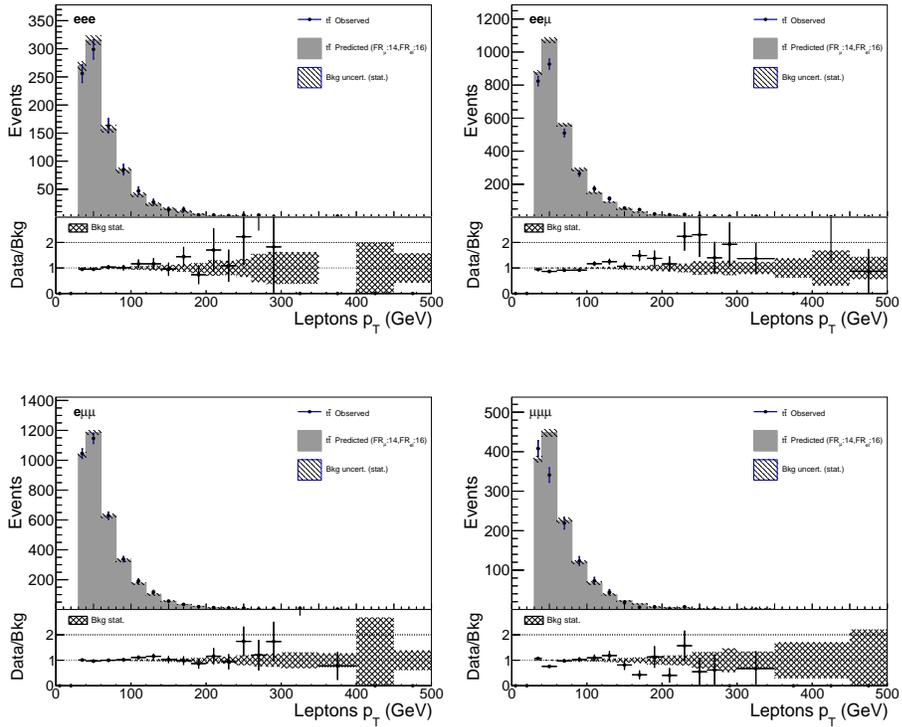


Figure 5.26: Lepton p_T distributions with the fake rates that corresponds the minimum χ^2 , in the four different trilepton categories, using $t\bar{t}$ powheg sample. 'Observed' events are trilepton events that pass the tight ID and 'Predicted' events are the result of using the matrix method.

Summary of fake rate measurement results

We used the control region, CR2, in data to extract the lepton fake rates for our analysis. We have cross checked the method using a separate control region, CR1, where we find that the values differ by 0.06 for electron fake rate and are identical for muon fake rate. Table 5.14(a) summarizes the results of the fake rate measurements in the control regions in data.

Using $t\bar{t}$ MC sample, we measured the fake rate using truth information for all events and for events with exactly 3 leptons. Using the MC fake rate measured in all events we predicted the number of prompt-prompt-nonprompt trilepton events where all leptons pass the tight ID. Comparing to the actual observed events, we observe a maximum yield discrepancy of 31%. Using the same χ^2 minimization method as we have used in data, we extracted lepton fake rates in $t\bar{t}$ trilepton events and compared it to MC fake rate obtained by simple counting in $t\bar{t}$ events with exactly 3 leptons. As the values are comparable, this serves as proof of concept of the method we apply in our analysis. Moreover, we repeat the χ^2 minimization method on $t\bar{t}$ MC and compare the case when CR2 selection criteria are applied to when signal region (SR) selection criteria are applied. We regard this comparison as a model for the possible differences in fake rate between CR2 and SR in data and we take these as a systematics uncertainty. In general, we found that the fake rate measurements in $t\bar{t}$ are comparable to the ones we measured data. Table 5.14(b)(c)(d) summarizes the fake rate studies using the $t\bar{t}$ MC sample.

Table 5.14: (a): Lepton fake rates measured in data in control region with 2 jets (CR2), in control region with 1 jet (CR1) and their differences. The uncertainties are statistical. (b): Deviation of yields between observed and predicted number of prompt-prompt-nonprompt $t\bar{t}$ MC trilepton events where all leptons pass the tight ID, considering all sources of nonprompt leptons. (c): Fake rate measurement using the χ^2 minimization in trilepton events using $t\bar{t}$ powheg sample compared to the fake rate measurement by its MC truth. (d): Fake rate measurement using the χ^2 minimization of the matrix method in trilepton events using $t\bar{t}$ powheg sample requiring CR2 and SR selections.

(a) Fake rate measurements in data

flavor	fake rates in CR2	fake rates in CR1	$\Delta(\text{fake rate})$ CR2 - CR1
e	0.20 ± 0.02	0.23 ± 0.02	0.03 ± 0.03
μ	0.14 ± 0.01	0.15 ± 0.02	0.01 ± 0.02

(b) Observed and prediction agreements from $t\bar{t}$ MC closure test.

Channel	$\Delta(\text{yield})$
eee	28%
$ee\mu$	31%
$e\mu\mu$	17%
$\mu\mu\mu$	20%

(c) Fake rates in $t\bar{t}$ MC

flavor	using χ^2 minimization	using MC truth	$\Delta(\text{fake rates})$
e	0.16	0.17	0.01
μ	0.14	0.13	0.01

(d) Fake rates in $t\bar{t}$ MC using the χ^2 minimization with CR2 and SR selections.

flavor	CR2 selection	SR selection	$\Delta(\text{fake rates})$
e	0.13	0.17	0.04
μ	0.11	0.11	-

In order to better understand the sources of uncertainties associated to the fake rate measurement, we factorize them as follows:

- Statistical: One standard deviation from the minimum of the χ^2 .
- Systematic (1): Difference of fake rate values between CR2 and SR, measured in MC $t\bar{t}$.
- Systematic (2): Remaining deviation of yields between data (subtracted by MC) and the data-driven background in CR2 (applying the measured fake rate).
- Systematic (3): Deviation of yields between events predicted by the matrix method and actual observed events, from the closure test.

The fake rate statistical uncertainty is shown in the first column of Table 5.14(a) derived from the one standard deviation from the minimum χ^2 in data, the fake rate systematic(1) uncertainty is taken from Table 5.14(d) derived using $t\bar{t}$ MC to model the discrepancy between the fake rates measured in control region and the signal region, systematic (2) is the taken from the difference in the yield between data, subtracted by MC background, and the yields from the data driven background in Table 5.9. The fake rate systematic(3) uncertainty is directly taken from Table 5.14(b) derived using $t\bar{t}$ MC closure test to model the discrepancy between prediction of yields by the matrix method and observed yields. We summarize the individual and combined effects as yield differences in Table 5.15. We note that the resulting yield uncertainties are not always symmetric.

The final uncertainties applied in the analysis are shown in the last two rows of Table 5.15 and again in the first three rows of Table 5.17 where we have summed in quadrature the statistical with the systematic(1) uncertainties and systematic(2) with systematic(3) uncertainties.

Table 5.15: Summary of uncertainties due to the fake rate measurement and the matrix method.

Type / Derived from	flavor	eee	$ee\mu$	$e\mu\mu$	$\mu\mu\mu$
Stat + Sys(1) / Data + MC(CR2-SR)	e	+30% -25%	+19% -17%	+14% -12%	0%
	μ	0%	4%	9%	+12% -11%
Sys(2) / $\frac{(Data-MC)}{ddbkg}$		35%	22%	5%	2%
Sys(3) / MC closure test		28%	31%	17%	20%
Sys(2) + Sys(3)		45%	38%	18%	20%

5.6 Systematic uncertainties

The systematic uncertainties included in the analysis and in the limit calculation are summarized in Table 5.16 for MC and in Table 5.17. The uncertainties rate and shape uncertainties unless noted otherwise.

Table 5.16: Summary of the systematic uncertainties on Monte Carlo samples

Category	uncertainty	Signal	Background
Lepton ID SF	2% / lepton	Yes	All
Lepton Iso SF	1% / lepton	Yes	All
Trigger SF	3% / trigger	Yes	All
Luminosity	2.6%	Yes	All
Pileup SF	4.6%	Yes	All
b tagging SF	$\pm\sigma$	Yes	All
b mistagging SF	$\pm\sigma$	Yes	All
PDF	$\pm\sigma$	Yes, shape+acceptance only	All
Renormalization scale	$\pm\sigma$	Yes, shape+acceptance only	All
Jet energy correction	$\pm\sigma$	Yes	All
Jet energy resolution	$\pm\sigma$	Yes	All

Table 5.17: Summary of the uncertainties on the data driven non prompt background. (*) Uncertainties are one-sided in the positive direction.

Category	eee	$ee\mu$	$e\mu\mu$	$\mu\mu\mu$
e FR	+30% -25%	+19% -17%	+14% -12%	0%
μ FR	0%	4%	9%	+12% -11%
Lepton FR	45%	38%	18%	20%
μ FR η dependence*	0%	12%	16%	33%
e PR measurement*	9%	5%	2%	0%
μ PR measurement*	0%	1%	2%	7%

Uncertainties due to jet energy scale (JEC), jet energy resolution (JER), b/light (mis-)tagging scale factors, and NNPDF3.0 parton distribution function (PDF) uncertainties are evaluated by raising and lowering the respective uncertainty by one standard deviation.

The uncertainties for the data-driven nonprompt background mainly come from the lepton fake rate measurements (shown in the first three rows of Table 5.17. These uncertainties were derived from measurements described in Section 5.5.3. In addition, we also looked at the effect of having an eta (pseudo-rapidity) dependence on the muon fake rate. We model the muon fake rate based on the eta dependence found by the 13 TeVsearch for vector-like-quark $X_{5/3}$ analysis [96]. We study

the effect by using a quadratic that (approximately) fit their muon fake rate eta dependence result:

$$f(\eta) = 0.026\eta^2 - 3.85 \times 10^{-18}\eta + f_\mu \quad (5.9)$$

, where η is the pseudo-rapidity, f_μ is the muon fake rate (mentioned in Section 5.5.2), and $f(\eta)$ is the eta dependent fake rate. Assuming this eta dependence, we found that it varies the data-driven background yield up to 33%, only in the positive direction. As for uncertainties due to the prompt rate measurements, we found that varying the electron (muon) prompt rate to unity, does not affect the fake rate measurements but the final data-driven background yield is affected up to 9% (7%), only in the positive direction. The reason for having one-sided variation is because the prompt rates were measured using triggers without lepton isolation criteria. Since our triggers require isolation criteria, we argue that it is reasonable to conclude that the prompt rate could only become slightly higher in value.

Table 5.17 summarizes the uncertainties applied in our analysis that are associated to the data-driven nonprompt background. The source and derivation of the values are described below.

- e/μ FR measurement : Obtained by varying the electron/muon fake rate by the statistical and sys(1) uncertainties as described in Section 5.5.3.
- Lepton FR : Obtained from the quadrature sum of sys(2) and sys(3), as described in Section 5.5.3.
- μ FR eta dependance : Assume μ fake rate is η dependent and follows Eq 5.9.
- e PR measurement: Assume electron prompt rate to be 1.0, instead of the values in Table 5.8.
- μ PR measurement: Assume μ prompt rate to be 1.0 instead of the value mentioned in Section 5.8

5.7 Kinematics distributions after selection

Combining the Monte Carlo background and data-driven nonprompt background, we show the kinematic distributions for our control region and signal region in the following section.

5.7.1 Kinematics distributions in control region

The plots in Figure 5.27 and Figure 5.28 show the kinematic distributions in the control region after the event selections. Table 5.18 shows the yields. We do not observe significant disagreement between data and the predicted background within the uncertainties.

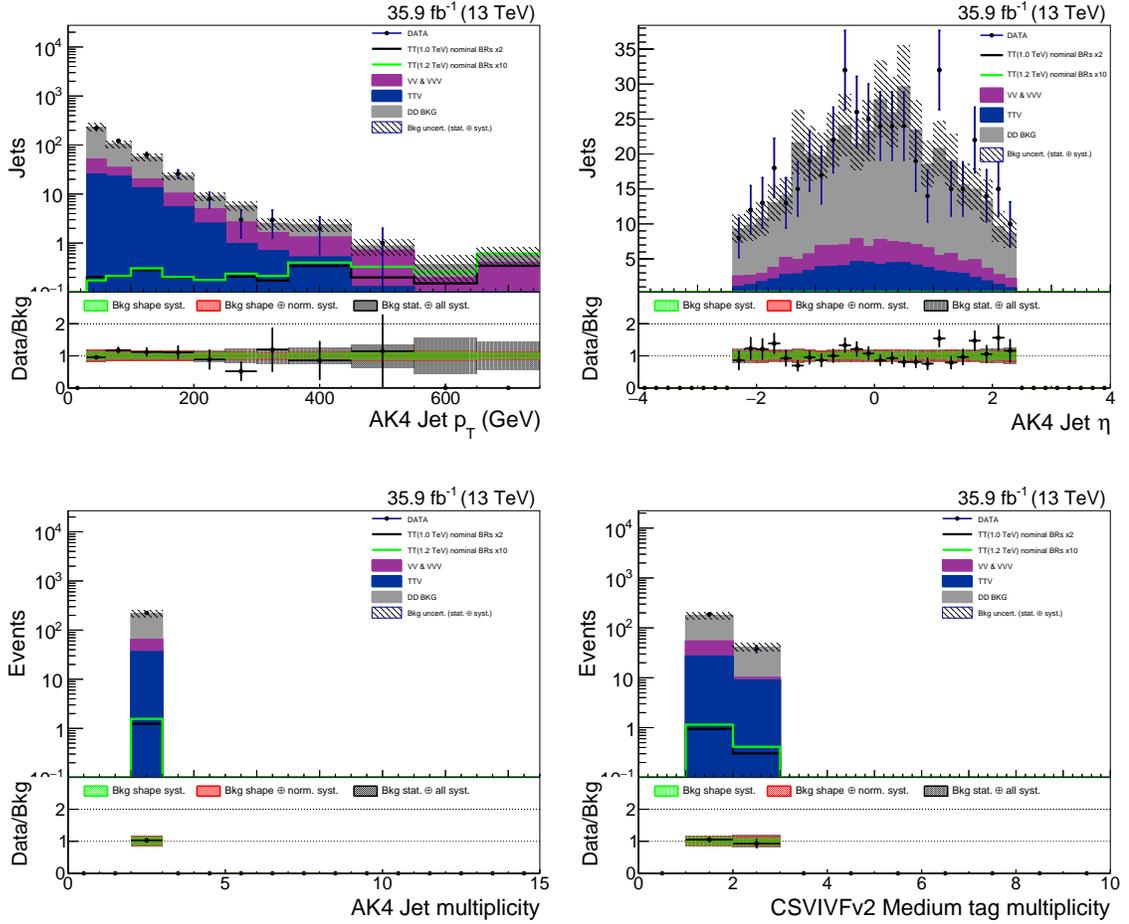


Figure 5.27: Predicted background distributions of AK4 Jet kinematic variables in the tripletonic final state after selection requirements are applied, in the control region. Top left: AK4 Jet p_T distribution. Top right: AK4 Jet η distribution. Bottom left: Number of AK4 Jets. Bottom right: Number of b-tagged AK4 Jets

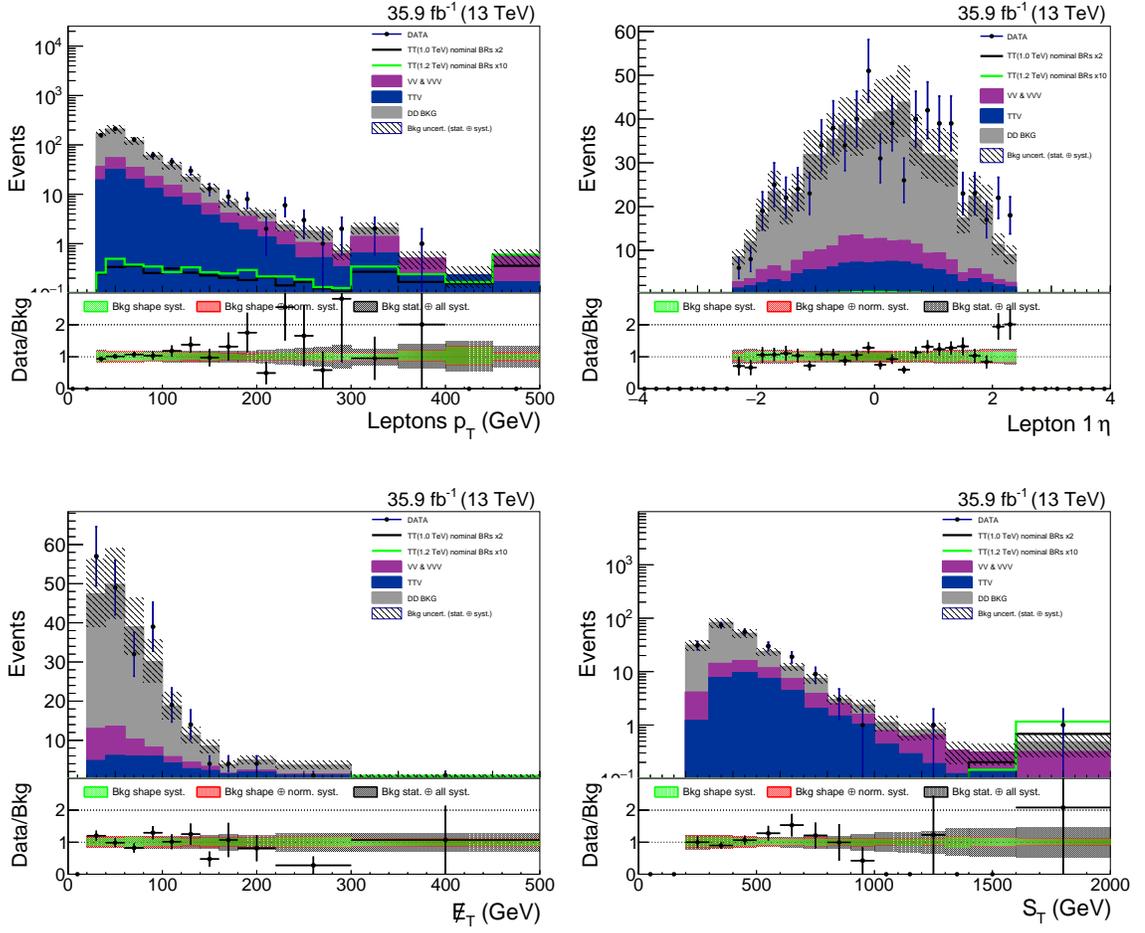


Figure 5.28: Predicted background distributions of kinematic variables in the tripletonic final state after selection requirements are applied, in the control region. Top left: Lepton p_T distribution. Top right: Lepton η distribution. Bottom left: p_T^{miss} distribution. Bottom right: S_T distribution.

Table 5.18: Number of events after selection requirements for the tripleton final state, in control region. The uncertainties includes both statistical and systematic.

Sample	eee	$ee\mu$	$e\mu\mu$	$\mu\mu\mu$	Total
Signal					
$T\bar{T}_{M1000}$	0.09 ± 0.02	0.16 ± 0.02	0.20 ± 0.03	0.16 ± 0.02	0.62 ± 0.09
$T\bar{T}_{M1200}$	0.03 ± 0.01	0.04 ± 0.01	0.05 ± 0.01	0.04 ± 0.01	0.16 ± 0.02
Background					
VV+VVV	4.97 ± 0.72	6.08 ± 0.85	7.85 ± 0.96	9.23 ± 1.3	28.09 ± 4.1
TTV	5.03 ± 0.76	10.43 ± 1.4	11.99 ± 1.6	8.33 ± 1.3	35.77 ± 6.1
ddbkg	19.65 ± 11.0	51.27 ± 23.0	60.13 ± 17.0	22.89 ± 9.6	153.9 ± 51.0
totBkg	29.65 ± 11.0	67.78 ± 23.0	79.96 ± 17.0	40.45 ± 9.8	217.8 ± 51.0
data	38	59	85	42	224
$\frac{\text{data}}{\text{Bkg}}$	1.28 ± 0.51	0.87 ± 0.32	1.06 ± 0.26	1.04 ± 0.3	1.03 ± 0.25

5.7.2 Kinematics distributions in signal region

The plots in Figure 5.29 and Figure 5.30 show the kinematic distributions in the signal region after the event selections. The central value of the yield of the predicted background is observed to slightly under-predict data by about 20%, however they are still in agreement within the uncertainties. The reason for this under-prediction can be explained by the fact that our fake rate was measured in the control region where it is expected to be lower than the actual fake rate in the signal region, as we have observed in the $t\bar{t}$ fake rate study in Table 5.14 (d). This expected fake rate discrepancy has been included in the systematic uncertainty.

Table 5.19 shows the yields. In addition, the yields used for calculating the data-driven background is shown in Table 5.20.

The signal efficiencies in for various signal decays are shown in Table 5.21.

Table 5.19: Number of events after selection requirements for the trilepton final state, in signal region. The uncertainties include both statistical and systematic.

Sample	eee	$ee\mu$	$e\mu\mu$	$\mu\mu\mu$	Total
Signal					
$T\bar{T}_{M_{1000}}$	1.60 ± 0.14	2.54 ± 0.18	3.32 ± 0.23	2.79 ± 0.23	10.2 ± 1.3
$T\bar{T}_{M_{1200}}$	0.40 ± 0.03	0.71 ± 0.05	0.90 ± 0.06	0.78 ± 0.06	2.78 ± 0.35
Background					
VV+VVV	4.32 ± 0.77	5.44 ± 0.78	6.52 ± 0.93	5.89 ± 0.89	22.1 ± 3.5
TTV	20.9 ± 2.9	31.9 ± 4.1	37.0 ± 4.7	35.8 ± 5.0	125.6 ± 21.0
ddbkg	19.2 ± 11.0	40.8 ± 18.0	50.6 ± 15.0	20.0 ± 8.4	130.6 ± 43.0
totBkg	44.35 ± 11.0	78.2 ± 19.0	94.14 ± 15.0	61.7 ± 9.8	278.3 ± 48.0
data	54.0	102	111	71	338
$\frac{data}{Bkg}$	1.22 ± 0.35	1.31 ± 0.34	1.18 ± 0.22	1.15 ± 0.23	1.214 ± 0.22

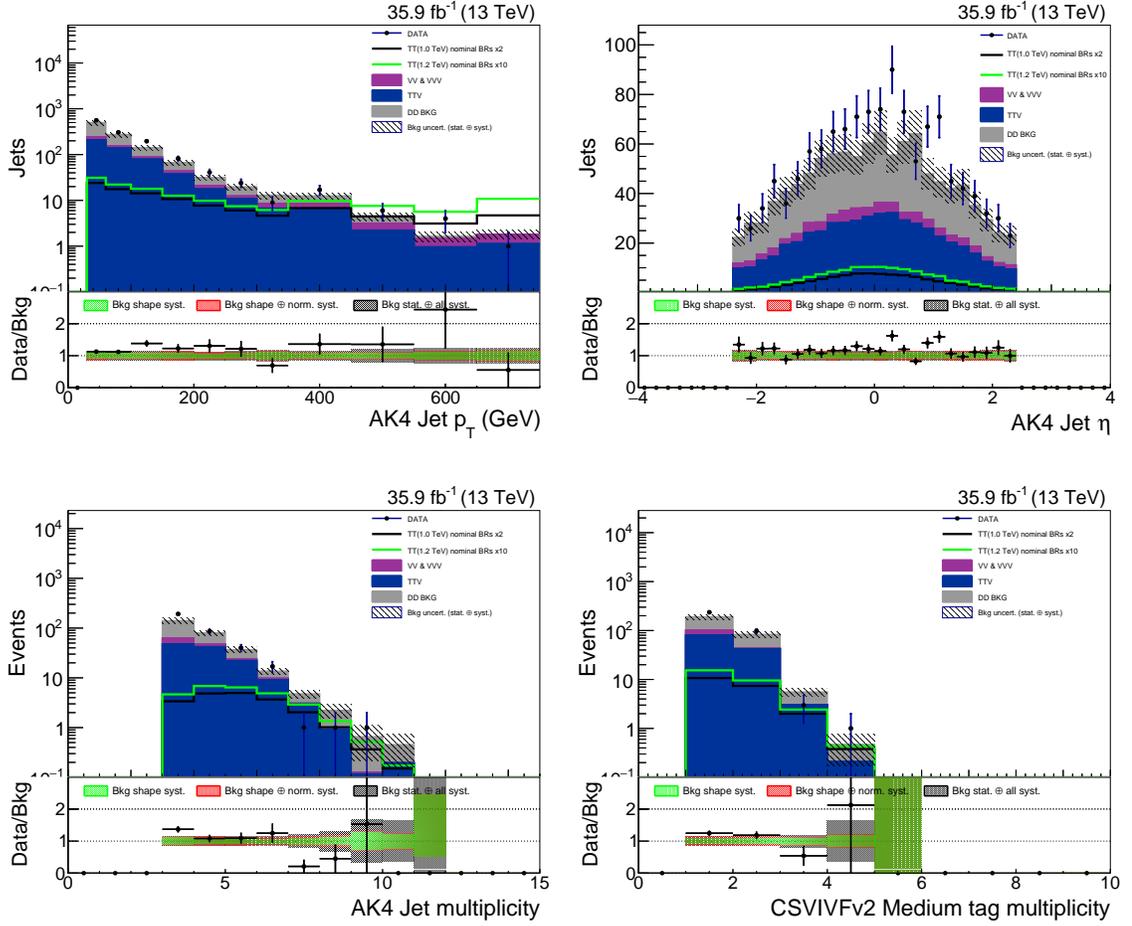


Figure 5.29: Predicted background distributions of AK4 Jet kinematic variables in the tripletonic final state after selection requirements are applied, in the signal region. Top left: AK4 Jet p_T distribution. Top right: AK4 Jet η distribution. Bottom left: Number of AK4 Jets. Bottom right: Number of b-tagged AK4 Jets

Table 5.20: Number of events with at least 1 loose lepton after the selection requirements in the signal region. These are inputs to construct the data-driven background. The uncertainties include both statistical and systematic. Note: ddbkg is the resulting data-driven background yield and the N's are counts events in data.

Nonprompt Background	eee	$ee\mu$	$e\mu\mu$	$\mu\mu\mu$	Total
N_{TTT}	54	102	111	71	338
N_{TTL}	86	91	179	136	720
N_{TLL}		126	102		
N_{LTT}		36	8		
N_{TLL}	10	23	42	25	144
N_{LTL}			42		
N_{LLT}			23		
N_{LLL}	3	4	2	-	9
ddbkg	12.7 ± 7.4	35 ± 13	43 ± 12	15.4 ± 8.8	106 ± 38

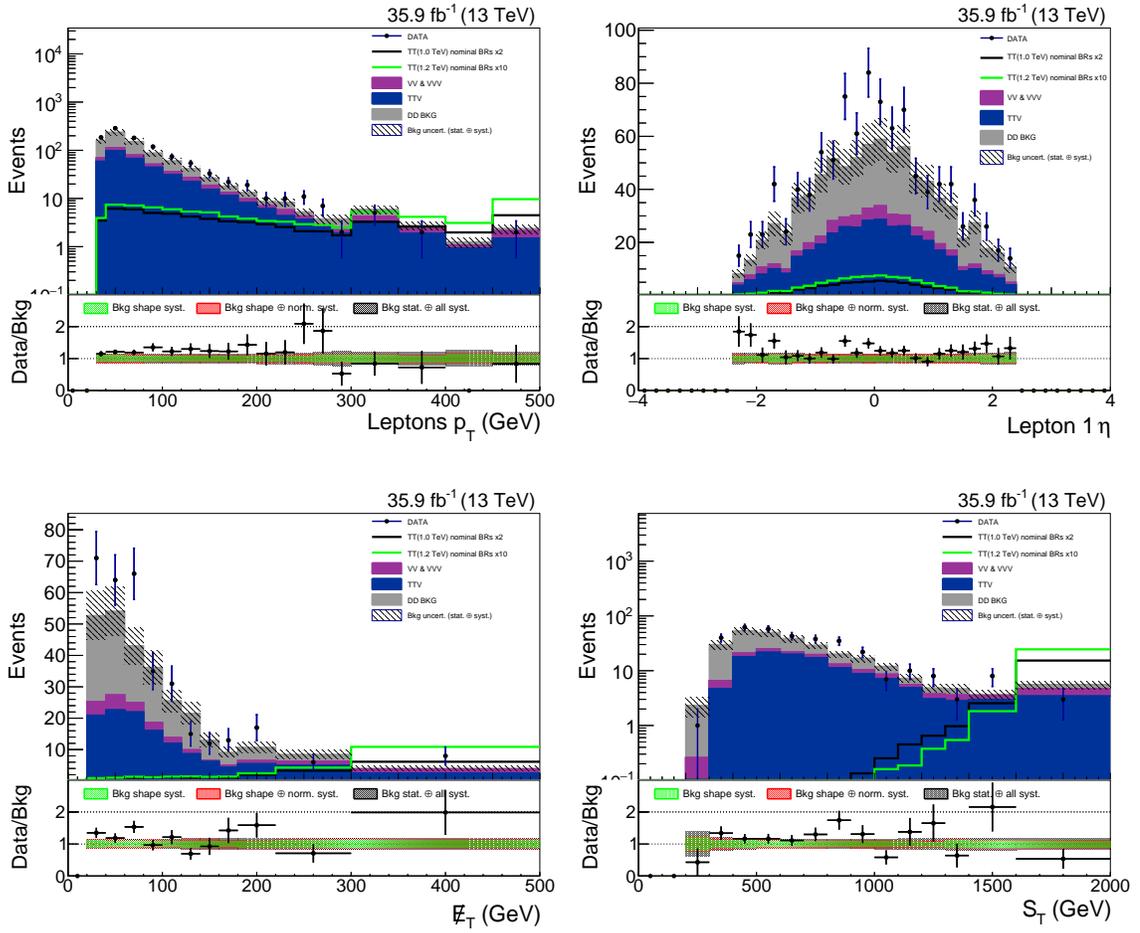


Figure 5.30: Predicted background distributions of kinematic variables in the tripletonic final state after selection requirements are applied, in the signal region. Top left: Lepton p_T distribution. Top right: Lepton η distribution. Bottom left: p_T^{miss} distribution. Bottom right: S_T distribution.

Table 5.21: Signal efficiencies, in percentages, after selection.

TT Mass	bWbW	tHtH	tZtZ	tZbW	tHbW	tZtH
800	0.000	0.716	1.893	0.833	0.265	1.346
900	0.003	0.733	1.954	0.879	0.279	1.414
1000	0.000	0.745	1.926	0.88	0.295	1.503
1100	0.003	0.817	1.938	0.846	0.302	1.523
1200	0.002	0.818	1.987	0.871	0.313	1.482
1300	0.007	0.778	1.940	0.841	0.294	1.481
1400	0.007	0.810	1.871	0.851	0.297	1.473
1500	0.003	0.789	1.931	0.814	0.281	1.450
1600	0.002	0.770	1.745	0.770	0.296	1.401
1700	0.005	0.729	1.749	0.726	0.301	1.337
1800	0.000	0.687	1.730	0.725	0.292	1.23

5.8 Final event selection and limits

We use the shape information from the ST distribution and the yields in the four categories eee , $ee\mu$, $e\mu\mu$, $\mu\mu\mu$ as a discriminant to produce our cross-section limit calculation with the Theta [119] statistical framework. We do not apply any further cuts on top of the selections described in Section 5.4.2.

Using a software called THETA program [120], we apply Bayesian statistics [121] to calculate 95% C.L. expected upper limits on the production cross section of $T\bar{T}$ at each simulated mass point. Statistical uncertainties are treated using the Barlow-Beeston lite method [122, 123]. Nuisance parameters are listed in Table 5.16 and Table 5.17. Normalization-only nuisance parameters are given log-normal priors, and shape uncertainties are treated using template morphing with Gaussian priors. The signal cross section is assigned a flat prior.

The observed 95%C.L. upper limit varies between 0.02 pb to 0.05 pb in the T mass range 800GeV to 1.8TeV. The observed 95%C.L. upper limit for cross section intersects the theory cross section at a mass of 1080 GeV for the singlet signal branching ratios, as shown in Figure 5.32. Figure 5.34 shows the template ST distributions after the limit fit and Figure 5.35 show the pulls by the nuisance parameters and their correlations.

The doublet and the $BR(T\rightarrow tZ)=100\%$, the upper limits are shown in Figure 5.33 and Figure 5.33 respectively, where the intersection with theory cross section are at 1223 GeV and 1287 GeV respectively.

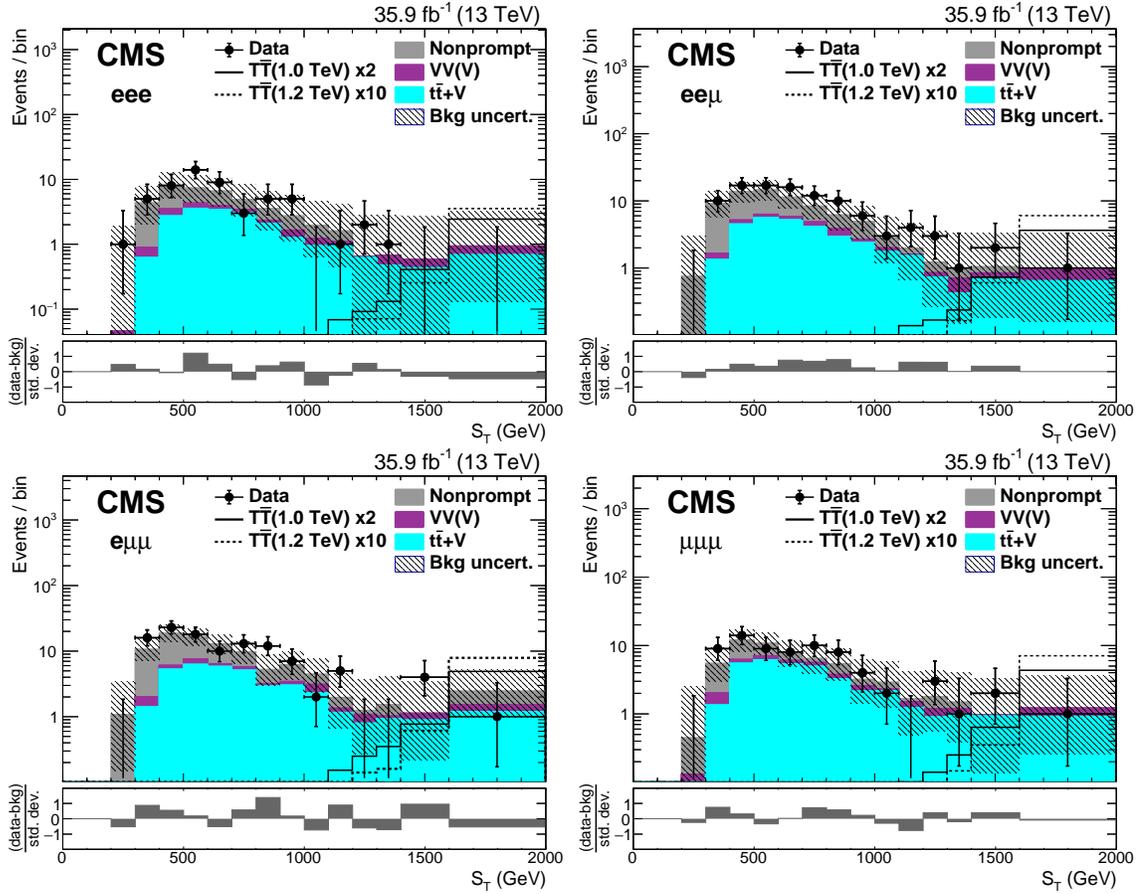


Figure 5.31: Predicted background distributions of final yield in the tripletonic final state after final selection requirements are applied, used as template to generate the limits. The lower panel shows the difference between data and background divided by the total uncertainty. [3]

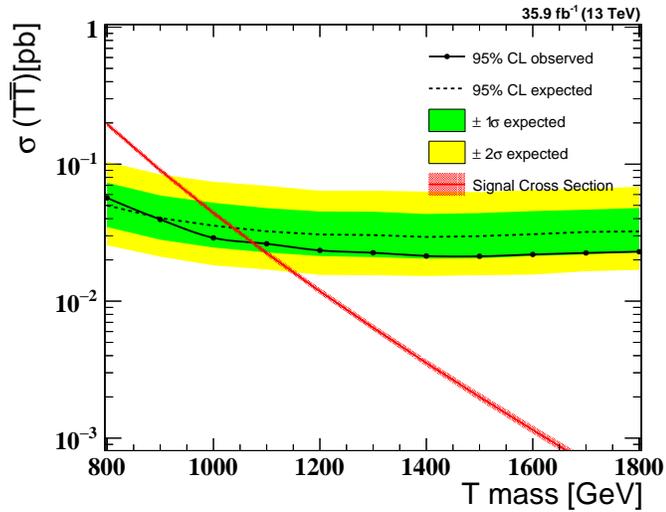


Figure 5.32: tripleton final state cross-section limits (with singlet branching fractions: 50% bW, 25% tH, 25% tZ), shown as a function of mass of T.

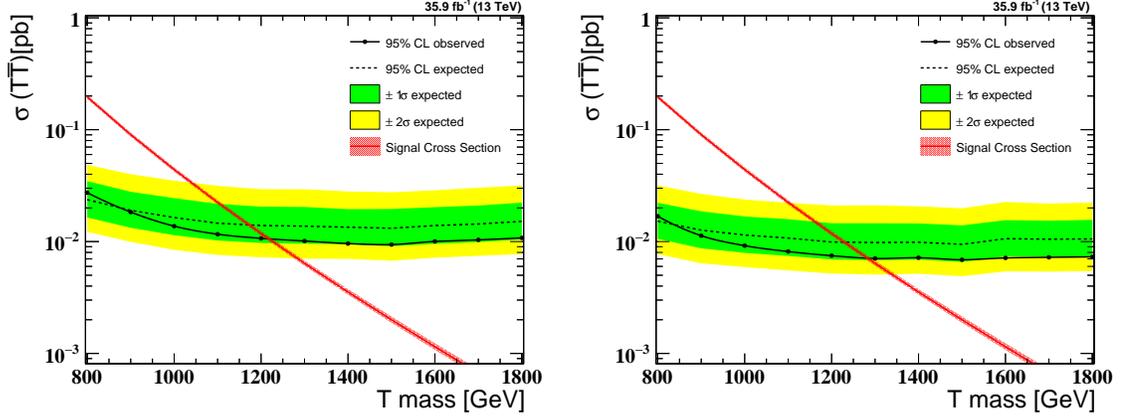


Figure 5.33: tripleton final state cross-section limits with the doublet branching fractions: 0% bW, 50% tH, 50% tZ (left) and with the branching fractions: 0% bW, 0% tH, 100% tZ (right), shown as a function of mass of T .

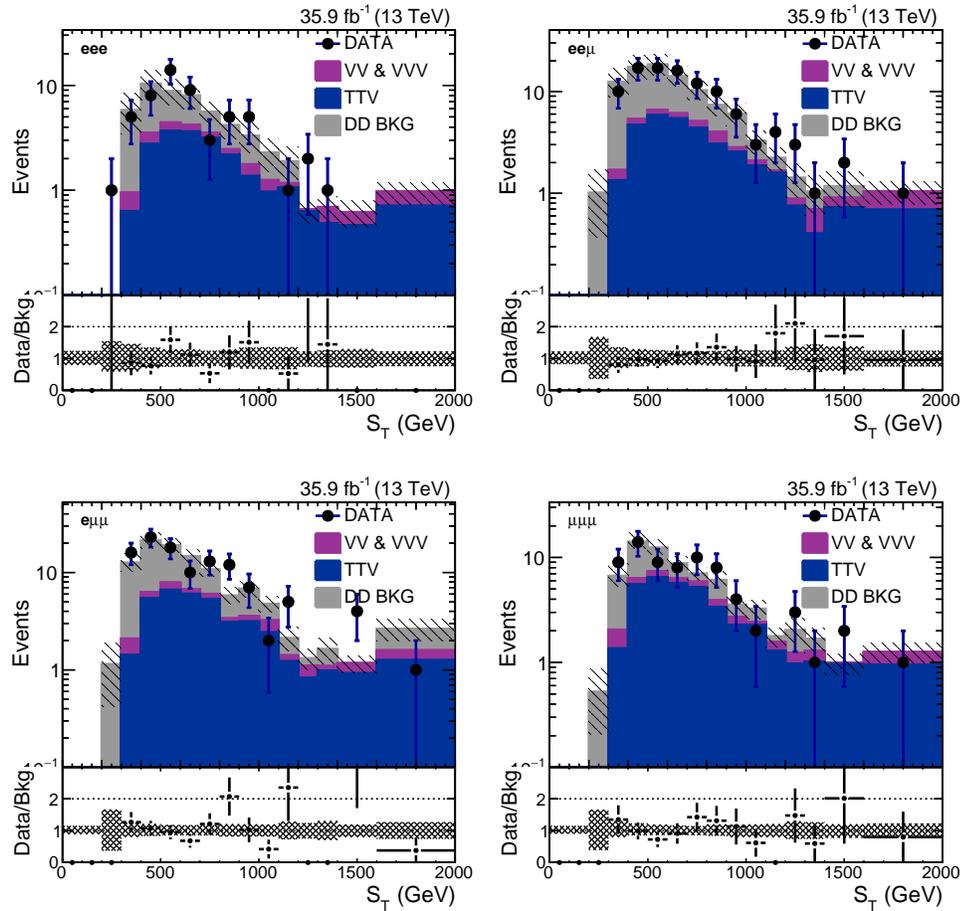


Figure 5.34: Distributions of S_T in the four flavor channels after the limit calculation fit. The uncertainties plotted include all uncertainties.

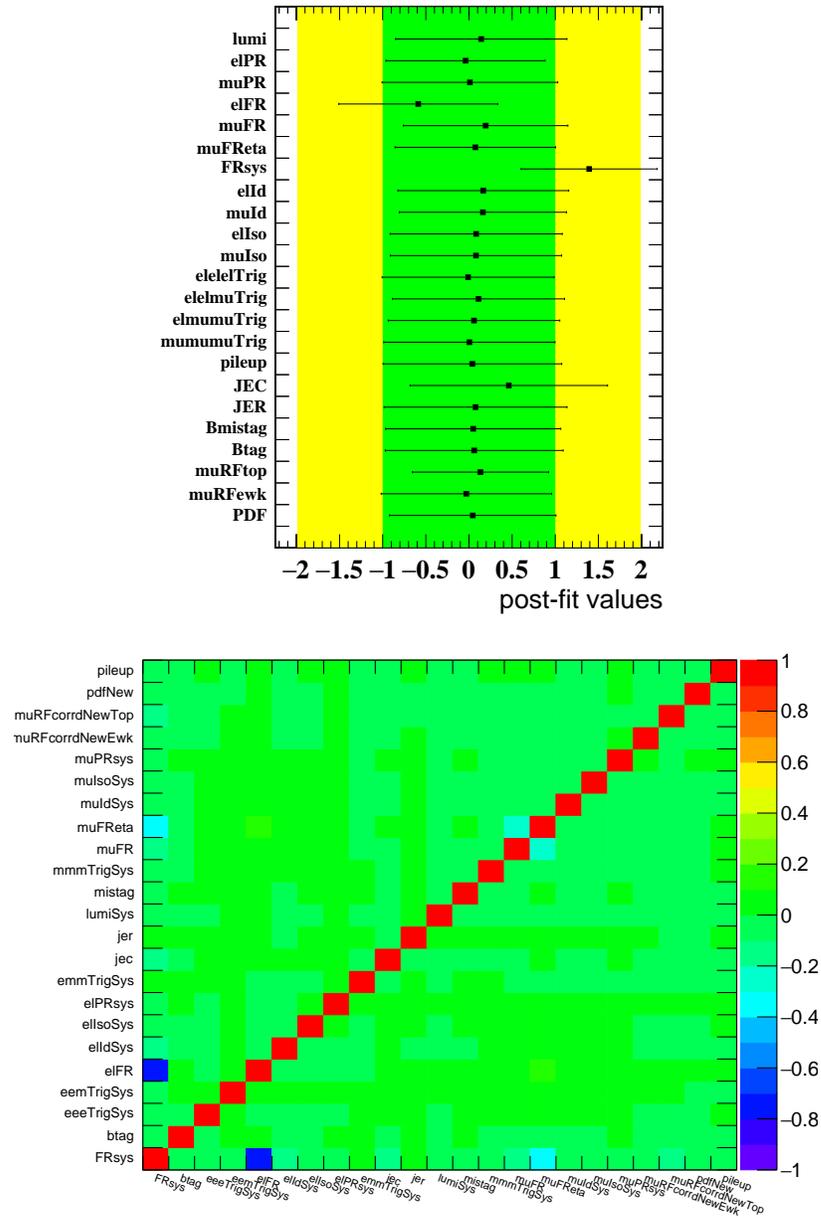


Figure 5.35: Post-fit pulls of the nuisance parameters (top) and the their correlations (bottom).

5.9 Same-sign dilepton final state

In addition to the trilepton channel, we perform measurement for the search for VLQ $T\bar{T} / B\bar{B}$ decaying to same-sign (SS) dilepton. This channel attempts to draw out a unique feature of VLQ signals, the presence of prompt SS dilepton pairs. In $T\bar{T}$ production SS lepton pairs are most common in events having a $T \rightarrow tH$ decay, with the Higgs boson decaying to a pair of W bosons, yielding 3 W bosons in the final state. In $B\bar{B}$ production SS lepton pairs are more frequent, arising from $B \rightarrow tW$ decays.

We perform the analysis using the framework which has been developed for the X5/3 pair decaying to SS dilepton search documented in a CMS physics analysis summary, B2G-16-019 [96]. We utilize the same general strategy, data samples, the MC background samples, triggers, object selections, scale factors, fake rates, and prompt rates, nonprompt background events modelling, as have been used by the X5/3 pair search. Most of the event selection are also identical except for the following two selections:

- Number of leptons is capped at exactly 2 (the same-sign pair) to be orthogonal with the trilepton selection
- Number of "constituents" (jets, additional leptons) is set ≥ 4 rather than ≥ 5 . Since additional leptons is zero this requires at least 4 jets in each event.

The SS dilepton channel is most sensitive to $T\bar{T}$ decaying to $tHbW$, and for $B\bar{B}$ decaying to $tWtW$.

The X5/3 pair SS dilepton analysis uses the cut-and-count strategy using the variable H_T^{lep} (the total transverse momentum of all the leptons and the jets in the event) as the main variable to discriminate between signal and background. We use the same cut-and-count strategy, and also with H_T^{lep} as the main discriminant.

As the X5/3 pair search, to model the backgrounds, we use the simulated di-boson (W^+W^+, WZ, ZZ), tri-boson (WWZ, ZZZ^5, WZZ), $tttt$ and $t\bar{t}$ events with an extra boson produced (ttW, ttZ, ttH). We employ the same methods to estimate the events containing nonprompt leptons as well as

⁵ZZZ is omitted in the tables since it produces 0 yield after final selection

the events with leptons with misidentified charges (ChargeMisID). We use identical dilepton triggers to select the events, and we also reconstruct the physics objects using identical methods and criteria. We also identically categorize the events as follows: ee , $e\mu$, $\mu\mu$. To note, the data set is divided in two sets: era B to early F, and late F to H, where the dataset corresponds to 35.9fb^{-1} of 25ns data, with the first triggering period containing 17.7fb^{-1} and the second containing 18.2fb^{-1} . Different set of triggers are used for each dataset (see Table 5.22). As a consequence, for both electrons and muons the leading lepton is required to have a pT above 40 GeV while the sub-leading is required to have a pT above 35 (30) GeV for the early (late) triggering era. Whereas AK4 jets are required to pT above 30 GeV for both eras. All plots and results shown here combines the two data sets.

Table 5.22: HLT trigger paths used for the trilepton analysis.

Event category	Trigger path, 2016 data, era: B to F(early)
ee	HLT_DoubleEle33_CaloIdL_GsfTrkIdVL_v*
$e\mu$	HLT_Mu30_Ele30_CaloIdL_GsfTrkIdVL_v*
$\mu\mu$	HLT_Mu30TkMu11_v*
Event category	Trigger path, 2016 data, era: F(late) to H
ee	HLT_DoubleEle37_Ele27_CaloIdL_GsfTrkIdVL_v*
$e\mu$	HLT_Mu(37/27)_Ele(27/37)_CaloIdL_GsfTrkIdVL_v*
$\mu\mu$	HLT_Mu30TkMu11_v*

As mentioned in the beginning of the section, we require *exactly* two tight (same-sign) leptons. The effect on applying this requirement on the estimated backgrounds is shown in Table 5.23. We then apply further requirements listed below:

- Quarkonia veto: $M_{ll} > 20$ GeV
- Associated Z-boson veto: veto any event where either of the leptons in the same-sign pair reconstructs to within 15 GeV of the mass of the Z-boson with any other lepton in the event which is not in the same-sign pair.
- Primary Z-boson veto: Invariant Dilepton Mass (M_{ll}) > 106.1 OR < 76.1 for dielectron channel only GeV
- Number of constituents ≥ 4 (not identical to the X5/3 pair search).

where the number of constituents, for our case, is effectively the number of AK4 jets in the event passing our jet selection since we do not have any other leptons not in the same-sign pair. We have performed a signal (T with mass 1 of TeV) to background sensitivity measurement to justify selecting at least four jets in the events where the total lepton and jets transverse momentum is at least 1200 GeV is the most optimal. See Table 5.24.

Table 5.23: Background yields for requiring two same-sign tight leptons and after restricting only 2 leptons in the event.

Sample	$N_{lep} \geq 2$	$N_{lep} = 2$	% decrease
ttZ	117.09 ± 17.84	76.69 ± 10.56	34.5
ttW	274.1 ± 58.04	241.37 ± 44.83	11.94
ttH	90.34 ± 28.39	73.13 ± 20.68	19.05
tttt	9.11 ± 4.63	7.14 ± 2.07	21.62
WZ	1820.58 ± 493.7	1569.18 ± 373.46	13.81
ZZ	357.16 ± 49.78	159.94 ± 27.50	55.22
WpWp	145.35 ± 74.96	145.28 ± 73.56	0.05
WWZ	19.88 ± 10.27	14.09 ± 4.35	29.12
WZZ	6.81 ± 3.52	5.03 ± 1.14	26.14
NonPrompt	10821.8 ± 5411.33	10309.70 ± 5155.28	4.73
ChargeMisID	22377.33 ± 6713.22	22377.33 ± 6713.22	0

Table 5.24: SS dilepton signal to background sensitivity measurement.

	MC	NonPrompt	ChargeMisID	Total Bkg	Sig 1 TeV	$\frac{S}{\sqrt{S+B}}$
$H_T^{lep} > 1200$ GeV						
+ nConst ≥ 5	14.97 ± 2.14	14.96 ± 7.81	4.08 ± 1.26	34.01 ± 8.20	5.47	0.87
+ nConst ≥ 4	19.98 ± 2.78	20.68 ± 10.67	6.75 ± 2.07	47.41 ± 11.22	6.58	0.90
+ nConst ≥ 3	24.72 ± 3.93	21.75 ± 11.22	8.90 ± 2.71	55.37 ± 12.19	6.97	0.88
$H_T^{lep} > 1100$ GeV						
+ nConst ≥ 4	27.99 ± 4.02	31.03 ± 15.85	10.06 ± 3.06	69.08 ± 16.63	7.11	0.81

Due to the change of the final selection cuts, we rederive several of the signal and MC background uncertainties:

- No change from X5/3: charge mis-ID, fake rate, luminosity, lepton ID/isolation/trigger efficiencies
- Added new: lepton reconstruction scale factor uncertainties of 1% per lepton, matrix element scale variation uncertainties for signal.
- Recalculated with method as used by X5/3 search : pileup (varies for background, 1% for signal), jet energy resolution (1% for backgrounds, 2% for signal)

- Recalculated with adjusted methods: JEC (now taking into account variation in nConst, ie. the number of AK4 jets), PDF (now calculated from RMS of total event yields)

Asymmetric values are used in limits calculation where needed. The recalculated systematic uncertainties values are summarized in Table 5.25.

Table 5.25: Systematic uncertainties, in percentages. Signal PDF and scale uncertainties are acceptance only. The values in the table are the maximum shifts from the nominal yield.

Sample	PileUp	PDF	scale	jet energy scale	jet resolution
ttZ	6	3	11	4	1
ttW	1	13	16	3	1
ttH	1	3	27	4	1
tttt	3	35	26	2	1
WpWp	1	35	35	9	1
WZ	10	11	15	9	1
ZZ	7	2	11	9	1
WWZ	6	2	20	10	1
WZZ	2	2	21	9	1
signal T (1 TeV)	1	≤ 8	≤ 1	≤ 6	1 – 2

5.9.1 Kinematic plots

We show several kinematics plots, removing the $n\text{Const}$ requirement and H_T^{lep} requirement, and instead just require at least two AK4 jets, shown in Figures 5.36-5.45.

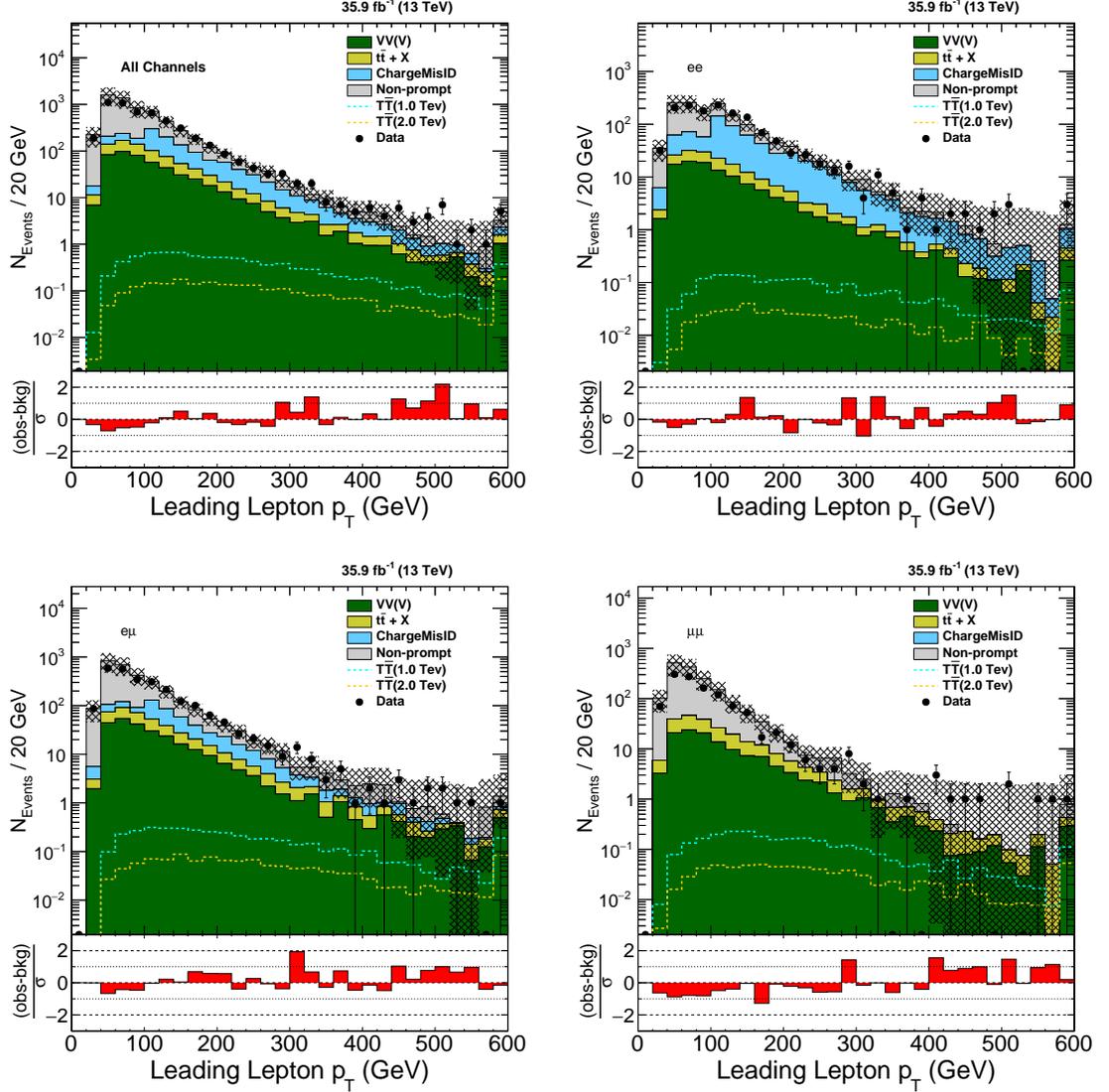


Figure 5.36: Distribution of leading Lepton p_T for the combined flavor channel (All) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty.

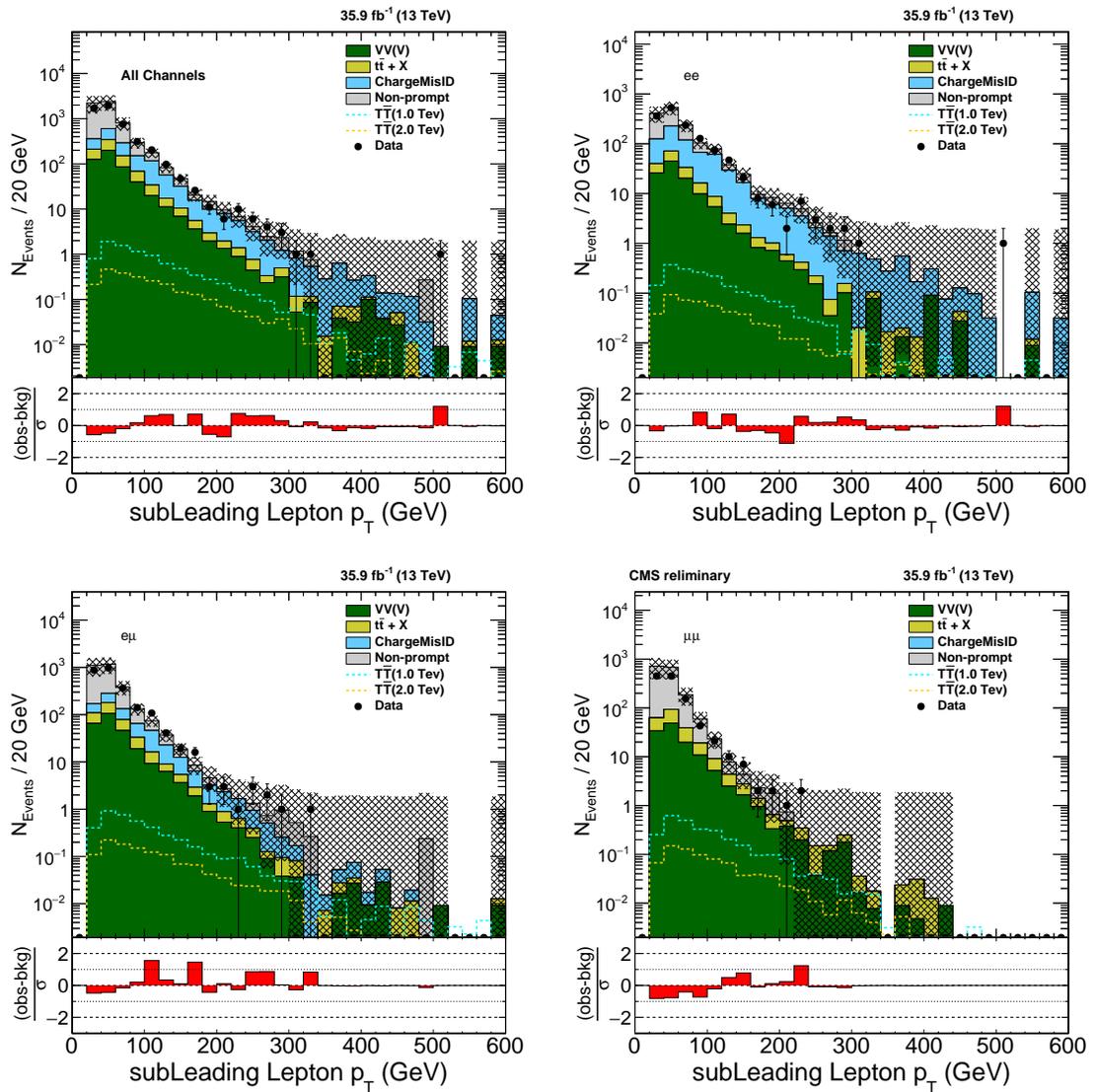


Figure 5.37: Distribution of sub-leading Lepton p_T for the combined flavor channel (All) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty.

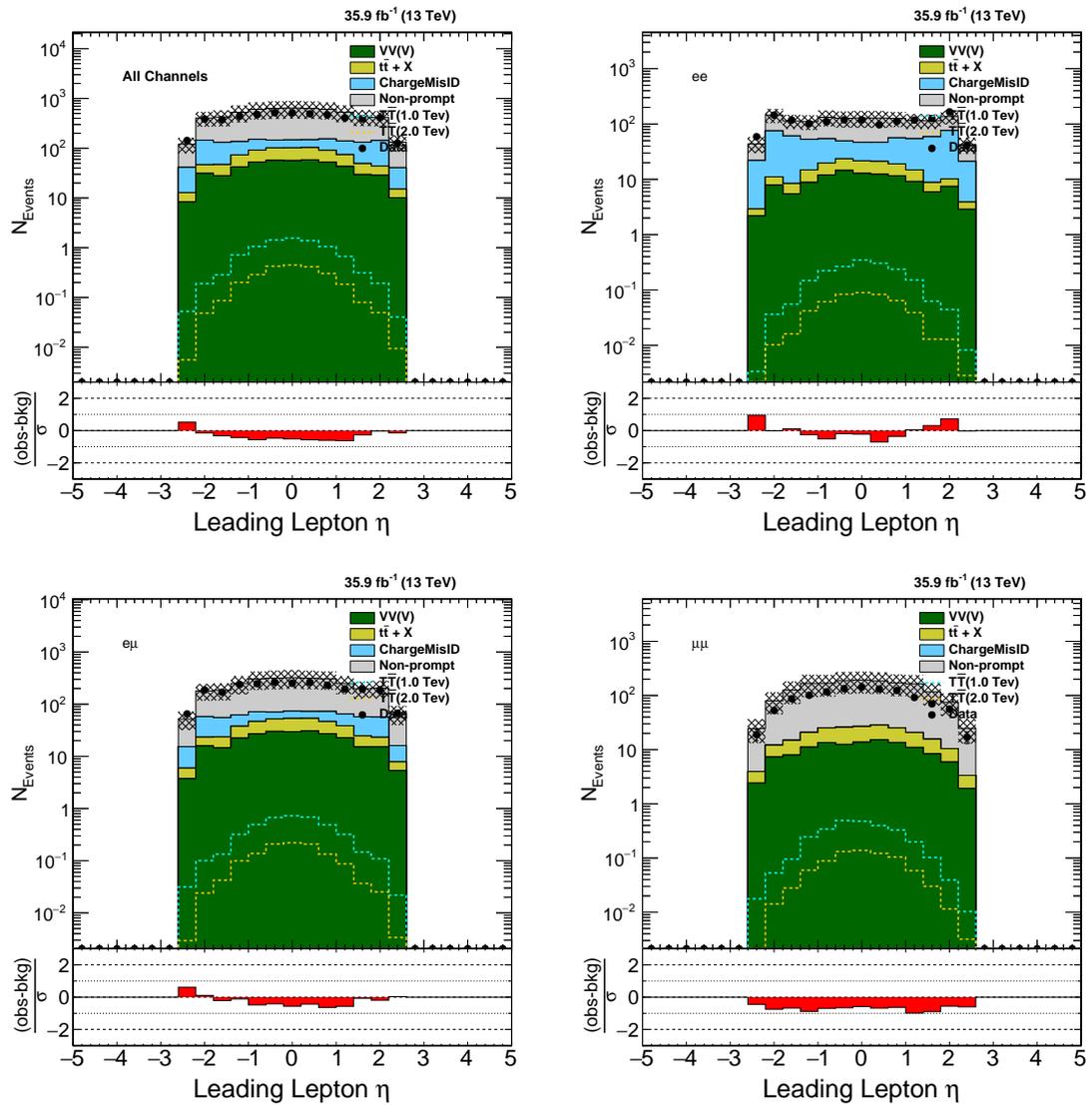


Figure 5.38: Distribution of leading lepton η for the combined flavor channel (All) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty.

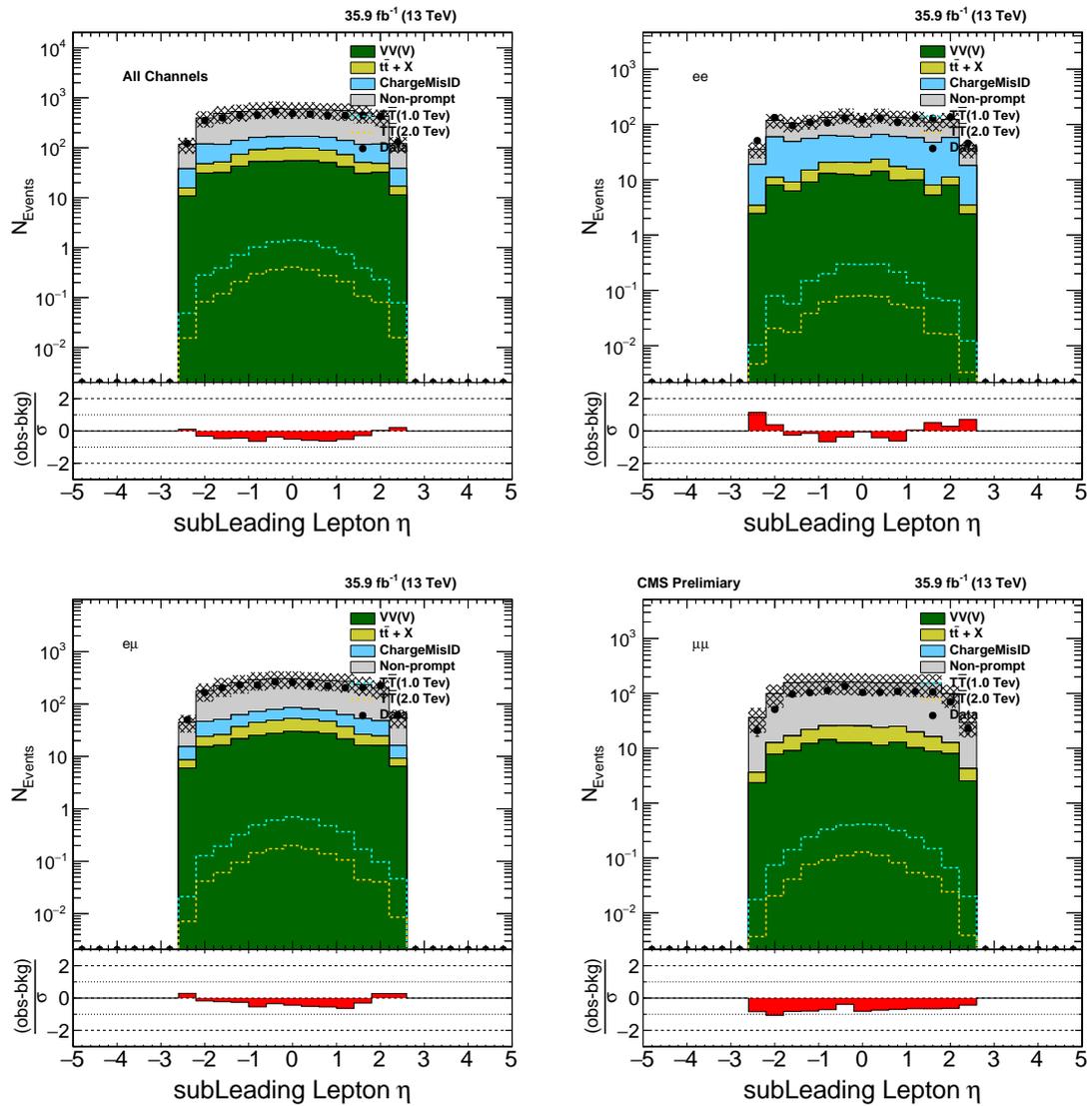


Figure 5.39: Distribution of Sub-leading lepton η for the combined flavor channel (All) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty.

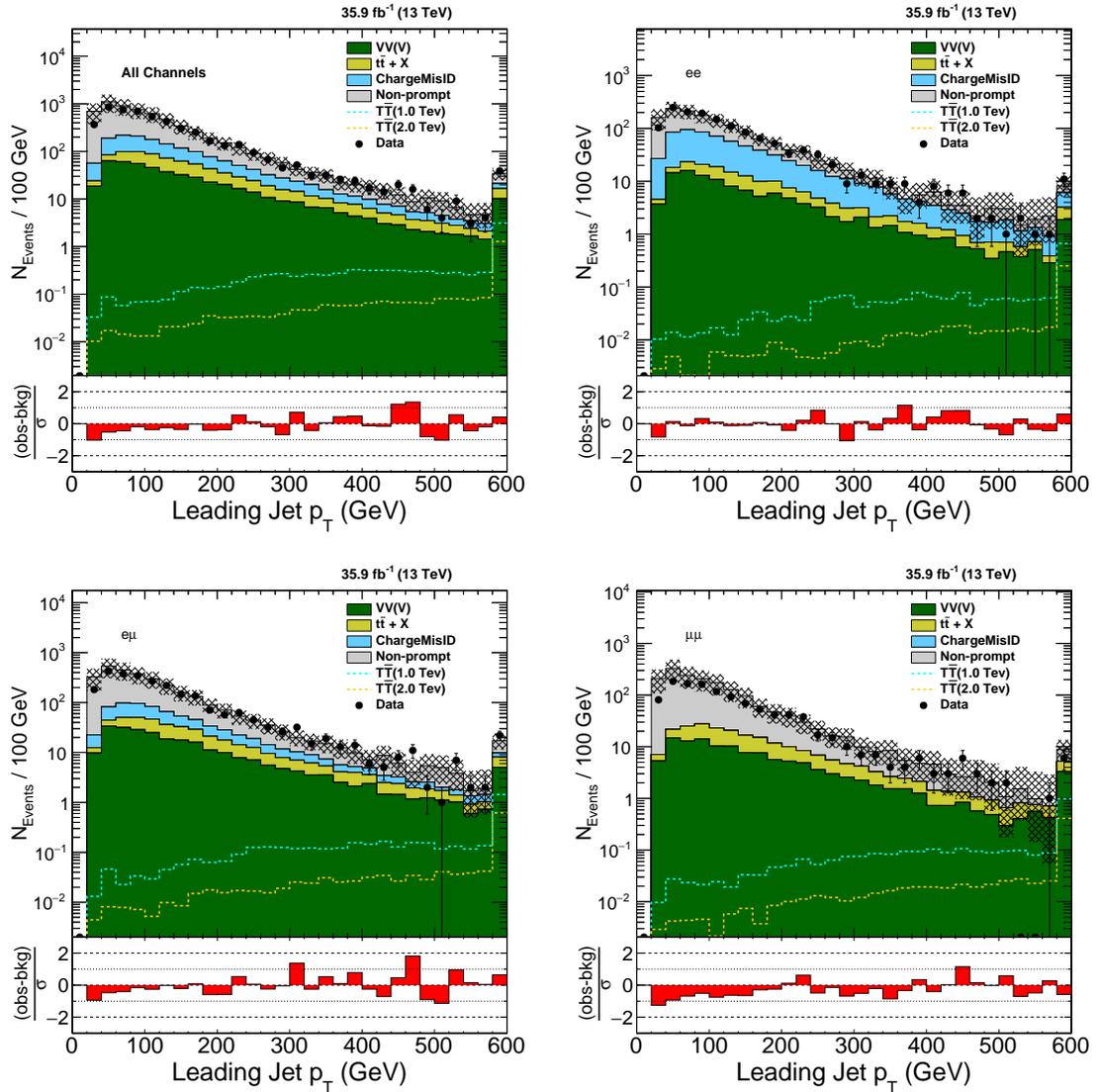


Figure 5.40: Distribution of leading Jet p_T for the combined flavor channel (All) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty.

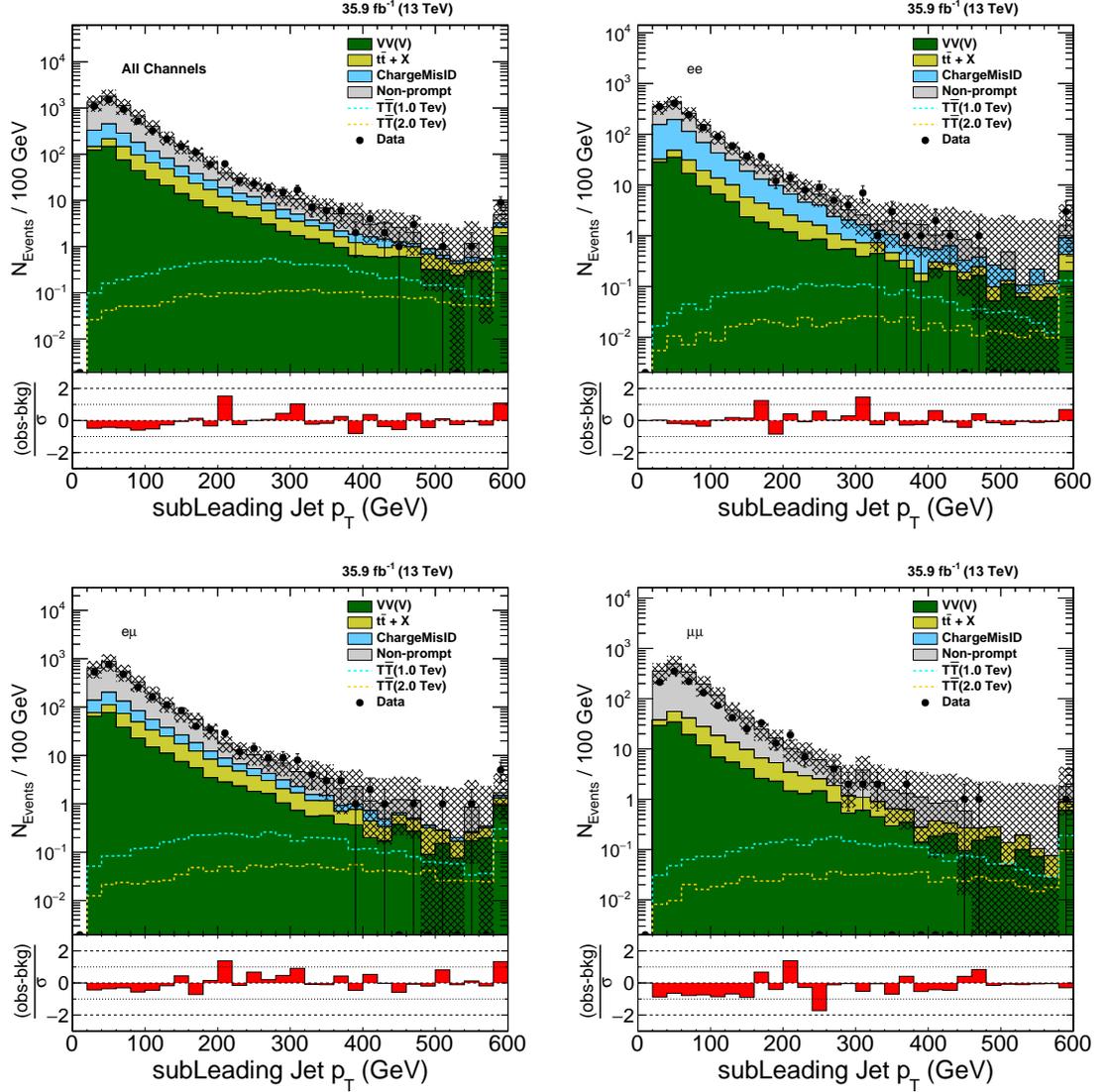


Figure 5.41: Distribution of sub-leading Jet p_T for the combined flavor channel (All) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty.

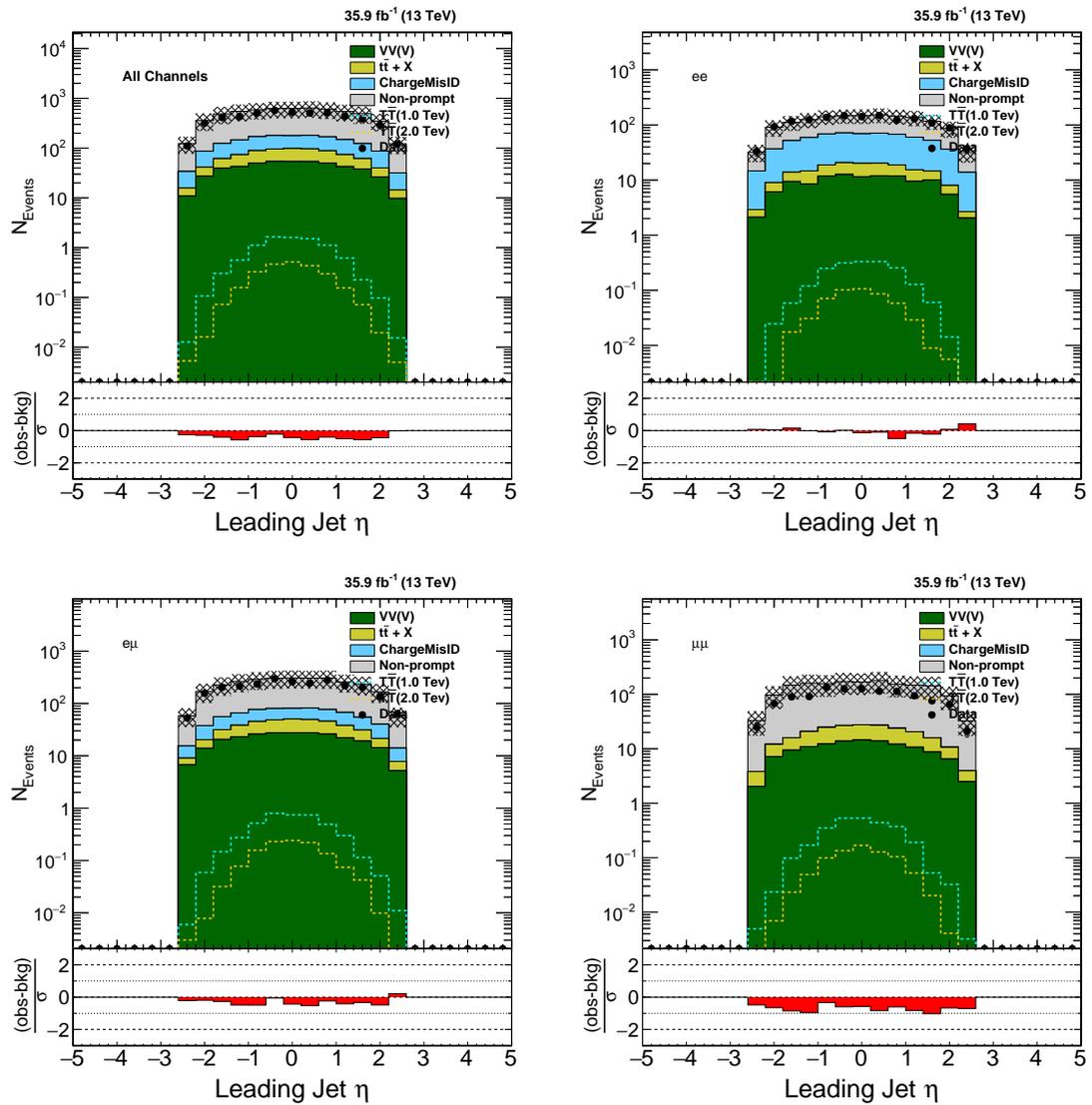


Figure 5.42: Distribution of leading Jet η for the combined flavor channel (All) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty.

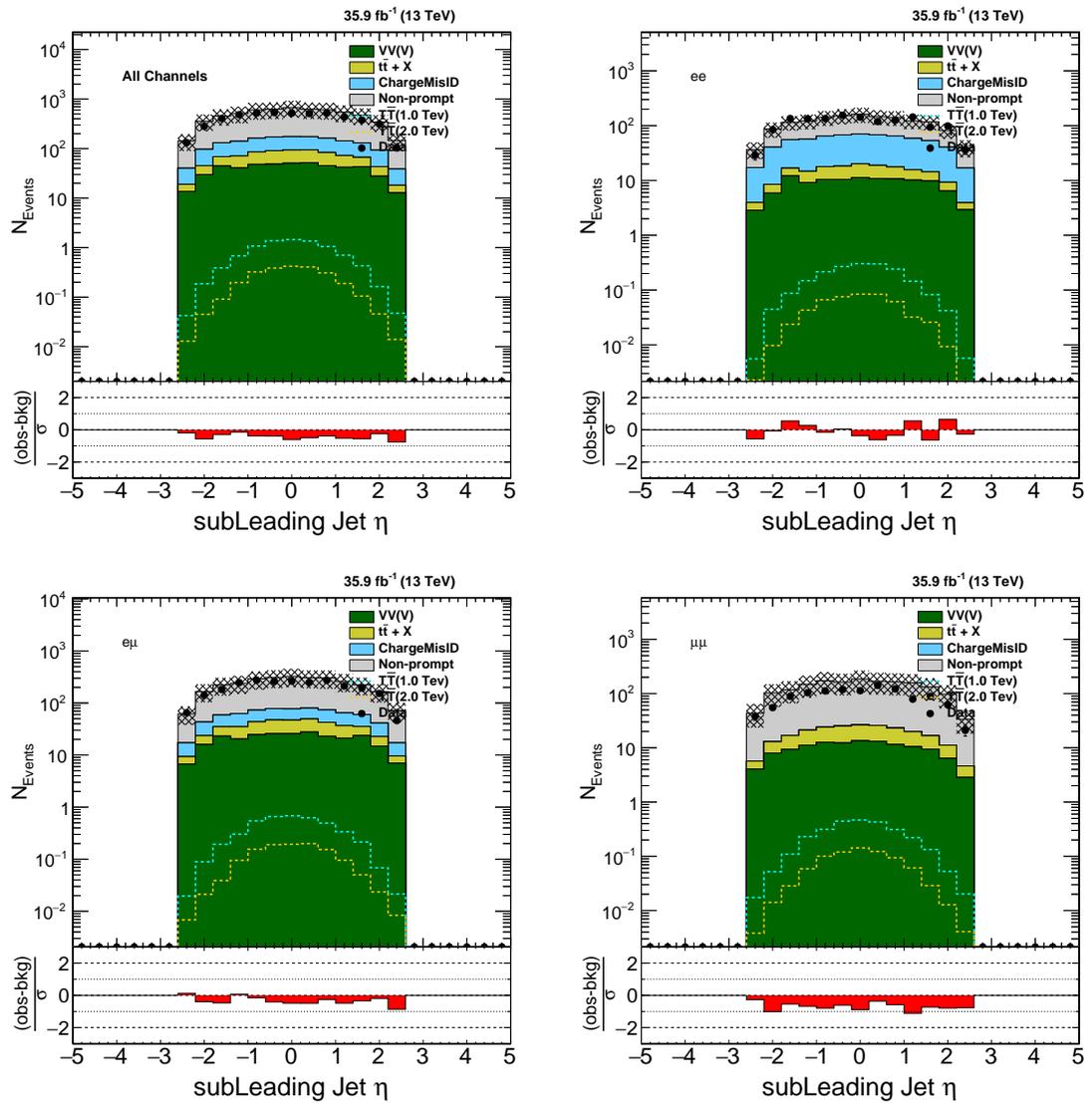


Figure 5.43: Distribution of sub-leading Jet η for the combined flavor channel (All) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty.

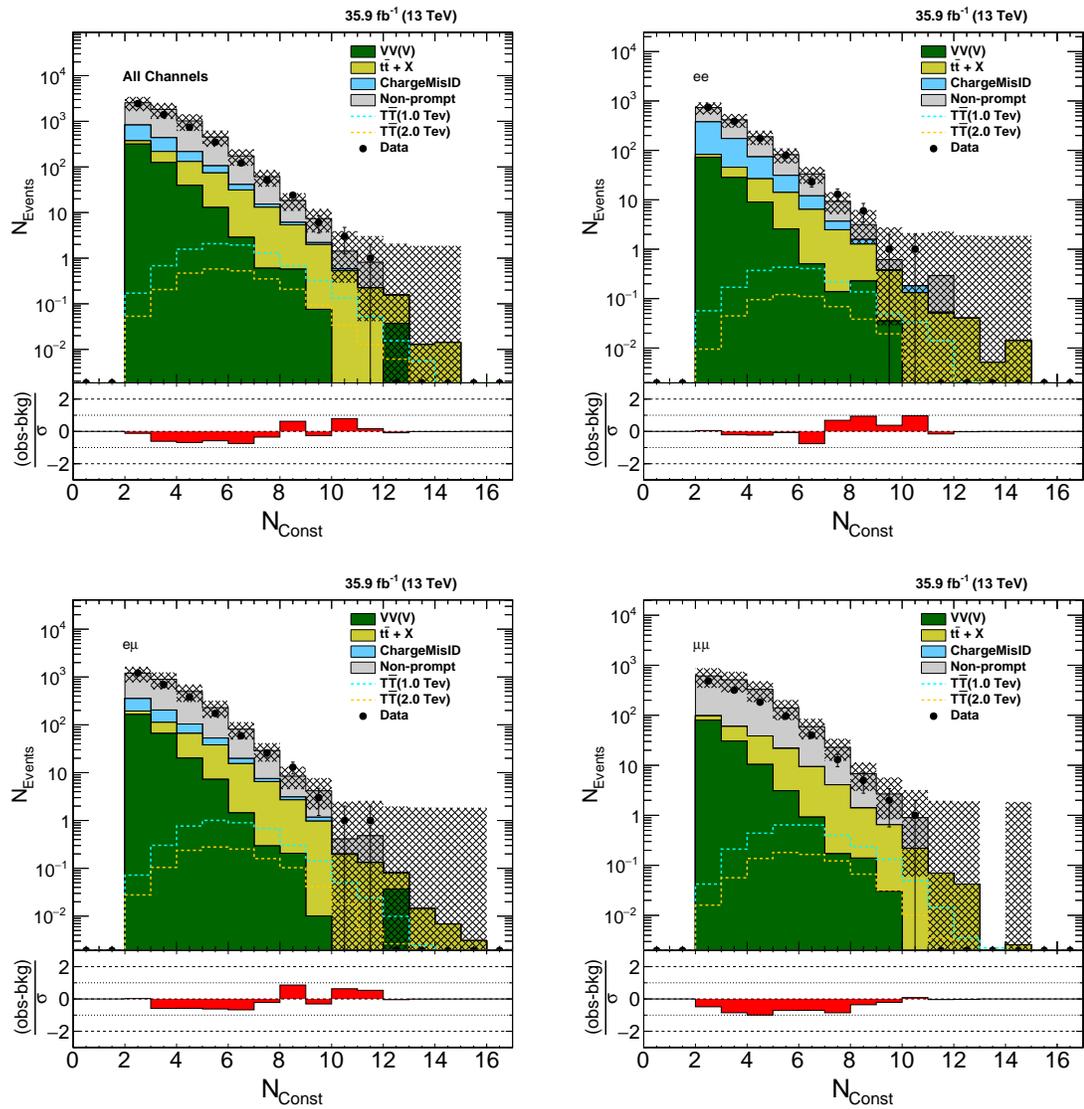


Figure 5.44: Distribution of $n\text{Const}$ for the combined flavor channel (*All*) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty.

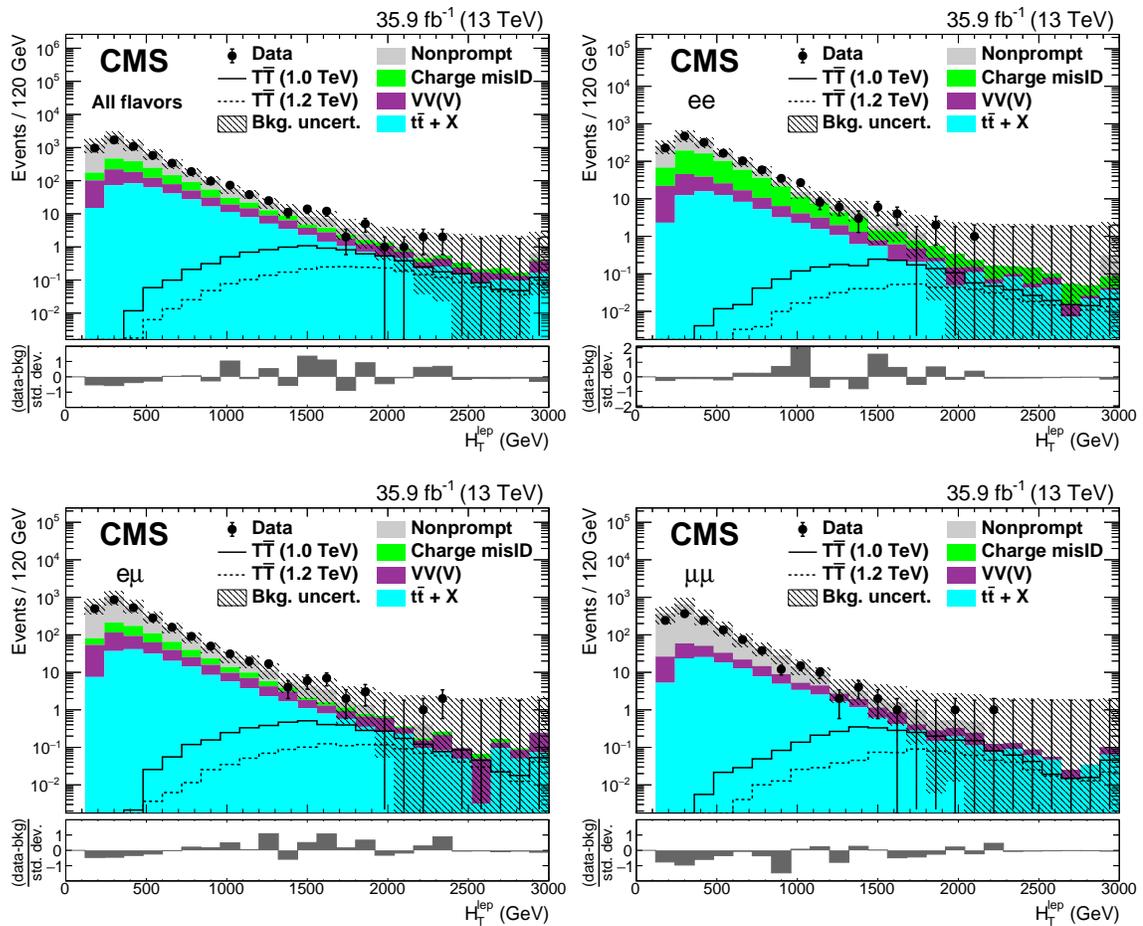


Figure 5.45: Distribution of H_T^{lep} for the combined flavor channel (All) and ee , $e\mu$, $\mu\mu$ flavor channels. Uncertainties in the plots include the statistical and systematic uncertainties. The lower panel shows the difference between data and background divided by the total uncertainty. [3]

5.9.2 Results

The final yields for the different channels are shown in Table 5.46. The yields for the different backgrounds and signal (for T of mass 1 TeV) are shown in Table 5.47 and Table 5.48 respectively. We also show the background rejection and signal efficiency in Table 5.49 and Table 5.50 respectively.

After applying all the selections, we count the final yields and use that information to calculate limits using Theta framework. We observe that the limits are a little under 800 GeV and 910 GeV for $T\bar{T}$ decaying to SS dilepton for the singlet and doublet branching ratios respectively. For $B\bar{B}$, we observe that the limits are 920 GeV and less than 800 GeV for $B\bar{B}$ decaying to SS dilepton for the singlet and doublet branching ratios respectively. The limits plots are shown in Figures 5.51 and 5.52. The corresponding postfit and nuisance correlation plot for $T\bar{T}$ singlet are shown in Figure 5.53.

In Figure 5.53, we show the postfit where we keep the electron fakerate nuisance separate from the muon fake rate nuisance for each of the lepton flavor channels. We assign electron fakerate nuisance to be 50% for the ee channel and muon fake rate nuisances to be 50% for the $\mu\mu$ channels, and for the $e\mu$ channel we assign both electron and muon fakerate nuisances to be 35% each. We observe that both electron and muon fake rate nuisances pull in the same direction with similar magnitudes and we observe no significant change in the resulting limits.

Figure 5.46: Summary of background yields and observed data after the full analysis selection for the 2016B-H datasets. Also shown are the number of expected events for a 1000 GeV VLQ T

	PSS MC	NonPrompt	ChargeMisID	Total Background	1000 GeV	Observed	$\frac{S}{\sqrt{S+B}}$
All	19.98 ± 2.78	20.68 ± 10.67	6.75 ± 2.07	47.41 ± 11.22	6.58 ± 0.36	52.00	0.90
ee	4.03 ± 0.59	4.65 ± 2.65	4.14 ± 1.29	12.82 ± 3.01	1.34 ± 0.08	12.00	0.36
$e\mu$	10.16 ± 1.41	10.58 ± 5.62	2.61 ± 0.81	23.35 ± 5.85	3.11 ± 0.18	31.00	0.61
$\mu\mu$	5.79 ± 0.86	5.44 ± 3.02	0.00 ± 0.00	11.23 ± 3.13	2.12 ± 0.12	9.00	0.58

Figure 5.47: Expected Number of Background Events vs. Analysis Cut for all channels combined, where NP is nonprompt and CMID is charge misidentified SS dilepton events.

Sample	SSDL	$(M_{dilep} > 20 \text{ GeV})$	EE-ZVeto	Lep1Pt > 40 GeV	nConst >= 4	$H_T^{lep} > 1200 \text{ GeV}$
ttZ	76.69 ± 11.05	76.05 ± 10.96	71.57 ± 10.31	70.71 ± 10.19	39.10 ± 5.66	1.99 ± 0.34
ttW	241.37 ± 51.74	239.13 ± 51.26	231.32 ± 49.59	228.62 ± 49.01	110.85 ± 23.77	8.95 ± 1.94
ttH	73.13 ± 20.42	71.81 ± 20.05	68.43 ± 19.10	67.07 ± 18.72	43.55 ± 12.16	2.07 ± 0.58
tttt	7.14 ± 3.15	7.09 ± 3.13	6.90 ± 3.04	6.86 ± 3.02	6.58 ± 2.90	1.62 ± 0.72
WZ	1569.18 ± 368.51	1545.07 ± 362.85	1460.31 ± 342.94	1416.61 ± 332.68	30.00 ± 7.13	1.43 ± 0.41
ZZ	159.94 ± 26.74	156.00 ± 26.08	143.66 ± 24.02	132.37 ± 22.13	1.97 ± 0.34	0.08 ± 0.02
WpWp	145.28 ± 74.26	145.12 ± 74.18	141.10 ± 72.13	139.37 ± 71.24	20.87 ± 10.67	3.28 ± 1.68
WWZ	14.09 ± 4.31	14.01 ± 4.29	13.35 ± 4.08	13.29 ± 4.06	2.71 ± 0.85	0.42 ± 0.15
WZZ	5.03 ± 1.20	5.00 ± 1.19	4.56 ± 1.08	4.48 ± 1.07	1.13 ± 0.28	0.15 ± 0.05
NP	10309.70 ± 5155.28	10131.61 ± 5066.23	7910.20 ± 3955.50	7409.45 ± 3705.12	1321.01 ± 660.84	20.68 ± 10.67
CMID	22377.33 ± 6713.22	22269.10 ± 6680.75	2688.90 ± 806.71	2589.06 ± 776.76	129.68 ± 38.94	6.75 ± 2.07

Figure 5.48: Expected Number of Signal Events vs. Analysis Cut for all channels combined

Sample	SSDL	$(M_{dilep} > 20 \text{ GeV})$	EE-ZVeto	Lep1Pt > 40 GeV	nConst >= 4	$H_T^{lep} > 1200 \text{ GeV}$
$T\bar{T}_{M1000} \text{ bWbW}$	0.17 ± 0.09	0.17 ± 0.09	0.17 ± 0.09	0.17 ± 0.09	0.09 ± 0.05	0.09 ± 0.05
$T\bar{T}_{M1000} \text{ tHbW}$	3.32 ± 1.68	3.32 ± 1.67	3.28 ± 1.65	3.28 ± 1.65	2.91 ± 1.47	2.40 ± 1.21
$T\bar{T}_{M1000} \text{ tHtH}$	1.56 ± 0.78	1.55 ± 0.78	1.53 ± 0.77	1.53 ± 0.77	1.49 ± 0.75	1.18 ± 0.60
$T\bar{T}_{M1000} \text{ tZbW}$	1.42 ± 0.72	1.42 ± 0.71	1.36 ± 0.69	1.36 ± 0.69	1.11 ± 0.56	0.93 ± 0.47
$T\bar{T}_{M1000} \text{ tZtH}$	2.08 ± 1.05	2.07 ± 1.04	2.03 ± 1.03	2.03 ± 1.02	1.94 ± 0.98	1.53 ± 0.77
$T\bar{T}_{M1000} \text{ tZtZ}$	0.63 ± 0.32	0.63 ± 0.32	0.60 ± 0.30	0.60 ± 0.30	0.54 ± 0.27	0.45 ± 0.23

Figure 5.49: Background rejection after applying SSDL and exactly two leptons requirements for all channels combined.

Sample	$(M_{dilep} > 20 \text{ GeV})$	EE-ZVeto	Lep1Pt > 40 GeV	nConst ≥ 4	HT > 1200 GeV
ttZ	0.0084	0.0667	0.0780	0.4902	0.9741
ttW	0.0093	0.0416	0.0528	0.5408	0.9629
ttH	0.0180	0.0642	0.0829	0.4045	0.9716
tttt	0.0065	0.0339	0.0394	0.0781	0.7735
WZ	0.0154	0.0694	0.0972	0.9809	0.9991
ZZ	0.0246	0.1017	0.1724	0.9877	0.9995
WpWp	0.0011	0.0287	0.0407	0.8563	0.9774
WWZ	0.0057	0.0528	0.0574	0.8080	0.9705
WZZ	0.0060	0.0948	0.1103	0.7755	0.9695
NonPrompt	0.0173	0.2327	0.2813	0.8719	0.9980
ChargeMisID	0.0048	0.8798	0.8843	0.9942	0.9997

Figure 5.50: Signal efficiency after applying SSDL and exactly two leptons requirements for all channels combined

Sample	$(M_{dilep} > 20 \text{ GeV})$	EE-ZVeto	Lep1Pt > 40 GeV	nConst ≥ 4	HT > 1200 GeV
$T\bar{T}$ M_{1000} bWbW	0.9759	0.9759	0.9759	0.5254	0.5167
$T\bar{T}$ M_{1000} tHbW	0.9973	0.9867	0.9857	0.8767	0.7219
$T\bar{T}$ M_{1000} tHtH	0.9971	0.9869	0.9843	0.9583	0.7599
$T\bar{T}$ M_{1000} tZbW	0.9970	0.9603	0.9603	0.7835	0.6536
$T\bar{T}$ M_{1000} tZtH	0.9971	0.9793	0.9776	0.9324	0.7364
$T\bar{T}$ M_{1000} tZtZ	0.9974	0.9487	0.9459	0.8569	0.7066

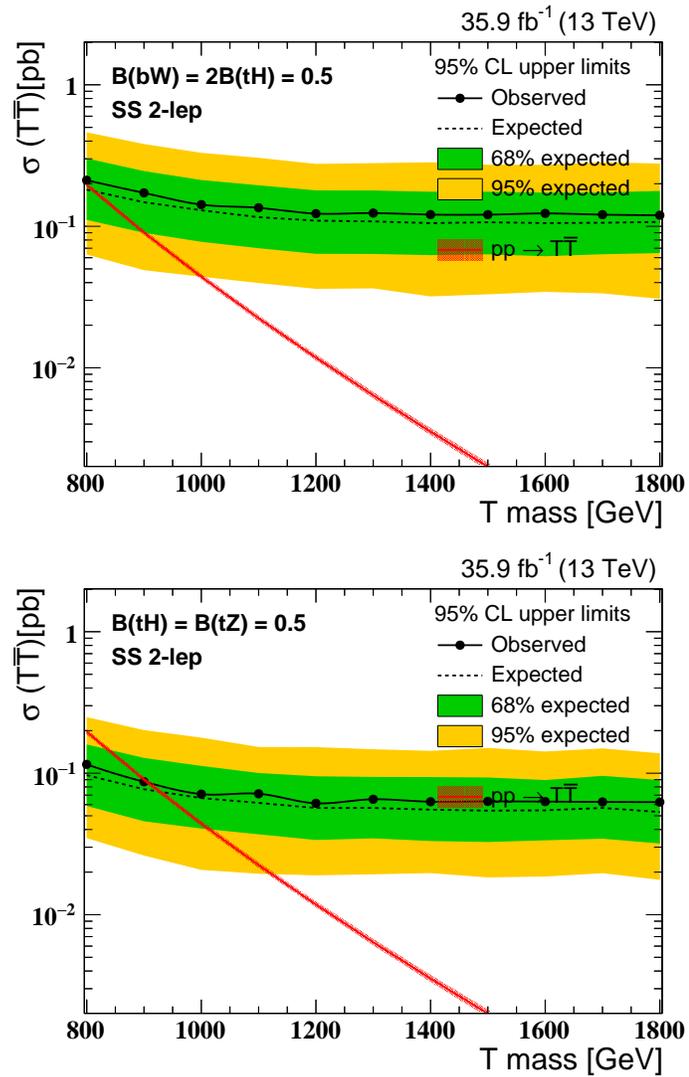


Figure 5.51: SS dilepton final state cross-section limits (top: with singlet branching fractions: 50% bW, 25% tH, 25% tZ), bottom: with doublet branching fractions: 0% bW, 50% tH, 50% tZ), shown as a function of mass of T .

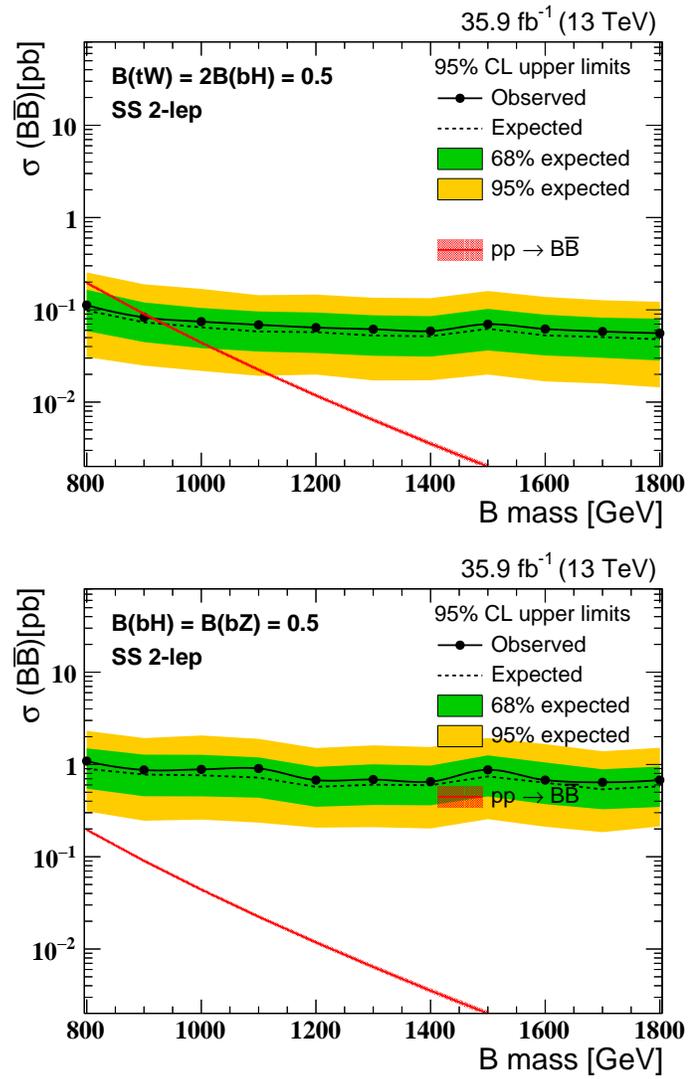


Figure 5.52: SS dilepton final state cross-section limits (top: with singlet branching fractions: 50% tW, 25% bH, 25% bZ), bottom: with doublet branching fractions: 0% tW, 50% bH, 50% bZ), shown as a function of mass of B.

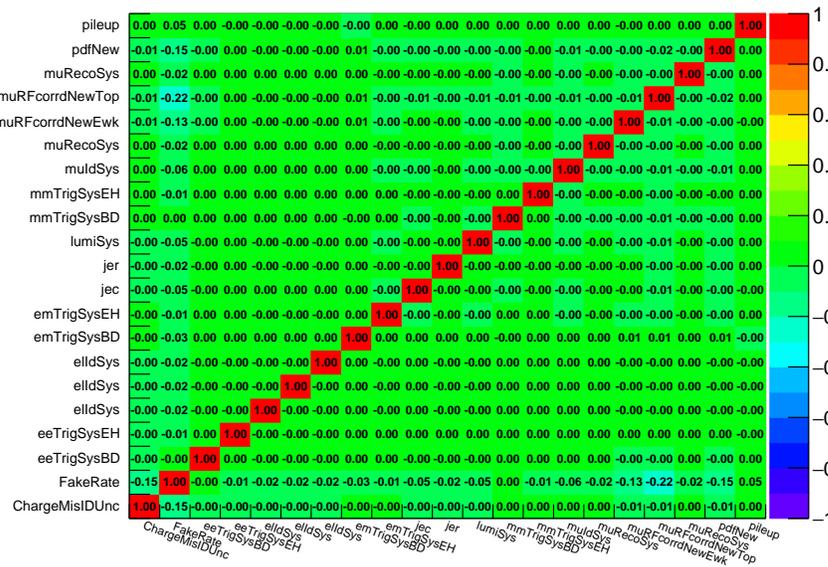
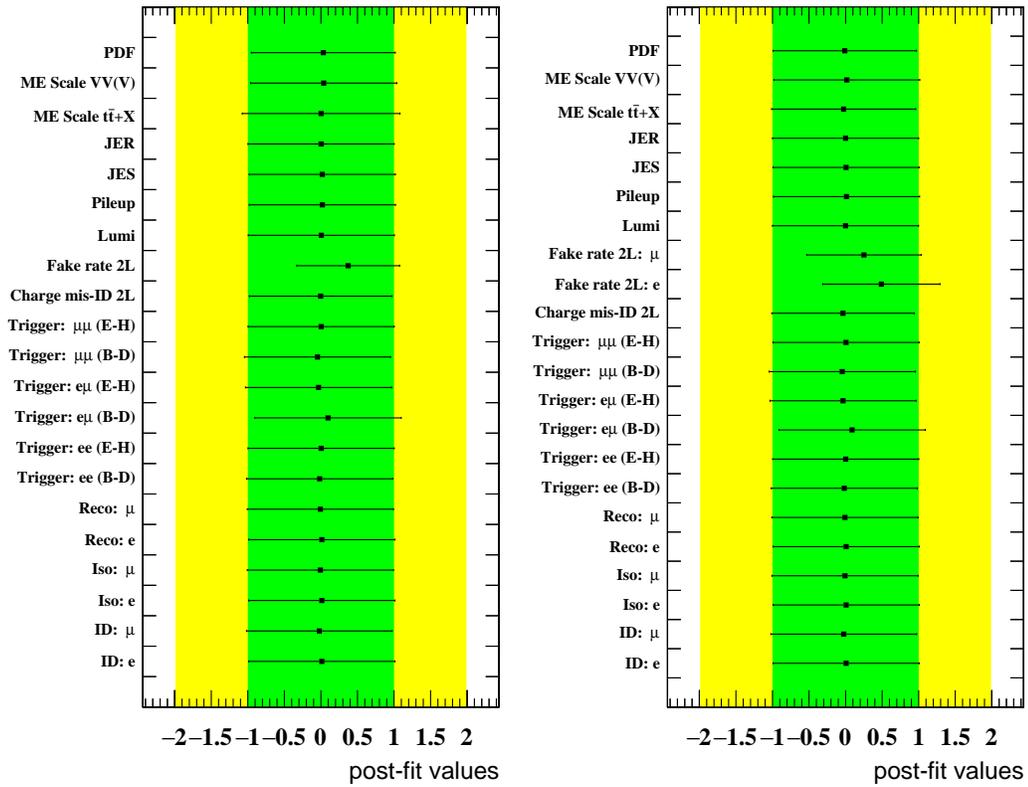


Figure 5.53: Postfit and correlation plots. The top right postfit plot corresponds to an alternate limit calculation setting where the electron and muon fake rate nuisances are kept separate.

5.10 Combined results: trilepton, SS dilepton, single-lepton

The equivalent VLQ $T\bar{T} / B\bar{B}$ search with exactly one electron/muon final states have been performed and documented in the following reference: [3].

Using the same method as used for the trilepton analysis, the combined results use the THETA program [120], and we calculate Bayesian credible intervals [121] to set 95% CL upper limits on the production cross section of $T\bar{T}$ at each simulated mass point, for various branching fraction scenarios. Limits are calculated in a simultaneous fit to binned marginal likelihoods from the $\min[M(l, b)]$ and S_T distributions for the 16 single-lepton signal-region categories, H_T distributions for the 6 single-lepton aggregate control regions, event yields for the SS dilepton channel, and S_T distributions for the 4 trilepton categories. Statistical uncertainties in the background estimates are treated using the Barlow–Beeston light method [122, 123]. Other systematic uncertainties are treated as nuisance parameters, as listed in Table 5.26. Normalization uncertainties are given log-normal priors, and shape uncertainties with shifted templates are given Gaussian priors with a mean of zero and width of one. The signal cross section is assigned a flat prior distribution.

Figure 5.54 shows 95% CL upper limits on the production of T and B quarks in the benchmark branching fraction scenarios. We exclude singlet T quark masses below 1200GeV (1160GeV expected), doublet T quark masses below 1280GeV (1240GeV expected), singlet B quark masses below 1170GeV (1130GeV expected), and doublet B quark masses below 940GeV (920GeV expected). Masses below 800GeV were excluded in previous searches. For T and B quark masses in the range 800–1800GeV, cross sections smaller than 30.4–9.4fb (21.2–6.1fb) and 40.6–9.4fb (101–49.0fb) are excluded for the singlet (doublet) scenario. Figure 5.55 shows the expected and observed limits for scans over many possible T and B quark branching fraction scenarios. Based on the branching fractions, lower limits on T and B quark masses range from 1140 to 1300GeV, and from 910 to 1240GeV.

Table 5.26: Summary of values for normalization uncertainties and dependencies for shape uncertainties. The symbol σ denotes one standard deviation of the uncertainty and “env” denotes an envelope of values. Background from opposite-sign dilepton events is denoted “OS”, background from nonprompt leptons is denoted “NP”, while other backgrounds modeled from simulation are denoted “MC”. For signals, theoretical uncertainties are labeled as “Shape” for shape-based searches, and “Accept.” for counting experiments. Additionally, “CR” denotes control region and “RMS” denotes root mean square. [3]

Source	Uncertainty	1ℓ		SS 2ℓ		$\geq 3\ell$	
		Sig	Bkg	Sig	Bkg	Sig	Bkg
Integrated luminosity	2.5%	Yes	MC	Yes	MC	Yes	MC
Reconstruction	1%	Yes	MC	Yes	MC	Yes	MC
Identification	2%(e), 3%(μ)	Yes	MC	Yes	MC	Yes	MC
Isolation (e, μ)	1%	Yes	MC	Yes	MC	Yes	MC
Trigger (e or μ)	$\pm\sigma(p_T, \eta)$	Yes	MC	—	—	—	—
Trigger ($\ell\ell$)	3%	—	—	Yes	MC	—	—
Trigger ($\ell\ell\ell$)	3%	—	—	—	—	Yes	MC
Charge misid. rate	30%	—	—	No	OS	—	—
ℓ misid. efficiency	50%	—	—	No	NP	—	—
ℓ misid. efficiency	4–30%	—	—	—	—	No	NP
μ misid. efficiency η dep.	12–33%	—	—	—	—	No	NP
NP method closure	17–31%	—	—	—	—	No	NP
NP method in CR	2–35%	—	—	—	—	No	NP
Prompt ℓ efficiency	2–9% (e), 1–7% (μ)	—	—	—	—	No	NP
Pileup	$\sigma_{\text{inel.}} \pm 4.6\%$	Yes	MC	Yes	MC	Yes	MC
Jet energy scale	$\pm\sigma(p_T, \eta)$	Yes	MC	Yes	MC	Yes	MC
Jet energy res.	$\pm\sigma(\eta)$	Yes	MC	Yes	MC	Yes	MC
H_T scaling	env(upper, lower fits)	No	W+jets	—	—	—	—
b tag: b	$\pm\sigma(p_T)$	Yes	MC	—	—	Yes	MC
b tag: light	$\pm\sigma$	Yes	MC	—	—	Yes	MC
W tag: τ_2/τ_1	$\pm\sigma$	Yes	MC	—	—	—	—
W tag: $\tau_2/\tau_1 p_T$	$\pm\sigma(p_T)$	Yes	MC	—	—	—	—
W/H tag: mass scale	$\pm\sigma(p_T, \eta)$	Yes	MC	—	—	—	—
W/H tag: mass res.	$\pm\sigma(\eta)$	Yes	MC	—	—	—	—
H tag: propagation	5%	Yes	MC	—	—	—	—
Renorm./fact. scale	env($\times 2, \times 0.5$)	Shape	MC	Accept.	MC	Shape	MC
PDF	RMS(replicas)	Shape	MC	Accept.	MC	Shape	MC
VV rate	15%	No	VV	—	—	—	—
Single tW rate	16%	No	tW	—	—	—	—

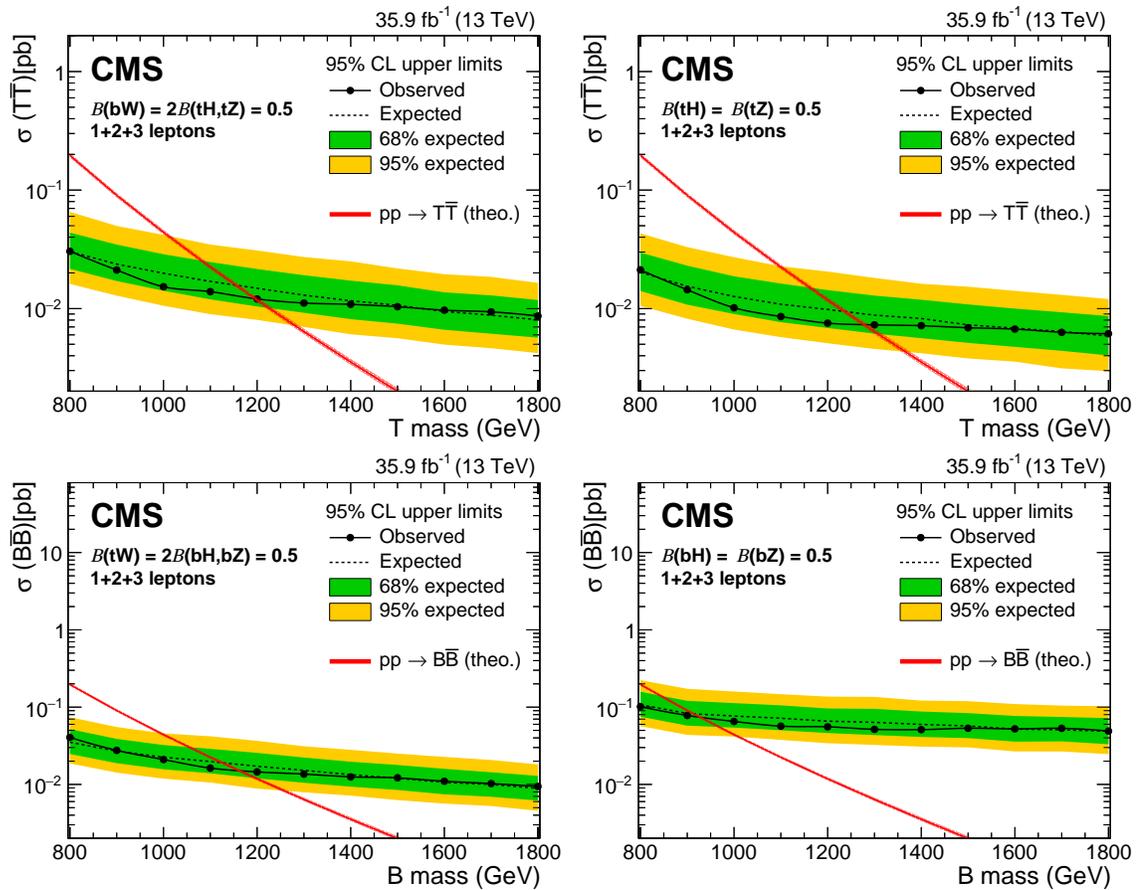


Figure 5.54: The 95% CL expected and observed upper limits on the cross section of $T\bar{T}$ (upper row) and $B\bar{B}$ (lower row) production after combining all channels for the singlet (left) and doublet (right) branching fraction scenarios. The predicted cross sections are shown by the red curve, with the uncertainty indicated by the width of the line. [3]

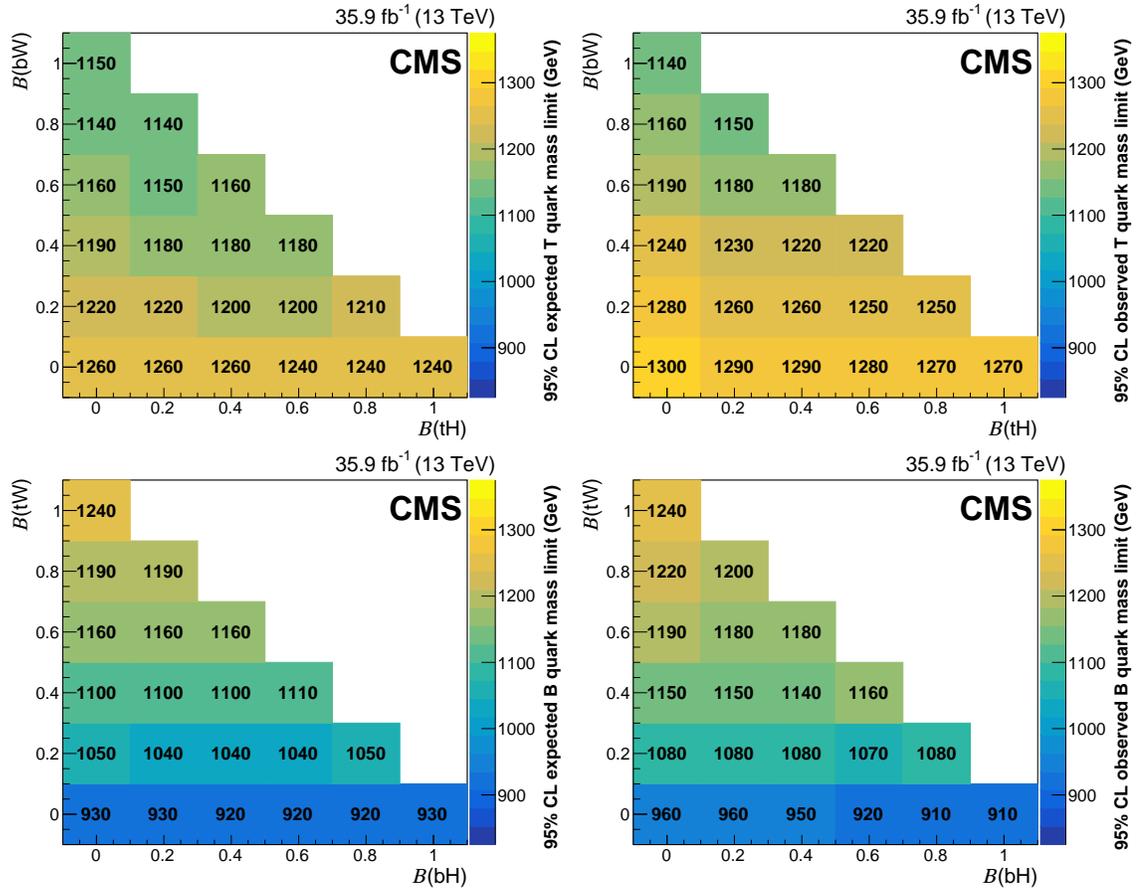


Figure 5.55: The 95% CL expected (left) and observed (right) lower limits on the T quark (upper row) and B quark (lower row) mass, expressed in GeV, after combining all channels for various branching fraction scenarios. [3]

Chapter 6

The XCone jet clustering algorithm and the search for VLQ

$$T\bar{T} / B\bar{B} \rightarrow e/\mu + \text{jets}$$

This chapter presents the XCone jet clustering algorithm, its use in CMS data, in the context of the search for VLQ $T\bar{T} / B\bar{B} \rightarrow e/\mu + \text{jets}$.

6.1 The XCone Jet Algorithm

In this section we introduce the XCone jet algorithm and present studies which explore the use of N -jettiness [19, 124] as way to determine the multiplicity of jets in an event. In the proceeding sections we present the data validation of the algorithm, and following that we present how XCone jets could be used in the search for top/bottom-like VLQ decaying to final states with one electron/muon.

6.1.1 N -Jettiness and XCone

N -jettiness is a value that measure likelihood of an event containing N jets, where N denotes the number of jets. More specifically, N -jettiness is a measure of how the particles are distributed along

given N axes. See equation in Figure 6.1 for the definition of N -jettiness. The XCone jet algorithm is an exclusive cone jet algorithm based on minimizing the event N -Jettiness, $\mathcal{T} = \min_{n_1, n_2, \dots, n_N} \tilde{\mathcal{T}}$, that returns a fixed number of jets.

We use the default N -Jettiness/XCone parameter settings as recommended by the original authors of XCone [19] (See Figure 6.2). We set the distance variable R to be 0.4 and in this study we will use PF [17] candidates with charge hadron subtraction as input to XCone.

$$\tilde{\mathcal{T}}_N = \sum_i \min \{ \rho_{\text{jet}}(p_i, n_1), \dots, \rho_{\text{jet}}(p_i, n_N), \rho_{\text{beam}}(p_i) \}$$

Figure 6.1: N -jettiness definition.

We use N -Jettiness with the following definitions (XCone Default) :

Name	Measure definitions	
	$\rho_{\text{jet}}(p_i, n_A)$	$\rho_{\text{beam}}(p_i)$
Conical Geometric	$\frac{p_{Ti}}{(2 \cosh y_i)^{\gamma-1}} \left(\frac{2 n_A \cdot p_i}{n_{TA} p_{Ti}} \frac{1}{R^2} \right)^{\beta/2}$	$\frac{p_{Ti}}{(2 \cosh y_i)^{\gamma-1}}$
XCone Default ($\beta = 2, \gamma = 1$)	$\frac{2 \cosh y_A}{R^2} n_A \cdot p_i$	p_{Ti}

```
//Njettiness Stuff - start
const int Nmax = 20;
Njettiness _njettiness(OnePass_GenET_GenKT_Axes(delta, power, XConeR), XConeMeasure(XConeBeta, XConeR));
```

Axis definition
Measure definition

- XCone R = 0.4
- XConeBeta = 2
- Delta = 1
- Power = 1/2

Figure 6.2: XCone definitions and parameters used in this study. [19]

We explore ways to determine jet multiplicity, N , based on the N -jettiness value. We use $t\bar{t}$ simulated events and vector-like quarks $T\bar{T} \rightarrow tHtH$ simulated events, where the final states are purely hadronic. We compare the jet multiplicity based on using N -jettiness and based on counting the number of AK4 jets clustered in the event. Then we produce XCone jets based on the resulting N we obtain.

6.1.2 Samples and Event Selections

We use Monte Carlo samples of vector-like quarks $T\bar{T}$ with different mass points¹ and Standard Model $t\bar{t}$ from the RunIISummer16MiniAODv2 campaign. The samples are tabulated in Table 6.1.

Table 6.1: MC samples used in this study.

Monte Carlo (MC) Dataset
TprimeTprime_M-800_TuneCUETP8M1_13TeV-madgraph-pythia8
TprimeTprime_M-1300_TuneCUETP8M1_13TeV-madgraph-pythia8
TprimeTprime_M-1800_TuneCUETP8M1_13TeV-madgraph-pythia8
TT_TuneCUETP8M2T4_13TeV-powheg-pythia8

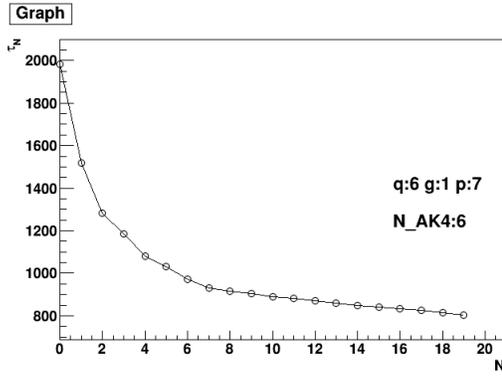
Event selection is performed by reading events from the miniAOD [104] file format with the LJMet software framework [105]. We select events where the final states are all hadronic by applying veto on events that contain (truth level) leptons based on the truth information. Throughout this study, we use AK4 to refer jets clustered using the anti-kT algorithm [112] with a distance parameter of $R=0.4$ and uses charge hadron subtracted PF candidates (PFchs). We also require that there at least be two central ($|\eta| < 2.4$) AK4 jets in the event, each with $p_T > 30$ GeV.

6.1.3 Determining Event Shape

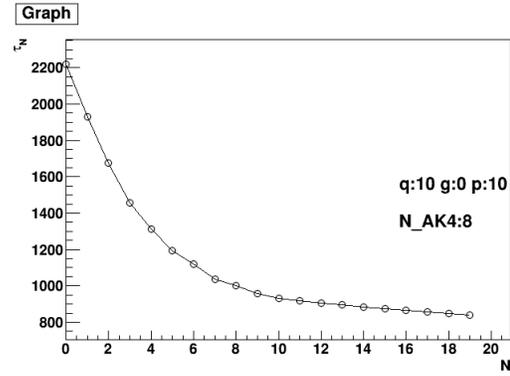
Given the values of the event N -jettiness, \mathcal{T}_N , plotted over some range of N we discuss ways to determine the jet multiplicity and shape of the event. Figure 6.3 shows the N -jettiness values plotted against N from 0 to 19 with distance parameter of $R = 0.4$ in $T'T' \rightarrow tHtH$ and $t\bar{t}$ events. The figures also show the number of generator level partons produced in the events. We observe that the N -jettiness plots initially decrease sharply with N and then it flattens at some higher N value. The N -jettiness plot by itself does not provide a clear and precise indication of the number of jets in the event and so we also look at other variables that are derived from N -jettiness: $\mathcal{T}_N - \mathcal{T}_{N-1}$, $\frac{\mathcal{T}_N}{\mathcal{T}_{N-1}}$ and $\frac{\mathcal{T}_N - \mathcal{T}_{N-1}}{\mathcal{T}_N}$. Typical plots of these variables for an event are shown in Figures 6.4, 6.5, 6.6 respectively.

Our strategy is to use the event N -jettiness information together with the number of partons

¹Higher T' mass point provides events with more boosted jets.

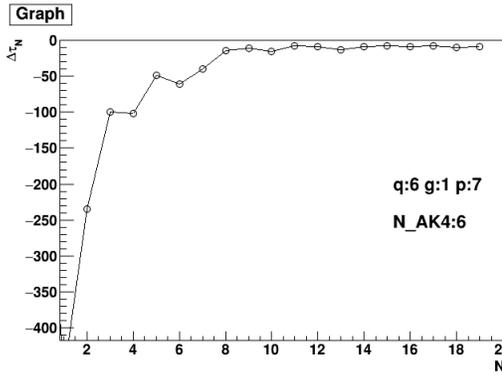


(a) N -jettiness over N , for a $t\bar{t}$ event

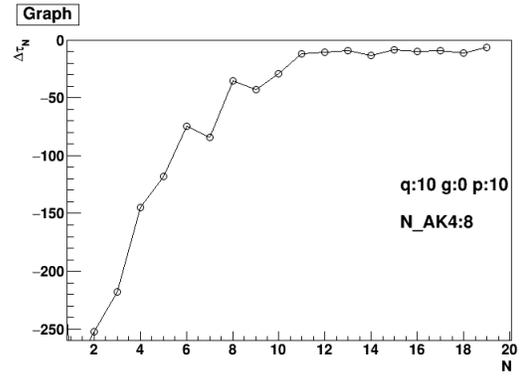


(b) N -jettiness over N , for a $T'T' \rightarrow tHtH$ event

Figure 6.3: N -jettiness plots for N from 0 to 19 with labels that shows the number of quarks (q), gluons (g), partons (p).

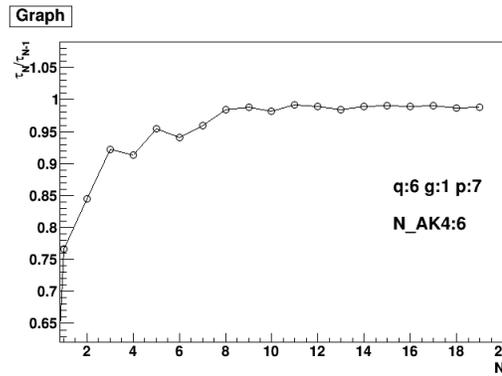


(a) $t\bar{t}$ event

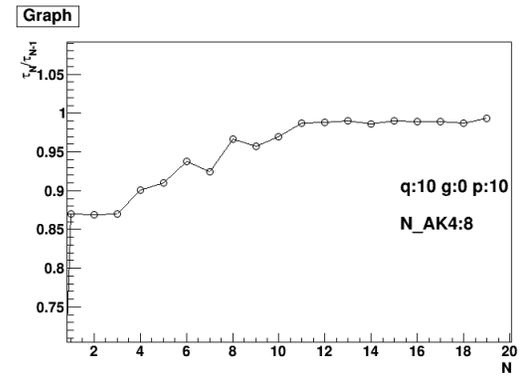


(b) $T'T' \rightarrow tHtH$ event

Figure 6.4: $\mathcal{T}_N - \mathcal{T}_{N-1}$, ($\Delta\mathcal{T}_N$), plots for N from 2 to 19 with labels that shows the number of quarks (q), gluons (g), partons (p).



(a) $t\bar{t}$ event



(b) $T'T' \rightarrow tHtH$ event

Figure 6.5: $\frac{\mathcal{T}_N}{\mathcal{T}_{N-1}}$ plots for N from 2 to 19 with labels that shows the number of quarks (q), gluons (g), partons (p).

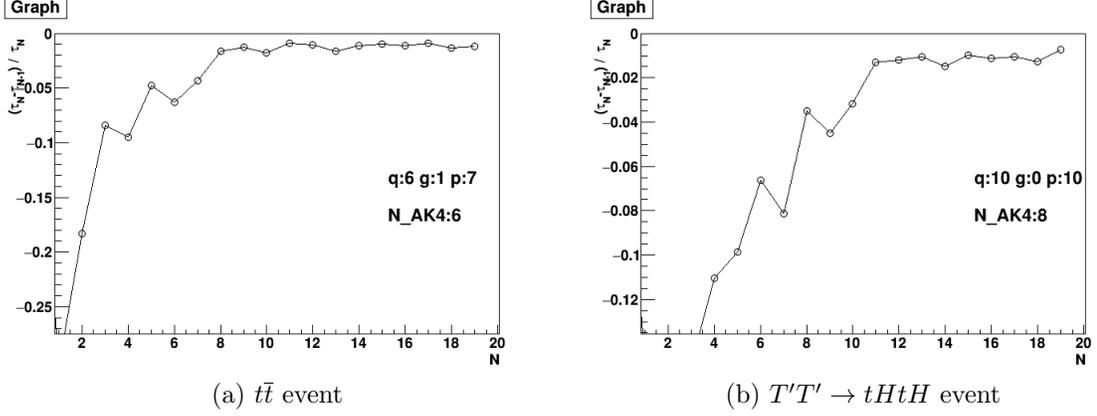


Figure 6.6: $\frac{\mathcal{T}_N - \mathcal{T}_{N-1}}{\mathcal{T}_N}$ plots for N from 2 to 19 with labels that shows the number of quarks (q), gluons (g), partons (p).

from truth information, as a guide to determine the "optimal" number of jets (to cluster). The plots shown in Figures 6.4, 6.5, 6.6 commonly show a sharp rise at low N followed by a plateau at some higher N . For $t\bar{t}$ the plateau occurs at lower N (≈ 6) in comparison to $T'T'$ (≈ 10). This motivates some cut-off or threshold scheme on these plots to determine the optimal N .

We consider applying a fixed threshold to determine the optimal N . For $\mathcal{T}_N - \mathcal{T}_{N-1}$, we apply a threshold at -30 . For $\frac{\mathcal{T}_N}{\mathcal{T}_{N-1}}$ we apply a threshold at 0.975 and for $\frac{\mathcal{T}_N - \mathcal{T}_{N-1}}{\mathcal{T}_N}$ we apply a threshold at -0.025 .

We scan each point and compare them to the threshold value. We consider two ways of scanning: forward scanning, starting from the lowest N point, or reverse scanning, starting from highest N . For forward scanning, the optimal N is the first point where the y-value is greater than the threshold. For reverse scanning, the optimal N is the point before the y-value is less than the threshold. If no points breaks the threshold, the threshold is lowered by 5 for $\mathcal{T}_N - \mathcal{T}_{N-1}$ and 0.025 for the other variables and then the scanning is performed once more. This process is repeated until the threshold is broken. See Figure 6.11 for an illustration of the step when the threshold is not broken at first scanning.

As each event will have an optimal N value, we plot the resulting distribution and see how 6-prong $t\bar{t}$ events compare with 10-prong ($T'T' \rightarrow tHtH$) events. Figure 6.7 shows AK4 multiplicity distributions jets. Figure 6.8 shows optimal N distributions based on $\mathcal{T}_N - \mathcal{T}_{N-1}$ distributions.

Figure 6.9 shows optimal N distributions based on $\frac{\mathcal{T}_N}{\mathcal{T}_{N-1}}$ distributions. Figure 6.10 shows optimal N distributions based on $\frac{\mathcal{T}_N - \mathcal{T}_{N-1}}{\mathcal{T}_N}$ distributions.

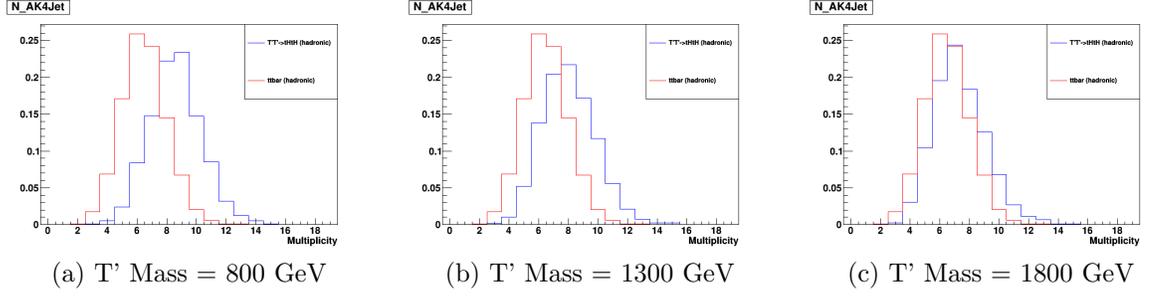


Figure 6.7: AK4 multiplicities in $t\bar{t}$ and $T'T'$.

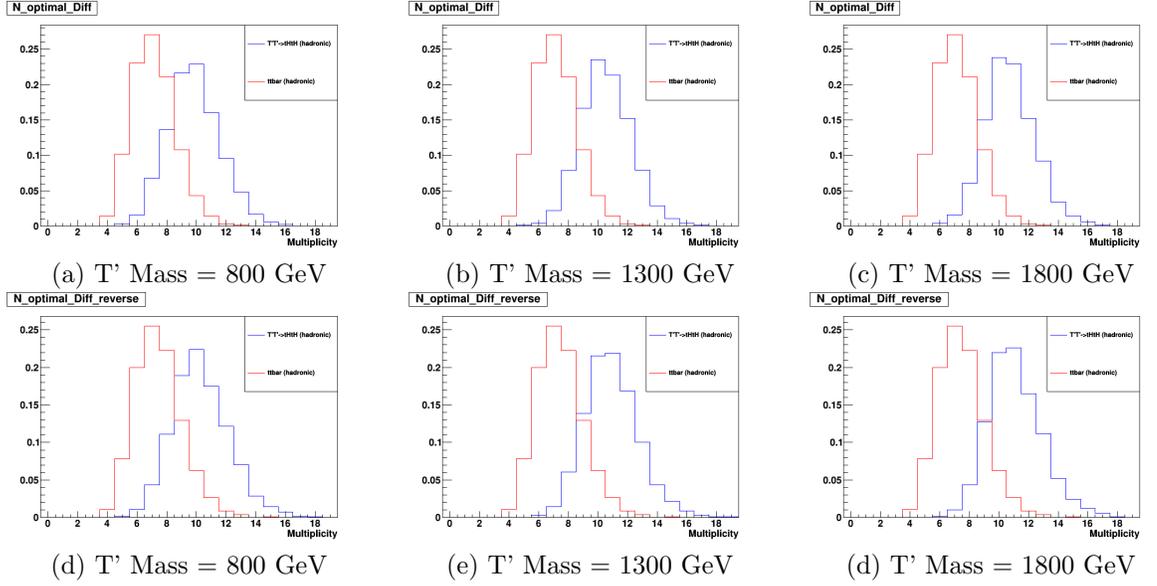


Figure 6.8: Optimal N in $t\bar{t}$ and $T'T'$ based on a forward, (a)-(c), and reverse, (d)-(f), scanning the $\mathcal{T}_N - \mathcal{T}_{N-1}$ distribution.

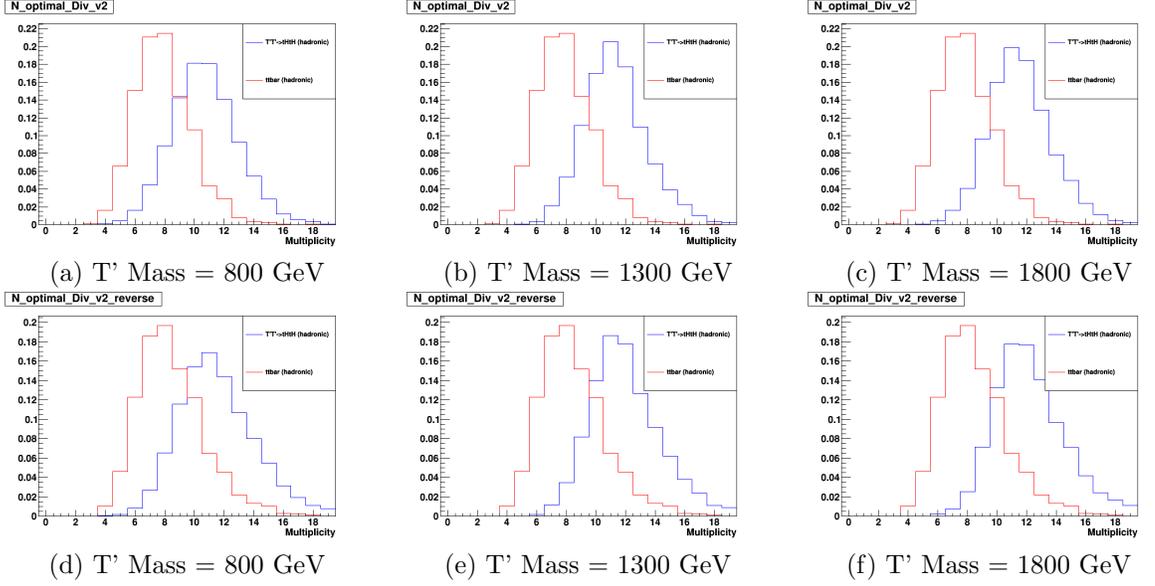


Figure 6.9: Optimal N in $t\bar{t}$ and $T'T'$ based on forward, (a)-(c), and reverse, (d)-(f), scanning the $\frac{\mathcal{T}_N}{\mathcal{T}_{N-1}}$ distribution.

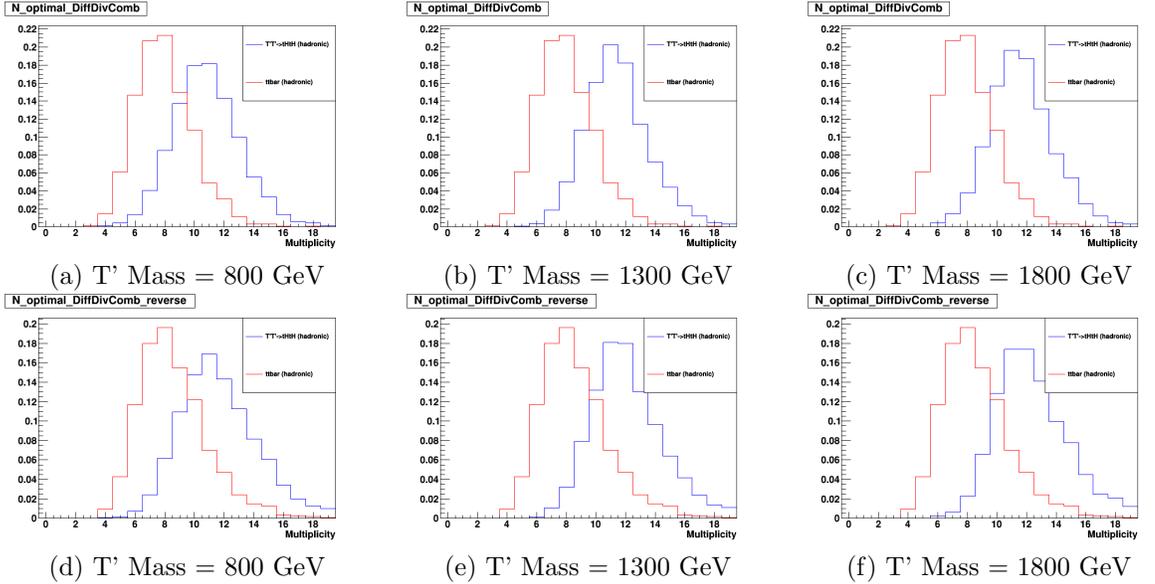


Figure 6.10: Optimal N in $t\bar{t}$ and $T'T'$ based on forward, (a)-(c), and reverse, (d)-(f), scanning the $\frac{\mathcal{T}_N - \mathcal{T}_{N-1}}{\mathcal{T}_N}$ distribution.

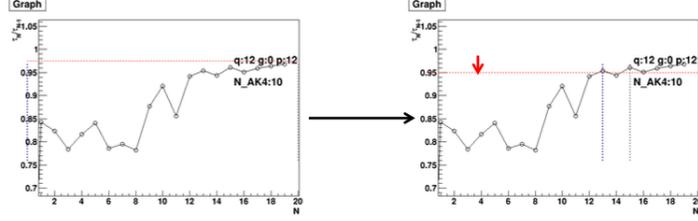


Figure 6.11: In some rare cases when the threshold (red line) is not broken (left), the threshold is lowered (right) and the vertical lines shows how forward (blue line) and reverse (black line) scanning give different result.

6.1.4 Results

Based on the optimal N distributions we plot ROC curves to quantify performance of each method to able to distinguish 6-prong $t\bar{t}$ events from 10-prong $T'T' \rightarrow tHtH$ events, shown in Figure 6.12.

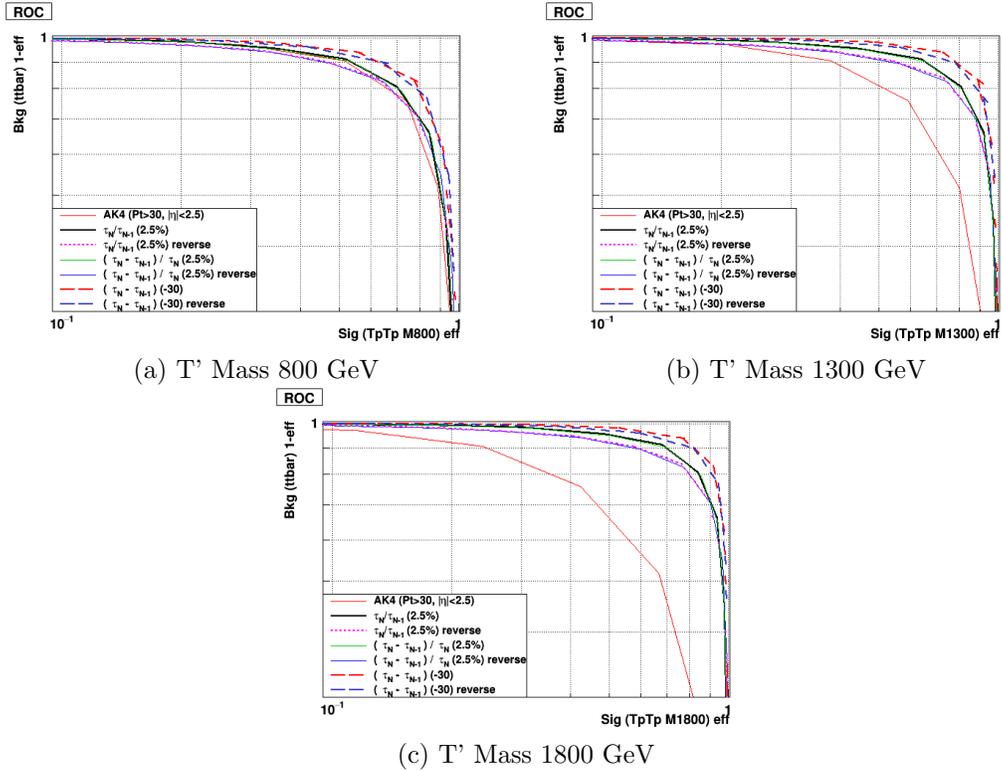


Figure 6.12: ROC curves comparing various methods to distinguish 6 prong $t\bar{t}$ events from 10-prong $T'T' \rightarrow tHtH$ events. The higher the T' mass the more collimated the jets are.

We conclude that the forward scan based on the $\tau_N - \tau_{N-1}$ with a threshold of -30 performs the best. And we observe that counting with AK4 has the least ability to distinguish $t\bar{t}$ events from

$T'T'$ events for all cases especially when the jets are more boosted. This illustrates the advantage of using N -jettiness over using AK4.

Using the resulting optimal N distribution based on forward scanning of $\mathcal{T}_N - \mathcal{T}_{N-1}$, we proceed to cluster N XCone jets and apply the same selections as applied to AK4. Figure 6.13 shows the p_T distribution of XCone jets in comparison to AK4 and Figure 6.14 shows XCone multiplicities in $t\bar{t}$ and $T'T'$ events. We plot ROC curve comparing jet counting using XCone with AK4, shown in Figure 6.15. We observe that jet counting using XCone performs better in distinguishing 6-prong $t\bar{t}$ from 10-prong $T'T'$ events, especially when the jets are more boosted.

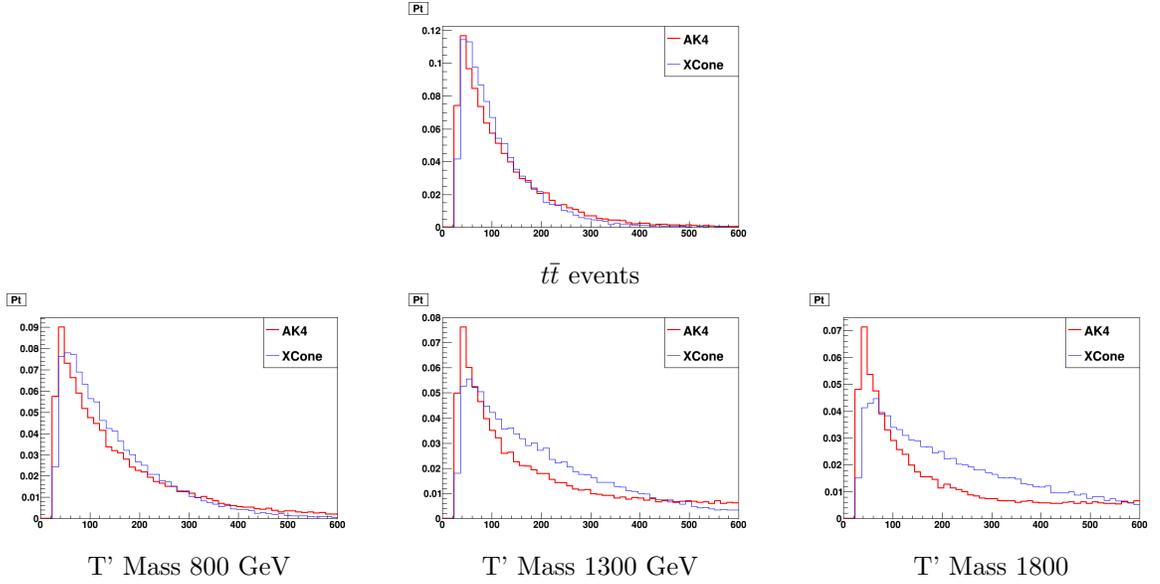


Figure 6.13: Jet p_T distribution (GeV) of AK4 and XCone jets in various samples.

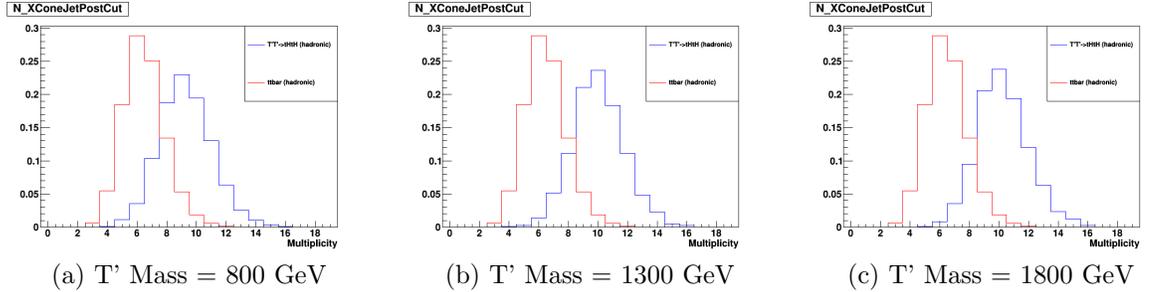


Figure 6.14: XCone multiplicities in $t\bar{t}$ and $T'T'$.

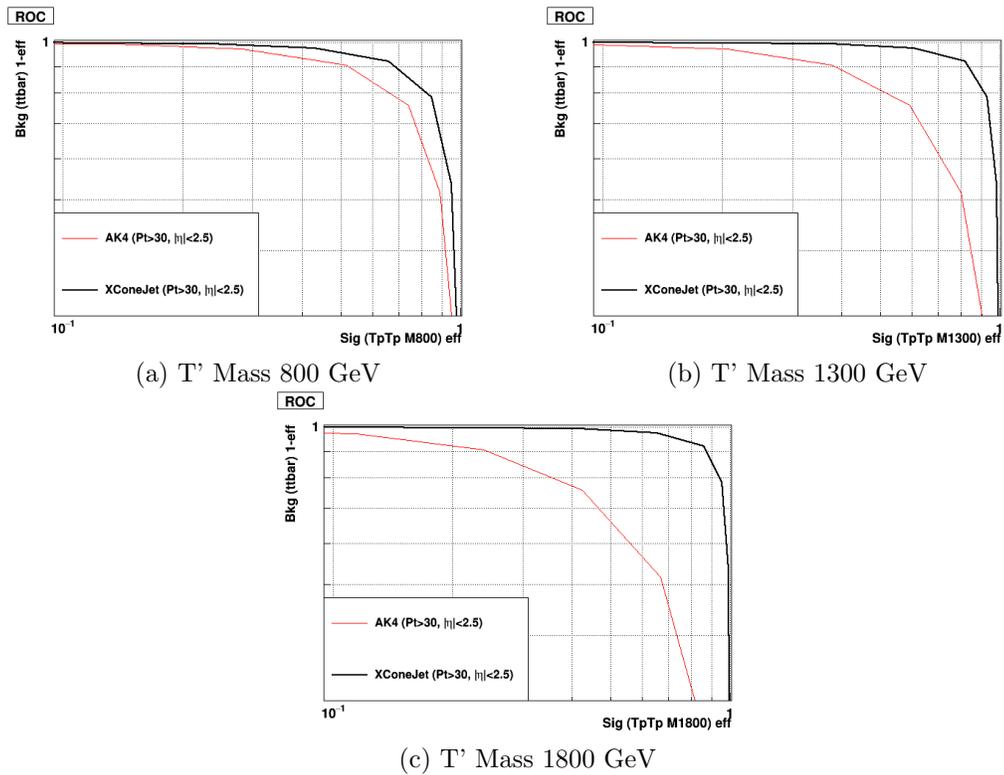


Figure 6.15: ROC curves comparing the performance of X Cone and AK4 to distinguish 6-prong $t\bar{t}$ events from 10-prong $T'T' \rightarrow tHtH$ events. The higher the T' mass the more collimated the jets are.

6.1.5 Summary: The XCone jet algorithm

We have performed a study that explores N -Jettiness and XCone to count the number of jets in an event. We used fully hadronic decays of MC $t\bar{t}$ (6-jet) and $T'T' \rightarrow tHtH$ (10-jet) events to compare the performance between using AK4 jets and using N -Jettiness. We also explore several ways to utilize N -Jettiness and we found that $\mathcal{T}_N - \mathcal{T}_{N-1}$ performs best. In comparison to using AK4 jets, N -Jettiness allows us to better distinguish $t\bar{t}$ events from $T\bar{T}$ events for the resolved case and even more so for the the boosted case.

6.2 Validation using 2016 CMS data

In this section we look at how N -jettiness and related variables in data compare with simulation, using Drell-Yan events and $t\bar{t}$ events.

The basic event and physics objects selections are identical to the ones described in Sec 5.3 of Chapter 5.

In addition, large radius jets are reconstructed with anti-kT distance parameter 0.8 (AK8 PUPPI [125]) and stored in the miniAOD format [104] if they have $p_T > 200$ GeV. For the mass, we use "Soft Drop" mass [126].

6.2.1 N -jettiness in Drell-Yan events

For Drell-Yan events we use the full 35.6fb^{-1} 2016 CMS *DoubleMuon* dataset and DYJetsToLL_M-50_TuneCUETP8M1.13TeV-madgraphMLM-pythia8 MC sample.

We require the events to satisfy the following selections criteria:

- Pass dimuon trigger: HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_(DZ)_v
- Require 2 muons that have opposite signs
- All muons pass the tight ID and isolation < 0.1
- Muon $p_T > 30$ GeV and $|\eta| < 2.4$ GeV

- $80 \text{ GeV} < M_{\mu\mu} < 100 \text{ GeV}$
- At least 1 AK4 Jet with $p_T > 30 \text{ GeV}$ and $|\eta| < 2.4 \text{ GeV}$

We first reconstruct the Z boson mass by measuring the invariant mass of the opposite sign muon pairs, shown in Figure 6.16 where we normalize MC yield to yield in Data. Then we plot various \mathcal{T}_N and $\mathcal{T}_N - \mathcal{T}_{N-1}$ distributions for data and MC shown in Figure 6.17 and Figure 6.18 respectively. We also plot jet multiplicity distributions and kinematics distributions comparing AK4 and X Cone jets in Figures 6.19 and 6.20. We only consider X Cone that pass the same p_T and η selection criteria as applied AK4.

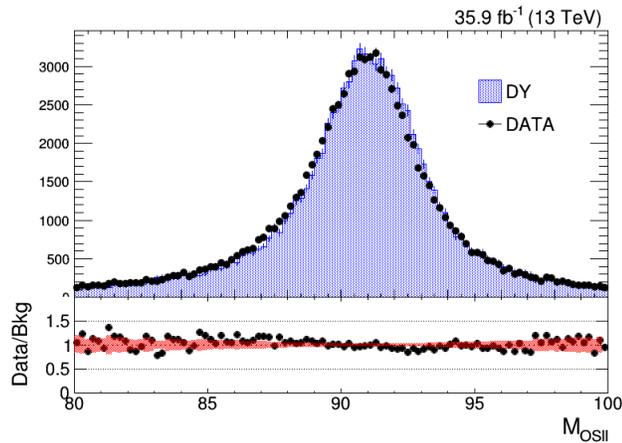


Figure 6.16: $M_{\mu\mu}$ distribution.

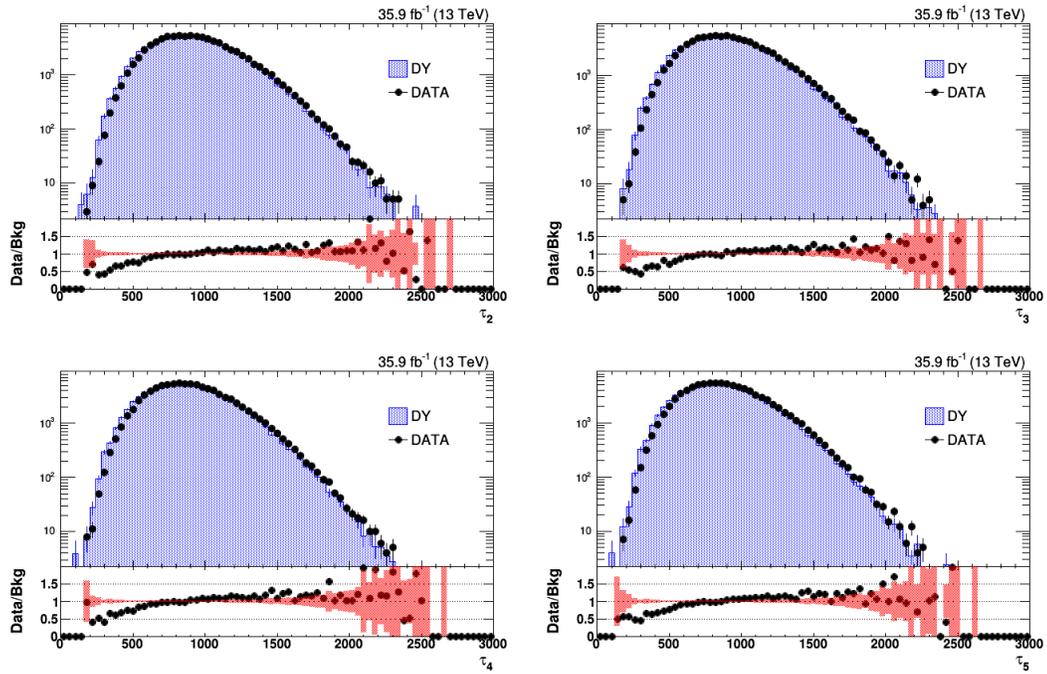
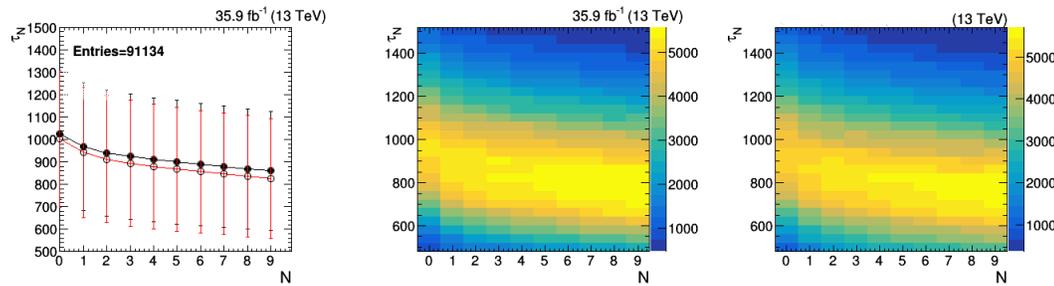
(a) \mathcal{T}_N distributions(b) \mathcal{T}_N plot for N from 0 to 9

Figure 6.17: Various \mathcal{T}_N plots and distributions for Drell-Yan events. The left most plot in (b) shows the mean of the distribution and the vertical bands are the standard deviation of \mathcal{T}_N distribution for a particular N

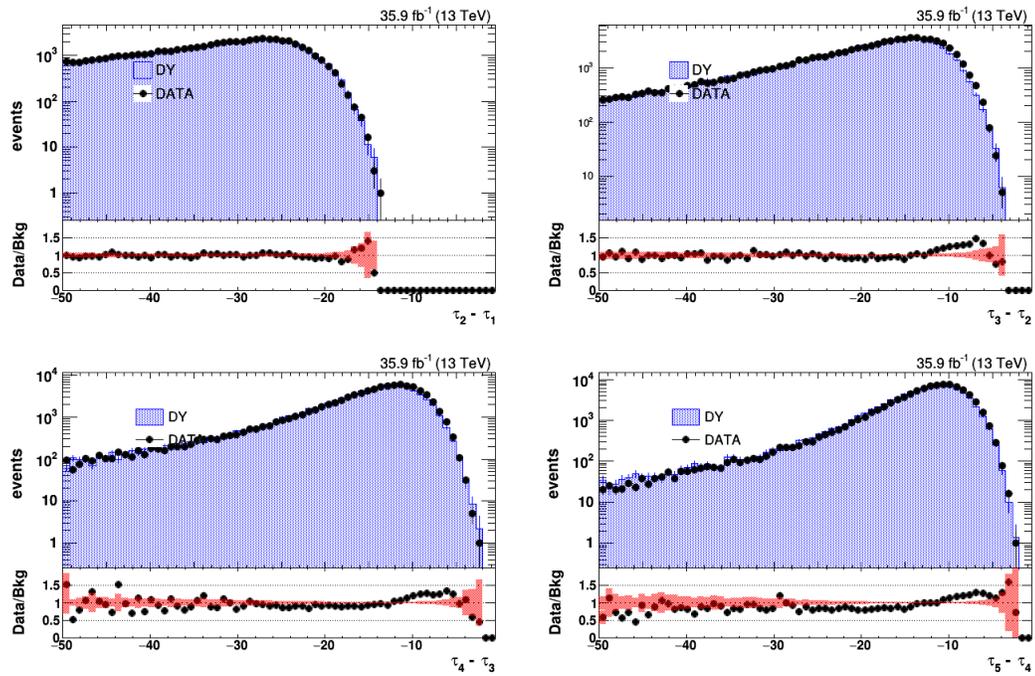
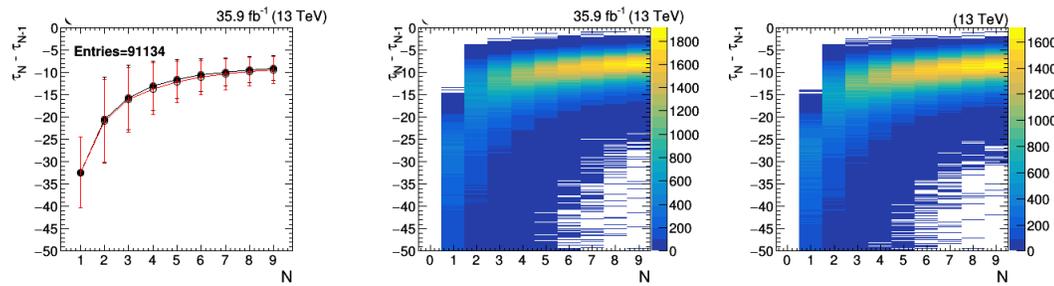
(a) $\tau_N - \tau_{N-1}$ distributions(b) $\tau_N - \tau_{N-1}$ plot for N from 0 to 9

Figure 6.18: Various $\tau_N - \tau_{N-1}$ plots and distributions for Drell-Yan events. The left most plot in (b) shows the mean of the distribution and the vertical bands are the standard deviation of $\tau_N - \tau_{N-1}$ distribution for a particular N

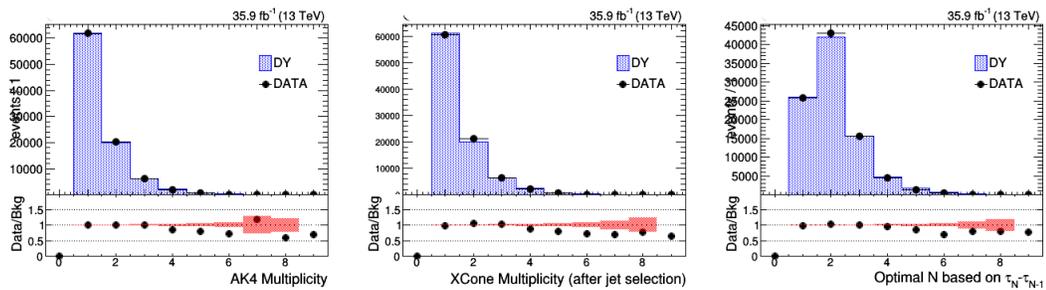


Figure 6.19: Jet multiplicity distribution.

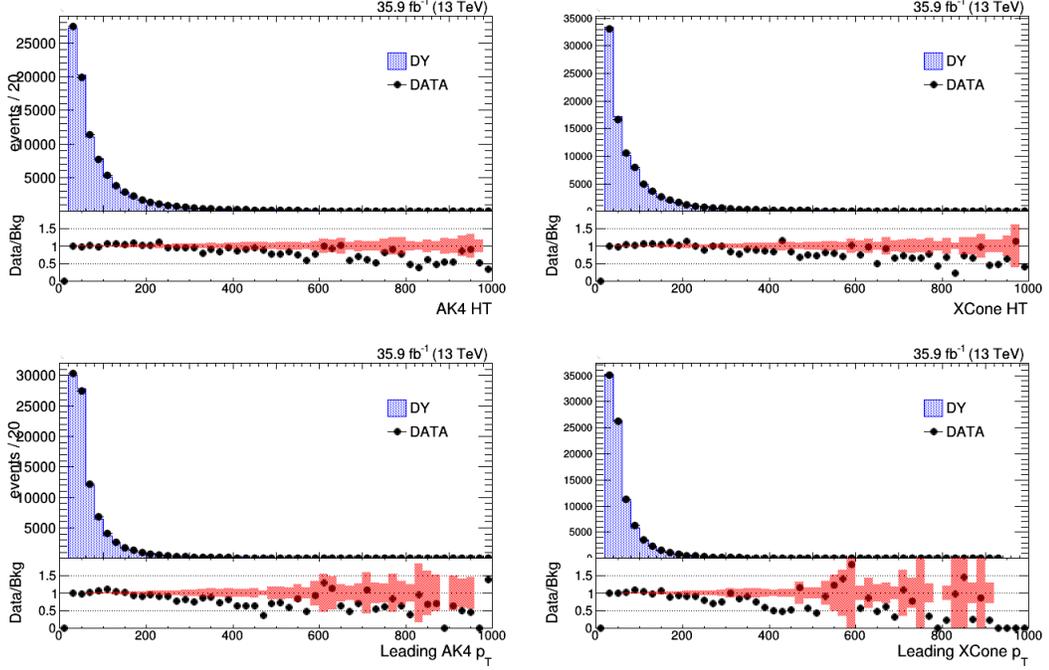


Figure 6.20: Kinematics distribution.

6.2.2 N -jettiness in semi leptonic $t\bar{t}$

For $t\bar{t}$ events we use full $35.6fb^{-1}$ 2016 CMS *SingleMuon* dataset and MC listed in Table 6.2.

Table 6.2: MC samples

Monte Carlo (MC) Dataset
TT_TuneCUETP8M2T4_13TeV-powheg-pythia8
ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M2T4
ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M2T4
DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8
WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8

We require the events to satisfy the following selection requirements:

- Pass single muon trigger: HLT_Mu50 OR HLT_TrkMu50
- Require exactly 1 muon in the event that pass the tight ID and isolation < 0.1
- Muon $p_T > 30$ GeV and $|\eta| < 2.4$ GeV
- At least 2 AK4 Jets with $p_T > 30$ GeV and $|\eta| < 2.4$ GeV

- At least 1 AK8 PUPPI Jet with $p_T > 200$ GeV and $|\eta| < 2.4$ GeV and $\Delta R(\mu, \text{jet}) > 1.0$
- AK8 PUPPI Soft Drop mass > 100 GeV
- At least 1 AK4 Jet $\Delta R(\mu, \text{AK4}) > 1.0$
- $\Delta R(\mu, \text{leading } p_T \text{ AK4}) > 1.0$
- at least 1 b-tagged AK4 jet.
- MET > 40 GeV
- $p_T(\mu + \text{MET}) > 200$ GeV

We plot the AK8 PUPPI Soft Drop mass and the invariant mass of the top three leading p_T $R=0.4$ jets where each jet satisfies $\Delta R(\mu, \text{jet}) > 0.1$. Figures 6.21, 6.24 show the plots of the jet masses with additional requirement on the p_T of the AK8 jet, from AK8 $p_T > 200$ GeV to AK8 $p_T > 800$ GeV in 200 GeV increments. As we increase the AK8 p_T requirement, we observe that X Cone is able to consistently reconstruct the top quark mass better than AK4. Whereas in the low AK8 p_T region, where AK8 does not capture all decays of the top quark, X Cone is able to reconstruct the top mass as well as AK4.

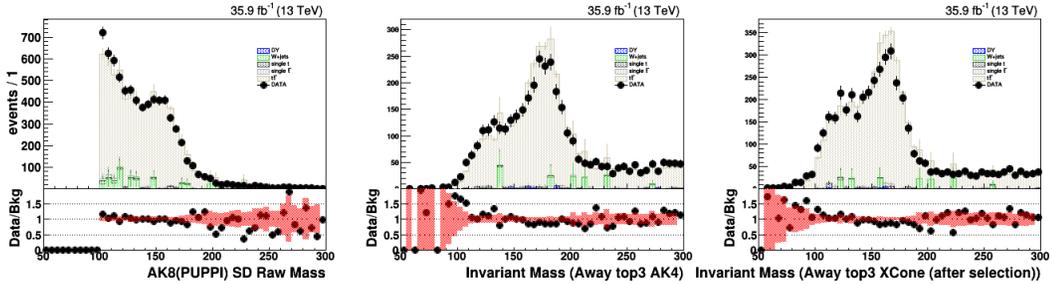


Figure 6.21: Jet mass distributions distribution with AK8 $p_T > 200$ GeV.

We plot various \mathcal{T}_N and $\mathcal{T}_N - \mathcal{T}_{N-1}$ distributions for data and MC shown in Figure 6.25 and Figure 6.26 respectively. We also plot jet multiplicity distributions and kinematics distributions comparing AK4 and X Cone jets in Figures 6.27 and 6.28 respectively. As before, we only consider X Cone that pass the same p_T and η selection criteria as applied AK4. For the plots mentioned in this paragraph, we require that the AK8 jet p_T is greater than 400 GeV.

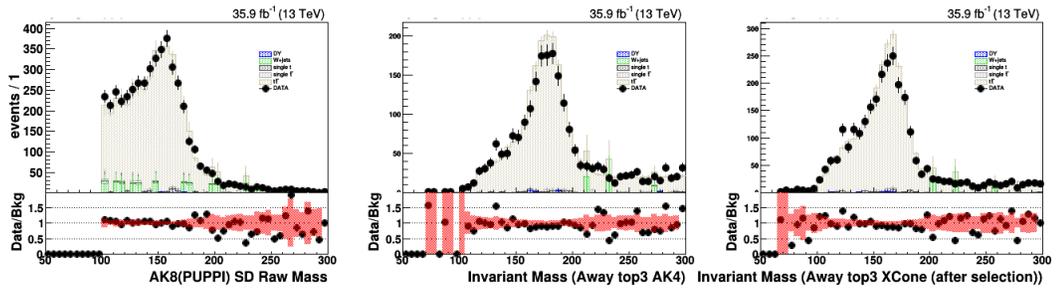


Figure 6.22: Jet mass distributions distribution with $AK8 p_T > 400$ Gev.

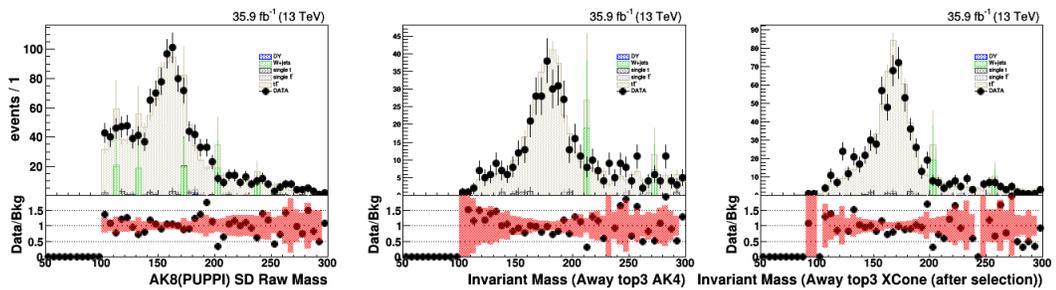


Figure 6.23: Jet mass distributions distribution with $AK8 p_T > 600$ Gev.

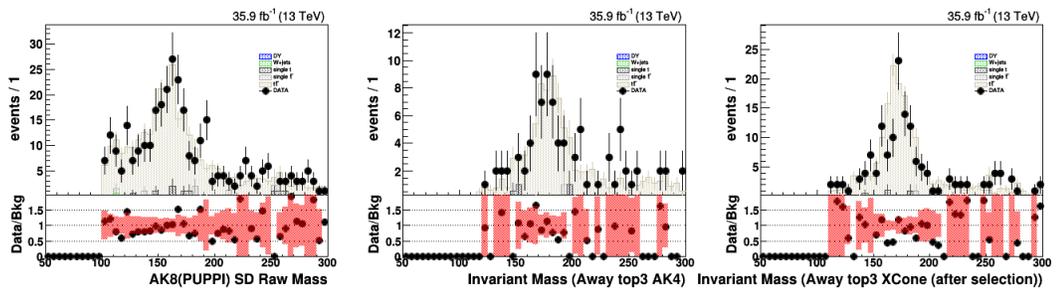


Figure 6.24: Jet mass distributions distribution with $AK8 p_T > 800$ Gev.

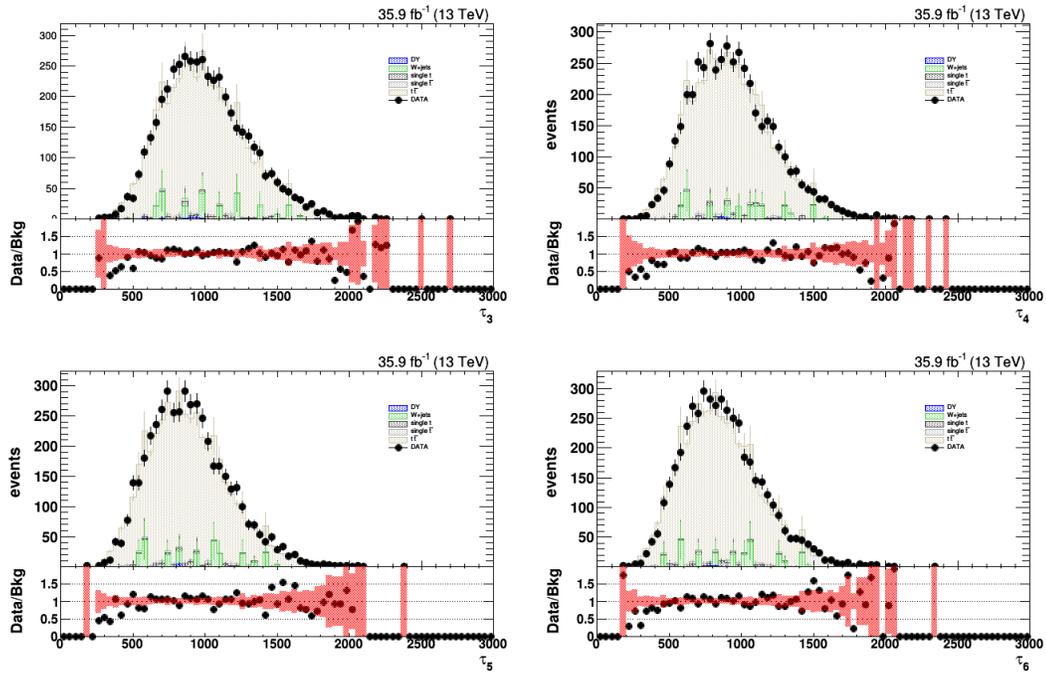
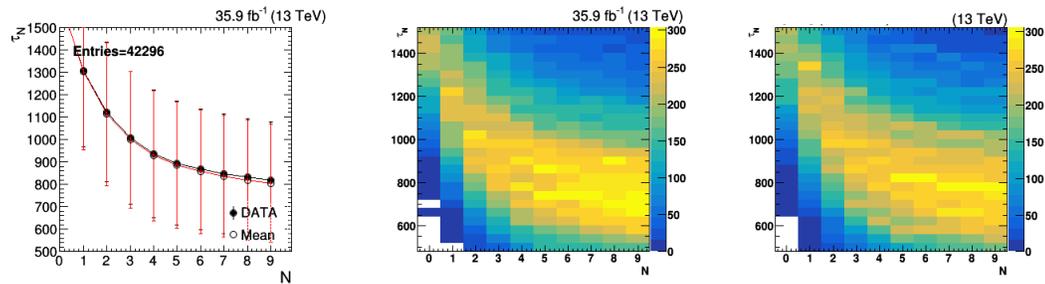
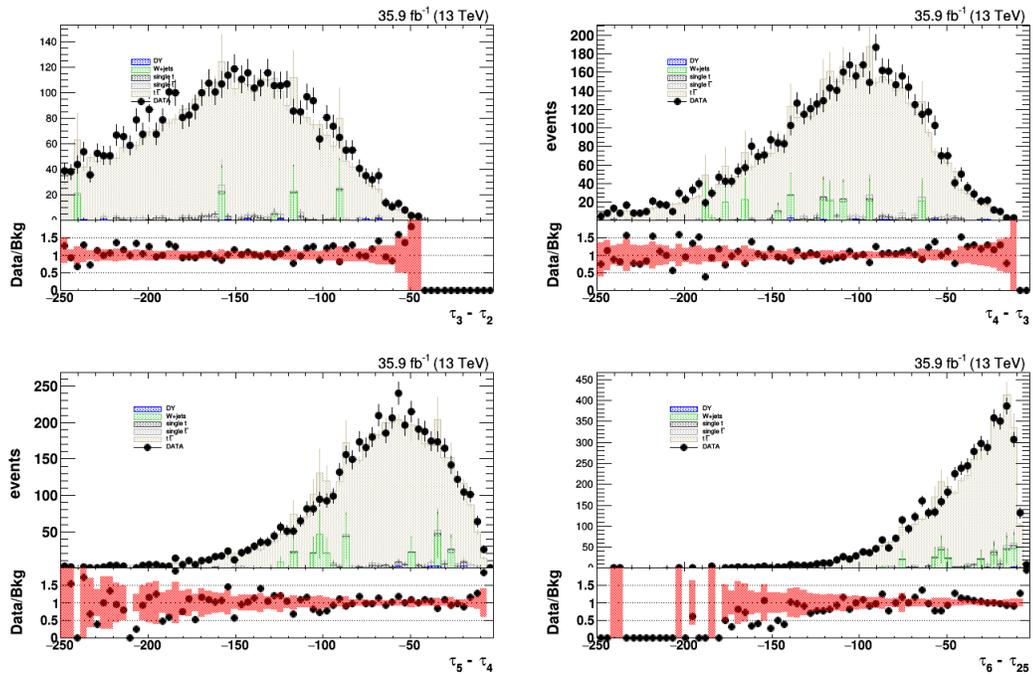
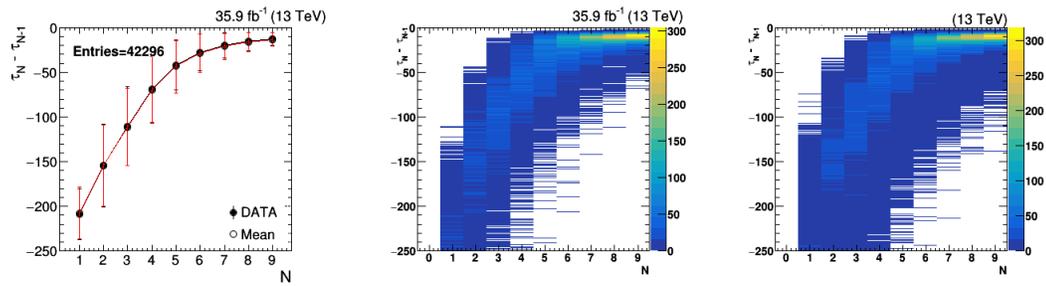
(a) \mathcal{T}_N distributions(b) \mathcal{T}_N plot for N from 0 to 9

Figure 6.25: Various \mathcal{T}_N plots and distributions for $t\bar{t}$ events. The left most plot in (b) shows the mean of the distribution and the vertical bands are the standard deviation of \mathcal{T}_N distribution for a particular N



(a) $\mathcal{T}_N - \mathcal{T}_{N-1}$ distributions



(b) $\mathcal{T}_N - \mathcal{T}_{N-1}$ plot for N from 0 to 9

Figure 6.26: Various $\mathcal{T}_N - \mathcal{T}_{N-1}$ plots and distributions for $t\bar{t}$ events. The left most plot in (b) shows the mean of the distribution and the vertical bands are the standard deviation of $\mathcal{T}_N - \mathcal{T}_{N-1}$ distribution for a particular N

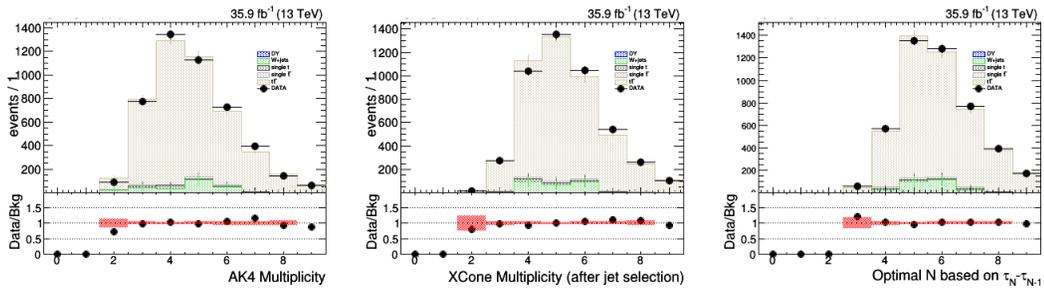


Figure 6.27: Jet multiplicity distribution.

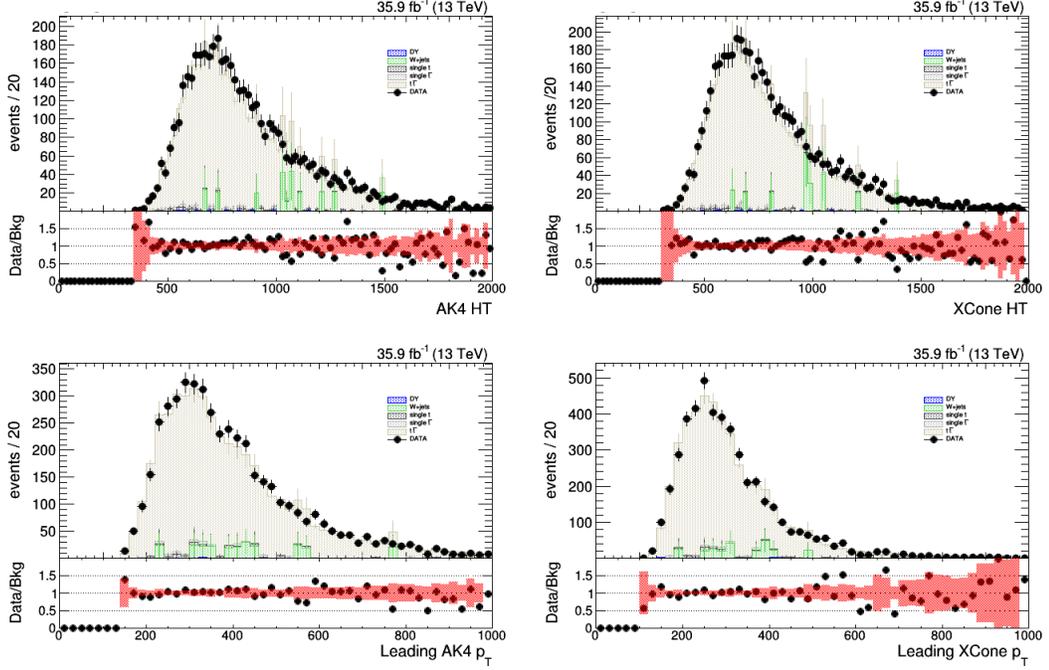


Figure 6.28: Kinematics distribution.

6.2.3 Public results for the BOOST 2017 conference

Based on the studies in the previous sections, the following figures were made into official CMS public results [127] intended for the BOOST 2017 conference in Buffalo, New York.

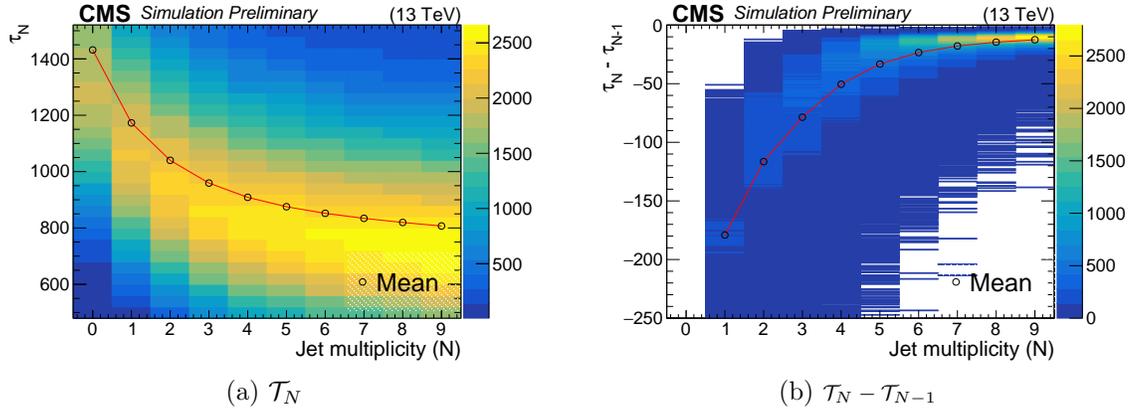


Figure 6.29: Distributions of \mathcal{T}_N and $\mathcal{T}_N - \mathcal{T}_{N-1}$, ($\Delta\mathcal{T}_N$), in semileptonic standard model $t\bar{t}$ MC plotted over a range of jet multiplicity for N from 0 to 9, and where the y-axis mean value is also shown.

Figure 6.29 shows semileptonic $t\bar{t}$ events with exactly one muon in the event and calculate the \mathcal{T}_N and $\mathcal{T}_N - \mathcal{T}_{N-1}$ values in order to determine the number of jets, given that we would want

to cluster X Cone jets with a cone measure of $R=0.4$. All plots shown in Figures 6.29-6.32 use 2016 13 TeV Monte-Carlo (MC) standard model $t\bar{t}$ samples simulated using POWHEG v2 where parton showering and the underlying event are simulated with PYTHIA 8.212 and where detector simulation is performed with GEANT4.

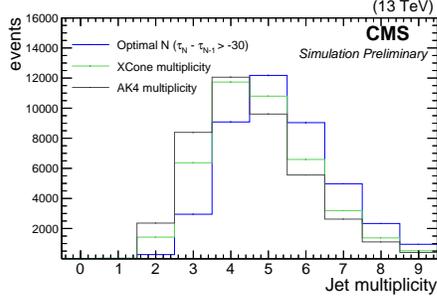


Figure 6.30: Distributions of Jet Multiplicities.

Figure 6.30 the jet multiplicity distributions of muonic semileptonic standard model $t\bar{t}$ MC events of AK4 jets, X Cone jets and the optimal jet multiplicity (N) based on the event N -jettiness. More specifically, we choose the smallest N that satisfy a threshold of $\mathcal{T}_N - \mathcal{T}_{N-1} > 30$ to select the optimal N for a given event. We then construct X Cone jets based on this optimal value. We only consider AK4 and X Cone jets with $p_T > 30$ GeV and that are central, $|\eta| < 2.5$. After jet selection X Cone multiplicity peak at 4 jets similar to AK4 multiplicity.

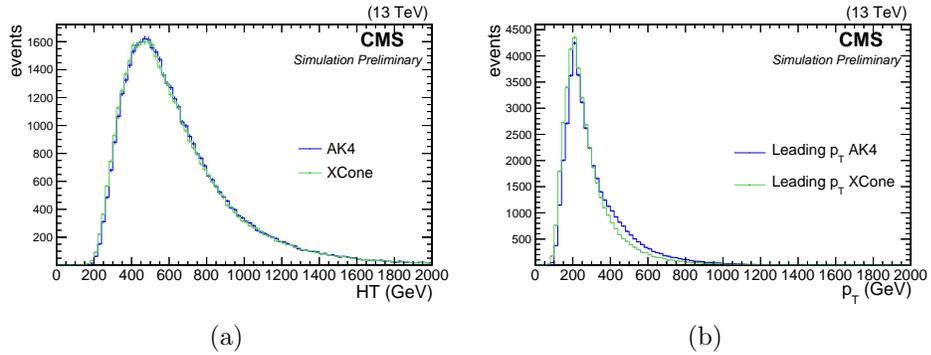


Figure 6.31: Kinematics distributions of AK4 and X Cone jets: p_T (left) and HT (right).

As sanity checks, in Figure 6.31 we plot the AK4 and X Cone p_T distributions and sum of AK4 and X Cone p_T distributions or HT distribution (right) of muonic semileptonic standard model $t\bar{t}$ MC events.

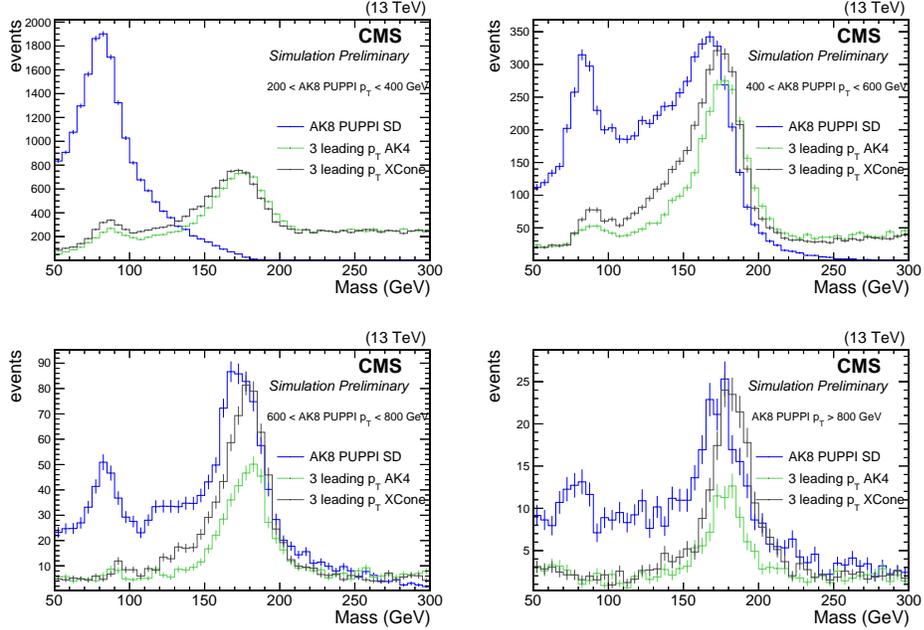


Figure 6.32: Comparisons of top mass reconstruction using AK8, AK4 and XCone jets.

Figure 6.32 shows top quark mass reconstruction in muonic semileptonic $t\bar{t}$ events, using AK8 PUPPI SoftDrop mass, invariant mass of the 3 leading p_T AK4 jets that are away from the muon, and invariant mass of the 3 leading p_T XCone jets that are away from the muon. In the non-boosted (low AK8 p_T) region, we observe that XCone constructs the top mass as well as AK4. As we go to the boosted (high AK8 p_T) region, we observe that XCone performs better than AK4. XCone jet energy has been scaled using AK4 jet energy corrections.

6.2.4 Summary: validation in data

We have performed validation studies using full 35.9 fb^{-1} 13 TeV 2016 DoubleMuon data in Drell-Yan and using 2016 SingleMuon data in $t\bar{t}$ events. In the case of $t\bar{t}$ events, we further show that we are able to reconstruct the top quark mass with better resolution using XCone jets than using AK4 jets in the boosted regime.

6.3 Utilizing X Cone jets in the search of $T\bar{T} / B\bar{B} \rightarrow e/\mu +$ jets

We now present studies on X Cone jets explicitly in the context of the search for a pair of top/bottom-like-VLQ decaying with exactly one electron/muon and jets (single lepton). The $T\bar{T} / B\bar{B} \rightarrow$ single lepton analysis using 2016 CMS data has been published by CERN and the full details of the search strategy is documented in the following paper [3]. We briefly summarize the selections, the event categories, and the main discriminants here.

The single lepton final state includes events with exactly one charged lepton, usually from the decay of a W boson in the bW channel or from the $t \rightarrow Wb$ decay in the tZ or tH channels. The search applies the following initial (pre)selections,

- One tight electron (muon) with $p_T > 60\text{GeV}$, $|\eta| < 2.5$ (2.4)
- No loose electrons (muons) with $p_T > 10\text{GeV}$, $|\eta| < 2.5$ (2.4)
- Three or more AK8 jets with $p_T > 200, 100, 50\text{GeV}$, $|\eta| < 2.4$
- $H_T > 400$ GeV
- $p_T^{\text{miss}} > 60$ GeV

and the following optimized final selections,

- 1 tight, isolated electron or muon with $p_T > 60$ GeV
- 0 loose leptons with $p_T > 10$ GeV
- ≥ 3 AK4 jets with $p_T > 300, 150,$ and 100 GeV
- An additional AK4 jet with $p_T > 30$ GeV in events with no tagged W or Higgs jet
- At least 1 b-tagged AK4 jet in events with 1 Higgs-tagged jet
- ≥ 2 AK8 jets, where AK8 is the anti-kt jets with distance parameter $R = 0.8$

- $p_T^{\text{miss}} > 75 \text{ GeV}$
- $\min[\Delta R(\text{leading AK8, other AK8})] < 3,$

The search categorizes events of the signal region into 16 categories based on lepton flavor (e, μ), the number of b-tagged AK4 jets (1, 2, 3+), the number of boosted W tagged jets (0, 1+), and the number of Higgs-tagged jets (1 subjet b tag or 2 subjet b tags). If an event has any AK8 jet passing the Higgs tag criteria and containing 2 b-tagged subjets, the event is categorized as “H2b”. If the event has no H2b tags, but has at least one AK8 just passing the Higgs tag criteria with 1 b-tagged subjet, the event is categorized as “H1b”. All events without H1b or H2b type Higgs tags are categorized according to W and b tag content. It is possible for a jet to be both W- and Higgs-tagged, but the search categorize first by the presence of Higgs tags so that no event falls in multiple categories.

As main discriminants, the search for $T\bar{T}$ in the single lepton final state uses S_T and the minimum mass constructed from the lepton (ℓ) and a b tagged jet, labeled $\min[M(\ell, b)]$. These mass variables are specifically powerful for the bW decay of the T quark, where background peaks at the top quark mass, it uses an alternate discriminant for the Higgs-tagged categories where the lepton is expected to come primarily from top quark decays rather than direct decays of the T quark. The search uses the S_T variable in these categories, which is the the scalar sum of H_T , lepton p_T , and p_T^{miss} .

For the XCone jets in this section, we consider $|\mathcal{T}_N - \mathcal{T}_{N-1}|$ threshold 30 GeV and we further select XCone jets using the same kinematics cuts applied to the AK4 jets.

6.3.1 Exploring observables based on XCone jets

We explore observables based on invariant masses involving XCone jets. Firstly we consider the following,

- maxMlep3XCone : $\max(M[\text{lep+closestXCone, 2nd XCone, 3rd XCone}])$
- Mlep3ClosestXCone : $M(\text{lep+closestXCone, 2nd nextClosestXCone, 3rd nextnextClosestXCone})$

where $maxMlep3XCone$ is constructed by taking the maximum value of the invariant mass of the lepton, the closest X Cone jet with the lepton, and two other X Cone jets in the event, and $Mlep3ClosestXCone$ is the invariant mass of the lepton and 3 closest X Cone jets to lepton.

We compare the above observables with S_T , $\min[M(l, b)]$, AK4 jets multiplicity, and X Cone jets multiplicity in $t\bar{t}$ (being the most significant background) and $T\bar{T}$ simulation samples. Figures 6.33, 6.34, and 6.35 show the comparison of the distributions for the two cases $T\bar{T} \rightarrow bWbW$, $T\bar{T} \rightarrow tZtZ$, $T\bar{T} \rightarrow tHtH$ where the first represents the least number of jets originating from signal, and the latter represents the most number of jets.

The $maxMlep3XCone$ observable combines the charged lepton with the three highest momentum jets, where one of the jet is the closest to the lepton. The idea is that in some instances the lepton-jet system would represent a leptonic top quark candidate, and the two high momentum jets would represent a vector boson candidate. And these together would represent a T quark candidate.

The $Mlep3ClosestXCone$ observable combines the charged lepton with the three closest jets. The idea here is that the decay products of T would decay into a lepton and jets which tend to be collimated, and tend to be closer to each other.

We observe that for $T \rightarrow bWbW$ $maxMlep3XCone$ has a similar distribution to S_T , and $Mlep3ClosestXCone$ behaves more similarly to $\min[M(l, b)]$. However, in both cases, neither of the newly defined observables performed better than the baseline discriminants. For $T \rightarrow tZtZ$ and $tHtH$ the new observables do not perform well to discriminate between signal and $t\bar{t}$. The performances of these observables can also be seen from the Receiver Operator Characteristic (ROC) plots, shown in Figures 6.36, 6.37 and 6.38 where the plots are separated into the 8 categories, and where electron and muons categories are combined. We observe that $\min[M(l, b)]$ is generally superior for the $T\bar{T} \rightarrow bWbW$ case, and S_T is generally superior for both the $T\bar{T} \rightarrow tZtZ$ and $T\bar{T} \rightarrow tHtH$ cases.

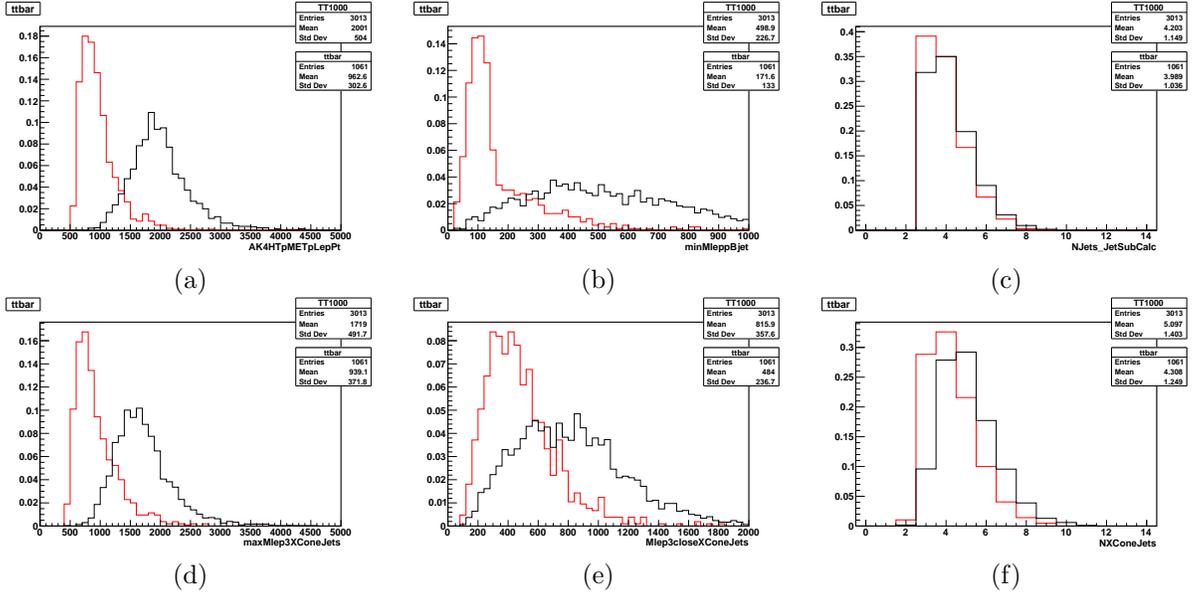


Figure 6.33: Comparison of XCones based observables in $t\bar{t}$ (in red) and $T\bar{T} \rightarrow bWbW$ (in black), where T has a mass of 1.0 GeV and preselection cuts are applied. S_T is labelled "AK4pHTpMETpLeptonPt", $\min[M(l, b)]$ is labelled "minMlepBjet", AK4 jets multiplicity is labelled "NJets_JetSubCalc", and XCones jets multiplicity is labelled "NXConeJets".

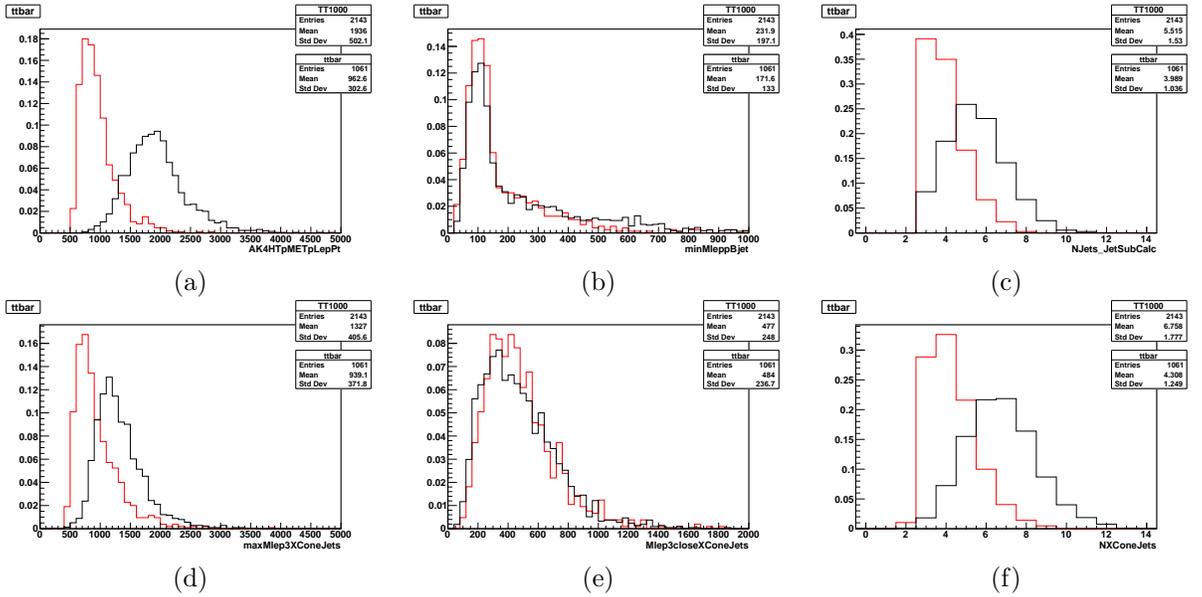


Figure 6.34: Comparison of XCones based observables in $t\bar{t}$ (in red) and $T\bar{T} \rightarrow tZt$ (in black), where T has a mass of 1.0 GeV and preselection cuts are applied. S_T is labelled "AK4pHTpMETpLeptonPt", $\min[M(l, b)]$ is labelled "minMlepBjet", AK4 jets multiplicity is labelled "NJets_JetSubCalc", and XCones jets multiplicity is labelled "NXConeJets".

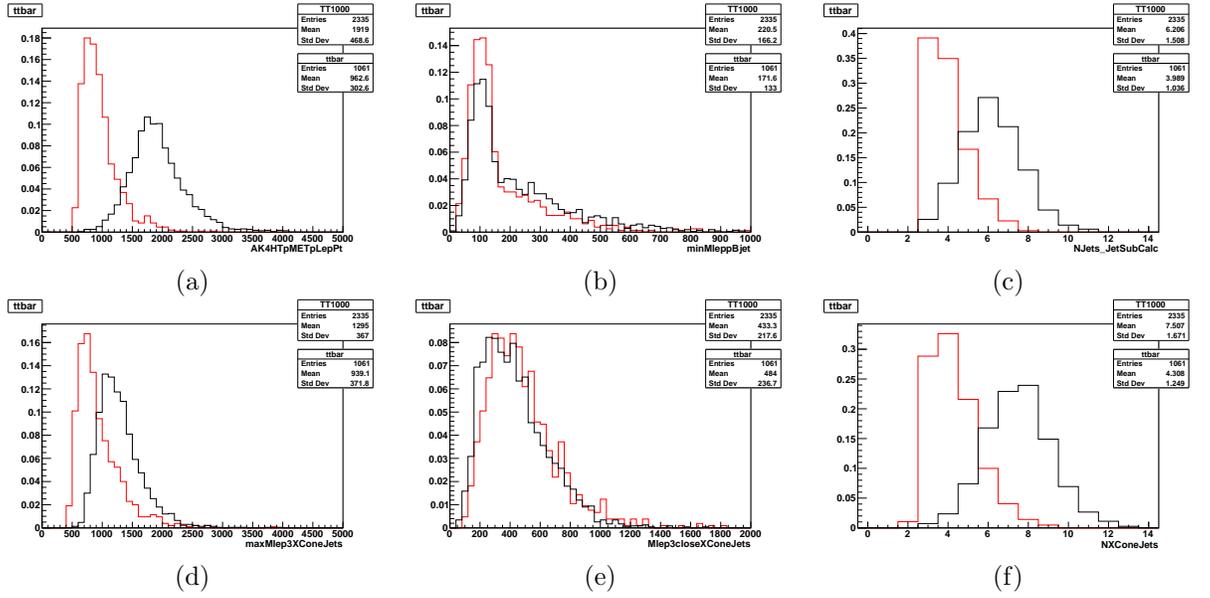
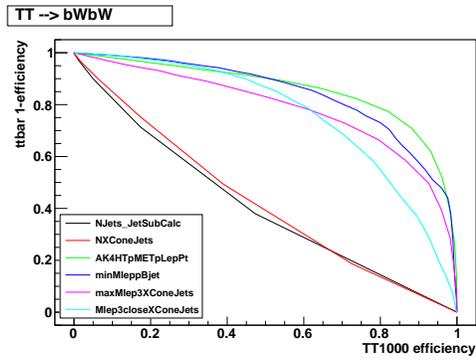
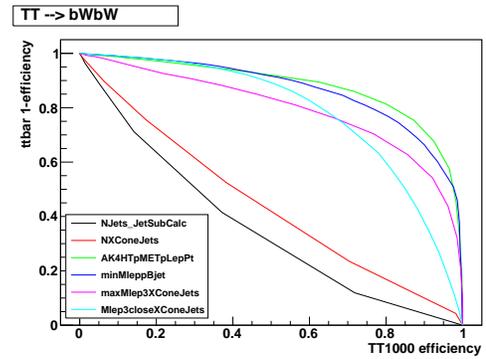


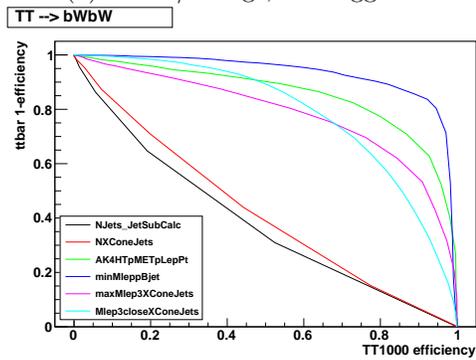
Figure 6.35: Comparison of XCone based observables in $t\bar{t}$ (in red) and $T\bar{T} \rightarrow tHtH$ (in black), where T has a mass of 1.0 GeV and preselection cuts are applied. S_T is labelled "AK4pHTpMETpLeptonPt", $\min[M(1,b)]$ is labelled "minMleppBjet", AK4 jets multiplicity is labelled "NJets_JetSubCalc", and XCone jets multiplicity is labelled "NXConeJets".



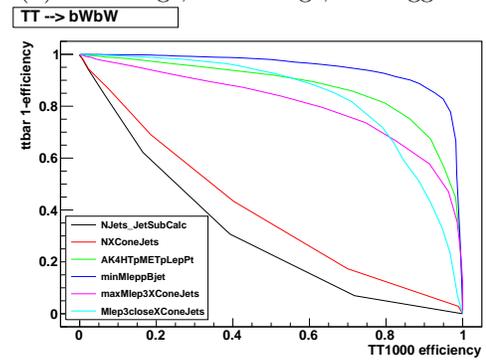
(a) No W/H-tags, 1 b-tagged AK4



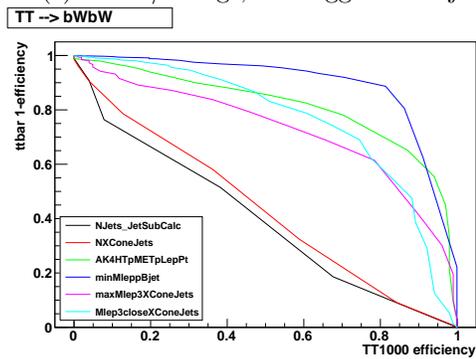
(b) No H-tags, 1+ W-tags, 1 b-tagged AK4



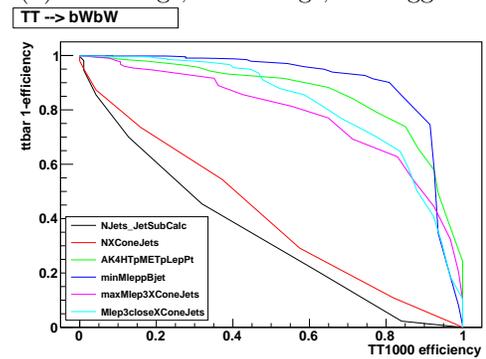
(c) No W/H-tags, 2 b-tagged AK4 jets



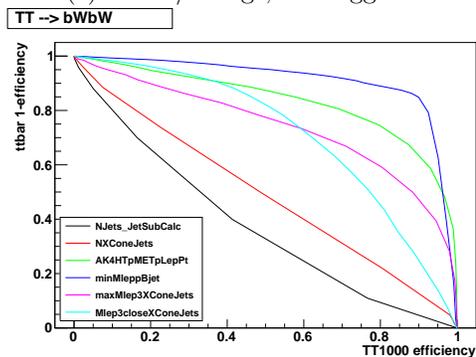
(d) No H-tags, 1+ W-tags, 2 b-tagged AK4



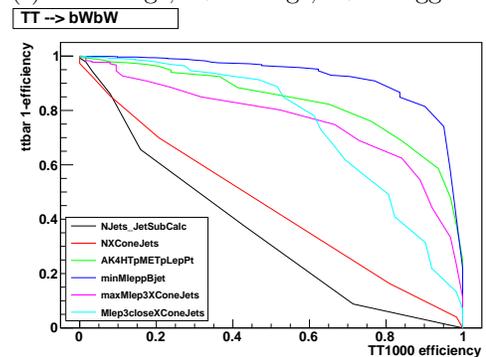
(e) No W/H-tags, 3 b-tagged AK4



(f) No H-tags, 1+ W-tags, 3+ b-tagged AK4

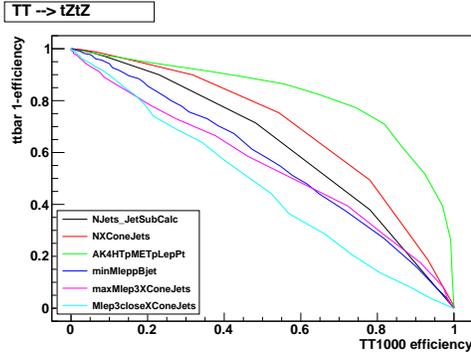


(g) 1 H-tagged AK8 jet, 1+ b-tagged AK8 subjet

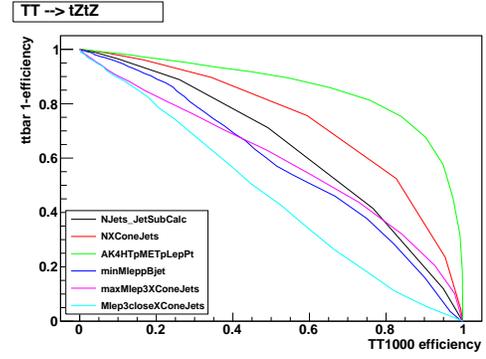


(h) 2 H-tagged AK8 jet, 1+ b-tagged AK8 subjet

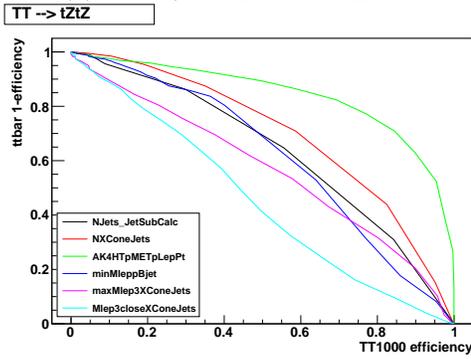
Figure 6.36: ROC curves for observables, in samples of $t\bar{t}$ and $T\bar{T} \rightarrow bWbW$, where T has a mass of 1.0 GeV. S_T is labelled "AK4pHTpMETpLeptonPt", $\min[M(l, b)]$ is labelled "minMleppBjet", AK4 jets multiplicity is labelled "NJets_JetSubCalc", and XCone jets multiplicity is labelled "NXConeJets".



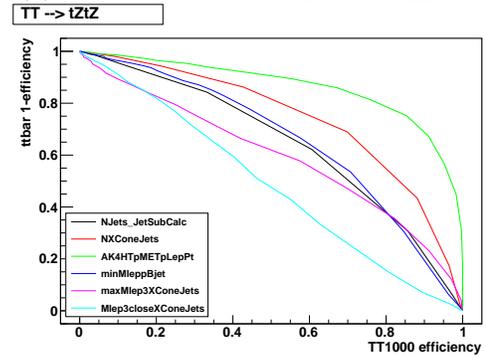
(a) No W/H-tags, 1 b-tagged AK4



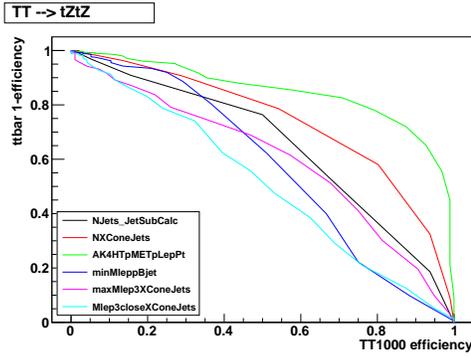
(b) No H-tags, 1+ W-tags, 1 b-tagged AK4



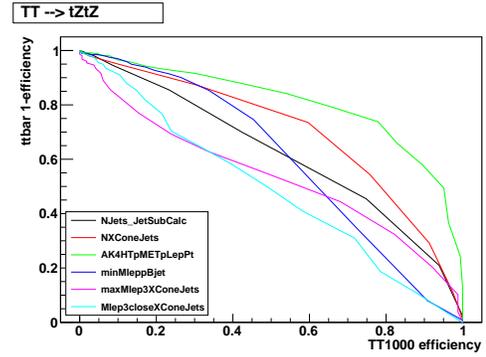
(c) No W/H-tags, 2 b-tagged AK4 jets



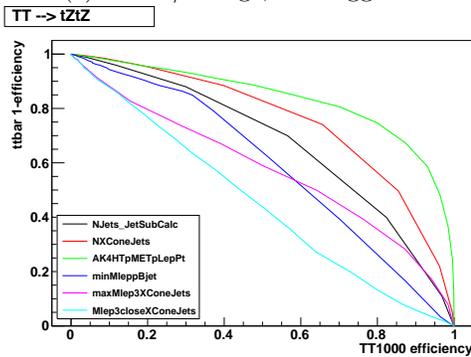
(d) No H-tags, 1+ W-tags, 2 b-tagged AK4



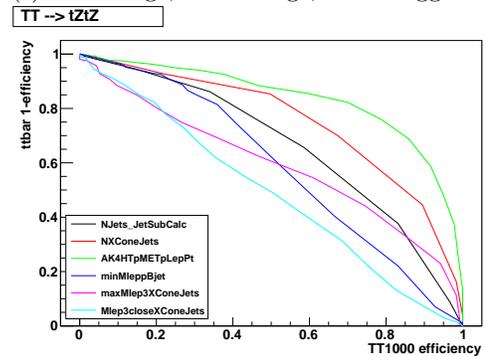
(e) No W/H-tags, 3 b-tagged AK4



(f) No H-tags, 1+ W-tags, 3+ b-tagged AK4

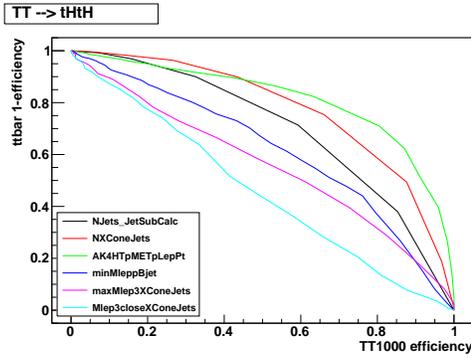


(g) 1 H-tagged AK8 jet, 1+ b-tagged AK8 subjet

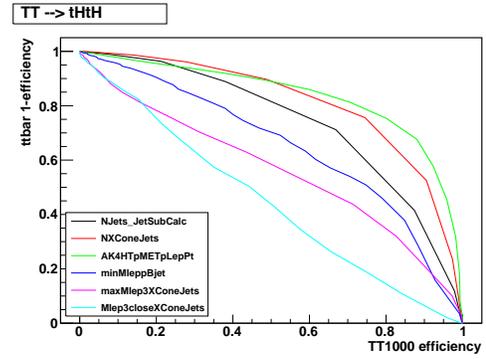


(h) 2 H-tagged AK8 jet, 1+ b-tagged AK8 subjet

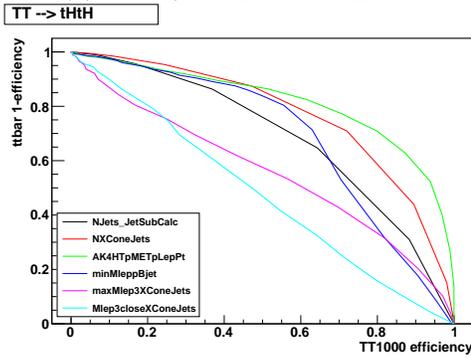
Figure 6.37: ROC curves for observables, in samples of $t\bar{t}$ and $T\bar{T} \rightarrow tZtZ$, where T has a mass of 1.0 GeV. S_T is labelled "AK4pHTpMETpLeptonPt", $\min[M(l, b)]$ is labelled "minMleppBjet", AK4 jets multiplicity is labelled "NJets_JetSubCalc", and Xcone jets multiplicity is labelled "NXConeJets".



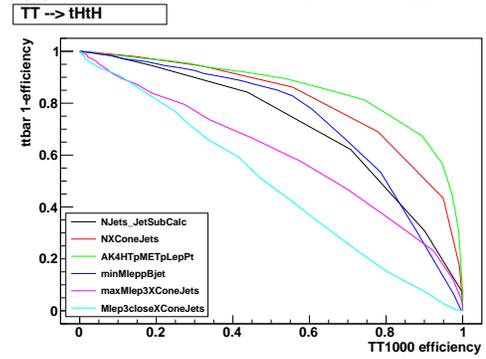
(a) No W/H-tags, 1 b-tagged AK4



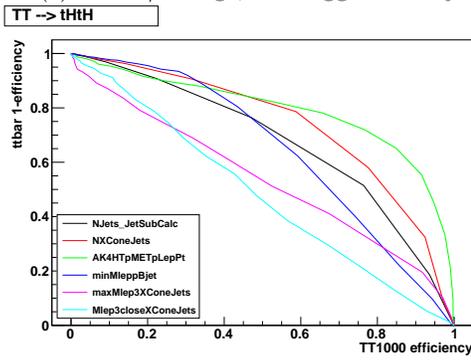
(b) No H-tags, 1+ W-tags, 1 b-tagged AK4



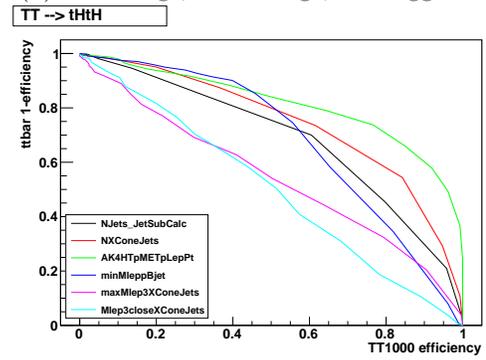
(c) No W/H-tags, 2 b-tagged AK4 jets



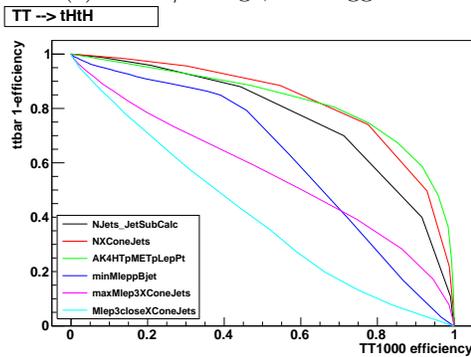
(d) No H-tags, 1+ W-tags, 2 b-tagged AK4



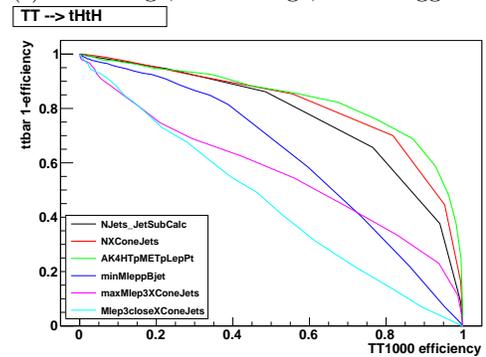
(e) No W/H-tags, 3 b-tagged AK4



(f) No H-tags, 1+ W-tags, 3+ b-tagged AK4



(g) 1 H-tagged AK8 jet, 1+ b-tagged AK8 subjet



(h) 2 H-tagged AK8 jet, 1+ b-tagged AK8 subjet

Figure 6.38: ROC curves for observables, in samples of $t\bar{t}$ and $T\bar{T} \rightarrow tHtH$, where T has a mass of 1.0 GeV. S_T is labelled "AK4pHTpMETpLeptonPt", $\min[M(l, b)]$ is labelled "minMleppBjet", AK4 jets multiplicity is labelled "NJets_JetSubCalc", and X Cone jets multiplicity is labelled "NXConeJets".

To further study these performances, we calculate the 95% CL production cross section upper limits² of $T\bar{T} / B\bar{B}$ for 5 signal decay scenarios: singlet, doublet, $\text{BR}(T \rightarrow bH)=100\%$, $\text{BR}(T \rightarrow tZ)=100\%$, $\text{BR}(T \rightarrow tH)=100\%$, using the following discriminants:

- S_T : use S_T for all categories,
- $\text{minMlbNXConeJets}S_T$: use XCone jet multiplicity for ≥ 3 b-tagged AK4 jets categories,
- minMlbNXConeJets : use XCone jet multiplicity for ≥ 3 b-tagged AK4 jets and H-tagged categories,
- $\text{minMlbNXConeJetsV2}$: , use XCone jet multiplicity for H-tagged categories,
- $\text{maxMlep3XCone}S_T$: use maxMlep3XCone for ≥ 3 b-tagged AK4 jets categories,
- maxMlep3XCone : use maxMlep3XCone for all categories.

Unless specified, the categories use the same discriminants as the baseline strategy. The results, shown in Table 6.3, shows that none of the newly proposed strategies are better than the baseline strategy, and we conclude that the newly defined observables do not performance better than the baseline observables.

Table 6.3: Comparison of VLQ T mass upper limits (in GeV) using various different discriminators. The limits are based on the 95% CL upper limits on the production cross section of $T\bar{T}$.

	singlet	doublet	tH	bW	tZ
minMlbST (baseline)	1086	1096	1184	1138	954
S_T	1082	1109	1191	1094	989
$\text{minMlbNXConeJets}S_T$	1078	1089	1185	1128	950
minMlbNXConeJets	1021	1053	1140	1121	917
$\text{minMlbNXConeJetsV2}$	1035	1057	1142	1122	918
$\text{maxMlep3XConeJets}S_T$	1029	1090	1183	996	952
maxMlep3XConeJets	969	992	1073	991	882

²Jet energy correction uncertainties were not included in these upper limits calculation.

Next, we construct observables with the idea that they would distinguish the T quark decaying to bW from background bW 's. Firstly, we perform "pseudo-b-tagging" of X Cone jets. An X Cone is pseudo-b-tagged if it is matched within $\Delta R < 0.01$ to a b-tagged AK4 jet. Additionally, we consider the two W boson candidates from the two possible values of the longitudinal momentum (p_z) of the neutrino candidates as constrained by the kinematics of the lepton, p_T^{miss} (representing p_T of the neutrino candidate) and the W boson mass,

$$|P_W|^2 = |P_{e/\mu} + P_\nu|^2 \quad (6.1)$$

$$M_W^2 = M_{e/\mu/\tau}^2 + M_\nu^2 + 2 \cdot (E_{e/\mu} E_\nu - \vec{p}_{e/\mu} \cdot \vec{p}_\nu) \quad ,$$

where P is 4-momentum and M is invariant mass, $E = \sqrt{|\vec{p}|^2 + M^2}$ is energy, and $\vec{p} = (p_T, p_z)$ is 3-momentum. As can be seen in the (quadratic) Eq 6.1, the only unknown is the p_z of the neutrino and thus the equation presents two solutions of it.

Based on pseudo-b-tagged X Cone jets and the W boson candidates we consider the invariant mass of the bW candidates:

$$\min [(M_1 + M_A), (M_2 + M_B)] = M_WcandBTtagXCone_XCone123, \quad (6.2)$$

where

$$M_1 = MassWcand1BtagXCone,$$

$$M_2 = MassWcand2BtagXCone,$$

$$M_A = Mass1XCone123,$$

$$M_B = Mass2XCone123, \quad (6.3)$$

and where M_1 is the invariant mass of the 1st W boson candidate and the closest pseudo-b-tagged X Cone, M_2 is the invariant mass of the 2nd W boson candidate and the closest pseudo-b-tagged X Cone, M_A is the invariant mass of M_1 and the 3 highest p_T X Cones excluding the corresponding pseudo-b-tagged X Cone, M_B is the invariant mass of M_2 and the 3 highest p_T X Cones excluding the corresponding pseudo-b-tagged X Cone. Furthermore, the 3 X Cone jets for M_A , M_B are permuted such that they satisfy being the minimum value requirements: $\min(M_1 - M_A)$, $\min(M_2 - M_B)$, respectively.

Figure 6.39 shows the distribution of $MassWcand1BtagXCone$, $MassWcand2BtagXCone$, $Mass1XCone123$, $Mass2XCone123$ in $t\bar{t}$ and $T\bar{T} \rightarrow bWbW$ samples, where the T has a mass of 1.0 GeV. Figure 6.40 shows the $M_WcandBTagXCone_XCone123$ distribution for $t\bar{t}$ and $T\bar{T} \rightarrow bWbW$, for T with masses 1.0, 1.4, and 1.8 GeV. The figures shows peaks approximately at the combined masses of $T\bar{T}$ (labelled by the black lines).

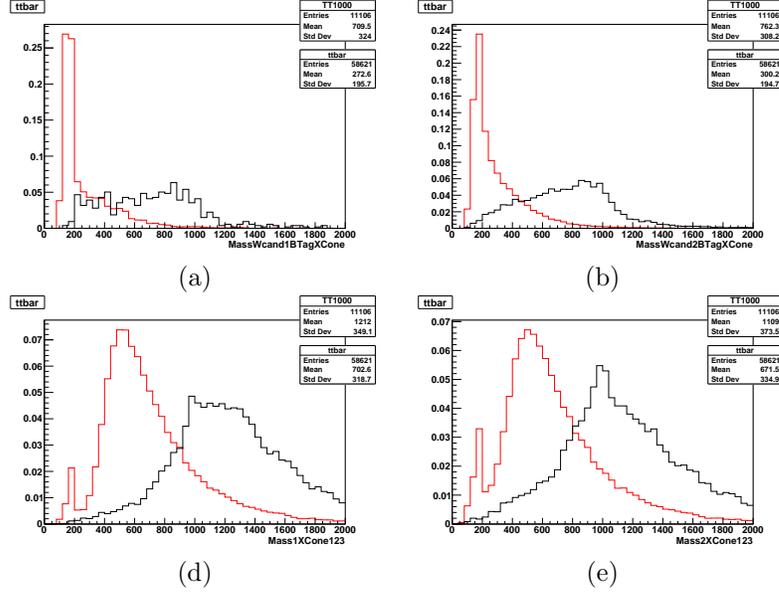


Figure 6.39: Distributions of various observables based on pseudo b-tagged X Cone jets, comparing $t\bar{t}$ (in red) and $T\bar{T} \rightarrow bWbW$ (in black), where T has a mass of 1.0 GeV, and only the preselection cuts are applied. (a) $MassWcand1BtagXCone$, (b) $MassWcand2BtagXCone$, (c) $Mass1XCone123$, (d) $Mass2XCone123$

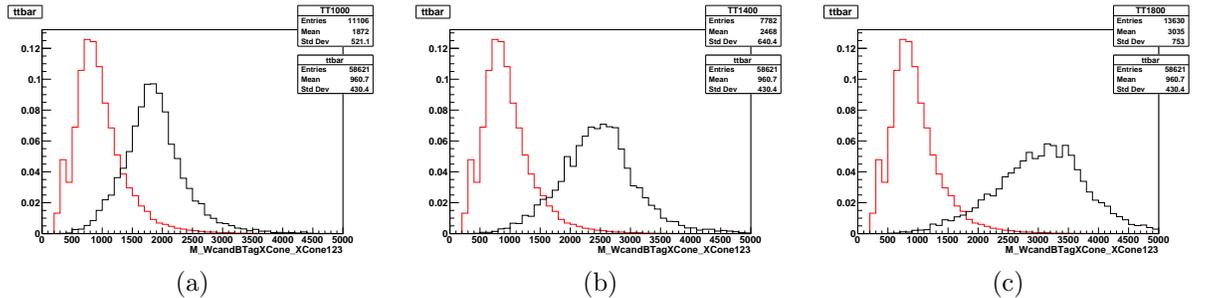


Figure 6.40: Distributions of the $M_WcandBTagXCone_XCone123$ observable, comparing $t\bar{t}$ (in red) and $T\bar{T} \rightarrow bWbW$ (in black), where T has a mass of (a) 1.0 GeV, (b) 1.4 GeV, (c) 1.8 GeV. Only the preselection cuts are applied.

Despite the $M_WcandBTagXCone_XCone123$ distribution showing clear discrimination between $t\bar{t}$ and $T\bar{T}$, its performance is generally not any better than the baseline observable, as shown

in Figures 6.41, 6.42, 6.43. The ROC curve plots suggests that $M_{-W} \text{ cand } B \text{ Tag } X \text{ Cone } X \text{ Cone } 123$ becomes better than $\min[M(1, b)]$ at T mass of 1.4 TeV, but not better than S_T for T with masses up to 1.8 TeV.

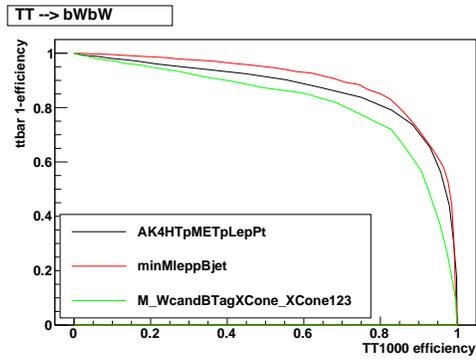
6.3.2 Optimizing selections using XCone jets multiplicity

We now look towards optimizing limit sensitivity by varying the XCone multiplicity requirement. We calculate the expected limits, using the baseline strategy, with requiring additionally several XCone jets in the event, and we consider 5 signal decay scenarios: singlet, doublet, $\text{BR}(T/B \rightarrow \text{bH}/\text{tH})=100\%$, $\text{BR}(T/B \rightarrow \text{tZ}/\text{bH})=100\%$, $\text{BR}(T/B \rightarrow \text{tH}/\text{bH})=100\%$. For this case we only apply the following uncertainties: luminosity, lepton identification, lepton isolation, lepton reconstruction, and the QCD renormalization scale uncertainties (see Table 5.26). Additionally we also consider the case where S_T is used as the sole discriminant.

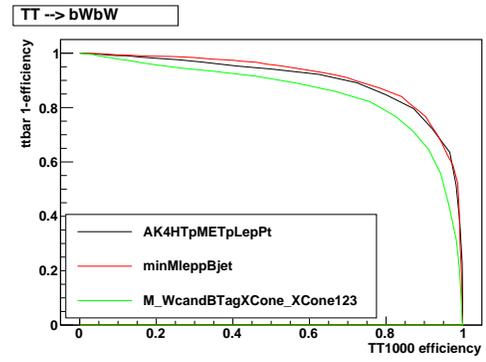
Specifically, we calculate the 95% CL expected upper limits as a function of the number XCone multiplicities and determine the multiplicity that produces the most sensitive limits for T/B masses from 0.8 to 1.8 GeV. Tables 6.4 and 6.5 show the baseline upper limits $T\bar{T}$ and $B\bar{B}$ production cross section (without any XCone multiplicity selection). The resulting optimal number XCone multiplicities and its improvement in comparison to the baseline limits are shown in Figures 6.44 for $T\bar{T}$ and in Figures 6.46 for $B\bar{B}$. Figures 6.45 and Figures 6.47 show the analogous results using solely S_T as discriminant.

Table 6.4: Baseline expected 95% CL upper limits (pb) for the production cross section of $T\bar{T}$, only considering luminosity, lepton identification, lepton isolation, lepton reconstruction, and QCD renormalization scale uncertainties.

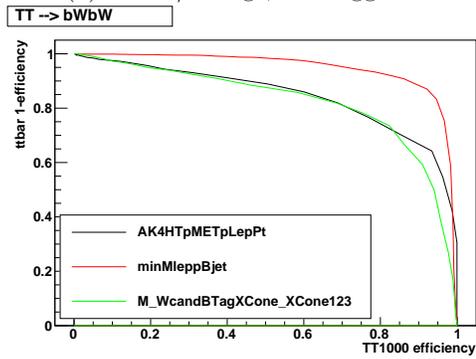
T mass (GeV)	bW	singlet	doublet	tZ	tH
800	0.019	0.036	0.056	0.108	0.037
900	0.016	0.027	0.037	0.069	0.024
1000	0.013	0.021	0.025	0.043	0.017
1100	0.011	0.016	0.019	0.036	0.012
1200	0.010	0.014	0.014	0.024	0.010
1300	0.009	0.012	0.011	0.019	0.008
1400	0.008	0.010	0.010	0.016	0.006
1500	0.008	0.009	0.008	0.013	0.006
1600	0.007	0.007	0.007	0.011	0.005
1700	0.006	0.007	0.006	0.009	0.004
1800	0.006	0.006	0.005	0.008	0.004



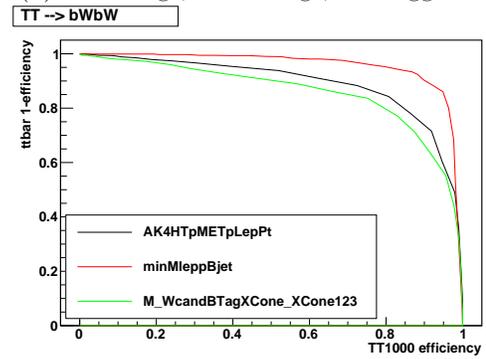
(a) No W/H-tags, 1 b-tagged AK4



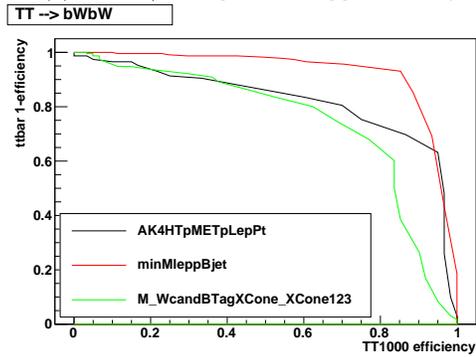
(b) No H-tags, 1+ W-tags, 1 b-tagged AK4



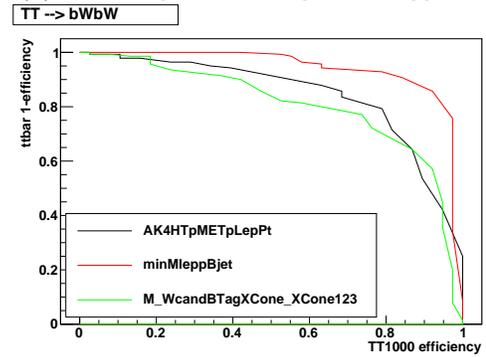
(c) No W/H-tags, 2 b-tagged AK4 jets



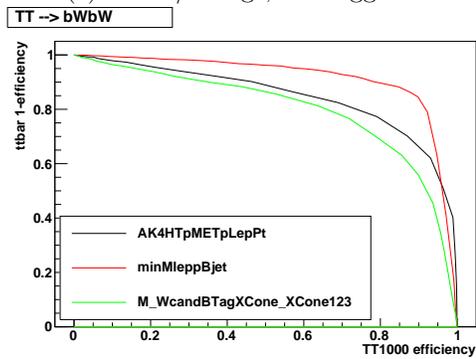
(d) No H-tags, 1+ W-tags, 2 b-tagged AK4



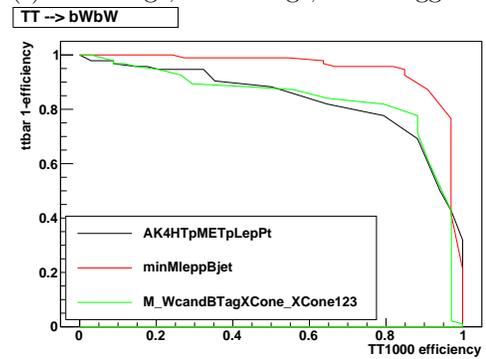
(e) No W/H-tags, 3 b-tagged AK4



(f) No H-tags, 1+ W-tags, 3+ b-tagged AK4

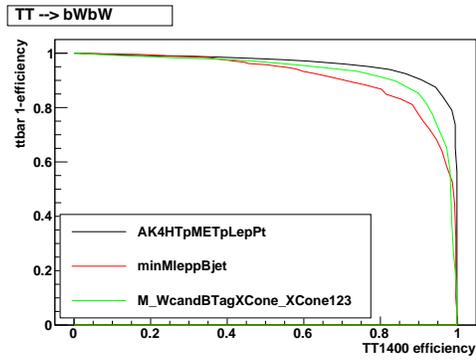


(g) 1 H-tagged AK8 jet, 1+ b-tagged AK8 subject

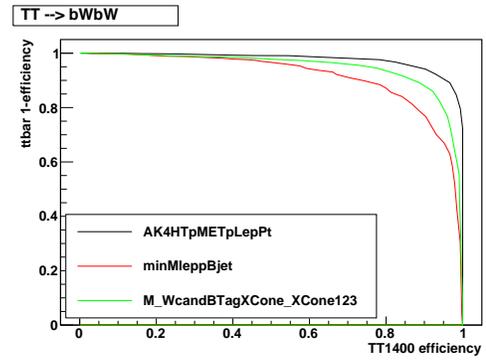


(h) 2 H-tagged AK8 jet, 1+ b-tagged AK8 subject

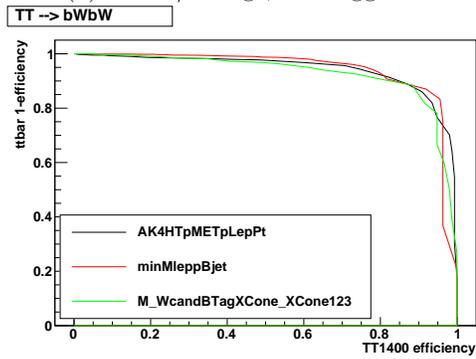
Figure 6.41: ROC curves for various observables, in samples of $t\bar{t}$ and $T\bar{T} \rightarrow bWbW$, where T has a mass of 1.0 GeV.



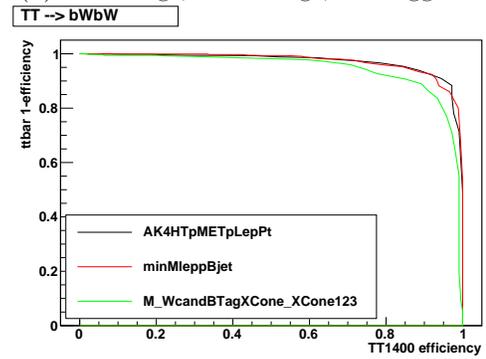
(a) No W/H-tags, 1 b-tagged AK4



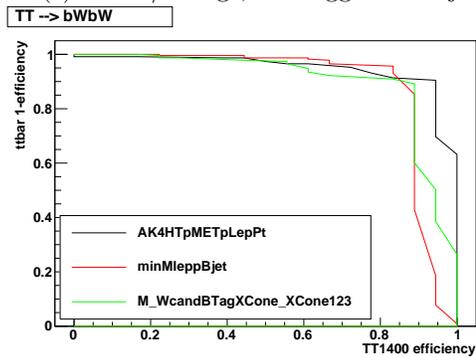
(b) No H-tags, 1+ W-tags, 1 b-tagged AK4



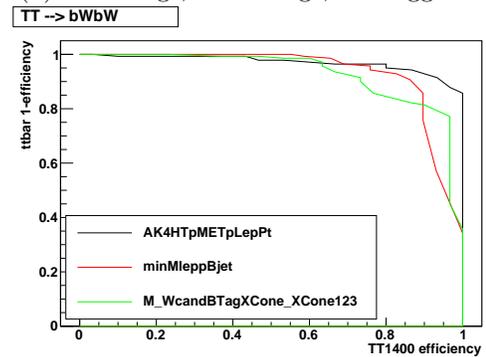
(c) No W/H-tags, 2 b-tagged AK4 jets



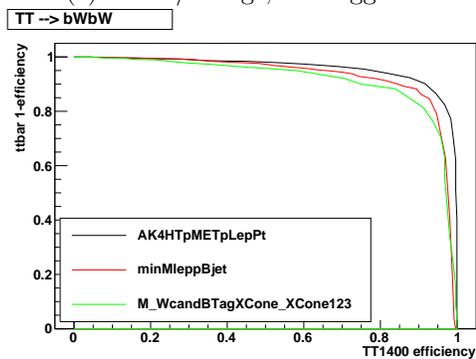
(d) No H-tags, 1+ W-tags, 2 b-tagged AK4



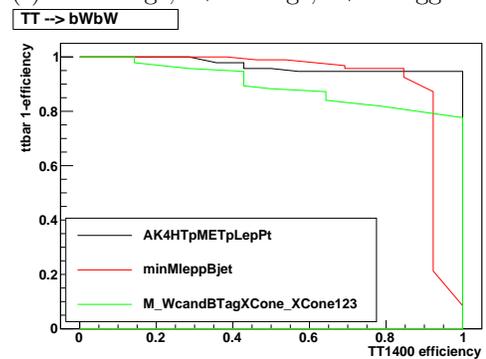
(e) No W/H-tags, 3 b-tagged AK4



(f) No H-tags, 1+ W-tags, 3 b-tagged AK4

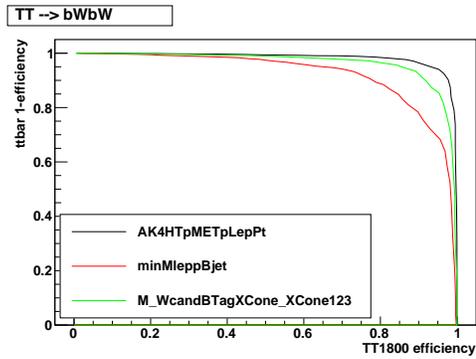


(g) 1 H-tagged AK8 jet, 1+ b-tagged AK8 subject

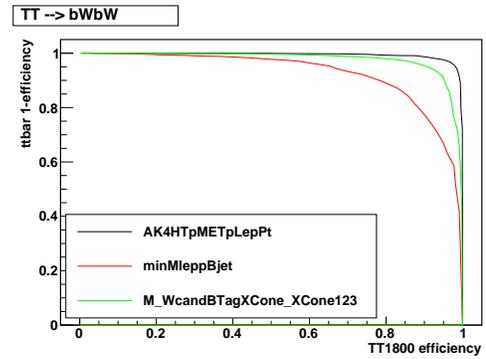


(h) 2 H-tagged AK8 jet, 1+ b-tagged AK8 subject

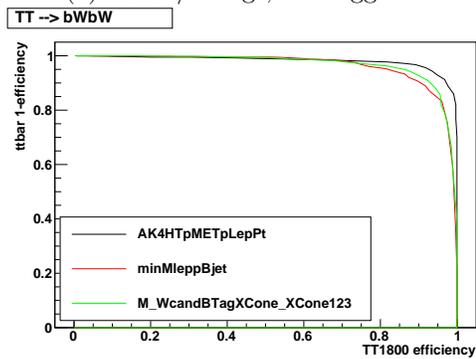
Figure 6.42: ROC curves for various observables, in samples of $t\bar{t}$ and $T\bar{T} \rightarrow bWbW$, where T has a mass of 1.4 GeV.



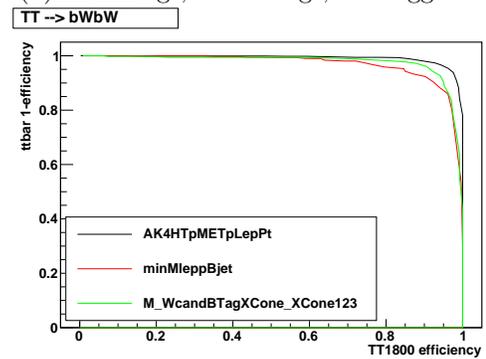
(a) No W/H-tags, 1 b-tagged AK4



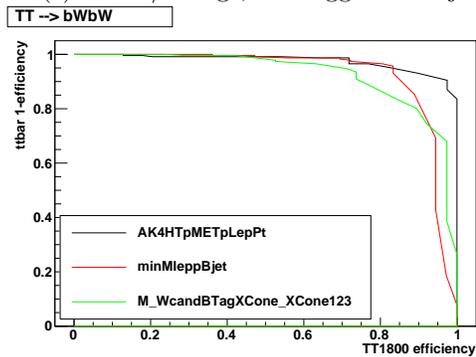
(b) No H-tags, 1+ W-tags, 1 b-tagged AK4



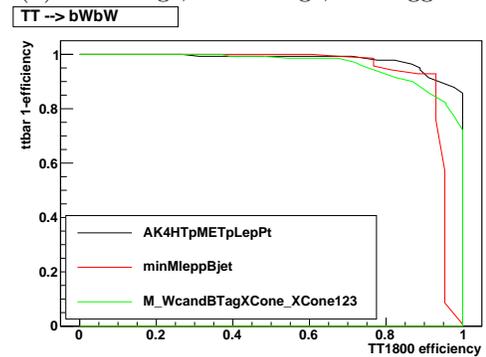
(c) No W/H-tags, 2 b-tagged AK4 jets



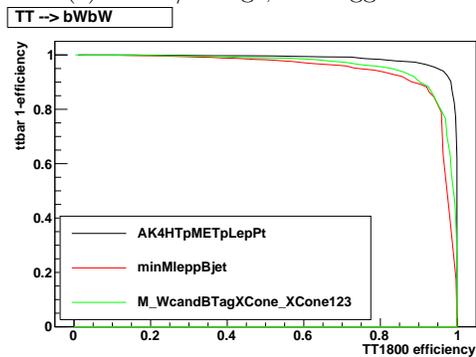
(d) No H-tags, 1+ W-tags, 2 b-tagged AK4



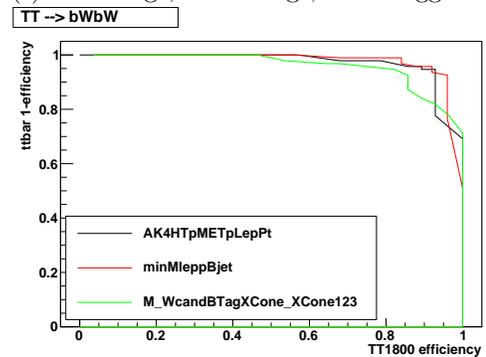
(e) No W/H-tags, 3 b-tagged AK4



(f) No H-tags, 1+ W-tags, 3 b-tagged AK4



(g) 1 H-tagged AK8 jet, 1+ b-tagged AK8 subjet



(h) 2 H-tagged AK8 jet, 1+ b-tagged AK8 subjet

Figure 6.43: ROC curves for various observables, in samples of $\bar{t}t$ and $T\bar{T} \rightarrow bWbW$, where T has a mass of 1.8 GeV.

Table 6.5: Baseline expected 95% CL upper limits (pb) for the production cross section of $B\bar{B}$, only considering luminosity, lepton identification, lepton isolation, lepton reconstruction, and QCD renormalization scale uncertainties.

B mass (GeV)	tW	singlet	doublet	bZ	bH
800	0.058	0.057	0.151	0.322	0.097
900	0.040	0.039	0.114	0.187	0.069
1000	0.027	0.029	0.089	0.147	0.058
1100	0.021	0.023	0.075	0.156	0.046
1200	0.017	0.019	0.062	0.114	0.037
1300	0.013	0.015	0.056	0.112	0.032
1400	0.010	0.013	0.052	0.101	0.033
1500	0.009	0.010	0.048	0.103	0.027
1600	0.007	0.009	0.043	0.083	0.025
1700	0.006	0.008	0.040	0.068	0.023
1800	0.005	0.007	0.036	0.065	0.021

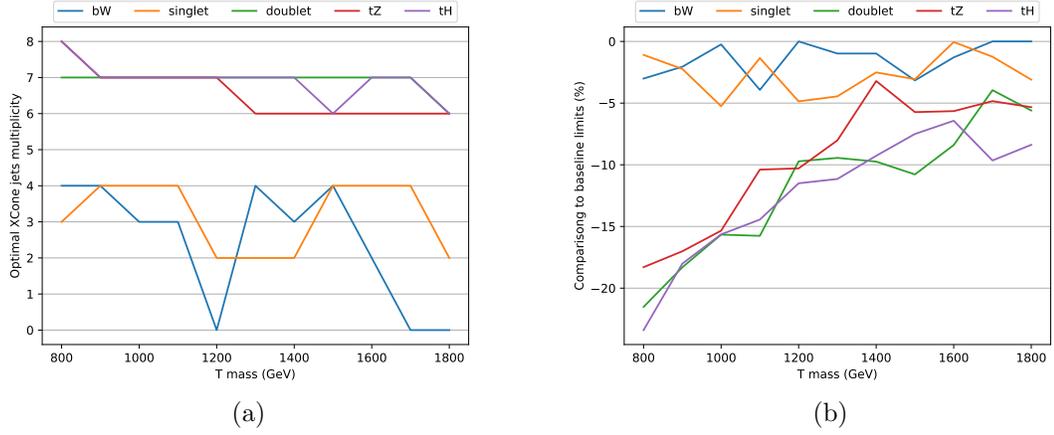


Figure 6.44: (a) Shows the optimal number of X Cone jet multiplicity selection using the $\min[M(l, b)]$, S_T discriminants in search for $T\bar{T}$, and (b) shows the corresponding limits improvement in comparison to the baseline limit shown in Table 6.4

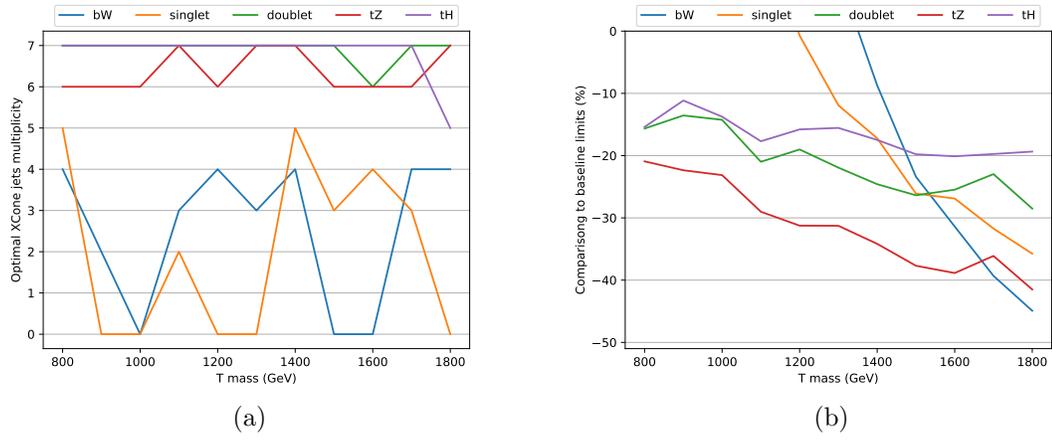


Figure 6.45: (a) Shows the optimal number of XCone jet multiplicity selection using only the S_T discriminant in search for $T\bar{T}$, and (b) shows the limits improvement in comparison to the baseline limit shown in Table 6.4

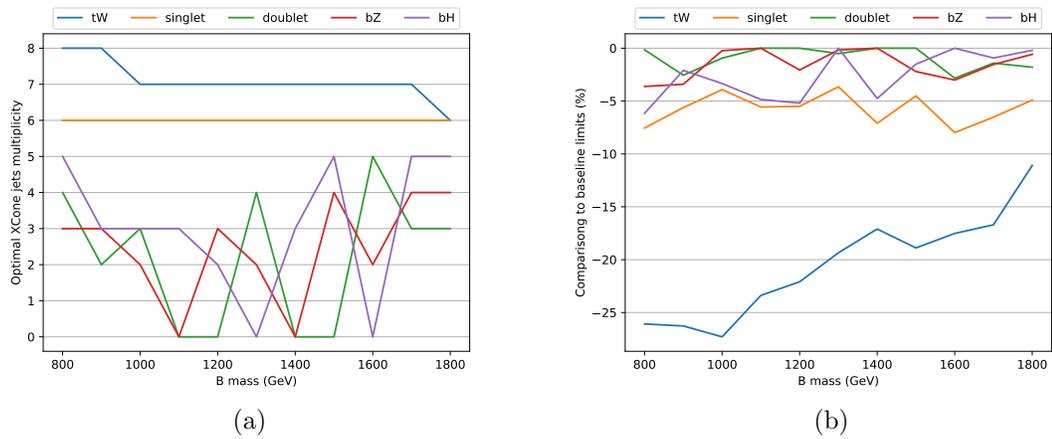


Figure 6.46: (a) Shows the optimal number of XCone jet multiplicity selection using the $\min[M(l, b)]$, S_T discriminants in search for $B\bar{B}$, and (b) shows the corresponding limits improvement in comparison to the baseline limit shown in Table 6.5

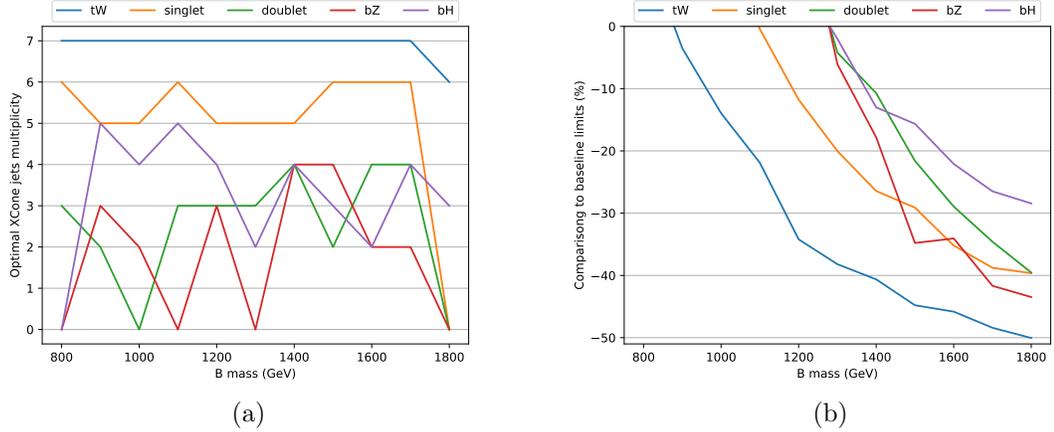


Figure 6.47: (a) Shows the optimal number of Xcone jet multiplicity selection using only the S_T discriminant in search for $B\bar{B}$, and (b) shows the limits improvement in comparison to the baseline limit shown in Table 6.4

We observe that for $T\bar{T}$, when bW is the dominant decay, requiring up to 4 Xcone jets using the baseline discriminants ($\min[M(l,b)]$, S_T) is optimal with improvements of a few percent in the limits, whereas when tZ,tH dominates the decay, requiring about 6 Xcone jets using the S_T discriminant is optimal with improvements of about 30%. On the other hand, we observe that for $B\bar{B}$, when bZ,bH is the dominant decay, requiring about 4 Xcone jets using the baseline discriminants ($\min[M(l,b)]$, S_T) is optimal with improvements of a few percent in the limits, whereas when tW dominates the decay, requiring about 6 Xcone jets using the S_T discriminant is optimal with limit improvements of about 30%.

To see the final impact on the mass upper limits, we calculate the expected limits and observed limits where we include all single lepton analysis systematic uncertainties (excluding the jet energy correction) listed in Table 5.26 and compare to the theoretical cross sections. The results are shown in Tables 6.6 and 6.7 for $T\bar{T}$, and in Tables 6.8 and 6.9 for $B\bar{B}$. The corresponding plots are shown in Figure 6.48 and Figure 6.49 for $T\bar{T}$ and $B\bar{B}$ respectively.

Table 6.6: Expected 95% CL upper limits for the mass of T (GeV) with XCone jet multiplicity requirements. The '*' labels the most sensitive limits for the particular signal decay scenario.

	T decay mode				
	$T \rightarrow bW$	singlet	doublet	$T \rightarrow tZ$	$T \rightarrow tH$
minMlbST					
0 XCone jets (baseline)	1138	1086*	1096	954	1184
3 XCone jets	1138	1081	1097	952	1183
4 XCone jets	1139*	1085	1097	953	1186
6 XCone jets	1047	1064	1099	970	1190
7 XCone jets	961	1040	1125	991	1199
ST					
3 XCone jets	1092	1076	1110	991	1192
4 XCone jets	1092	1076	1112	991	1192
6 XCone jets	971	1057	1122	997	1195
7 XCone jets	839	1023	1134*	1015*	1208*
Improvement	0.1%	0%	3.5%	6.4%	2.0%

Table 6.7: Observed 95% CL upper limits for the mass of T (GeV) with XCone jet multiplicity requirements. The '*' labels the most sensitive (expected) limits for the particular signal decay scenario.

	T decay mode				
	$T \rightarrow bW$	singlet	doublet	$T \rightarrow tZ$	$T \rightarrow tH$
minMlbST					
0 XCone jets (baseline)	1141	1099*	1123	996	1192
3 XCone jets	1140	1096	1125	1003	1191
4 XCone jets	1134*	1094	1119	976	1190
6 XCone jets	872	1004	1099	936	1191
7 XCone jets	805	1013	1127	890	1204
ST					
3 XCone jets	1115	1077	1081	890	1154
4 XCone jets	1116	1074	1076	889	1153
6 XCone jets	1071	1085	1111	1036	1179
7 XCone jets	1003	1091	1162*	1105*	1220*
Improvement	-0.6%	0%	5.7%	10.9%	2.3%

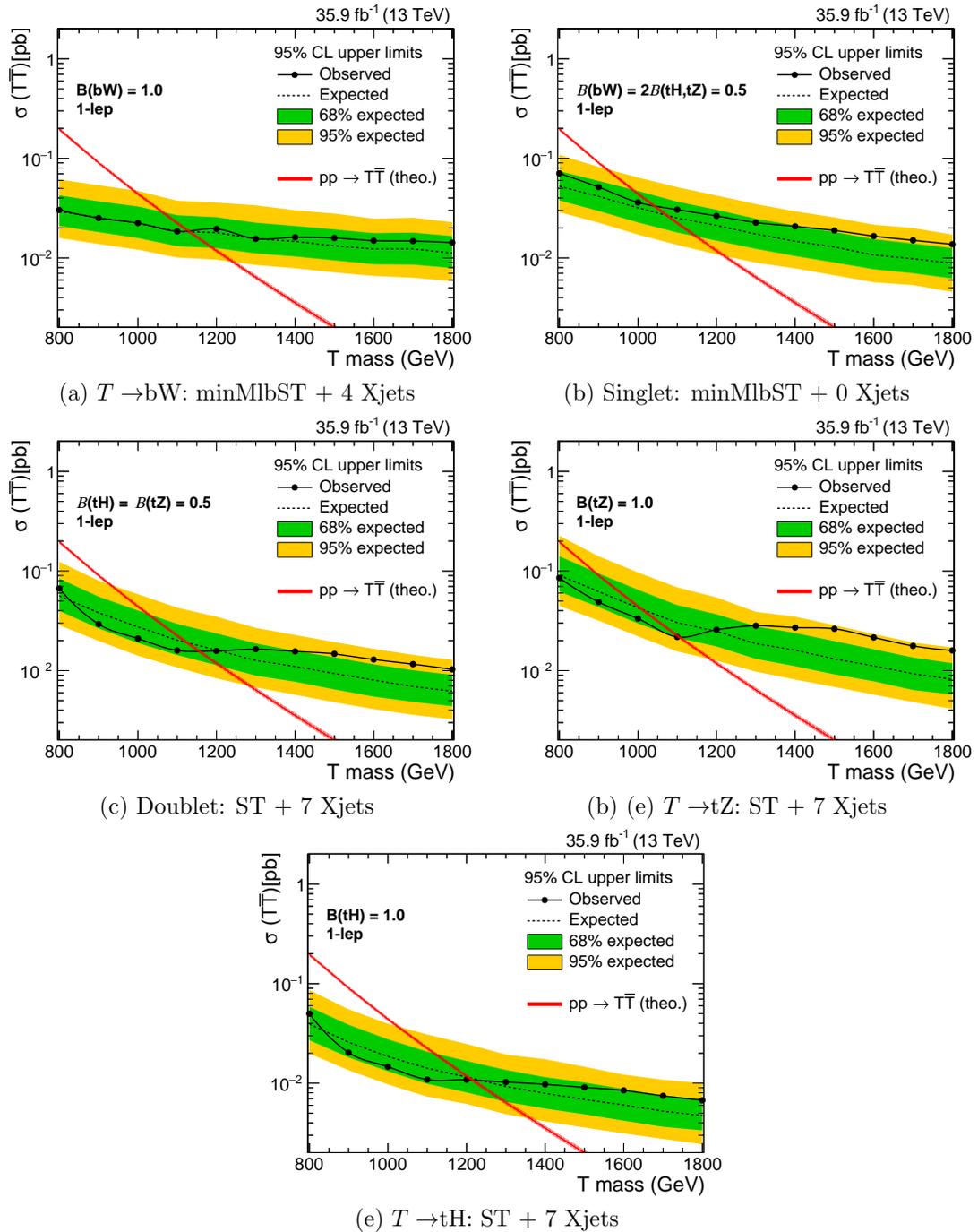


Figure 6.48: The 95% CL upper limits for the mass of T (GeV), which corresponds to the optimal choice of Xcone jet multiplicity requirement as labelled '*' in Tables 6.6 and 6.7

Table 6.8: Expected 95% CL upper limits for the mass of B (GeV) with XCone jet multiplicity requirements. The '*' labels the most sensitive limits for the particular signal decay scenario.

	B decay mode				
	$B \rightarrow tW$	singlet	doublet	$B \rightarrow bZ$	$B \rightarrow bH$
minMlbST					
0 XCone jets (baseline)	1051	1007	< 800	< 800	894
3 XCone jets	1052	1014	< 800	< 800	895*
5 XCone jets	1062	1018	< 800	< 800	890
6 XCone jets	1078	1020*	< 800	< 800	872
7 XCone jets	1100	1010	< 800	< 800	< 800
ST					
3 XCone jets	1103	1007	< 800	< 800	846
5 XCone jets	1102	1011	< 800	< 800	849
6 XCone jets	1107	1020	< 800	< 800	844
7 XCone jets	1131*	994	< 800	< 800	< 800
Improvement	7.6%	1.3%	—	—	0.1%

Table 6.9: Observed 95% CL upper limits for the mass of B (GeV) with XCone jet multiplicity requirements. The '*' labels the most sensitive (expected) limits for the particular signal decay scenario.

	B decay mode				
	$B \rightarrow tW$	singlet	doublet	$B \rightarrow bZ$	$B \rightarrow bH$
minMlbST					
0 XCone jets (baseline)	973	997	< 800	< 800	876
3 XCone jets	970	998	< 800	< 800	879*
5 XCone jets	919	947	< 800	< 800	863
6 XCone jets	858	905*	< 800	< 800	837
7 XCone jets	979	947	< 800	< 800	< 800
ST					
3 XCone jets	1122	1022	< 800	< 800	870
5 XCone jets	1123	1025	< 800	< 800	870
6 XCone jets	1114	1011	< 800	< 800	869
7 XCone jets	1164*	1083	< 800	< 800	< 800
Improvement	19.3%	-9.22%	—	—	0.3%

6.3.3 Summary and discussion: XCone jets in $T\bar{T} / B\bar{B} \rightarrow$ single lepton

We explored several new observables constructed based on the XCone jet. We have not found any of these observables to be performing better than the observables used in the default baseline search strategy. However we observe that the observable $M_{WcandBTagXCone_XCone123}$ could potentially be a useful observable since the distribution produces a peak approximately at the masses of the T . This mass information has yet to be exploited in the baseline search.

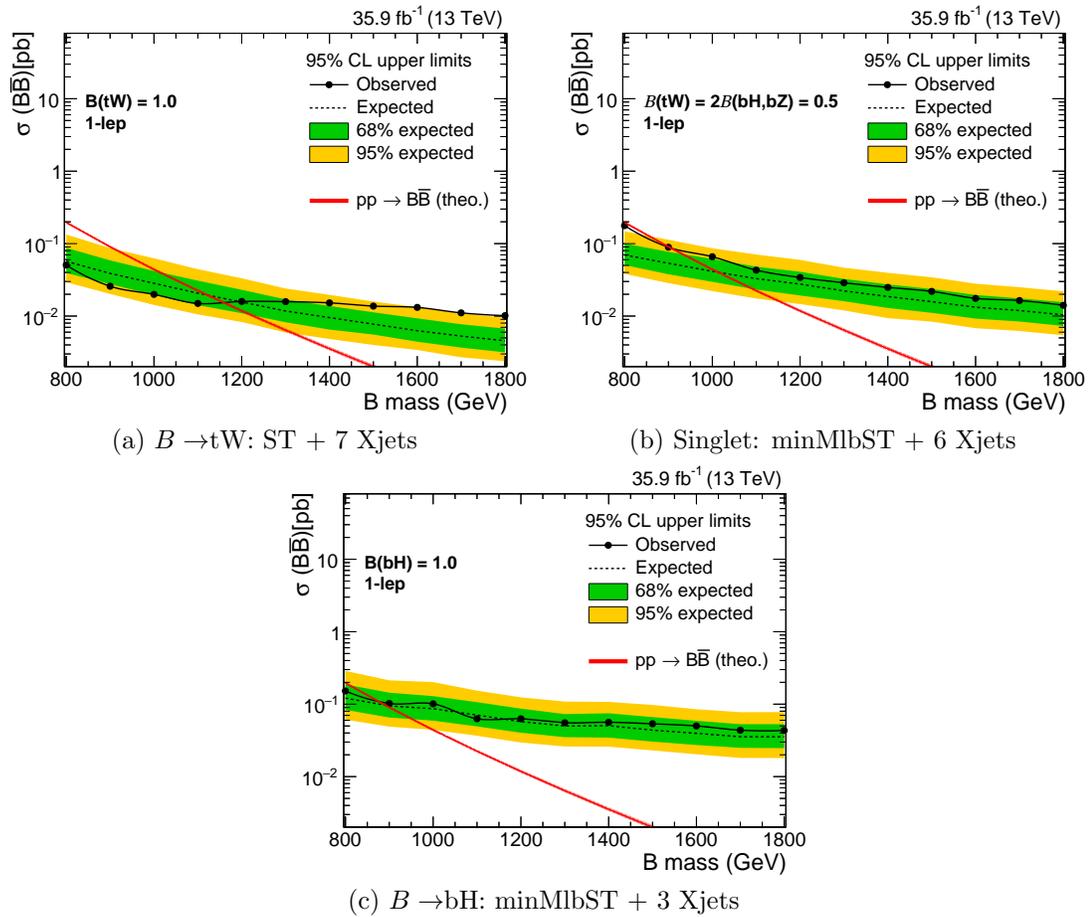


Figure 6.49: The 95% CL upper limits for the mass of B (GeV), which corresponds to the optimal choice of Xcone jet multiplicity requirement as labelled '*/' in Tables 6.8 and 6.9

As most T/B decays will produce more jets than any other SM processes, it is expected that requiring a high number of jet multiplicity would increase the signal discrimination against background processes. T/B are heavy resonances that would produce high momentum jets as decay products, such that the jets would be collimated. The advantage of using the X Cone jet algorithm is that it would be able to distinguish these collimated jets. Requiring X Cone multiplicity requirements, we observe that some of the upper limits for the mass of the T/B are improved, especially for the final states with high number of jet multiplicity. Specifically, the observed mass limit for the T quark could be improved by up to 10% for the $\text{BR}(T \rightarrow tH)=100\%$ scenario, and for the B quark could be improved by up to about 20% for the $\text{BR}(T \rightarrow tW)=100\%$ scenario.

Chapter 7

Summary and conclusions

We have discussed the standard model theory of particle physics as the current best theory that describes known fundamental matter and forces. We briefly reviewed the mathematical description of the theory, as well as the limitation and extensions which predicts the existence new type of fundamental particles called vector-like-quarks.

Following that, we have described the CMS detector at CERN's LHC, which the data used in this thesis was collected, and how events and the physics objects are reconstructed.

We have analyzed 2016 data to search for evidence of $T\bar{T}$ production in events with three or more charged leptons and two leptons with the same charge. In LHC Run 1 heavy T quarks with mass less than 696 GeV were excluded for the nominal branching ratios, and with 35.9 fb^{-1} of integrated luminosity at a center-of-mass energy of 13 TeV we observe improved sensitivity. In the trilepton final state the heavy T quark mass exclusion limit reaches 1080 GeV for the singlet branching ratios. In the same-sign dilepton final state the limit reaches 910 GeV for the doublet branching ratios. Combining the trilepton, SS dilepton and the single-lepton channels, we exclude T (B) quarks at 95% confidence level with masses below 1200 (1170) GeV in the singlet branching fraction scenario and 1280 (940) GeV in the doublet branching fraction scenario. For other branching fraction scenarios this search excludes T (B) quark masses below 1140–1300 GeV(910–1240 GeV). This represents an improvement in sensitivity of typically 200–600 GeV, compared to previous CMS

results. These results are the strongest exclusion limits to date for T quarks with $\mathcal{B}(tZ)$ greater than ≈ 0.5 and for B quarks with $\mathcal{B}(tW)$ less than ≈ 0.6 .

Lastly, we have described the XCone jet clustering algorithm and explored its use in the search for a pair of top/bottom-like-VLQ decaying to final states with one electron/muon, which results in improvements of exclusion limits up to about 15% compared to the baseline strategy.

Appendix A

Identification of high momentum

Higgs decaying to b-quark pairs

(boosted double-b tagging)

In the early stages of my PhD research, I collaborated with other members of the CMS experiment to develop and commission a method for identifying jets with large a distance parameter (Anti-kT algorithm with $\Delta R = 0.8$) originating from high momentum Higgs boson decaying to b quark pairs. This method is referred to as the *boosted double-b tagger*. This chapter summarizes the studies of the boosted double-b tagger. However, as mentioned in Section 4.4.8, this method is not utilized in the final strategy used in this thesis. My contribution to this study is mainly on the efficiency and scale factor measurements. All results presented in this chapter has been referenced from the published CMS documentation, Ref [20].

A.1 Introduction

We present a novel approach to identifying boosted $H \rightarrow b\bar{b}$ candidates which tries to fully exploit the presence of two b quarks inside an anti-kT jet with large radius of $R=0.8$ and their topology in

relation to the fact that the b hadron flight directions are strongly correlated with the energy flows of the two subjets. To discriminate $b\bar{b}$ originated from a heavy resonance from QCD jets initiated by single partons, we have developed a dedicated multivariate (MVA) tagging algorithm, named “double-b tagger”, implemented and optimized using the TMVA package [128]. To reconstruct b hadron decay vertices, we apply the Inclusive Vertex Finder (IVF) algorithm [129, 130] which identifies secondary vertices independently of the jet clustering. We reconstruct the decay chains of the two b hadrons by associating reconstructed secondary vertices to the subjet axes represented by τ -axes defined in later in this chapter. No other substructure variable or quantity is employed. We find that this novel approach greatly improves the ability to identify boosted Higgs bosons with respect to previously used methods.

A.2 Event samples, reconstruction and wide jet identification

Simulated Monte Carlo (MC) samples of heavy resonances decaying to two Higgs bosons ($X \rightarrow HH$) have been used as source of H jets. This simple topology is optimal for this study since no other objects are present in the final state and as the mass of the resonance increases, the H bosons are produced with larger boost. An example is KK-Graviton of signal [131] which is produced through gluon fusion and has spin 2. Several mass points are considered in order to cover a large enough phase space to study the p_T dependency, (800GeV- 3.5 TeV).

QCD multijets events, used in Section A.3 and A.4, are simulated using PYTHIA [132] for different \hat{p}_T bins and combined together to cover a broad kinematic range.

Top quark pair events are simulated with the next-to-leading-order generator POWHEG v2 [133, 134, 135, 136]. This generator is also used for the electroweak production of single top quarks in the tW channel [137]. The MC@NLO generator is used for the s- and t-channel processes of single top quark production [138] and for the Z+jets backgrounds. The generation of the W+jets was performed with MADGRAPH [139]. The MLM matching scheme is used, allowing up to four additional partons

in the matrix element [140]. All samples are interfaced to PYTHIA for the showering.

All events are generated using the parton distribution functions (PDF) from the NNPDF 3.0 PDF sets [141], while for the showering the underlying event tune CUETP8M1 [142] is used. To simulate accurately the LHC luminosity conditions during the 2015 data taking period, additional pp interactions overlapping with the event of interest in the same bunch crossing, denoted as pileup events, are added in the simulated samples to reproduce the pileup distribution measured in data.

Data corresponding to an integrated luminosity of 2.6 fb^{-1} at $\sqrt{s} = 13 \text{ TeV}$ with 25 ns bunch spacing in 2015 are used. They have been collected with single jet triggers with p_T threshold of 200, 260, 320 and 400 GeV in order to measure the double-b tagger efficiency. All triggers except the one with the highest threshold have been prescaled to limit the trigger rates, which means that the event samples they recorded correspond to a lower integrated luminosity. Triggers with different p_T thresholds are combined to gain efficiency, taking trigger prescale factors into account. Apart from the prescaling, the trigger efficiency is more than 99% in the phase space selected for this study. Collision events recorded with a single muon trigger, requiring $p_T(\mu) > 45 \text{ GeV}$ and $|\eta(\mu)| < 2.1$ are used for the mistagging measurement from top quark jets.

Stable particles are identified with the particle-flow (PF) algorithm [143, 144] that reconstructs each individual particle with an optimized combination of information from the various elements of the CMS detector.

Events are required to have at least one reconstructed vertex consistent with a pp interaction. The vertex with the highest sum of the transverse momentum squared of the associated physics objects is considered to be the primary interaction vertex.

Muons are reconstructed within $|\eta| < 2.4$ by selection criteria based on the compatibility of the track reconstructed by means of the silicon tracker only and of the combination of the hits in both the silicon tracker and the muon spectrometer [82]. Additional requirements are based on the compatibility of the trajectory with the primary vertex and on the number of hits observed in the tracker and muon systems. The muon isolation requirement is computed using the reconstructed tracks within $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\Phi)^2} < 0.3$ from the muon direction, excluding the muon itself.

Jets are reconstructed from particle-flow candidates using the anti- k_T clustering algorithm [?], with a distance parameter of $R=0.8$ (AK8), as implemented in the FASTJET package [145, 146]. Jet energy corrections, as a function of pseudorapidity and transverse momentum of the jet, are applied [147]. Jet identification criteria are also applied to reject fake jets from detector noise and jets originating from primary vertices not associated with the hard interaction [148]. We select jets in the event requiring $|\eta| < 2.4$, so that they fall within the tracker acceptance.

A.3 Double-b tagger algorithm

Several observables exploiting the distinctive properties of b hadrons are employed as input variables for the CSVv2 [149] algorithm used in the CMS collaboration. Following that example we have adapted their definition to deal with the $b\bar{b}$ topology. We substitute the jet axis information with the two τ -axes to resolve the two b hadron decay chains we expect for the $H \rightarrow b\bar{b}$ signal.

A.3.1 Discriminating variables

We present here the discriminating variables that are used as input to the MVA algorithm to distinguish between the signal $H \rightarrow b\bar{b}$ jets and the background from inclusive QCD jets. The variables rely on reconstructed tracks, secondary vertices (SV) as well as the two-SV system. Since the angular separation between the decay products of a resonance depend on the momentum and the mass of the resonance, in order to keep the algorithm as general as possible, one of the guiding principle in the selection of input variables is that the variables do not have strong dependence on the jet p_T and the jet mass.

Tracks with $p_T > 1\text{GeV}$ are associated to jets in a cone $\Delta R < 0.8$ around the jet axis, where the jet axis is defined by the primary vertex and the direction of the jet momentum. Then we associate each track to the closest τ -axis. The distance of a track to the τ -axis is defined as the distance of closest approach of the track to the axis. In order to reject tracks from pileup this quantity is required to be less than $700\ \mu\text{m}$. The point on the track that is closest to the τ -axis must be within 5 cm of the primary vertex. The contamination from decay products of long-lived particles, e.g. neutral

kaons, is reduced by removing pairs of tracks compatible with the kaon masses within 30 MeV. The impact parameter, IP, of a track with respect to the primary vertex is used to distinguish the decay products of a b hadron from prompt tracks. The IP is calculated in three dimensions and the impact parameter significance, SIP, is defined as the ratio of the IP to its estimated uncertainty. Several input variables related to the presence and properties of secondary vertices coming from b hadron decay have been investigated. Using tracks with $p_T > 0.8\text{GeV}$, secondary vertices are identified through the Inclusive Vertex Finder (IVF) [129, 130] algorithm. This algorithm is not seeded from tracks associated to the reconstructed jets, but it uses as input the collection of reconstructed tracks in the event. The reconstructed secondary vertices are associated to jets in a cone $\Delta R < 0.7$ and then to the closest τ -axis within that jet. For each τ -axis, track momenta of the constituent tracks from all the SVs associated to a given τ -axis are added to compute the SV mass and the SV transverse momentum for that τ -axis.

The input variables to the double-b tagger MVA discriminant are:

- The first four SIP values for selected tracks ordered in decreasing SIP;
- For each τ -axis we consider the first two SIP values for their respective associated tracks ordered in decreasing SIP, to further discriminate against single b quark and light flavor jets from QCD when one or both SV are not reconstructed due to IVF inefficiencies;
- The measured IP significance in the plane transverse to the beam axis, 2D SIP, of the first two tracks (first track) that raises the SV invariant mass above the bottom (charm) threshold of 5.2 (1.5) GeV;
- The number of SV associated to the jet;
- The significance of the 2D distance between the primary vertex and the secondary vertex, flight distance, for the SV with the smallest 3D flight distance uncertainty, for each of the two τ -axes;
- The ΔR between the SVs with the smallest 3D flight distance uncertainty and its τ -axis, for each of the two τ -axes;

- The relative pseudorapidity, η_{rel} , of the tracks from all SVs with respect to their τ -axis for the three leading tracks ordered in increasing η_{rel} , for each of the two τ -axes;
- The total SV mass, defined as the total mass of all SVs associated to a given τ -axis, for each of the two τ -axes;
- The ratio of the total SV energy, defined as the total energy of all SVs associated to a given τ -axis, and the total energy of all the tracks associated to the fat jet that are consistent with the primary vertex, for each of the two τ -axes;
- The information related to the two-SV system, the z variable, defined as:

$$z = \Delta R(\text{SV}_0, \text{SV}_1) \cdot \frac{p_{\text{T}}(\text{SV}_1)}{m(\text{SV}_0, \text{SV}_1)} \quad (\text{A.1})$$

where SV_0 and SV_1 are SVs with the smallest 3D flight distance uncertainty. The z variable helps rejecting the $b\bar{b}$ background from gluon splitting relying on the different kinematic properties compared to the $b\bar{b}$ pair from the decay of a massive resonance.

We select as discriminating variables all those with enough classifier separation (a default output of TMVA), that show small correlation with the other inputs and improve the QCD background discrimination by at least 5%. In total 27 variables are used as input to the multivariate discriminant. The most discriminating variables are the SIP for the most displaced tracks, the vertex energy ratio for SV_0 , and the 2D SIP for the first track above bottom threshold. In Fig. A.1 distributions for some discriminating input variables are shown for the signal $\text{H} \rightarrow b\bar{b}$ jets and the background QCD jets. In particular $g \rightarrow b\bar{b}$ and single b quark production are shown separately as well as light flavor jet contribution. The secondary vertex multiplicity and the vertex energy ratio for SV_0 , along with SIP of the first track above bottom threshold show a good separation between the $\text{H} \rightarrow b\bar{b}$ jets and different QCD jet components. The z variable shows good discrimination against the $g \rightarrow b\bar{b}$ contribution.

Several variables related to the presence and properties of soft leptons arising from the b hadron decay have also been investigated. Despite a small gain in performance, the soft lepton variables

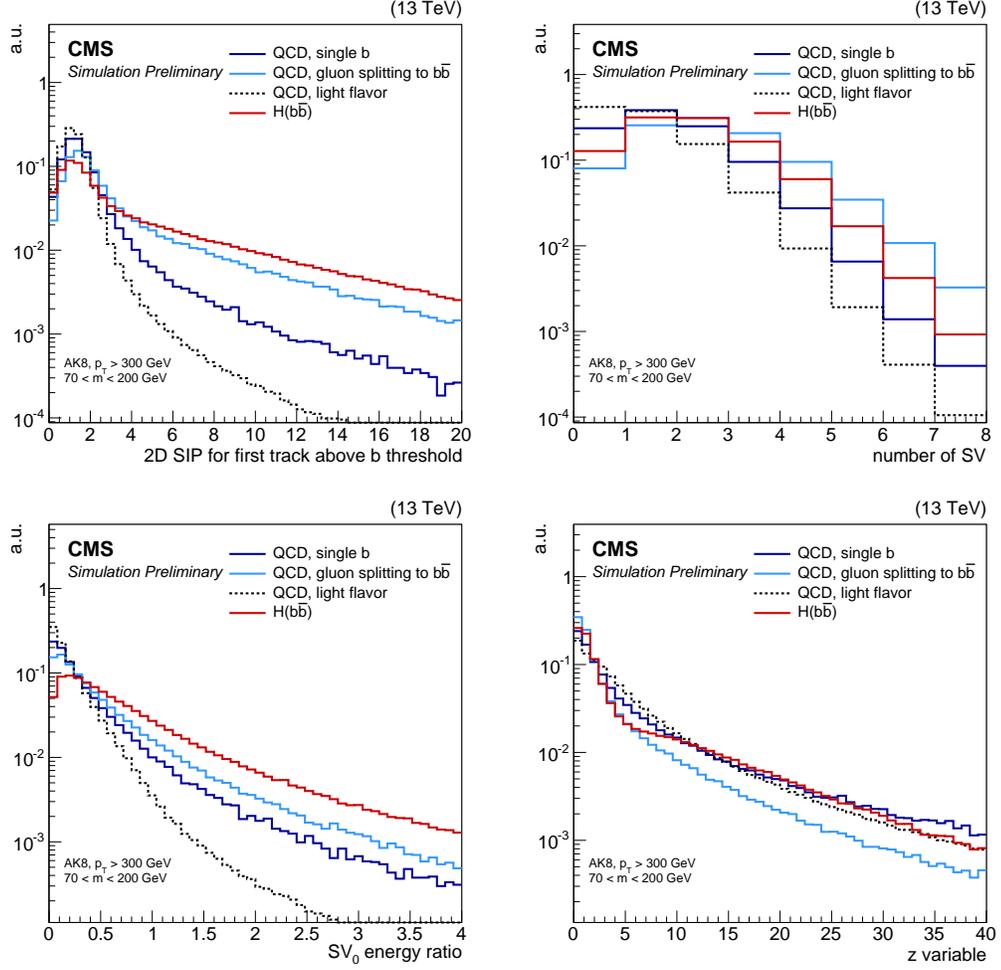


Figure A.1: Distributions of 2D IP significance for the most displaced track raising the SV invariant mass above the bottom quark threshold, number of secondary vertices associated to the AK8 jet, the vertex energy ratio for SV_0 , and the z variable. Comparison between $H \rightarrow b\bar{b}$ jets from simulated samples of KK-Graviton decaying to HH and QCD jets containing zero, one or two b quarks are used. AK8 jets are selected with $p_T > 300$ GeV and pruned jet mass $70 < m < 200$ GeV. The distributions are normalized to unit area. [20]

were excluded from the final list of input variables since they could introduce undesired biases in the performance measurement in data where μ -tagged jets from QCD multijets events are used.

A.4 Efficiency measurement in 2015 data

The efficiency of the double-b tagger is measured in the data sample consisting of high p_T jets enriched in $b\bar{b}$ from gluon splitting. In order to select topologies as similar as possible to a signal jet, we require an AK8 jet with $p_T > 300$ GeV and pruned mass > 50 GeV. We ask the jet to be matched to at least two muons, each with $p_T > 7$ GeV and $|\eta| < 2.4$. Each pruned subjet is required to have at least one muon among its constituents and within $\Delta R < 0.4$ from the subjet axis (“double-muon tagged”). An alternative selection that requires at least one muon is also examined as cross check for the measurement (“single-muon tagged”). While this single-muon selection allows for a larger dataset in which to perform the tagger efficiency measurement, the gluon splitting topology in this inclusive phase space is less signal like relative to the double-muon selection. Thus, to maximize the similarity between the $g \rightarrow b\bar{b}$ and the $H \rightarrow b\bar{b}$ topology, the measurement is performed requiring double-muon tagged jets.

The comparison between the data and the simulated samples of the variables that are used as inputs to the double-b tagger shows good agreement, as can be seen Fig. A.2. In Fig. A.3 we report also the double-b tagger output in data and simulated events. The total number of entries in the simulation is normalized to the observed number of entries in data. Overall the agreement between data and simulation is fairly good.

The efficiency of the double-b tagger is measured in data and MC for three different operating points. The measurement relies on the Jet Probability (JP) discriminant, for which the expected simulated distributions (“templates”) are different for the various jet flavors. The fraction of b (from gluon splitting) jets is estimated by fitting the data distribution of the JP variable with the templates. This so-called Lifetime Tagging (LT) method [87] is also used to perform the measurement of the b jet identification efficiency scale factors for the standard anti- k_T $R = 0.4$ (AK4) jets [149].

The QCD MC sample is split into events containing b quark jets arising from gluon splitting

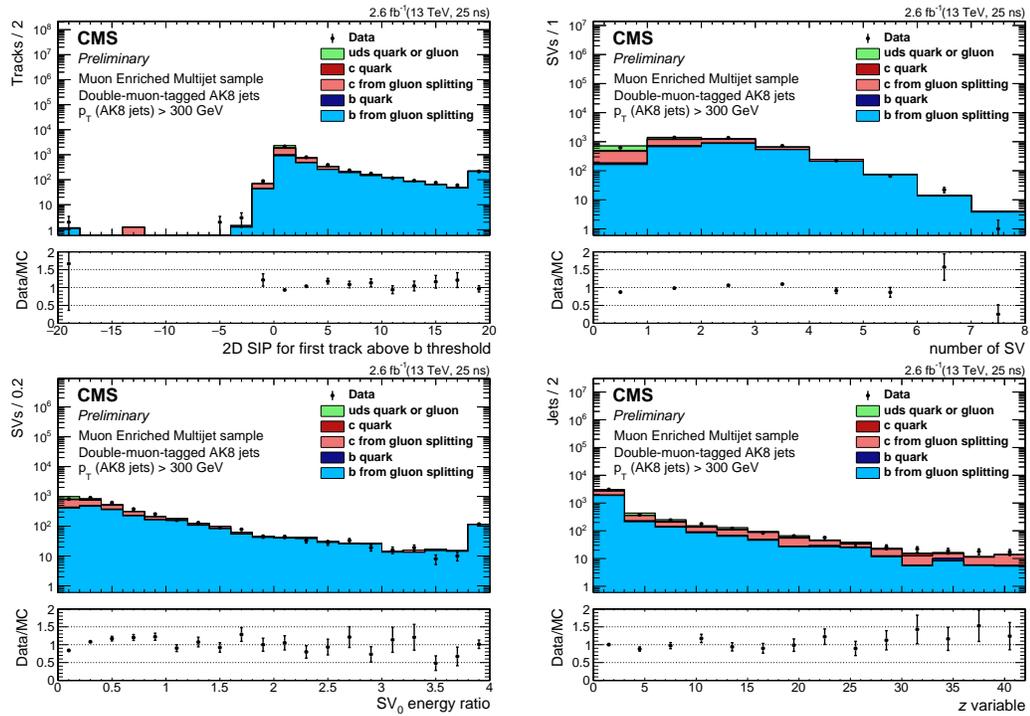


Figure A.2: Distributions of 2D IP significance for the most displaced track raising the SV invariant mass above the b quark threshold, number of secondary vertices associated to the AK8 jet, the vertex energy ratio for SV_0 , and the z variable. Data and simulated events are shown for the double-muon tagged jets selection. Simulated events are normalized to the yield observed in data, the overflow is in last bin. The bottom panel in each figure shows the ratio of the number of events observed in data to that of the MC prediction. [20]

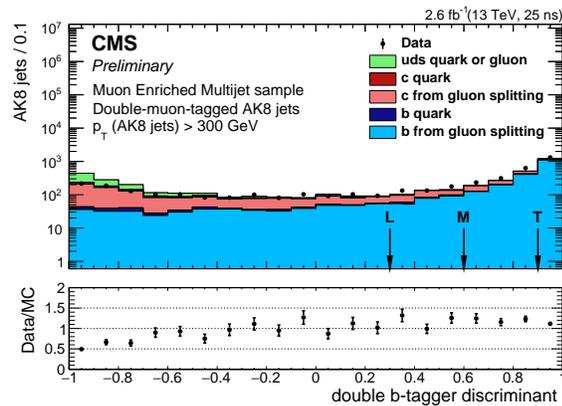


Figure A.3: Double-b tagger discriminant distribution in data and simulated samples for the double-muon tagged jets selection. Simulated events are normalized to the yield observed in data. The loose, medium and tight operating points are also reported. The bottom panel shows the ratio of the number of events observed in data to that of the MC prediction. [20]

and those (from b, c, light parton) which are not associated to this process, by requiring at least two generator level b hadrons clustered inside the jet. An example of fitted distributions for the JP discriminant in data is presented in Fig. A.4.

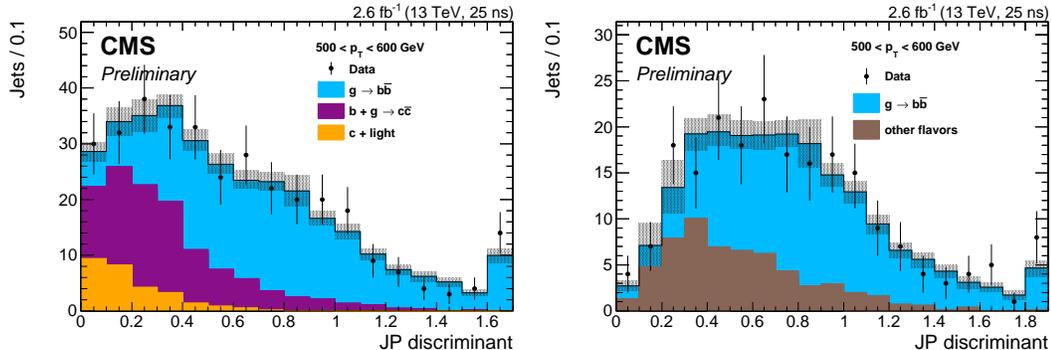


Figure A.4: Comparison of the JP discriminant distribution for the data and the sum of the fitted templates for all selected jets (left) and those jets passing the loose double-b tagger requirement (right) with p_T between 500 and 600 GeV. The shaded area represents the statistical and systematic (refer to the text for details) uncertainties on MC templates. Double-muon tagged AK8 jets are used for this measurement. The overflow is included in the last bin. [20]

The resulting data/MC efficiency scale factors (SFs) are presented in Fig. A.5 and listed in Tables A.1–A.3 for the double-muon tagged selection. The measurement is done for jets with p_T up to 700 (500) GeV for loose and medium (tight) operating points, which is driven by the size of the available data sample. Jets with larger p_T are included in the last p_T bin with an additional contribution up to $\simeq 20\%$ to the total number of jets selected in this bin.

As several background processes are being varied as a combined template in the fit procedure, the results could be sensitive to the prediction of the flavor composition of this background sample. The uncertainty on the scale factor due to the template definition is estimated by conservatively varying the normalization of each background contribution by $\pm 50\%$. As a cross check, the scale factor derivation is also performed by using all the background contributions as individual templates in the fit. The background template normalization variation contributes up to 5% as a systematic uncertainty on the scale factor.

Uncertainties on jet energy scale (JES) corrections are included as shape systematics on the JP discriminant and their impact on the scale factor measurement is negligible. Systematic uncertainties

due to bad modeling of track multiplicity and the b fragmentation function contribute 5% and 2% at most, respectively. Those associated to pileup, c quark fragmentation function, uncertainties on the fragmentation rate of a c quark to various D mesons, the branching ratios for c hadrons to muons, the K_S and Λ production fraction are found to be negligible.

We estimate the effect of the residual shape differences in the double-b tagger discriminant distribution between simulated $H \rightarrow b\bar{b}$ and $g \rightarrow b\bar{b}$ jet topologies. We compute a set of weights in order to match gluon splitting to H jets for the vertex energy ratio for SV_0 and the z variable distributions. Then, these weights are applied to both data and simulated events and the SFs are measured again. We found the SFs computed with and without applying these weights to agree within the uncertainty, validating the assumption of the gluon splitting being a good proxy for the signal in the selected phase space.

The SFs derived using double-muon and single-muon tagged jets are compatible, though the double-muon SFs have larger uncertainties, due to the limited size of the data sample. In both cases the Data/MC SFs are compatible with unity within uncertainties.

Table A.1: Loose double-b tag efficiency (ϵ) and Data/MC efficiency ratio (SF). Uncertainties are both statistical and systematic for the SF and data efficiency, while for the MC efficiency only the statistical uncertainty is reported. Jets with $p_T > 700$ GeV are included in the last bin.

p_T (GeV)	300 - 400	400 - 500	500 - 600	600 - 700
ϵ (Data)	0.79 ± 0.07	0.78 ± 0.09	0.70 ± 0.14	0.66 ± 0.17
ϵ (MC)	0.83 ± 0.01	0.79 ± 0.01	0.77 ± 0.01	0.68 ± 0.01
SF	0.95 ± 0.08	0.98 ± 0.12	0.91 ± 0.18	0.97 ± 0.25

Table A.2: Medium double-b tag efficiency (ϵ) and Data/MC efficiency ratio (SF). Uncertainties are both statistical and systematic for the SF and data efficiency, while for the MC efficiency only the statistical uncertainty is reported. Jets with $p_T > 700$ GeV are included in the last bin.

p_T (GeV)	300 - 400	400 - 500	500 - 600	600 - 700
ϵ (Data)	0.70 ± 0.07	0.70 ± 0.09	0.60 ± 0.12	0.58 ± 0.12
ϵ (MC)	0.75 ± 0.01	0.70 ± 0.01	0.64 ± 0.01	0.55 ± 0.01
SF	0.92 ± 0.09	0.99 ± 0.12	0.94 ± 0.19	1.05 ± 0.21

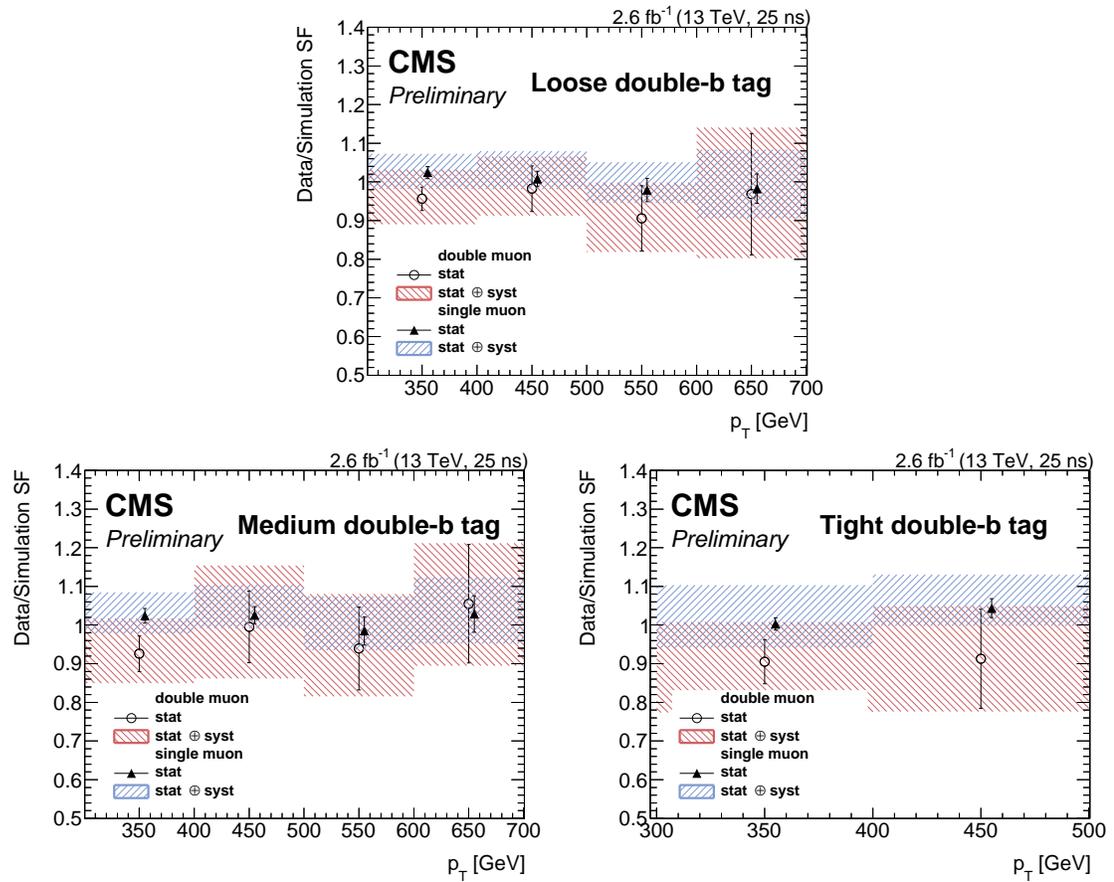


Figure A.5: Data/MC efficiency ratio (SF) for loose, medium and tight double-b tagger requirement obtained with single and double-muon tagged selections. Central values of scale factors are artificially shifted along the x-axis for better visibility. [20]

Table A.3: Tight double-b tag efficiency (ϵ) and Data/MC efficiency ratio (SF). Uncertainties are both statistical and systematic for the SF and data efficiency, while for the MC efficiency only the statistical uncertainty is reported. Jets with $p_T > 500$ GeV are included in the last bin.

p_T (GeV)	300 - 400	400 - 500
ϵ (Data)	0.43 ± 0.04	0.36 ± 0.05
ϵ (MC)	0.47 ± 0.01	0.39 ± 0.01
SF	0.90 ± 0.09	0.91 ± 0.14

Appendix B

Shower Deconstruction algorithm as a high momentum $H \rightarrow b\bar{b}$ tagger

In the early stages of my PhD research, I explored several methods for identifying jets originating from a high momentum Higgs boson to a b quark pair decay. One of the methods uses the *Shower Deconstruction* algorithm. This chapter summarizes preliminary studies of the algorithm using CMS simulated samples. However, as mentioned in Section 4.4.8, this method is not utilized in the final strategy used in this thesis and all the results presented in this chapter are not official CMS results.

B.1 Introduction

As the LHC Run II operates at 13 TeV, we expect to see particles produced with higher energies creating decay products in the form of jets that are more boosted and more collimated in comparison to Run I. Shower Deconstruction (SD) [150] is an algorithm developed for the purpose of discriminating jets from a certain desired physics process, which in our case is a Higgs to b-quark pair decay, versus jets from QCD (and $t\bar{t}$) background in these higher energy boosted scenarios. In this note, we present the first dedicated study on the performance of SD using CMS simulated data as a Higgs decaying to b quark pair jet tagger.

For the samples, we use the simulated events of Radion decaying to a pair of SM Higgs and for background samples we use simulated $t\bar{t}$ jets and QCD events. From Section B.2 up to Section B.4 we use 8 TeV samples listed in Table B.1, and in subsequent Sections in the study we use 13 TeV samples listed in Table B.2.

Table B.1: 8 TeV simulated samples used

Signal	RadionToHH_4b_M800_13TeV_TuneZ2star_8TeV-Madgraph_pythia6/AODSIM/PU_S10_START53_V19-v1
Background	TTJets_MassiveBinDECAY_TuneZ2star_8TeV-madgraph-tauola/AODSIM/PU_S10_START53_V7C-v1
Background	/QCD_Pt-300to470_TuneZ2star_8TeV_pythia6/Summer12_DR53X-PU_S10_START53_V7A-v2/AODSIM
Background	/QCD_Pt-470to600_TuneZ2star_8TeV_pythia6/Summer12_DR53X-PU_S10_START53_V7A-v2/AODSIM
Background	ZPrimeToTTJets_M1000GeV_W10GeV_TuneZ2s tar_8TeV-madgraph-tauola/AODSIM/ PU_S10_START53_V7A-v1

Table B.2: 13 TeV simulated samples used

Signal	/Rad_HHto4b_M800_13TeV/cvnerier-AODSIM-00b0dfdbad9f076cf4490daebae9e3db/USER
Signal	/Rad_HHto4b_M800_13TeV/cvnerier-AODSIM-23a47987951ce9b9de505095a9a0b8a7/USER
Signal	/Rad_HHto4b_M1000_13TeV_PHYS14_25_V1/cvnerier-AODSIM-23a47987951ce9b9de505095a9a0b8a7/USER
Signal	/Rad_HHto4b_M1200_13TeV_PHYS14_25_V1/cvnerier-AODSIM-23a47987951ce9b9de505095a9a0b8a7/USER
Signal	/Rad_HHto4b_M1400_13TeV_PHYS14_25_V1/cvnerier-AODSIM-23a47987951ce9b9de505095a9a0b8a7/USER
Signal	/Rad_HHto4b_M1600_13TeV_PHYS14_25_V1/cvnerier-AODSIM-23a47987951ce9b9de505095a9a0b8a7/USER
Signal	/Rad_HHto4b_M1800_13TeV_PHYS14_25_V1/cvnerier-AODSIM-23a47987951ce9b9de505095a9a0b8a7/USER
Signal	/Rad_HHto4b_M2000_13TeV_PHYS14_25_V1/cvnerier-AODSIM-23a47987951ce9b9de505095a9a0b8a7/USER
Background	ZPrimeToTTJets_M1000GeV_W10GeV_Tune4C_13TeV-madgraph-tauola/Phys14DR-PU20bx25_PHYS14_25_V1-v1/AODSIM
Background	ZPrimeToTTJets_M2000GeV_W20GeV_Tune4C_13TeV-madgraph-tauola/Phys14DR-PU20bx25_PHYS14_25_V1-v1/AODSIM
Background	QCD_Pt-300to470_Tune4C_13TeV_pythia8/Phys14DR- PU20bx25_trkalmb_castor_PHYS14_25_V1-v2/AODSIM
Background	QCD_Pt-470to600_Tune4C_13TeV_pythia8/Phys14DR- PU20bx25_trkalmb_castor_PHYS14_25_V1-v2/AODSIM
Background	QCD_Pt-600to800_Tune4C_13TeV_pythia8/Phys14DR- PU20bx25_trkalmb_castor_PHYS14_25_V1-v1/AODSIM
Background	QCD_Pt-800to1000_Tune4C_13TeV_pythia8/Phys14DR- PU20bx25_trkalmb_castor_PHYS14_25_V1-v2/AODSIM
Background	QCD_Pt-1000to1400_Tune4C_13TeV_pythia8/Phys14DR- PU20bx25_trkalmb_castor_PHYS14_25_V1-v1/AODSIM
Background	QCD_Pt-1400to1800_Tune4C_13TeV_pythia8/Phys14DR- PU20bx25_trkalmb_castor_PHYS14_25_V1-v1/AODSIM
Background	QCD_Pt-2400to3200_Tune4C_13TeV_pythia8/Phys14DR- PU20bx25_trkalmb_PHYS14_25_V1-v1/AODSIM
Background	QCD_Pt-3200_Tune4C_13TeV_pythia8/Phys14DR- PU20bx25_trkalmb_PHYS14_25_V1-v1/AODSIM

This note is organized as follows. Firstly we will briefly elaborate on the method of Shower Deconstruction and study its main behaviors as sanity checks in Section B.2, then in Section B.3 we look at the efficiency of SD, in Section B.5 we study the microjet b-tagging and calculations of tag and fake rate and we explore various jet clustering algorithms for the microjet. In Section B.6 we attempt to optimize SD by varying the input parameters values. Briefly in Section B.7 we discuss how SD shapes the mass distribution of jets. Then in Section B.8 we compare performance of SD

with other jet substructure variables, namely τ_2/τ_1 [151] and q-jet volatility [152].

B.2 Shower Deconstruction

Shower Deconstruction (SD) is a jet substructure analysis algorithm [150] which assigns a jet with a value, denoted by χ , which is the ratio of two probabilities: the probability that, given a signal hypothesis (Higgs to b-quark pair decay), the final state or jet will have a certain substructure configuration, and the probability that, given a background hypothesis (light quarks and gluon decays), the jet has this same substructure configuration.

The analytic functions used to calculate these probabilities follows that of PYTHIA [153] [132] and HERWIG [154]. However, a very much simplified parton shower algorithm is used. The probabilities are calculated by considering "shower histories" which are tree Feynman diagrams for how final state partons could have evolved from a hard scatter. For signal events, the final states are assumed to have come from the decay of a Higgs boson and some additional parton showers from initial state radiation. Whereas for background events, the final states are assumed to have come from a parton shower induced by a high p_T parton and also parton showers starting from initial state radiation (including radiation from the underlying event). The probabilities are then constructed based on these shower histories. For more detailed description of the equations and the deconstruction methods used in calculating the probabilities, one can refer to [150].

Ideally, a jet reconstructed from an actual Higgs to b-quark decay would be calculated to have a higher probability as coming from a desired signal than from background process. Whereas if the jet actually was not reconstructed from a Higgs to b-quark pair decay, it would score a lower probability of it coming from a signal process. In this way, χ would ideally be a higher number for jets reconstructed from a signal process and lower for jets reconstructed from a background process. Therefore, χ can be used as discriminating variable between signal and background.

More specifically, SD firstly takes a large cone size jet (fatjet), in our case Anti- k_T [84] fatjets, and reclusters their constituents into smaller cone size jets (microjets) (and we keep ones with minimum p_T of 15 GeV). Here, we choose the microjets, for most of our study, to be reclustered using k_T

algorithm [155] with distance parameter $R = 0.15$. Each microjet will be assigned a b-tag value from matching it to a secondary vertex (SV) and the SV's are reconstructed using Inclusive Vertex Finder (IVF) algorithm [156]. The b-tag value can be either 1 (successful), 0 (b-tagging not attempted), or -1 (failed). In our study, we attempt to b-tag all microjets. Then, SD takes this collection of microjets, with specific kinematics and b-tag value configuration, as input and calculate the χ value for each fatjet.

In the following subsections, we look at several characteristics of Shower Deconstruction. Namely, we perform sanity checks, see how the distributions of χ discriminate between signal and background, and look also at the result of our microjet b-tagging scheme.

B.2.1 Sanity Check (using 8 TeV samples and Anti- k_T $R=1.2$ fatjets)

We begin by using jet and SD input parameters similar to the study conducted by the authors of SD [150]. Namely, we use Anti- k_T fatjets with distance parameter $R = 1.2$ (AK12) and smaller cone size k_T $R = 0.15$ microjets, with SD input parameters $m_H = 125$ GeV, $\Delta m_H = 10$ GeV, tag rate = 70%, fake rate = 1%. For the rates, we start from idealized values based on [87].

For jet selection, we consider p_T ranges as shown in Figure B.1 (left). The corresponding mass distribution is also shown in Figure B.1 (right). We feed these jets into the SD machinery and we look at the resulting log of χ distribution, shown in Figure B.3, of signal and background samples where the integral of the distributions are normalized to unity. The signal can be clearly seen to have a distinct peak compared to the backgrounds ($t\bar{t}$ and QCD), which shows that, as a proof-of-concept, SD method indeed produce discrimination between signal and background.

However, generally not all the fatjets have a valid χ or in other words some fatjets will be rejected by SD. This means that the jets have configurations such that SD is not able to calculate the probabilities. These jets are assigned $\chi = 0$. Another reason why SD might return a $\chi = 0$ is when the jet mass is outside a certain set mass window, and when the jet transverse momentum is far away from the boosted regime. Figure B.2 shows the p_T and mass distribution of the jets which do pass SD, ie. having $\chi > 0$ (about 17% in this case). We observe that the SD accepts only

boosted jets ($p_T > 200$ for AK12)¹ and tends to accept signal fatjets with mass around 125 GeV. These observations tell us that SD performs consistently with what we would expect of a boosted Higgs tagger.

Next, we will discuss the microjet b-tagging scheme.

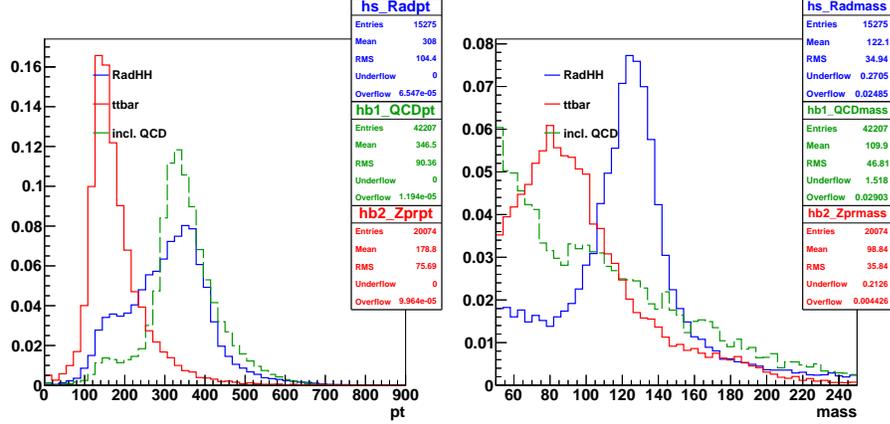


Figure B.1: p_T (left) and mass (right) distributions of signal Radion(M800) $\rightarrow HH$ jets, and backgrounds $t\bar{t}$ jets and QCD jets

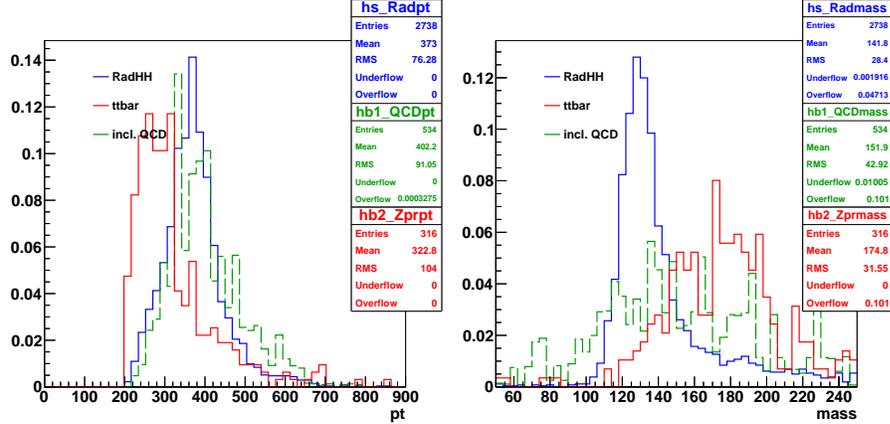


Figure B.2: p_T (left) and mass (right) distributions of signal Radion(M800) $\rightarrow HH$ jets, and backgrounds $t\bar{t}$ jets and QCD jets, which have valid SD discriminator χ value, ie. $\chi > 0$

B.2.2 Microjet b-tagging

To perform optimally as a Higgs to b quark pair tagger, SD requires the b-tag value of each microjets as inputs, and this is not provided internally by SD. Instead, the user must perform the

¹For future studies, it would be interesting to see the efficiency by considering only fatjets with $p_T > 200$ GeV in order to decouple the SD algorithm's internal p_T threshold and other effects.

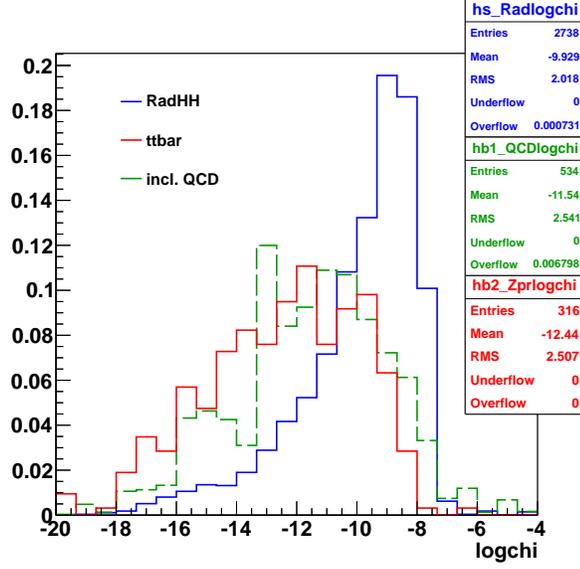


Figure B.3: Log χ distributions of signal Radion(M800) $\rightarrow HH$ jets, and backgrounds $t\bar{t}$ jets and QCD jets

b-tagging on the microjets. For this, we match the microjets to IVF SV based on the $\Delta R(\text{microjet}, \text{IVF SV}) = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ distance. We consider a successful match if this distance is less than the microjet cone size, ie. in this case $\Delta R < 0.15$. If an IVF SV is found to be matched with a microjet then it is assigned a b-tag value of 1, if not then it is assigned a b-tag value of -1 , and if IVF SV matching is not attempted then it will be assigned a b-tag value 0. In our study, we attempt to match all of our microjets and so all microjets will either have 1 or -1 as b-tag value.

Figure B.4 shows the distribution of number of b-tagged microjets. It is consistent with what we expect for Higgs decaying in to two b-quarks, ie. the signal distribution having more 2 b-tagged microjets than background. This also reflects that our microjet b-tagging scheme works as intended.

Figure B.5 shows the distribution of the $\Delta R(\text{Microjet}, \text{MC truth b-quark})$ and $\Delta R(\text{Microjet}, \text{IVF SV})$ of the signal sample. Based on the latter, there are 54% of microjets that are matched to an IVF SV, and based on the former there 67% of microjets that are matched to generator level b-quarks.

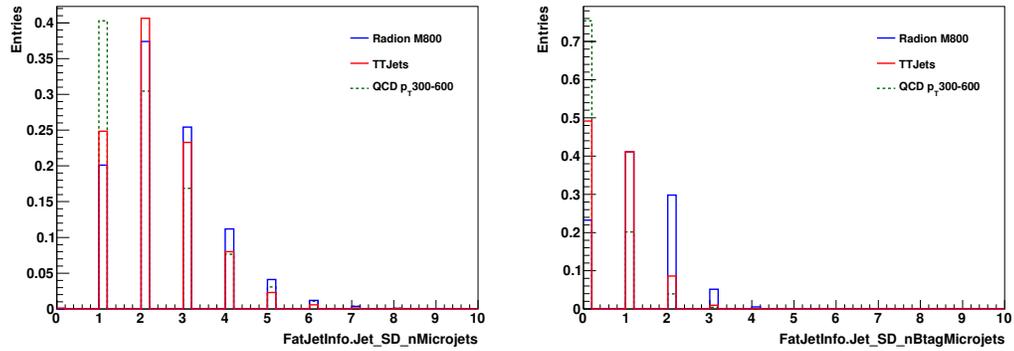


Figure B.4: Number of microjets (left) and number of b-tagged microjets (right) in signal and background fatjets.

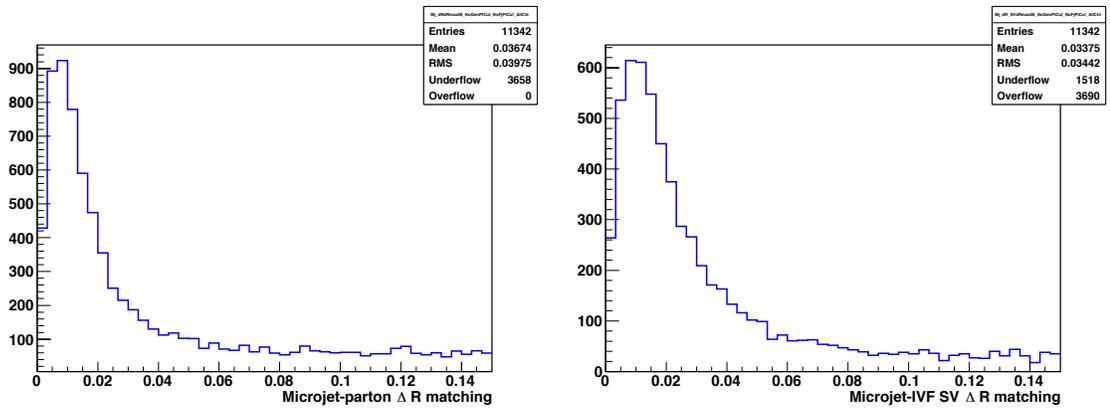


Figure B.5: Distributions of $\Delta R(\text{Microjet}, \text{MC truth b-quark})$ (left) and $\Delta R(\text{Microjet}, \text{IVF SV})$ (right) of signal sample.

B.3 Efficiency (using 8 TeV samples and Anti-kT R=1.2 fat-jets)

In this section we would like to study the inherent efficiency of the SD algorithm. We begin by matching the fatjets with generator level particles and see how many of the matched fatjets has a valid χ value. Here, as a first step, we only consider $t\bar{t}$ as background. We match signal (background) fatjets with generator level Higgs (top) particle then see how much of the sample is accepted by SD, ie. fraction of fatjets that have non-zero χ values. Here, we use the same input parameters as in Section B.2.1

In order to have higher statistics, we switch to using $Z' \rightarrow t\bar{t}$ sample with mass point of 1000 GeV. This gives us $t\bar{t}$ sample with p_T mostly within the boosted region for Higgs to $b\bar{b}$ jet, ie. $p_T > 200$ GeV for R=1.2 fatjets [157]. Figure B.6 shows the $\log \chi$ and p_T distribution of the signal and the (high p_T) background sample to be used in this section.

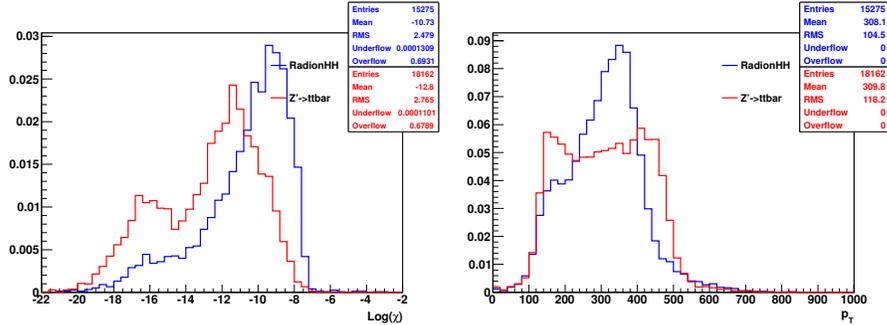


Figure B.6: $\log \chi$ (left) and p_T (right) distributions of signal Radion (M800) $\rightarrow HH$ jets, and backgrounds $Z' \rightarrow t\bar{t}$ jets

We aim to obtain as much pure sample of boosted Higgs (double b) fatjet as possible and so we impose a matching criteria $\Delta R(\text{fatjet, gen}) < 0.05$. We see in the top right plot of Figure B.7 that the resulting matched fatjets are mostly within the boosted regime. From the number of entries in the plots, we also observe that the fatjet reconstruction efficiency is about 43%. After feeding the matched fatjets into the SD algorithm, we observe that SD is able to calculate χ , coincidentally also, about 43% of the matched fatjets. Here, the total efficiency is therefore 18.5%.

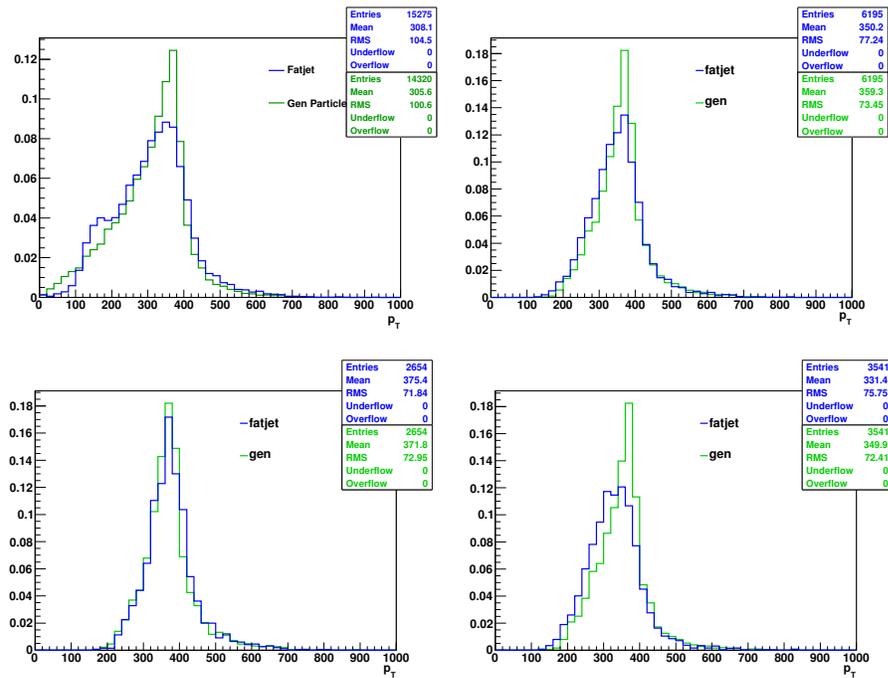


Figure B.7: (Top left) p_T distribution of unmatched fatjet and generator Higgs particle. (Top right) p_T distribution of matched fatjet and generator Higgs particle with $\Delta R(\text{fatjet}, \text{gen Higgs}) < 0.05$. (Bottom left) p_T distribution of matched fatjets and generator Higgs particles with $\Delta R(\text{fatjet}, \text{gen Higgs}) < 0.05$ and accepted by SD. (Bottom right) p_T distribution of matched fatjets and generator Higgs particles with $\Delta R(\text{fatjet}, \text{gen Higgs}) < 0.5$ and rejected by SD.

Furthermore, we also study how the $\log \chi$ distribution changes as we vary the ΔR of the matching between fatjet and generator particle. We select boosted jets, ie. we require jets with $p_T > 200$, and we plot a signal efficiency versus background rejection (1-efficiency) curve based on the $\log \chi$ distributions, shown in Figure B.8, of signal against $Z' \rightarrow t\bar{t}$ background with three different ΔR 's: $\Delta R < 0.05$ (tight), $\Delta R < 0.2$ (medium), $\Delta R < 1.2$ (loose). The numerator is the surviving number of jets given a χ selection, and the denominator is the total number of jets after a matching. We emphasize here that the denominators used for calculating the efficiencies depend on which ΔR is being considered.

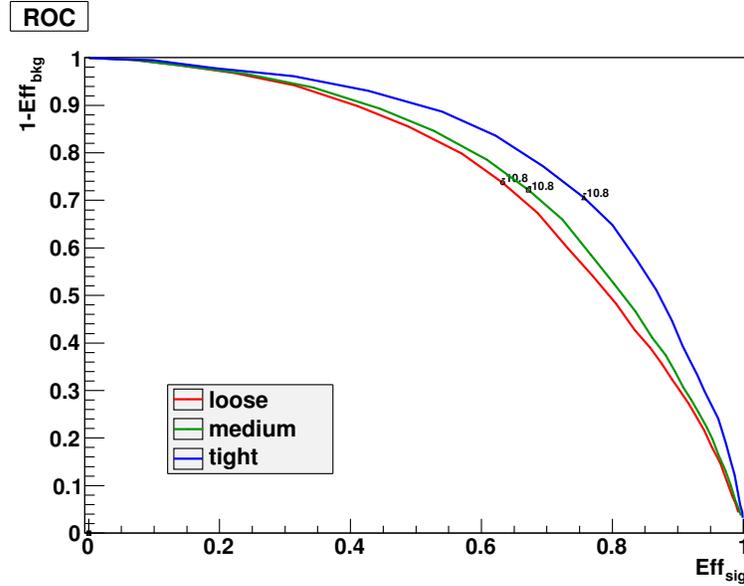


Figure B.8: Log χ performance curve of signal against $Z' \rightarrow t\bar{t}$ background with three different ΔR 's: $\Delta R < 0.05$ (tight), $\Delta R < 0.2$ (medium), $\Delta R < 1.2$ (loose). The performance of cutting at $\log \chi = -10.8$ is plotted on each line.

The efficiency versus rejection curves in Figure B.8 shows that with more stringent matching, ie. the purer the samples, the more fatjets SD can accept out of the total number of fatjets in the corresponding samples. Another thing to note is that the increase of efficiency from the medium to the tight ΔR is much higher than the increase from the loose to tight ΔR .

B.3.1 Number of microjets and b-tagged microjets

To further understand SD algorithm, we look at the number microjets and number of b-tagged microjets within a fatjet and which of those fatjets are accepted and rejected by SD. We can see the distribution of the number of microjets per fatjet in Figure B.9 and the distribution for the number of b-tagged microjets per fatjet in Figure B.10.

As a consistency check, we observe that SD only returns valid χ values ($\chi > 0$) for fatjets that have at least 2 microjets (seen in middle plot in Figure B.9) and that SD tends to accept fatjets that have at least 2 b-tagged microjets from signal, and fatjets that have at least 1 b-tagged in $Z' \rightarrow t\bar{t}$ background (seen in middle plot in Figure B.10).

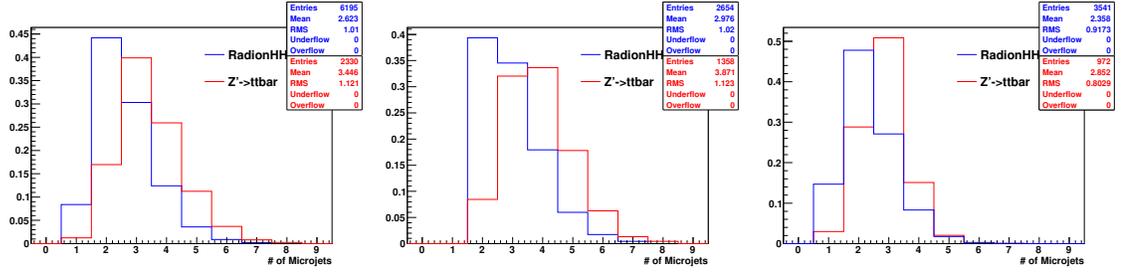


Figure B.9: (Left) Distribution of the number of microjets in signal and background fatjets. (Middle) Distribution of the number of microjets in signal and background fatjets that have a valid SD interpretation or $\chi > 0$. (Right) Distribution of the number of microjets in signal and background fatjets that have an invalid SD interpretation or $\chi = 0$.

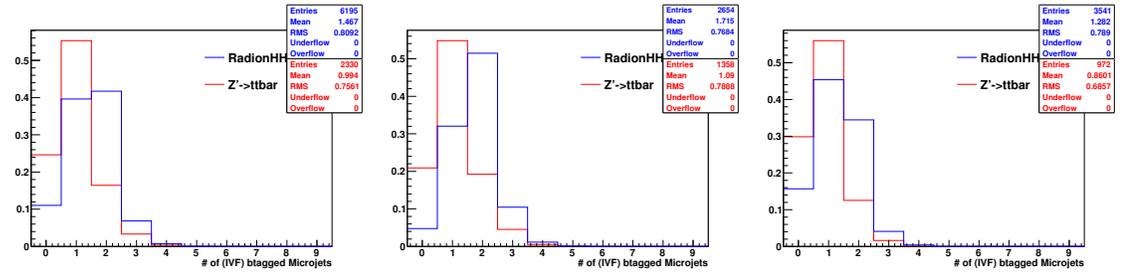


Figure B.10: (Left) Distribution of the number of (IVF) b-tagged microjets in signal and background fatjets. (Middle) Distribution of the number of (IVF) b-tagged microjets in signal and background fatjets that have a valid SD interpretation or $\chi > 0$. (Right) Distribution of the number of (IVF) b-tagged microjets in signal and background fatjets that have an invalid SD interpretation or $\chi = 0$.

B.4 Efficiency (using 13 TeV samples, AK8 fatjets)

We continue our SD studies using the following jet configurations: Anti- k_T with distance parameter $R = 0.8$ fatjets² (AK8) and k_T microjets with distance parameter $R = 0.15$. And, we use the following SD input parameters: $\Delta m_H = 20$ GeV, b-tag rate = 70%, fake rate = 1%. From here on, we use 13 TeV simulated samples.

With a smaller fatjet cone size, we observe that SD now accepts fatjets with $p_T > 300$ GeV (seen in middle plot of Figure B.12 and Figure B.13), which is consistent with the expected boosted regime using $R=0.8$ jets for Higgs jet [157].

We also observe that fatjets with two microjets and two b-tagged microjets fill in mostly at the peak value of the $\log \chi$ distribution for signal, as seen in Figure B.11, consistent with signal fatjet having two b-jets.

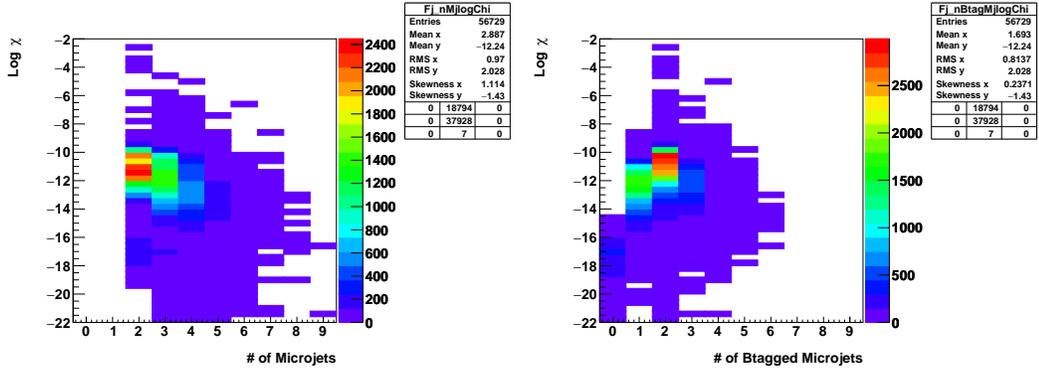


Figure B.11: (Left) 2D histogram of number of microjets versus $\log \chi$ distribution of signal sample. (Right) 2D histogram of number of b-tagged microjets versus $\log \chi$ distribution of signal sample. Signal fatjets are matched with generator level Higgs with $\Delta R < 0.05$ and with $p_T > 300$ GeV selection.

Furthermore, in Figure B.14 we can see how the p_T distribution progressively change³ as we perform fatjet truth matching (with $\Delta R < 0.05$) and subsequently apply Shower Deconstruction. By the number of entries, we observe that up to 60% of the matched fatjets has a valid χ value and they are mostly in the boosted region.

²We switch to this cone size to synchronize with the standard jet cone size used in CMS

³Again, as with the similar 8 TeV study, it would be interesting for a future study to decouple the SD algorithm's internal p_T cut from other effects by considering only fatjets with $p_T > 300$ GeV initially. Similarly, for microjet and b-tagged microjet efficiencies.

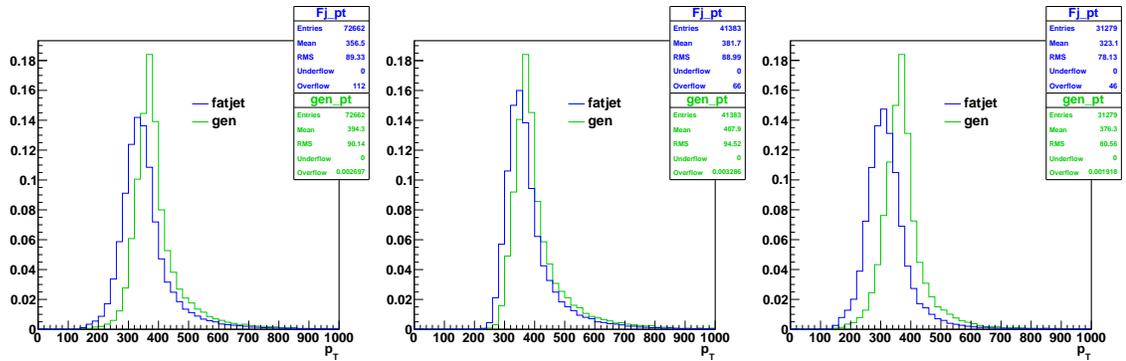


Figure B.12: (Left) Signal p_T distribution of matched fatjets with generator Higgs, with $\Delta R < 0.05$. (Middle) p_T distribution of those that are accepted by SD. (Right) p_T distribution of those that are rejected by SD.

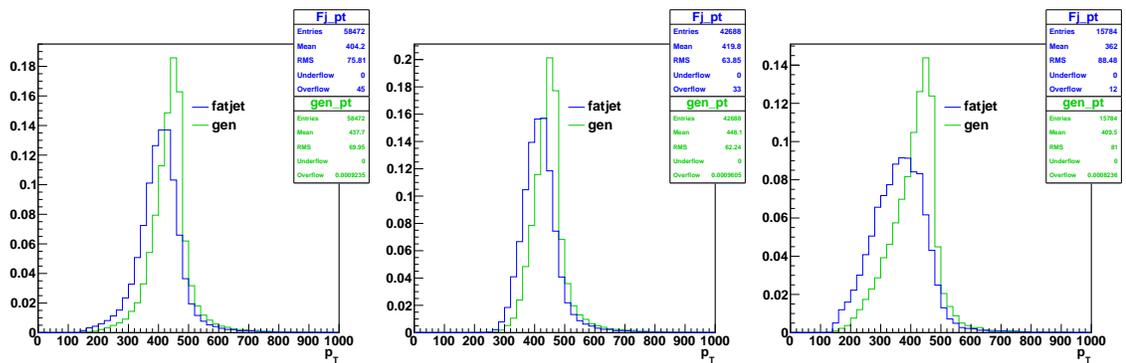


Figure B.13: (Left) $Z' \rightarrow t\bar{t}$ background p_T distribution of matched fatjets with generator top quarks, with $\Delta R < 0.05$. (Middle) p_T distribution of those that are accepted by SD. (Right) p_T distribution of those that are rejected by SD.

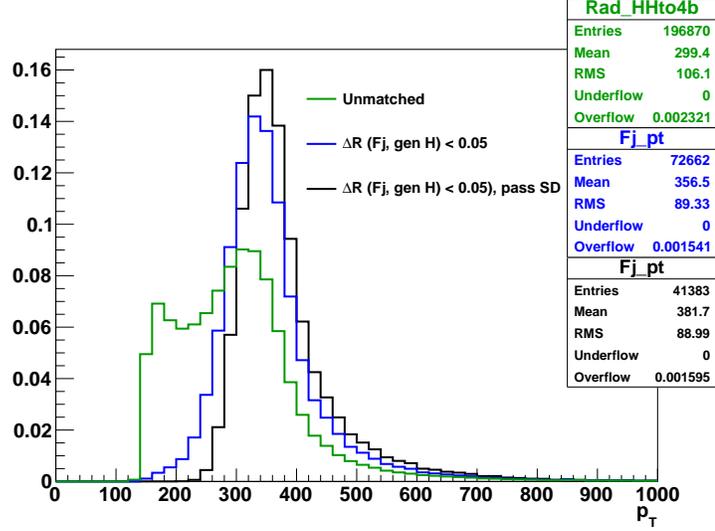


Figure B.14: Signal p_T distribution before matching with generator Higgs, after matching, and after matching and applying SD algorithm (using $\Delta m_H = 20$ GeV).

B.4.1 SD efficiency using selected fatjets

Earlier in this Section, we have observed that SD performs better with purer signal sample (ΔR (fatjet, gen Higgs) < 0.05), accepts boosted ($p_T > 300$) fatjets with at least 2 microjets, and it tends to accept more the ones with at least 2 b-tagged microjets for signal. Now we look at the efficiency of SD given that the fatjets meet all these conditions.

Figure B.15 shows the p_T distribution of fatjets passing the selection criteria and Figure B.16 shows the efficiency versus rejection curve (with three values of Δm_H). From the latter plot we observe that SD is able to calculate about 90% of the fatjets using $\Delta m_H=30$ GeV. Moreover, this value gives the most optimal efficiency in comparison to $\Delta m_H=20$ GeV and $\Delta m_H=10$ GeV. We will look at the optimization more in Section B.6. Before doing so, we look at the microjet b-tagging rate.

B.5 Microjet b-tag rate and fake rate

In this section, we calculate the microjet b-tag rate and fake rates. As mentioned in the latter part of Section B.2, the microjet b-tagging are essentially how well it is matched with a IVF SV using

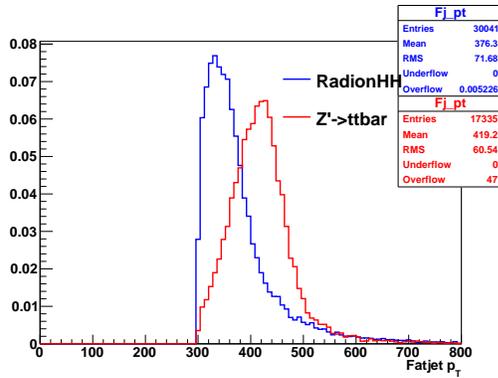


Figure B.15: p_T distribution of matched fatjets ($\Delta R(\text{fatjet}, \text{gen}) < 0.05$), have minimum p_T of 300 GeV and those that have at least 2 microjets and at least 2 b-tagged microjets.

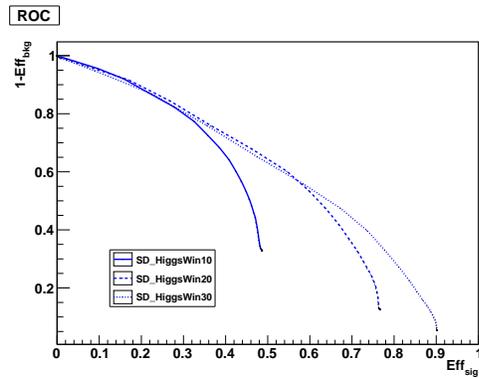


Figure B.16: Performance curve or ROC of $\log \chi$ distribution of signal against $Z' \rightarrow t\bar{t}$ background with various Higgs mass windows as input parameter to SD, where the denominator is the number of entries in Figure B.15.

ΔR . We first calculate using the default k_T jet algorithm [155] and later also look at how fake rate changes with different jet algorithm to cluster the microjets.

B.5.1 Tag rate and fake rate using k_T microjets

The calculations are conducted as follows. For the tag rate, we select signal fatjets (matched with generator level Higgs) and match the microjets to generator level b-quarks (and later on as a second calculation we match the microjets to generator level b-hadrons). Then we count the number of matched microjets that are successfully b-tagged, ie. matched with a secondary vertex, with an extra strict selection of ΔR (microjet, IVF SV) < 0.04 , and hence obtaining the fraction of microjets that are correctly b-tagged. For the fake rate, we select QCD background fatjets with u/d/s/g - flavors, where the flavors are determined by ghost hadron clustering method [158], and match the microjets with the corresponding u/d/s/g generator level particles and then count the number of matched microjets that are b-tagged hence obtaining the fraction the microjets that are incorrectly b-tagged.

For the tag rate calculation, matching the microjets with b-quarks we obtain the tag rate to be about 70%, shown by the number of entries in Figure B.17 where p_T distribution of the microjets are shown. Whereas matching the microjets with b-hadrons we obtain the tag rate to be about 60%, shown by the number of entries in Figure B.18.

For the fake rate, we perform the calculations for $Z' \rightarrow t\bar{t}$ and QCD samples. For the latter we calculate for two p_T bins, 300-470 GeV and 470-600 GeV. We find that the tag rate is about 18% for $Z' \rightarrow t\bar{t}$, 26% for the lower p_T bin QCD and 28% for the higher p_T bin QCD, which can be inferred from the number of entries in the left and middle plots of Figure B.19, Figure B.20 and Figure B.21 respectively.

To further confirm on these fake rate calculation results we look at the number of reconstructed IVF secondary vertices in u/d/s/g flavored fatjets, shown in Figure B.22. We find that there are about 38%, 43%, and 49% of u/d/s/g flavored fatjets which have at least one IVF secondary vertices for $Z' \rightarrow t\bar{t}$, QCD300-470, and QCD470-600 samples respectively.

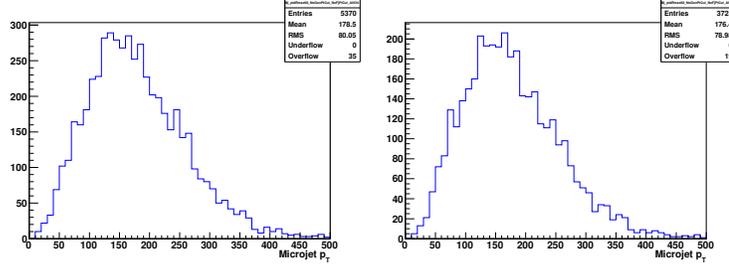


Figure B.17: (Left) B-quark matched microjets p_T distribution of signal sample. (right) B-quark matched microjet p_T distribution of signal sample that are IVF b-tagged.

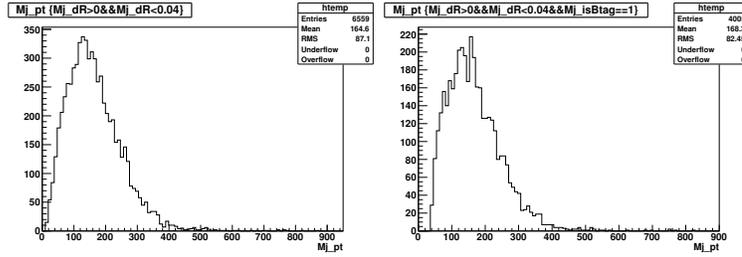


Figure B.18: (Left) B-hadron matched microjets p_T distribution of signal sample. (right) B-hadron matched microjet p_T distribution of signal sample that are IVF b-tagged.

The high number of incorrectly b-tagged microjets can be explained in the following way. Considering u/d/s/g fatjets in the background samples have on average 2 microjets (based on what is shown in Figure B.23), if we suppose that there are 100 fatjets then there are a total of 200 microjets and about $\sim 40\%$ (40) of the fatjets have at least one IVF SV, and about $\sim 20\%$ (40) of the microjets are incorrectly b-tagged. Here we have the number of fatjets with at least one IVF SV is equal to the number of incorrectly b-tagged microjets. This implies that on average there is one incorrectly b-tagged microjets for every fatjets with a least one IVF SV. Hence, we suspect that the high number of IVF SV reconstructed in u/d/s/g fatjets is ultimately the reason why the fake rate is relatively high.

Furthermore, we graph the bin by bin ratio of the p_T distribution of middle plots and left plots of Figure B.19, Figure B.20, and Figure B.21 and we observe that the fake rates are p_T dependent, as can be seen in the plots on the right in Figure B.19, Figure B.20, and Figure B.21.

We expect that reducing the actual fake rate will be increase the overall performance of SD but unfortunately we have not determined a b-tagging method for these microjets that will result in a

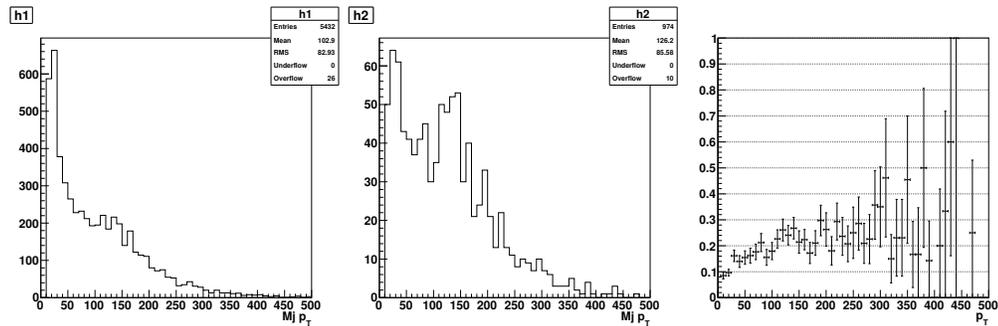


Figure B.19: (Left) p_T distribution of matched $Z' \rightarrow t\bar{t}$ microjets. (Middle) p_T distribution of matched $Z' \rightarrow t\bar{t}$ microjets that are IVF b-tagged. (Right) fake rate p_T dependance obtained by graphing the ratio of middle plot against the left plot.

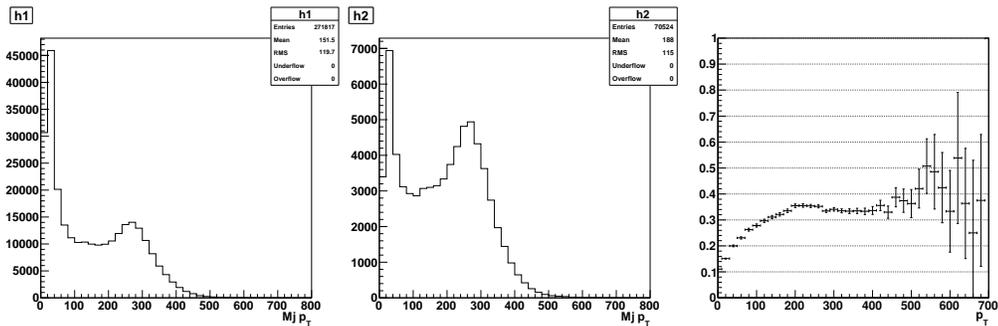


Figure B.20: (Left) p_T distribution of matched QCD300-470 microjets. (Middle) p_T distribution of matched QCD300-470 microjets that are IVF b-tagged. (Right) fake rate p_T dependance obtained by graphing the ratio of middle plot against the left plot.

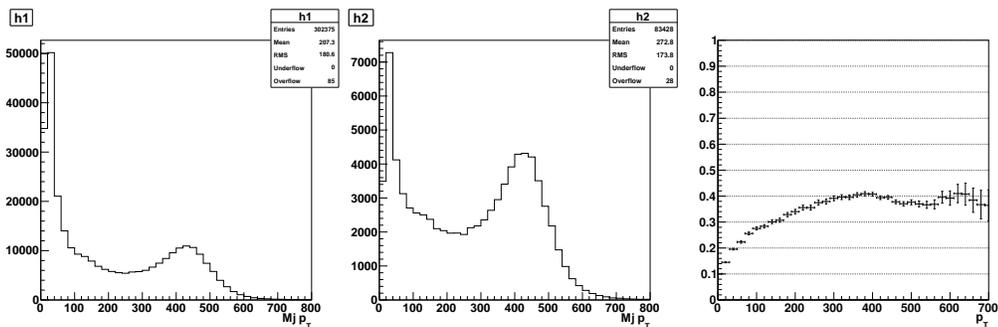


Figure B.21: (Left) p_T distribution of matched QCD470-600 microjets. (Middle) p_T distribution of matched QCD470-600 microjets that are IVF b-tagged. (Right) fake rate p_T dependance obtained by graphing the ratio of middle plot against the left plot.

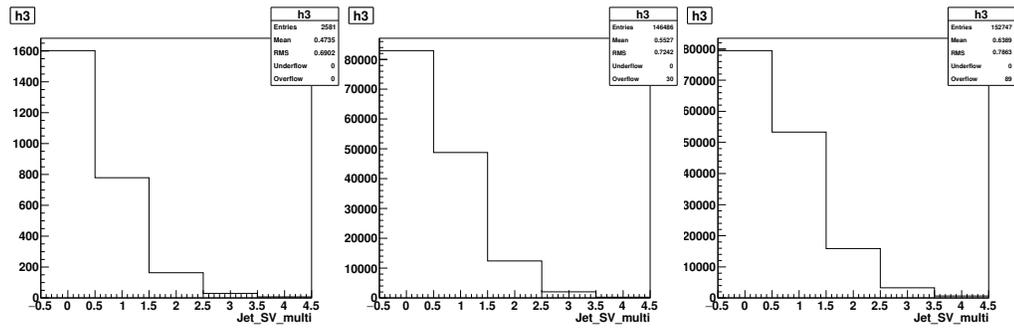


Figure B.22: Number of IVF SV among u/d/s/g flavoured fatjets in $Z' \rightarrow t\bar{t}$ sample (left), QCD300-470 sample (middle), and QCD470-600 sample (right)

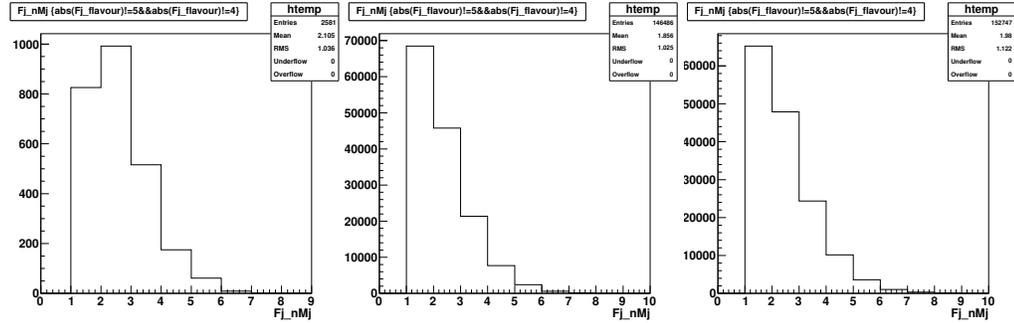


Figure B.23: Number of microjets among u/d/s/g flavoured fatjets in $Z' \rightarrow t\bar{t}$ sample (left), QCD300-470 sample (middle), and QCD470-600 sample (right)

lower fake rate. We leave this issue for future studies.

The results for the calculation of tag rates, fake rates (and rate of IVF SV) found in the microjets is summarized Table B.3, B.4 respectively.

Matched particle	b-quark	B-hadron
Tag rate	70%	60%

Table B.3: The tag rate calculation results.

Sample	$Z' \rightarrow t\bar{t}$	QCD $300 < p_T < 470$ GeV	QCD $470 < p_T < 600$ GeV
Fake rate	18%	26%	28%
Rate for IVF SV	38%	43%	49%

Table B.4: The fake rate and rate of IVF SV found in the microjets.

B.5.2 Comparing k_T , AK and CA microjets

We now look at the effect of using three different jet algorithms to cluster the microjets, on the b-tag fake rate. The algorithms considered are k_T [155], Anti- k_T (AK) [84], and *Cambridge/Aachen* (CA) [159]. Here we use only QCD470-600 to represent the background sample.

We observe that there are no significant changes in fake rate nor in the number of IVF SV between k_T , AK, CA algorithms, as can be seen in Figures B.24 and B.25. Therefore we expect no improvement in performance using other jet clustering algorithms to cluster the microjet. We therefore choose to consider only k_T clustering algorithms for the rest of our study.

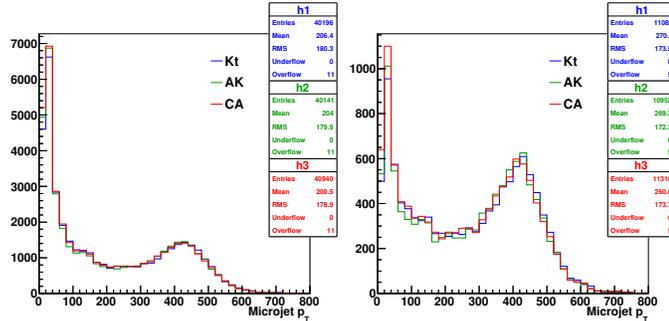


Figure B.24: (Left) p_T distribution QCD470-600 microjets clustered from u/d/s/g flavoured fatjets, using various jet algorithms. (Right) p_T distribution QCD470-600 b-tagged microjets clustered from u/d/s/g flavoured fatjets, using various jet algorithms.

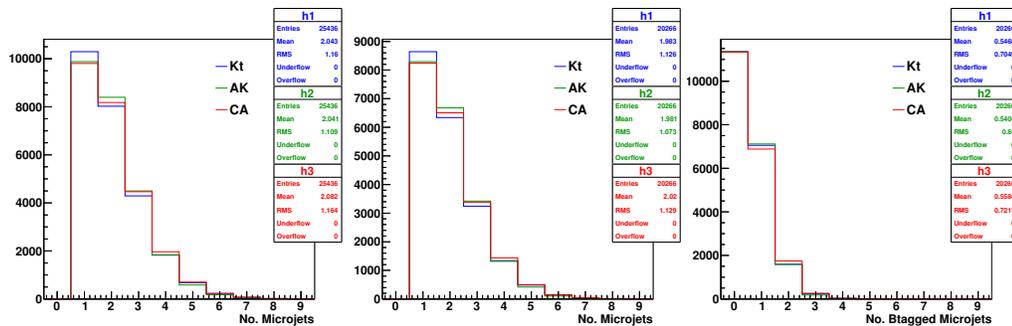


Figure B.25: (Left) Number of microjets from fatjets in QCD470-600 sample. (Center) Number of microjets from u/d/s/g fatjets in QCD470-600 sample. (Right) Number of b-tagged microjets from u/d/s/g flavored fatjets in QCD470-600 sample.

B.6 Optimizing performance

After the studies presented in previous sections, we now look to optimize SD by studying how its performance changes as we vary the input parameters. The parameters we vary include the following: the microjet b-tagging tag-rate, the fake rate, the Higgs mass window (Δm_H) and lastly the microjet cone size. Through out most of this section we select fatjets with pruned mass between 80 and 150 GeV, when comparing Δm_H we choose pruned mass between 50 and 200 GeV. For all cases we consider two different p_T regions. One region is between 300 GeV and 470 GeV (lower p_T) and the other is between 470 GeV and 600 GeV (higher p_T). We also define two different QCD backgrounds where one is QCD from single b quark (QCDsingle-b) and the other is QCD from gluon splitting to a pair of b quarks (QCDgsp).

We produce signal efficiency versus background rejection (or 1–efficiency) curves to quantify the performances, given a set of input parameter values. Depending on the p_T categories, the denominators used to calculate the efficiencies are the number of fatjets that pass the selection criteria we determined in the previous paragraph, namely the ones that pass the mass and p_T selections.

B.6.1 Microjet b-tag rate

We found in Section B.5 that tag rate calculation using b-quark matching is about 70% and using b-hadron matching is about 60%. Figure B.26 shows the performance curve and we observe that

there are no significant improvements by changing the value from 70% to 60%. And so we conclude that 70% is already the optimal choice.

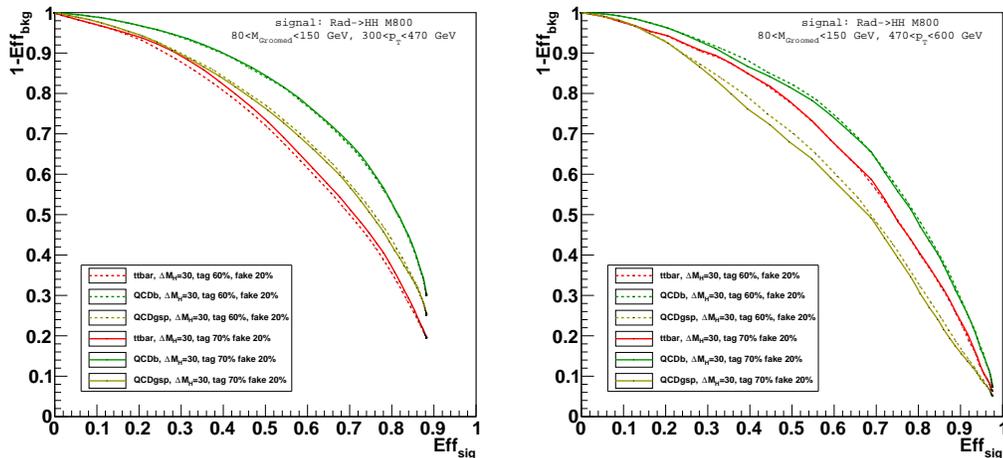


Figure B.26: Signal efficiency versus background rejection curve comparing different tag rate input parameter for lower p_T (left) and higher p_T (right) fatjets.

Next we vary the microjet b-tag fake rate.

B.6.2 Microjet b-tag fake rate

Based on calculations in Section B.5, we vary the fake rate input parameter from 1% (default) to 20%. We observe a slight improvement in the performance, shown in Figure B.27, particularly in the middle to low efficiency region of the lower p_T fatjets. We also observe that despite the fake rate being relatively high, SD's performance increases using the more accurate value of fake rate as an input parameter.

The increase in performance is consistently seen against all backgrounds. We conclude that fake rate of 20% is a more optimal choice.

B.6.3 Higgs mass window

Next, we look at varying the Higgs mass window for values 20 GeV, 30 GeV and 50 GeV while considering fatjets with pruned mass of between 50 and 200 GeV and kT R=0.20 microjets. Setting Δm_H to be 30 GeV in comparison to 20 GeV increases the performance particularly for the higher

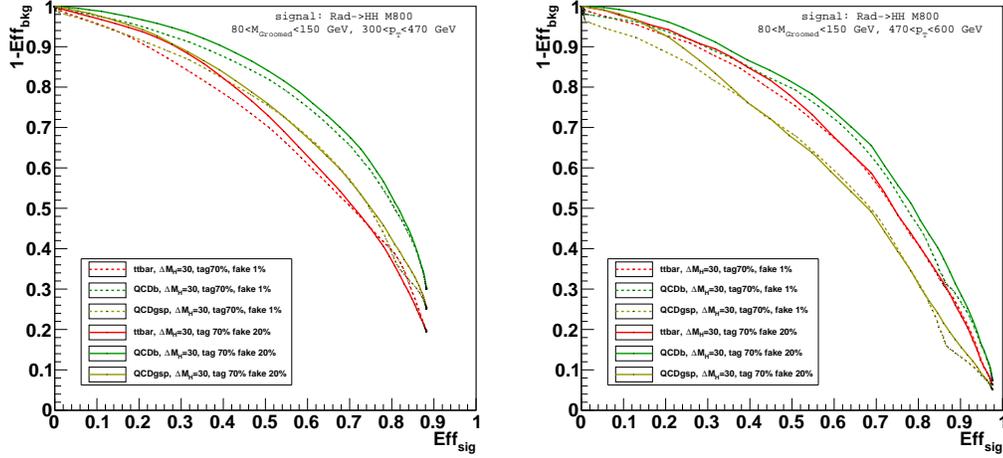


Figure B.27: Signal efficiency versus background rejection curve comparing different b-tag fakerate input parameter for lower p_T (left) and higher p_T (right) fatjets.

signal efficiency region, and it extends the maximum signal efficiency from around 0.7 to 0.9 for the lower p_T fatjets, as shown in the left plot of Figure B.28. This is consistent with Figure B.16 in Section B.4.1. There, it was clear that $\Delta m_H = 10$ GeV performs worst, so we do not reconsider it here.

When we extend the mass window parameter to 50 GeV, the performance becomes worse compared to 30 GeV, as shown in Figure B.29. The maximum efficiency for 50 GeV expectedly goes higher than 30 GeV, however the overall performance for most of the efficiency regions are worse.

These observations in performances are consistently seen against all backgrounds for both lower and higher p_T fatjets. We conclude that Higgs mass window of 30 GeV is the optimal choice.

B.6.4 Microjet cone size

Lastly, we look at how the performance of SD is affected by varying the cone size of the microjet and we consider three values of cone sizes: $R = 0.1, 0.15, 0.2$. We keep microjet b-tag rate, fake rate, and Higgs mass window parameters constant and set them to be the optimal values as we have determined in previous sections.

Microjet cone size of $R = 0.15$ is consistently more optimal in comparison to $R = 0.10$, shown in Figure B.30. Whereas, $R = 0.20$ is overall more optimal in comparison to $R = 0.15$ as can be seen in

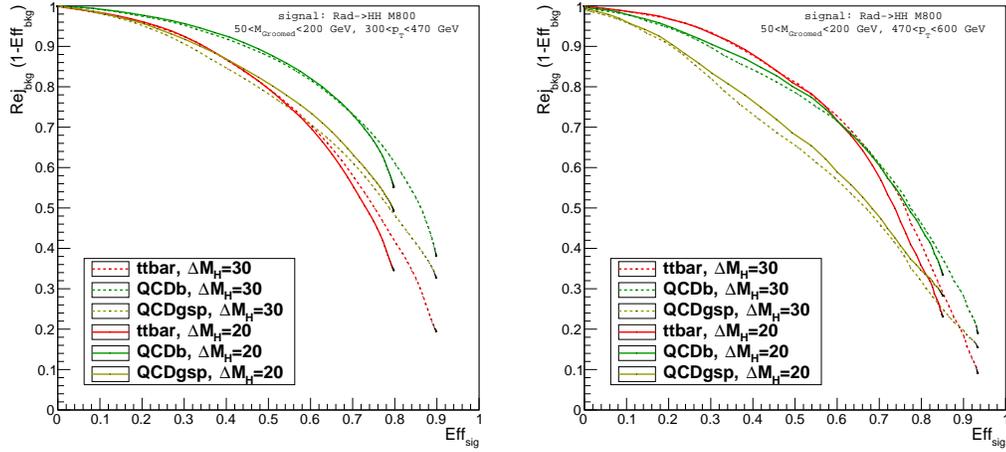


Figure B.28: Signal efficiency versus background rejection curve comparing different higgs mass window input parameters of 20 GeV and 30 GeV for lower p_T (left) and higher p_T (right) fatjets.

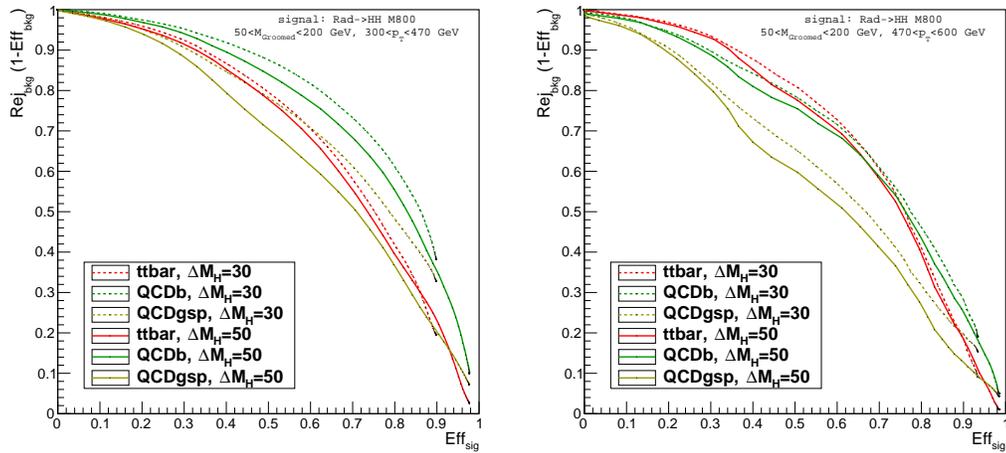


Figure B.29: Signal efficiency versus background rejection curve comparing higgs mass window input parameters of 30 GeV and 50 GeV for lower p_T (left) and higher p_T (right) fatjets.

Figure B.31. In particular this can be seen in the for lower p_T fatjets in Figure B.31 where signal efficiency is slightly extended. Overall, we conclude that the optimal microjet cone size is 0.20.

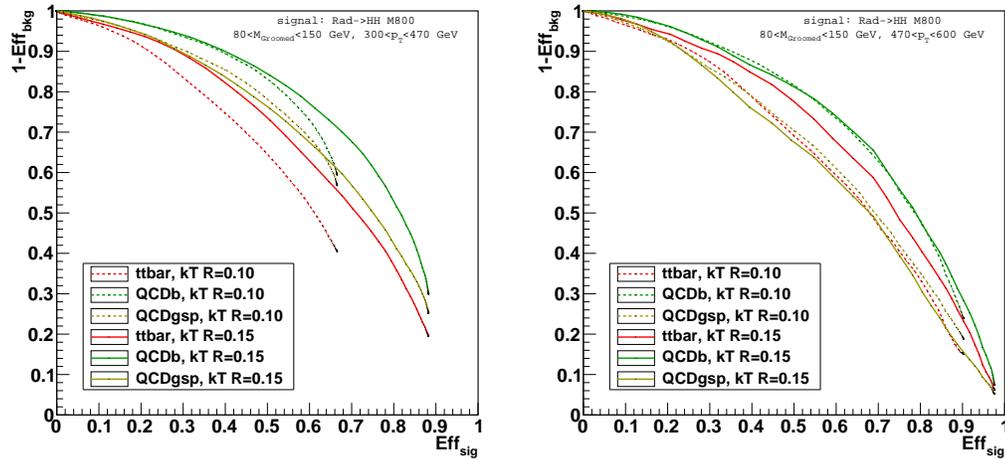


Figure B.30: Signal efficiency versus background rejection curve comparing microjet cone size $R=0.10$ with $R=0.15$ for lower p_T (left) and higher p_T (right) fatjets.

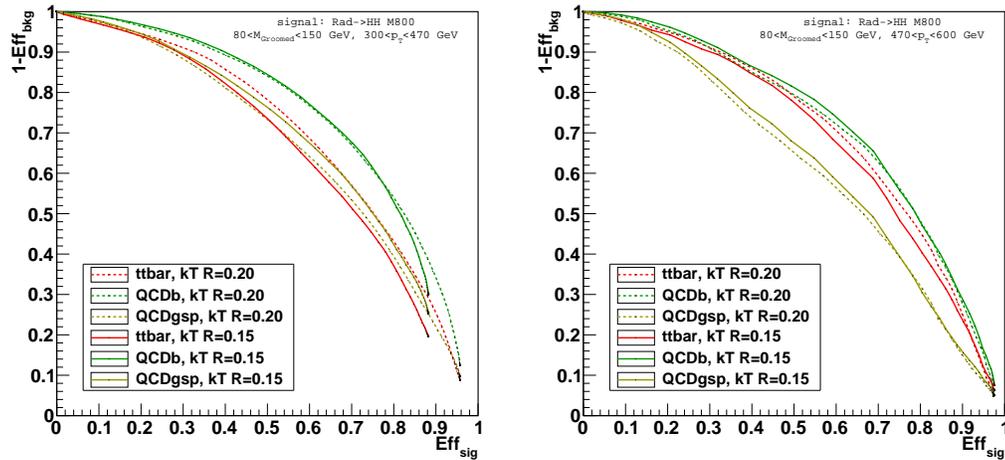


Figure B.31: Signal efficiency versus background rejection curve comparing microjet cone size $R=0.20$ with $R=0.15$ for lower p_T (left) and higher p_T (right) fatjets.

The summary of the configuration variables studied and the resulting optimal values are summarized in Table B.5

Table B.5: Configurations varied for optimization, and values looked at in this study. The values underlined are the observed to be the optimal values. For the microjet cone size the optimal value depends on the background.

Parameters	Values
Microjet b-tag rate	<u>70%</u> , 60%
Microjet b-tag fake rate	1%, <u>20%</u>
Higgs mass window (Δm_H)	20 GeV, <u>30 GeV</u> , 50 GeV
Microjet jet cone size	0.10, 0.15, <u>0.20</u>

B.6.5 Discussion on discrimination between different backgrounds

It is worth mentioning that for the higher p_T fatjets, $Z' \rightarrow t\bar{t}$ background is consistently better discriminated in comparison to QCDgsp background through out the optimization study. Whereas for lower p_T fatjets, QCD in general is expectedly better discriminated in comparison to $Z' \rightarrow t\bar{t}$. We have not further investigated why SD behave in this way and we leave this for future studies.

B.7 Mass distribution shaping

In this section, we briefly discuss how mass distribution of the jets is affected by SD and we compare with τ_2/τ_1 [151] and q-jet volatility [152]. Here, we use all the samples listed in Table B.2, except for the Radion sample listed in the first row⁴.

We know that SD takes mass of the signal particle, in this case the Higgs boson, as an input parameter. Therefore we would like to know whether there are any dramatic changes to the distribution of the mass once SD algorithm is applied to the jets.

To see how mass distribution is affected, we select fatjets with $p_T > 300$ GeV and mass range of between 80 GeV and 155 GeV, then we plot four mass distributions of signal jets and background jets: without any further cuts, with an $\chi > 0$ cut (which means that SD has been able to calculate a χ value), with a $\tau_2/\tau_1 < 0.5$, with a q-vol < 0.07 cut⁵. Figure B.32 shows the resulting plots for QCD (inclusive), $Z' \rightarrow t\bar{t}$ background as well as signal.

⁴This is 50ns Radion sample, where as all others listed in this Table B.2 are 25ns samples

⁵Note for future studies, a better comparison would be to choose cuts that would result in roughly the same value of signal or background efficiencies across these different methods.

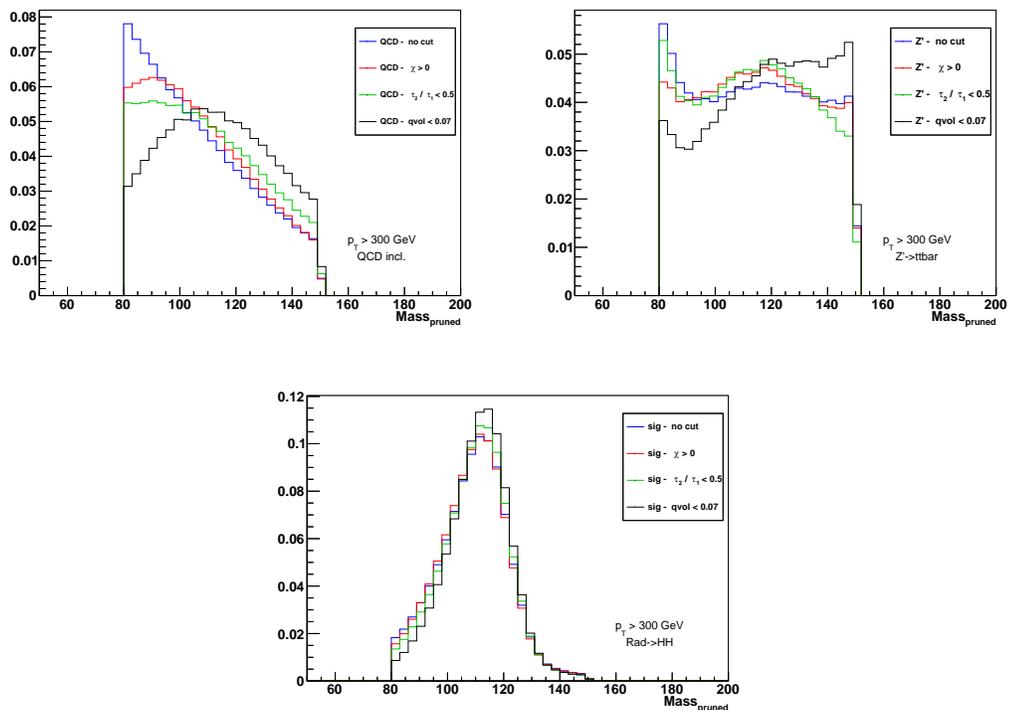


Figure B.32: Mass distribution of signal (bottom) and QCD (top left) and $Z' \rightarrow t\bar{t}$ (top right) background before and after SD, τ_2/τ_1 , q-vol selections.

From the top plots in Figure B.32, we can see that background mass distributions are not heavily shaped by applying SD and τ_2/τ_1 . Whereas q-vol seem to cause the most shaping to the mass distribution of background jets. For signal mass distribution, shown in the bottom plot of Figure B.32, all three variables seem to have little affect to the mass distribution. We conclude that, SD does not cause any dramatic change to the mass distribution in comparison to τ_2/τ_1 and q-vol.

B.8 Performance comparison

Using the optimized input parameters, in this section we compare performance of SD with the following jet substructure variables, τ_2/τ_1 [151] and q-jet volatility [152]. Unlike in Section B.6, we treat QCD as inclusive rather than splitting them into categories. In this comparison study we use all the samples listed in Table B.2, except for the Radion sample listed in the first row⁶.

⁶This is 50ns Radion sample, where as all others listed in this Table B.2 are 25ns samples

Given now we have a new set of (optimized) input parameters, we consider a new set of fatjet selections. Having to use Higgs mass window of 30 GeV, we select jets with pruned mass between 95 and 155 GeV. And for the transverse momentum we select p_T from 300 GeV up to 10000 GeV. As a last criteria, in order to have a direct comparison between the substructure variables, we select jets with valid SD discriminator value, ie. $\chi > 0$. The p_T distributions of signal and backgrounds passing these selection are shown in Figure B.33.

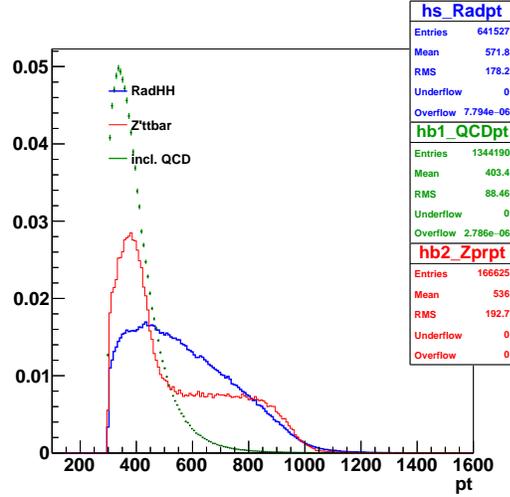


Figure B.33: p_T distribution of signal and backgrounds after selection of $p_T > 300$ GeV, $95 < m_{pruned} < 155$, $\chi > 0$.

We further divide the p_T into 7 bins: 300-400 GeV, 400-500 GeV, 500-600 GeV, 600-800 GeV, 800-1000 GeV, 1000-10000 GeV, and 300-10000 GeV. The following 7 plots in Figure B.34 show the comparison between the three jet substructure discrimination variables.

As shown in Figure B.34 and Figure B.35, SD generally outperforms τ_2/τ_1 and q-jet volatility in all of the p_T categories and for all backgrounds. In addition, based on these plots we pick three working points that represents signal efficiencies of 90%, 70%, 50% and determine the background rejection values for SD, τ_2/τ_1 and q-jet volatility. We put the resulting values in Table B.6 (for QCD inclusive background) and Table B.7 (for $Z' \rightarrow t\bar{t}$) according to the p_T bins. From these tables we consistently observe that SD generally outperforms the other two jet substructure variables, in terms of the background rejection level.

To see the performance comparison from another perspective, we also looked at fixing the

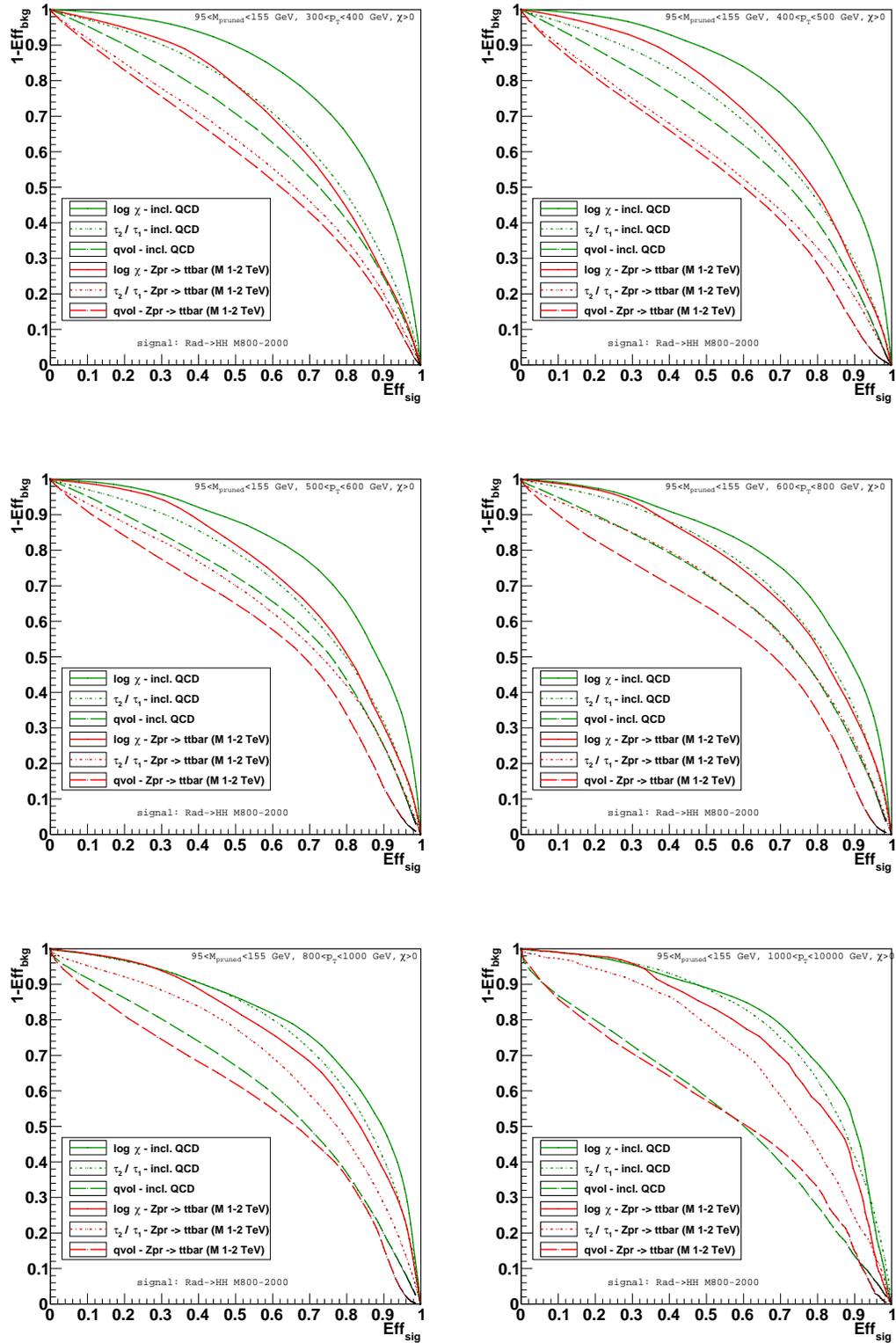


Figure B.34: Signal efficiency versus background rejection (1-efficiency) of SD in comparison to τ_2/τ_1 and q-jet volatility, in p_T bins.

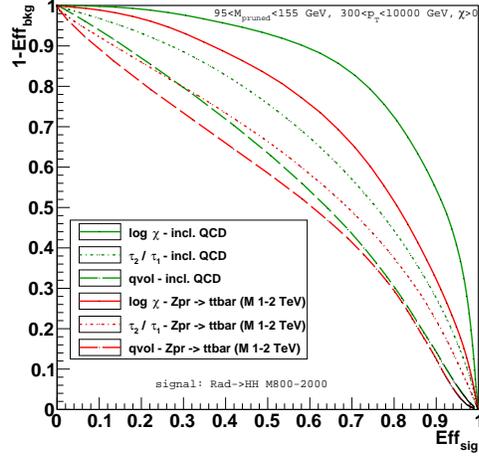


Figure B.35: Signal efficiency versus background rejection (1-efficiency) of SD in comparison to τ_2/τ_1 and q-jet volatility, inclusive in p_T ranging from 300 GeV to 10000 GeV

p_T (GeV)	$\varepsilon_s = 90\%$			$\varepsilon_s = 70\%$			$\varepsilon_s = 50\%$		
	SD	τ_2/τ_1	q-vol	SD	τ_2/τ_1	q-vol	SD	τ_2/τ_1	q-vol
300-400	46.5	30	24.5	76	61	53	89.5	79	70.5
400-500	46	29	22.5	76	58.5	52.5	89	77	69.5
500-600	46.5	31.5	25	76	62	57	88	79.5	72.5
600-800	47	35.5	24.5	75	67	57	87	83	73
800-1000	49	40.5	20	75.5	71.5	49.5	86.5	86	67
1000-10000	50	41.5	13.5	88.5	75	40	85	85	58
300-10000	54	28	14	83	57	43.5	93	76	63.6

Table B.6: QCD inclusive background rejection (%) for SD, τ_2/τ_1 and q-jet volatility using working points with signal efficiencies being, $\varepsilon_s = 90\%$, $\varepsilon_s = 70\%$, and $\varepsilon_s = 50\%$. The values are obtained based on the plots in Figure B.34 and Figure B.35.

p_T (GeV)	$\varepsilon_s = 90\%$			$\varepsilon_s = 70\%$			$\varepsilon_s = 50\%$		
	SD	τ_2/τ_1	q-vol	SD	τ_2/τ_1	q-vol	SD	τ_2/τ_1	q-vol
300-400	25.5	20	17.5	58	46	43	79	63.5	60
400-500	27.5	19.5	11.5	61	43.5	41	80	60.5	58
500-600	30.5	25	13	64	53.5	48	81.5	70	65
600-800	33.5	26	13	65	56	48	82	73.5	64
800-1000	37.5	28	15.5	68	59	46.5	82.5	77.5	62
1000-10000	36.5	24	15	70	58.5	43	84	79	57.5
300-10000	32	21.5	12	65.5	49	41	83	66	58

Table B.7: $Z' \rightarrow t\bar{t}$ background rejection (%) for SD, τ_2/τ_1 and q-jet volatility using working points with signal efficiencies being $\varepsilon_s = 90\%$, $\varepsilon_s = 70\%$, and $\varepsilon_s = 50\%$. The values are obtained based on the plots in Figure B.34 and Figure B.35.

background rejection and see how much signal efficiency we get. The result is shown in Figure B.36 for QCD background and B.37 for $Z' \rightarrow t\bar{t}$ background. For QCD background, we observe that at 10% rejection level SD performs slightly better at $p_T < 800$ GeV and τ_2/τ_1 performs better than SD at higher p_T . At 50% rejection level, SD consistently performs best. At 90% rejection level, SD performs better than τ_2/τ_1 by much larger margin at the lower p_T bins and at higher bins they the signal efficiency levels are comparable. At all rejection levels, q-jet performs the worst. For $Z' \rightarrow t\bar{t}$ background, SD consistently performs best for most of cases. Overall, SD generally outperforms the other two jet substructure discriminants.

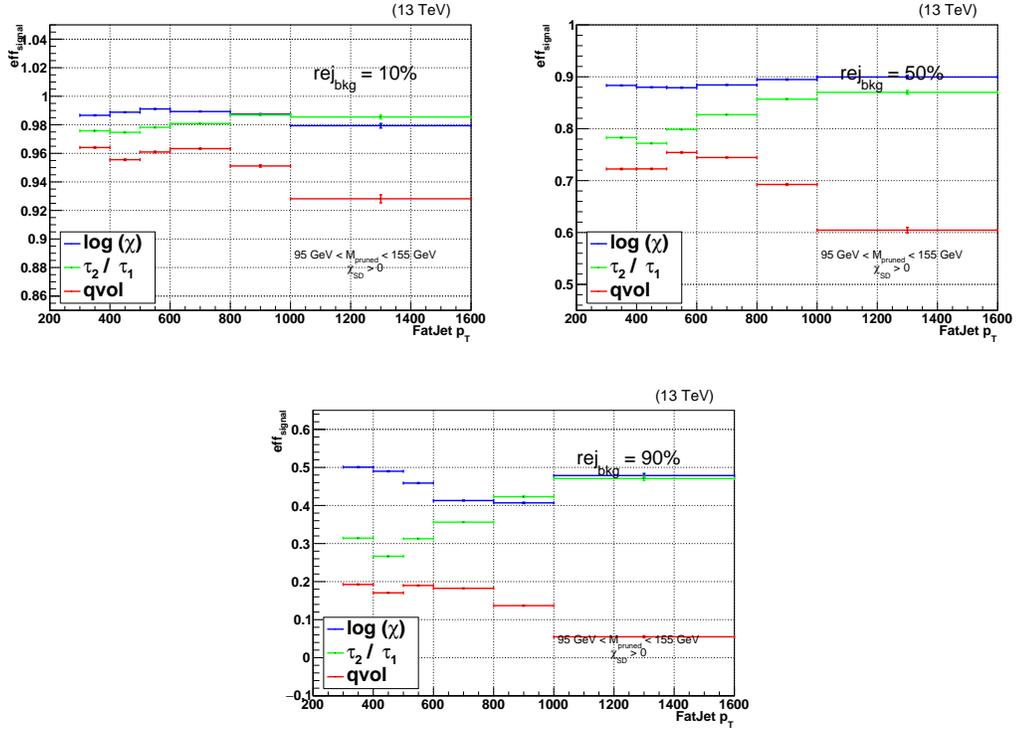


Figure B.36: Signal efficiency versus p_T for a given QCD (inclusive) background rejection based on $\log(\chi)$, τ_2/τ_1 and q-vol selections

Furthermore, SD is observed to be relatively stable against change in p_T as can be seen in Figure B.38.

We conclude that, after optimization, overall SD is superior⁷ in performance in comparison to τ_2/τ_1 and q-jet volatility.

⁷However, we note that there is possibly a bias for SD against the other jet substructure methods due to the fact that no b-tagging procedure has been performed in these other methods. We leave the bias issue to a future study.

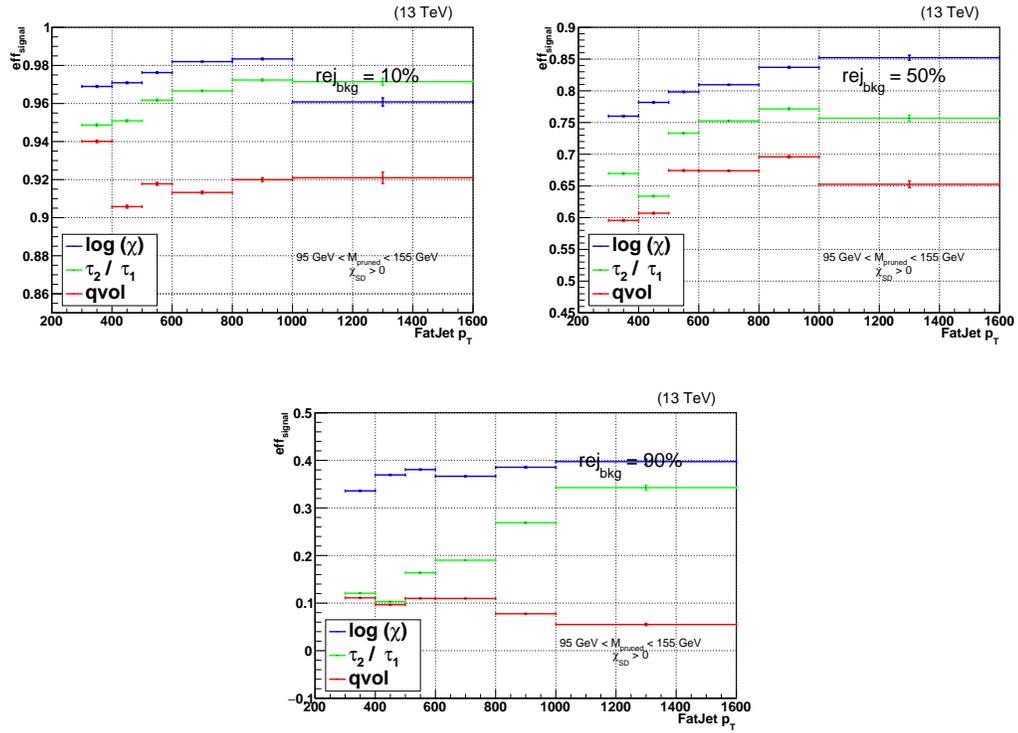


Figure B.37: Signal efficiency versus p_T for a given $Z' \rightarrow t\bar{t}$ background rejection based on $\log(\chi)$, τ_2/τ_1 and q-vol selections

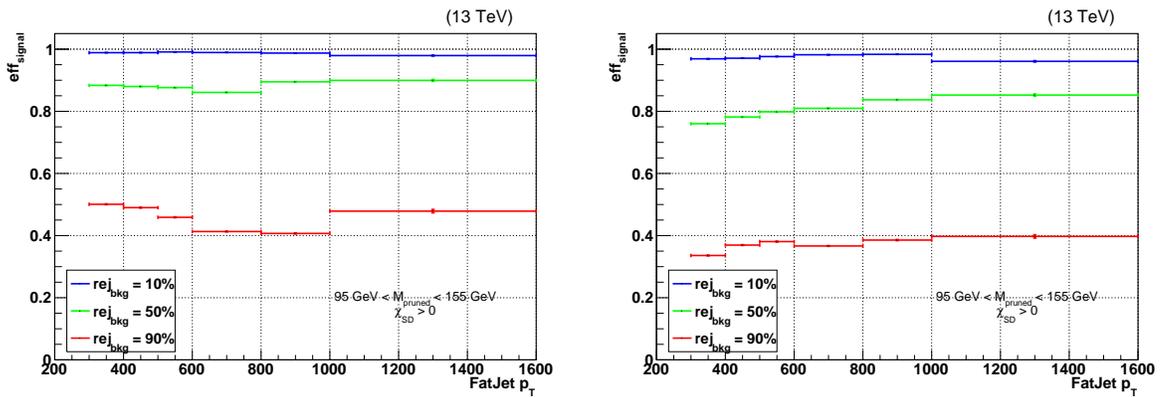


Figure B.38: Signal efficiency versus p_T for a given QCD inclusive (left) and $Z' \rightarrow t\bar{t}$ (right) background rejection based on a $\log(\chi)$ selection

B.9 Conclusion

We have presented a study on how Shower Deconstruction, as a Higgs to double b-quark tagger, performs on CMS simulated samples.

Using AK12 jets, SD only accepts fatjets with $p_T > \sim 200$ GeV and using AK8 fatjets SD accepts fatjets with $p_T > \sim 300$ GeV. We also find that the purer the sample SD performs better. Studying its characteristics, SD requires at least two microjets, and tends to want two b-tagged microjets. Selecting AK8 fatjets that satisfy $p_T > 300$ GeV, have at least two b-tagged microjets, have purer sample ($\Delta R < 0.05$), SD efficiency can be up to 90%.

We also calculate the IVF SV matching microjet b-tag and fake rate. For microjet b-tag, we find that microjet b-tag rate using b-quark matching is about 70% and using b-hadron is about 60%. Microjet b-tag fake rate is calculated to be in the order of 20% due to the high number of IVF SV reconstructed in u/d/s/g fatjets and this rate is dependent on p_T . In addition, we found no significant difference between using k_T , AK, CA jet algorithms for clustering the microjets in terms of the fake rate.

As for the optimization, first we look at the tag rate input parameter. Keeping the value at 70% is already optimal despite the calculated b-tag rate using b-hadron method is 60%. For the fake rate optimization, setting the value to be closer to the actual fake rate value (of order 20%) actually increases performance and is the optimal parameter value. Furthermore, widening the Higgs mass window parameter to 30 GeV is the optimal choice. Lastly, we found that overall the optimal microjet cone size is $R=0.20$.

Furthermore, we observe that, in comparison to τ_2/τ_1 and the q-jet volatility, applying SD does not appear to dramatically change the mass distribution of jets.

For the performance comparison with other jet substructure variables, the τ_2/τ_1 and the q-jet volatility, SD performs best and its performance is stable against p_T .

Appendix C

The Cosmic Rack

During the course of my PhD research, I have worked on the maintenance of the cosmic rack (c-rack) at CERN as service work to the CMS collaboration under the tracker data acquisitions (DAQ) group. This chapter documents the studies I have conducted with the c-rack.

The c-rack consists of genuine CMS silicon strip modules of the tracker outer barrel (TOB) subdetector. It provides all the necessary equipment and connections for the CMS-like operation of the subdetector operation. It has been used as part of the TOB integration and verification effort, and also provide a unique possibility to do other tracker studies such as configuration, track reconstruction, cluster latency studies.

The c-rack can mount up to 10 layers of modules, which can hold two TOB rods each. Each TOB rod can host 6 modules measuring only one coordinate (the r-phi modules) or 12 modules arranged in 6 pairs measuring both coordinates (the stereo modules). In the stereo modules one sensor is rotated by a small angle, which enables measurement of both coordinates.

As triggers, there are two large plastic scintillators mounted on top and on bottom of the c-rack, which covers the active area of of the rods. Each scintillator is equipped with Photo Multiplier Tubes (PMTs) on both sides. The trigger rate is approximately 1 Hz, translating to about 1 hit per module every 25s [21].

Figure C.1 show sketches of the c-rack at the design stage. During the maintenance period of

2016 to 2018, the c-rack was configured differently compared to how it was initially configured.

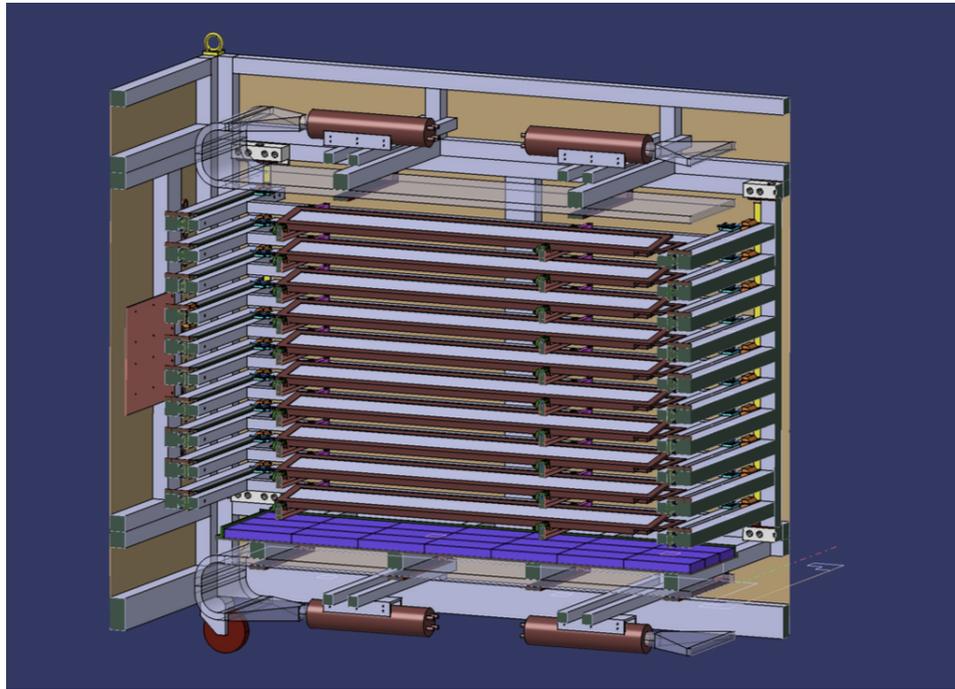
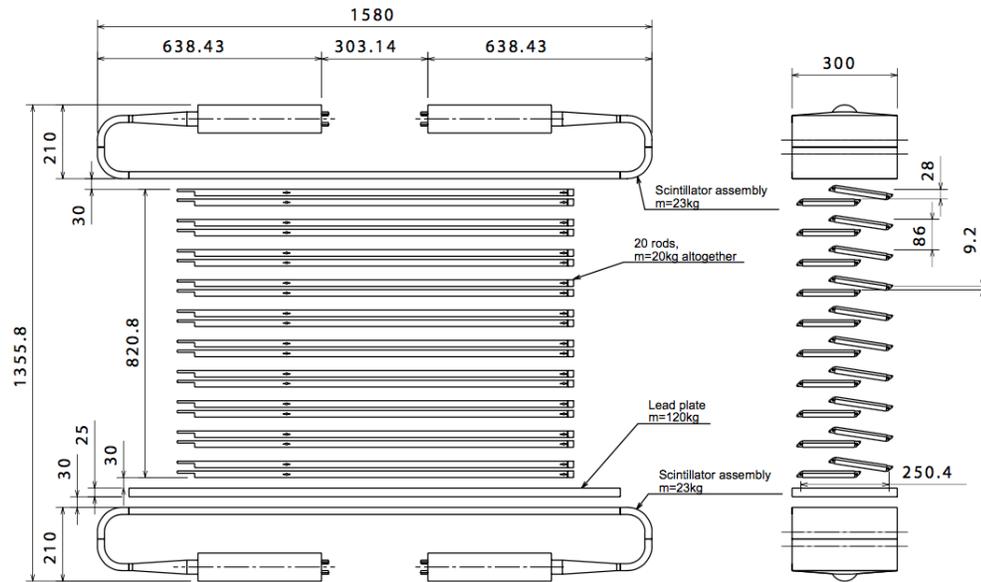


Figure C.1: Sketches of the c-rack at design stage. [21]

C.1 Hardware identification maps and visual mapping

One of the work that has been done is re-establishing the maps between the various hardware identification numbers and the geometrical coordinates. The work involves reconfiguring and updating the CMS software with the correct information of the c-rack and also providing the correct mapping of various hardware identification numbers. The identification number includes the module number, the electronics identification: Communication and Control Unit (CCU) and Detector Control Unit (DCU) numbers.

Previously, the coordinates of the c-rack was set to be between -50cm and 50cm in the z coordinate. For convenience, we have shifted the coordinates by 50cm and have it to be between 0 and 100cm . However the CMS software (CMSSW), version 9 or later, has a particular way of generating the module numbers such that the shifting in z has shuffled the module numbers. Work has been done to reorganize the mapping of the module numbers with the geometrical coordinate such that they are in ascending order corresponding to the increasing z . The result is shown in Figure C.2. There are 84 modules registered by CMSSW geometry software modules.

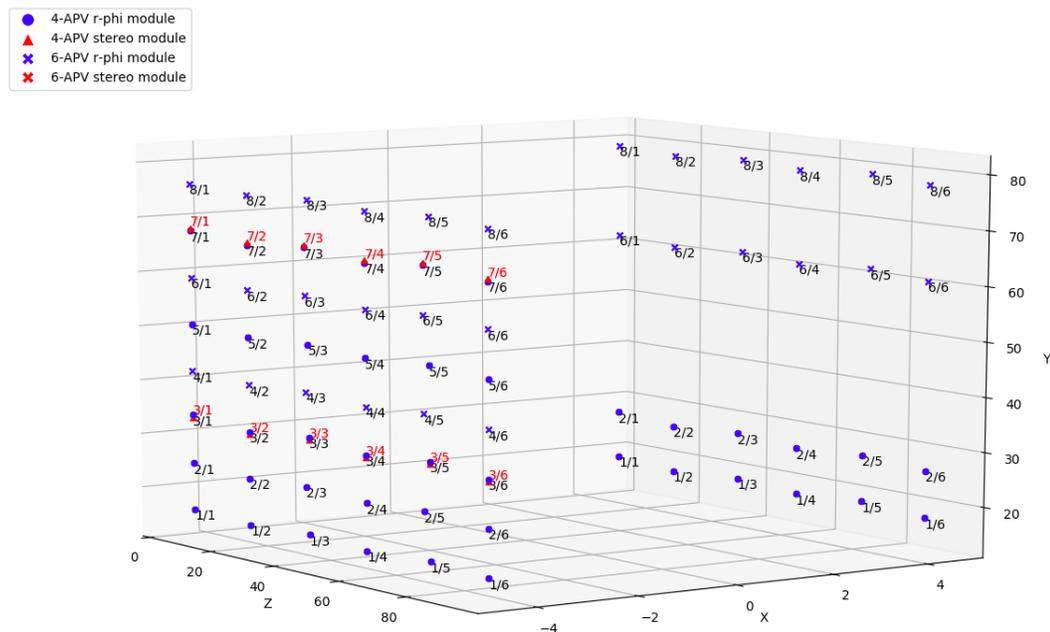


Figure C.2: Visualisation of the c-rack modules in x,y,z coordinates as registered by the CMS software, showing the layer number and the module number (layer/module number).

We performed a physical inspection of the c-rack on May 23, 2018 and we found that the geometrical layout was largely consistent with the spatial coordinates registered in CMSSW, except for the two bottom layers that were at that moment not registered in CMSSW. Furthermore the coordinates of which detector module contained 4 or 6 APV-modules, and which layers were double-sided, were also consistent. Figure C.3 and Table C.1 shows the state of the c-rack in mid-2018 as determined by the physical inspection.

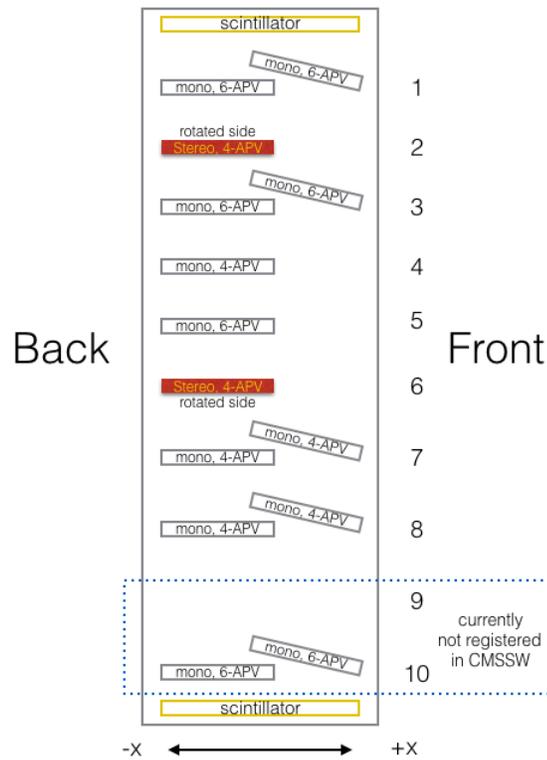


Figure C.3: State of the c-rack on May 23rd, 2018.

Table C.1: State of the c-rack on May 23rd, 2018.

Layer (physical)	Layer (CMSSW)	Rod	Module type	APV type	Additional comments
1	8	Front	mono	6	--
1	8	Back	mono	6	--
2	7	Front	--	--	--
2	7	Back	stereo	4	tilted module on top side.
3	6	Front	mono	6	--
3	6	Back	mono	6	--
4	5	Front	--	--	--
4	5	Back	mono	4	--
5	4	Front	--	--	--
5	4	Back	mono	6	--
6	3	Front	--	--	--
6	3	Back	stereo	4	tilted module on bottom side.
7	2	Front	mono	4	--
7	2	Back	mono	4	--
8	1	Front	mono	4	--
8	1	Back	mono	4	--
9	--	Front	--	--	--
9	--	Back	--	--	--
10	--	Front	mono	6	--
10	--	Back	mono	6	--

However we found that there were only 66 modules that were registered in the tracker database, and furthermore there were only 62 modules that were registered in the FED cabling check procedure (in CMSSW). Figure C.4 shows the modules the 62 connected modules.

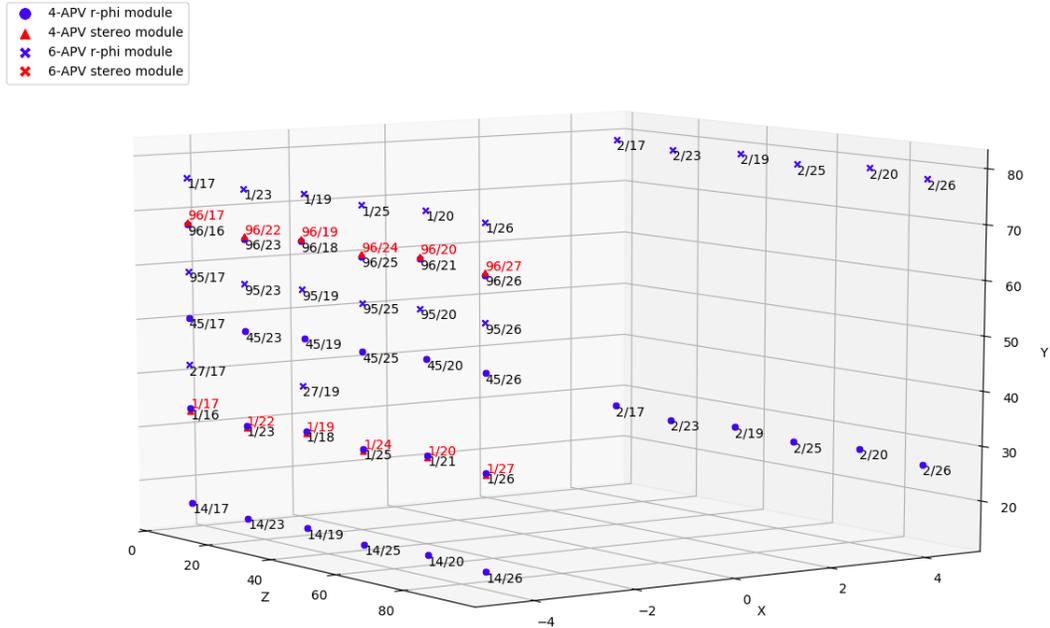


Figure C.4: Visualisation of the c-rack modules in x,y,z coordinates, showing CCU address and channel (Address/channel).

C.2 Cluster latency measurement

The APV samples the signal of the silicon strips that it reads out and saves the readings into the pipeline with a 40 Mhz rate. The latency in the tracker is the time distance between the write and the read pointer in the APV. In order words, it is roughly the time distance between the time an event fired a trigger and the trigger signal reaching the APV, such that the APV records the correct events (muons) that the PMTs are triggering on. The latency depends on a host of things such as cable length, time spent in electronics, etc. Specifically, we measure the total charge of clusters of hits across all strips at 25 ns bunch crossing (bx) time steps. For this reason it is called the cluster latency measurement. We scan latency time from 50 to 80 bx units and measure the charge of the clusters across all strips, recording about 300 events per time steps. This measurement was done in 2016 and the results is shown in Figure C.5.

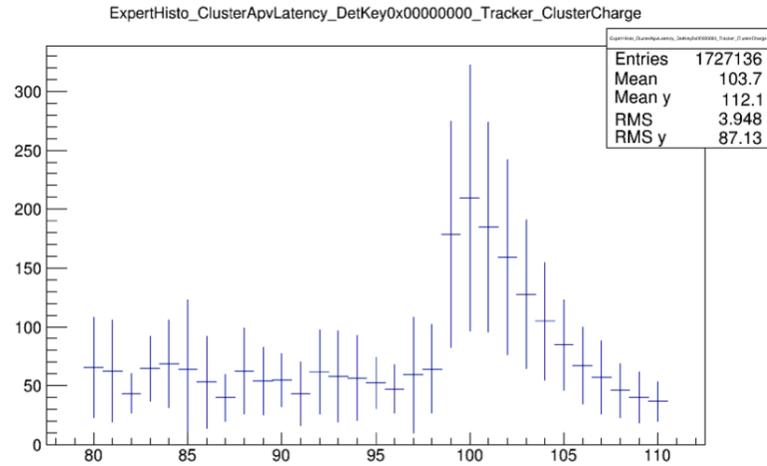


Figure C.5: Cluster latency measurement of 50 to 80 bx units of 300 events per time step.

In order to measure the precise latency time, we invert the x-axis and fit with a CR-RC pulse shape function. The fit returns peak value of 59.6 ± 0.6 as shown in Figure C.6.

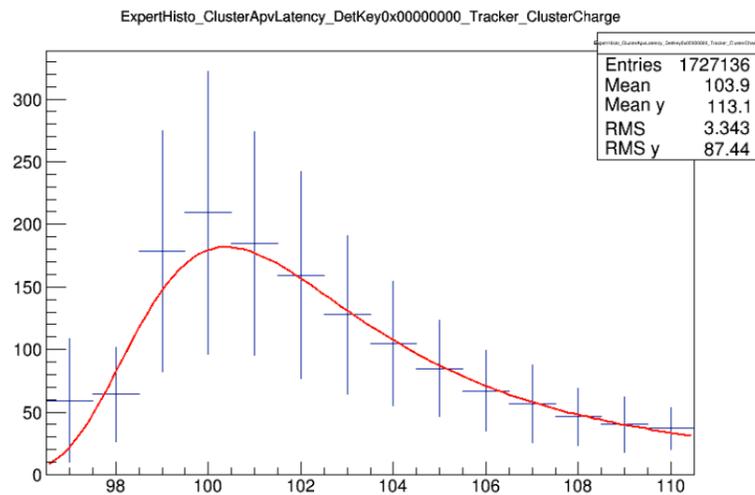


Figure C.6: Cluster latency measurement where the x-axis has been inverted in order to fit using the CR-RC pulse shape function.

Bibliography

- [1] Y. Okada and L. Panizzi, “LHC signatures of vector-like quarks,” *Adv. High Energy Phys.* **2013** (2013) 364936. [arXiv:1207.5607 \[hep-ph\]](#).
- [2] L. Panizzi, “Vector-like quarks: t' and partners,” *Nuovo Cim.* **C037** (2014) no. 02, 69–79. [\[doi\]](#).
- [3] CMS Collaboration, A. M. Sirunyan *et al.*, “Search for vector-like T and B quark pairs in final states with leptons at $\sqrt{s} = 13$ TeV,”. [arXiv:1805.04758 \[hep-ex\]](#).
- [4] Wikipedia: Standard_Model. [\[link\]](#).
- [5] J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, “A Historical Profile of the Higgs Boson. An Updated Historical Profile of the Higgs Boson,”. [\[link\]](#). Comments: 22 pages, 5 figures, update of arXiv:1201.6045, to be published in the volume.
- [6] S. Myers, “The Large Hadron Collider 2008-2013,” *International Journal of Modern Physics A* **28** (2013) no. 25, 1330035. <http://www.worldscientific.com/doi/pdf/10.1142/S0217751X13300354>.
- [7] AC Team, “Computer-generated diagram of an LHC dipole. Schma d’un dipole du LHC.” Sep, 1998.
- [8] L. R. Evans and P. Bryant, “LHC Machine,” *JINST* **3** (2008) S08001. 164 p. <https://cds.cern.ch/record/1129806>. This report is an abridged version of the LHC Design Report (CERN-2004-003).

- [9] T. Sakuma and T. McCauley, “Detector and event visualization with SketchUp at the CMS experiment,” Tech. Rep. CMS-CR-2013-379. arXiv:1311.4942, CERN, Geneva, Oct, 2013. <https://cds.cern.ch/record/1626816>. Comments: 5 pages, 6 figures, Proceedings for CHEP 2013, 20th International Conference on Computing in High Energy and Nuclear Physics.
- [10] CMS Collaboration, S. Chatrchyan *et al.*, “The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC,” *JINST* **8** (2013) P11002. [arXiv:1306.6905](https://arxiv.org/abs/1306.6905) [[physics.ins-det](https://arxiv.org/abs/1306.6905)].
- [11] Wikipedia: Pseudorapidity. <https://en.wikipedia.org/wiki/Pseudorapidity>.
- [12] CMS Collaboration, S. Chatrchyan *et al.*, “The CMS experiment at the CERN LHC,” *JINST* **3** (2008) S08004. [[doi](https://doi.org/10.1088/1748-0221/3/S08/S08)].
- [13] CMS Collaboration, S. Chatrchyan *et al.*, “Description and performance of track and primary-vertex reconstruction with the CMS tracker,” *JINST* **9** (2014) no. 10, P10009. [arXiv:1405.6569](https://arxiv.org/abs/1405.6569) [[physics.ins-det](https://arxiv.org/abs/1405.6569)].
- [14] C. Biino, “The CMS Electromagnetic Calorimeter: overview, lessons learned during Run 1 and future projections,” *Journal of Physics: Conference Series* **587** (2015) no. 1, 012001. [[link](https://arxiv.org/abs/1503.07541)].
- [15] CMS Collaboration, S. Chatrchyan *et al.*, “Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data,” *JINST* **5** (2010) T03012. [arXiv:0911.4991](https://arxiv.org/abs/0911.4991) [[physics.ins-det](https://arxiv.org/abs/0911.4991)].
- [16] CMS Collaboration, M. S. Kim, “CMS reconstruction improvement for the muon tracking by the RPC chambers,” *PoS RPC2012* (2012) 045. [arXiv:1209.2646](https://arxiv.org/abs/1209.2646) [[physics.ins-det](https://arxiv.org/abs/1209.2646)]. [JINST8,T03001(2013)].

- [17] CMS Collaboration, A. M. Sirunyan *et al.*, “Particle-flow reconstruction and global event description with the CMS detector,” *JINST* **12** (2017) no. 10, P10003. [arXiv:1706.04965](#) [[physics.ins-det](#)].
- [18] CMS Collaboration, A. M. Sirunyan *et al.*, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV,” *Submitted to: JINST* (2017) . [arXiv:1712.07158](#) [[physics.ins-det](#)].
- [19] I. W. Stewart, F. J. Tackmann, J. Thaler, C. K. Vermilion, and T. F. Wilkason, “XCone: N-jettiness as an Exclusive Cone Jet Algorithm,” *JHEP* **11** (2015) 072. [arXiv:1508.01516](#) [[hep-ph](#)].
- [20] CMS Collaboration, “Identification of double-b quark jets in boosted event topologies,” Tech. Rep. CMS-PAS-BTV-15-002, CERN, Geneva, 2016. [[link](#)].
- [21] C. Bloch, D. Abbaneo, and C. W. Fabjan, “Studies for the Commissioning of the CERN CMS Silicon Strip Tracker,” 2008. [[link](#)]. Presented on 16 Jan 2008.
- [22] CMS Collaboration, S. Chatrchyan *et al.*, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Physics Letters B* **716** (2012) no. 1, 30 – 61. [[link](#)].
- [23] ATLAS Collaboration, G. Aad *et al.*, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Physics Letters B* **716** (2012) no. 1, 1 – 29. [[link](#)].
- [24] CMS Collaboration, CMS Collaboration, “Search for pair-produced vector-like B quarks in proton-proton collisions at $\sqrt{s} = 8$ TeV.” Submitted to *Phys. Rev. D*, 2015.
- [25] CMS Collaboration, S. Chatrchyan *et al.*, “Inclusive search for a vector-like T quark with charge $\frac{2}{3}$ in pp collisions at $\sqrt{s} = 8$ TeV,” *Phys. Lett.* **B729** (2014) 149–171. [arXiv:1311.7667](#) [[hep-ex](#)].

- [26] **ATLAS** Collaboration, G. Aad *et al.*, “Search for pair production of heavy top-like quarks decaying to a high- W boson and a b quark in the lepton plus jets final state at with the ATLAS detector,” *Physics Letters B* **718** (2013) no. 475, 1284 – 1302. [\[link\]](#).
- [27] M. Thomson, *Modern particle physics*. Cambridge University Press, New York, 2013. [\[link\]](#).
- [28] W. Hollik, “Quantum field theory and the Standard Model,” [\[link\]](#). Comments: 44 pages, Lectures given at the 2009 European School of High-Energy Physics, Bautzen, Germany, 14-27 Jun 2009.
- [29] M. D. Schwartz, *Quantum Field Theory and the Standard Model*. Cambridge University Press, 2014. [\[link\]](#).
- [30] M. E. Peskin and D. V. Schroeder, *An introduction to quantum field theory*. Westview, Boulder, CO, 1995. [\[link\]](#). Includes exercises.
- [31] R. M. Godbole, “Field Theory and the Electro-Weak Standard Model,” in *Proceedings, 2nd Asia-Europe-Pacific School of High-Energy Physics (AEPSHEP 2014): Puri, India, November 04-17, 2014*, pp. 1–61. 2017. [arXiv:1703.04978 \[hep-ph\]](#).
- [32] M. Srednicki, *Quantum Field Theory*. Cambridge Univ. Press, Cambridge, 2007. [\[link\]](#).
- [33] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, “Experimental Test of Parity Conservation in Beta Decay,” *Phys. Rev.* **105** (Feb, 1957) 1413–1415. [\[link\]](#).
- [34] **CMS** Collaboration, S. Chatrchyan *et al.*, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Phys. Lett.* **B716** (2012) 30–61. [arXiv:1207.7235 \[hep-ex\]](#).
- [35] **ATLAS** Collaboration, G. Aad *et al.*, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys. Lett.* **B716** (2012) 1–29. [arXiv:1207.7214 \[hep-ex\]](#).

- [36] **Particle Data Group** Collaboration, C. Patrignani *et al.*, “Review of Particle Physics,” *Chin. Phys.* **C40** (2016) no. 10, 100001. [\[doi\]](#).
- [37] M. Gell-Mann, “Symmetries of Baryons and Mesons,” *Phys. Rev.* **125** (Feb, 1962) 1067–1084. [\[link\]](#).
- [38] S. L. Glashow, “Partial-symmetries of weak interactions,” *Nuclear Physics* **22** (1961) no. 4, 579 – 588. [\[link\]](#).
- [39] S. Weinberg, “A Model of Leptons,” *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266. [\[link\]](#).
- [40] A. Salam and J. C. Ward, “Weak and electromagnetic interactions,” *Il Nuovo Cimento (1955-1965)* **11** (Feb, 1959) 568–577. [\[link\]](#).
- [41] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons,” *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323. [\[link\]](#).
- [42] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons,” *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509. [\[link\]](#).
- [43] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global Conservation Laws and Massless Particles,” *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587. [\[link\]](#).
- [44] H. Yukawa, “On the Interaction of Elementary Particles I,” *Proc. Phys. Math. Soc. Jap.* **17** (1935) 48–57. [\[doi\]](#). [Prog. Theor. Phys. Suppl.1,1(1935)].
- [45] **Super-Kamiokande** Collaboration, Y. Fukuda *et al.*, “Evidence for oscillation of atmospheric neutrinos,” *Phys. Rev. Lett.* **81** (1998) 1562–1567. [arXiv:hep-ex/9807003](#) [\[hep-ex\]](#).
- [46] **SNO** Collaboration, Q. R. Ahmad *et al.*, “Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory,” *Phys. Rev. Lett.* **89** (2002) 011301. [arXiv:nucl-ex/0204008](#) [\[nucl-ex\]](#).

- [47] **KamLAND** Collaboration, K. Eguchi *et al.*, “First results from KamLAND: Evidence for reactor anti-neutrino disappearance,” *Phys. Rev. Lett.* **90** (2003) 021802.
[arXiv:hep-ex/0212021](#) [[hep-ex](#)].
- [48] S. F. King, “Neutrino mass models,” *Rept. Prog. Phys.* **67** (2004) 107–158.
[arXiv:hep-ph/0310204](#) [[hep-ph](#)].
- [49] S. Bilenky, “Neutrino oscillations: From a historical perspective to the present status,” *Nuclear Physics B* **908** (2016) 2 – 13. [[link](#)]. Neutrino Oscillations: Celebrating the Nobel Prize in Physics 2015.
- [50] N. Cabibbo, “Unitary Symmetry and Leptonic Decays,” *Phys. Rev. Lett.* **10** (Jun, 1963) 531–533. [[link](#)].
- [51] M. Kobayashi and T. Maskawa, “CP Violation in the Renormalizable Theory of Weak Interaction,” *Prog. Theor. Phys.* **49** (1973) 652–657. [[doi](#)].
- [52] S. Weinberg, “The cosmological constant problem,” *Rev. Mod. Phys.* **61** (Jan, 1989) 1–23.
[\[link\]](#).
- [53] C. Cski and P. Tanedo, “Beyond the Standard Model,” [[link](#)]. 119 pages, 16 figures, minor revisions and corrections from published version. Proceedings of the 2013 European School of High-Energy Physics, Paradfurdo, Hungary, 5-18 June 2013, edited by M. Mulders and G. Perez, CERN-2015-004 (CERN, Geneva, 2015), ISBN: 9789290834205. v2: typos corrected, references updated.
- [54] J. A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer, and M. Prez-Victoria, “Handbook of vectorlike quarks: Mixing and single production,” *Phys. Rev.* **D88** (2013) no. 9, 094010.
[arXiv:1306.0572](#) [[hep-ph](#)].
- [55] A. D. Simone, O. Matsedonskyi, R. Rattazzi, and A. Wulzer, “A First Top Partnew Hunter’s Guide,” *JHEP* **1304** (2013) . [[link](#)].

- [56] M. J. Dugan, H. Georgi, and D. B. Kaplan, “Anatomy of a composite Higgs model,” *Nuclear Physics B* **254** (1985) 299 – 326. [\[link\]](#).
- [57] K. Agashe, R. Contino, and A. Pomarol, “The Minimal composite Higgs model,” *Nucl. Phys.* **B719** (2005) 165–187. [arXiv:hep-ph/0412089](#) [\[hep-ph\]](#).
- [58] R. Contino, L. Da Rold, and A. Pomarol, “Light custodians in natural composite Higgs models,” *Phys. Rev.* **D75** (2007) 055014. [arXiv:hep-ph/0612048](#) [\[hep-ph\]](#).
- [59] C. Anastasiou, E. Furlan, and J. Santiago, “Realistic Composite Higgs Models,” *Phys. Rev.* **D79** (2009) 075003. [arXiv:0901.2117](#) [\[hep-ph\]](#).
- [60] A. Carmona, M. Chala, and J. Santiago, “New Higgs Production Mechanism in Composite Higgs Models,” *JHEP* **07** (2012) 049. [arXiv:1205.2378](#) [\[hep-ph\]](#).
- [61] M. Gillioz, R. Grober, C. Grojean, M. Muhlleitner, and E. Salvioni, “Higgs Low-Energy Theorem (and its corrections) in Composite Models,” *JHEP* **10** (2012) 004. [arXiv:1206.7120](#) [\[hep-ph\]](#).
- [62] I. Antoniadis, K. Benakli, and M. Quiros, “Finite Higgs mass without supersymmetry,” *New J. Phys.* **3** (2001) 20. [arXiv:hep-th/0108005](#) [\[hep-th\]](#).
- [63] Y. Hosotani, S. Noda, and K. Takenaga, “Dynamical gauge-Higgs unification in the electroweak theory,” *Phys. Lett.* **B607** (2005) 276–285. [arXiv:hep-ph/0410193](#) [\[hep-ph\]](#).
- [64] B. Grinstein, M. Redi, and G. Villadoro, “Low Scale Flavor Gauge Symmetries,” *JHEP* **11** (2010) 067. [arXiv:1009.2049](#) [\[hep-ph\]](#).
- [65] D. Guadagnoli, R. N. Mohapatra, and I. Sung, “Gauged Flavor Group with Left-Right Symmetry,” *JHEP* **04** (2011) 093. [arXiv:1103.4170](#) [\[hep-ph\]](#).
- [66] N. Arkani-Hamed, A. G. Cohen, E. Katz, and A. E. Nelson, “The Littlest Higgs,” *JHEP* **07** (2002) 034. [arXiv:hep-ph/0206021](#) [\[hep-ph\]](#).

- [67] P. W. Graham, A. Ismail, S. Rajendran, and P. Saraswat, “A Little Solution to the Little Hierarchy Problem: A Vector-like Generation,” *Phys. Rev.* **D81** (2010) 055016. [arXiv:0910.3020 \[hep-ph\]](#).
- [68] S. P. Martin, “Extra vector-like matter and the lightest Higgs scalar boson mass in low-energy supersymmetry,” *Phys. Rev.* **D81** (2010) 035004. [arXiv:0910.2732 \[hep-ph\]](#).
- [69] K. S. Babu, I. Gogoladze, M. U. Rehman, and Q. Shafi, “Higgs Boson Mass, Sparticle Spectrum and Little Hierarchy Problem in Extended MSSM,” *Phys. Rev.* **D78** (2008) 055017. [arXiv:0807.3055 \[hep-ph\]](#).
- [70] J. A. Aguilar-Saavedra, “Mixing with vector-like quarks: constraints and expectations,” *EPJ Web Conf.* **60** (2013) 16012. [arXiv:1306.4432 \[hep-ph\]](#).
- [71] F. del Aguila, J. A. Aguilar-Saavedra, and R. Miquel, “Constraints on top couplings in models with exotic quarks,” *Phys. Rev. Lett.* **82** (1999) 1628–1631. [arXiv:hep-ph/9808400 \[hep-ph\]](#).
- [72] A. Hoecker, “Physics at the LHC Run-2 and Beyond,” [\[link\]](#). Comments: Lecture notes from the 2016 European School of High-Energy Physics, 15-28 June 2016, Skeikampen, Norway (61 pages, 56 figures).
- [73] CERN, “The accelerator complex,” <https://cds.cern.ch/record/1997193>.
- [74] **ATLAS** Collaboration, G. Aad *et al.*, “The ATLAS Experiment at the CERN Large Hadron Collider,” *JINST* **3** (2008) S08003. [\[doi\]](#).
- [75] **ALICE** Collaboration, K. Aamodt *et al.*, “The ALICE experiment at the CERN LHC,” *JINST* **3** (2008) S08002. [\[doi\]](#).
- [76] **LHCb** Collaboration, A. A. Alves, Jr. *et al.*, “The LHCb Detector at the LHC,” *JINST* **3** (2008) S08005. [\[doi\]](#).
- [77] **CMS** Collaboration, *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997. [\[link\]](#).

- [78] CMS Collaboration, A. K. Nayak, “Reconstruction of physics objects in the CMS detector,” *PoS CHARGED2012* (2012) 010. [arXiv:1310.7408 \[hep-ex\]](#).
- [79] USCMS, ECAL/HCAL Collaboration, S. Abdullin *et al.*, “The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c,” *Eur. Phys. J.* **C60** (2009) 359–373. [\[doi\]](#). [Erratum: *Eur. Phys. J.*C61,353(2009)].
- [80] CMS Collaboration, D. Acosta, “CMS Trigger Improvements Towards Run II,” *Nucl. Part. Phys. Proc.* **273-275** (2016) 1008–1013. [\[doi\]](#).
- [81] R. Fruhwirth, “Application of Kalman filtering to track and vertex fitting,” *Nucl. Instrum. Meth.* **A262** (1987) 444–450. [\[doi\]](#).
- [82] CMS Collaboration, S. Chatrchyan *et al.*, “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV,” *JINST* **7** (2012) P10002. [arXiv:1206.4071 \[physics.ins-det\]](#).
- [83] CMS Collaboration, V. Khachatryan *et al.*, “Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV,” *JINST* **10** (2015) no. 06, P06005. [arXiv:1502.02701 \[physics.ins-det\]](#).
- [84] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm,” *JHEP* **04** (2008) 063. [arXiv:0802.1189 \[hep-ex\]](#).
- [85] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet user manual,” *Eur. Phys. J. C* **72** (2012) 1896. [arXiv:1111.6097 \[hep-ph\]](#).
- [86] CMS Collaboration, “Determination of jet energy calibration and transverse momentum resolution in CMS,” *Journal of Instrumentation* **6** (Nov., 2011) 11002. [arXiv:1107.4277 \[physics.ins-det\]](#).
- [87] CMS Collaboration, S. Chatrchyan *et al.*, “Identification of b-quark jets with the CMS experiment,” *JINST* **8** (2013) P04013. [arXiv:1211.4462 \[hep-ex\]](#).

- [88] N. Arkani-Hamed, A. G. Cohen, and H. Georgi, “Electroweak symmetry breaking from dimensional deconstruction,” *Phys.Lett.* **B513** (2001) 232–240. [arXiv:hep-ph/0105239 \[hep-ph\]](#).
- [89] M. Schmaltz and D. Tucker-Smith, “Little Higgs review,” *Ann.Rev.Nucl.Part.Sci.* **55** (2005) 229–270. [arXiv:hep-ph/0502182 \[hep-ph\]](#).
- [90] The ALEPH, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak Working Group, “Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP,”. [hep-ex/1302txyz](#).
- [91] **CMS B2G PAG** Collaboration, “TWiki: B2G Monte Carlo, Heavy exotic quarks pair production,” <https://twiki.cern.ch/twiki/bin/view/CMS/B2GMonteCarlo> . [\[link\]](#).
- [92] F. del Aguila, L. Ametller, G. L. Kane, and J. Vidal, “Vector Like Fermion and Standard Higgs Production at Hadron Colliders,” *Nucl. Phys.* **B334** (1990) 1–23. [\[doi\]](#).
- [93] S. Bhattacharya, T. Speer, M. Luk, A. Avetisyan, S. Malik, U. Heintz, and M. Narain, “Inclusive Search for Heavy Vector-Like top quarks in multiple-lepton final states at $\sqrt{s}=8$ TeV (AN-2013/006),” . [\[link\]](#).
- [94] S. Bhattacharya, T. Speer, M. Luk, A. Avetisyan, S. Malik, U. Heintz, and M. Narain, “Inclusive search for top partners in single- and multiple-lepton final states at $\sqrt{s}=8$ TeV (B2G-12-015),” . [\[link\]](#).
- [95] H. Bakhshian, A. Fahim, A. Calderon, P. Lobelle, J. Marco, P. Martinez, T. Rodrigo, R. Vilar, G. Cappello, M. Chiorboli, A. Tricomi, J. Cuevas, J. Fernandez, L. Lloret, G. Dissertori, P. Milenovic, F. Moortgat, P. Nef, L. Pape, F. Ronga, L. Sala, B. Stieger, K. Theofilatos, D. Treille, and M. Weber, “Computing the contamination from fakes in leptonic final states, CMS Internal Document: AN-2010/261,”. [\[link\]](#).

- [96] **CMS Collaboration**, S. Chatrchyan *et al.*, “Search for heavy vector-like quarks decaying to same-sign dileptons,” *CMS Physics Analysis Summary CMS-PAS-B2G-16-019* (2017) . [\[link\]](#).
- [97] **GEANT4 Collaboration**, S. Agostinelli *et al.*, “GEANT4: A simulation toolkit,” *Nucl. Instrum. Meth.* **A506** (2003) 250–303. [\[doi\]](#).
- [98] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, “MadGraph 5: going beyond,” *Journal of High Energy Physics* **2011** (2011) 1–40. [\[link\]](#).
- [99] T. Sjöstrand, S. Mrenna, and P. Skands, “PYTHIA 6.4 physics and manual,” *JHEP* **05** (2006) 026. [arXiv:hep-ph/0603175](#) [\[hep-ph\]](#).
- [100] T. Sjöstrand, S. Mrenna, and P. Skands, “A Brief Introduction to PYTHIA 8.1,” *Comput. Phys. Comm.* **178** (2008) . [\[link\]](#).
- [101] O. Iorio and A. Jafari, “Single-top t-channel: Run II working meeting,”. [\[link\]](#).
- [102] **CMS Higgs PAG Collaboration**, “TWiki: Summary table of samples produced for the 1 Billion campaign, with 25ns bunch-crossing,” <https://twiki.cern.ch/twiki/bin/view/CMS/SummaryTable1G25ns> . [\[link\]](#).
- [103] **CMS Monte Carlo Generator Group Collaboration**, “TWiki: Standard Model Cross Sections for CMS at 13 TeV,” <https://twiki.cern.ch/twiki/bin/view/CMS/StandardModelCrossSectionsat13TeVInclusive> . [\[link\]](#).
- [104] **CMS miniAOD Working Group Collaboration**, “TWiki: Workbook MiniAOD,” <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMiniAOD> . [\[link\]](#).
- [105] **CMS LJMet Working Group Collaboration**, “LJMet for Run 2,” <https://twiki.cern.ch/twiki/bin/view/CMS/LjmetComRun2> . [\[link\]](#).
- [106] M. Cacciari, G. P. Salam, and G. Soyez, “The catchment area of jets,” *JHEP* **04** (2008) 005. [arXiv:0802.1188](#) [\[hep-ph\]](#).

- [107] **CMS E-gamma POG** Collaboration, “TWiki: MVA Electron Identification Run2,” [\[link\]](#).
- [108] **CMS Muon POG** Collaboration, “SWGGuide: Muon ID Run2,” <https://twiki.cern.ch/twiki/bin/view/CMS/SWGGuideMuonIdRun2> . [\[link\]](#).
- [109] C. Richardson, “Status Updates for X53 to Same-sign dileptons using 2016 data. CMS VHF Group Meeting, Jan 18, 2017,” . [\[link\]](#).
- [110] **CMS Muon POG** Collaboration, “Muon T&P Instructions for Run-II,” <https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonTagAndProbeTreesRun2> . [\[link\]](#).
- [111] **CMS Muon POG** Collaboration, “Muon dedicated studies. Results on the full 2016 data,” <https://twiki.cern.ch/twiki/bin/view/CMS/MuonWorkInProgressAndPagResults> . [\[link\]](#).
- [112] M. Cacciari, G. P. Salam, and G. Soyez, “The Anti-k(t) jet clustering algorithm,” *JHEP* **0804** (2008) 063. [arXiv:hep-ph/0802.1189](#) [\[hep-ph\]](#).
- [113] **CMS JetMET POG** Collaboration, “Jet Identification,” <https://twiki.cern.ch/twiki/bin/view/CMS/JetID> . [\[link\]](#).
- [114] **CMS** Collaboration, A. M. Sirunyan *et al.*, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV,” *JINST* **13** (2018) P05011. [arXiv:1712.07158](#) [\[physics.ins-det\]](#).
- [115] P. Verwilligen, “Muons in the CMS High Level Trigger System,” *Nuclear and Particle Physics Proceedings* **273** (2016) 2509 – 2511. [\[link\]](#).
- [116] S. P. Martin, “A Supersymmetry primer,” [arXiv:hep-ph/9709356](#) [\[hep-ph\]](#). [Adv. Ser. Direct. High Energy Phys.18,1(1998)].
- [117] **CMS** Collaboration, A. M. Sirunyan *et al.*, “Search for supersymmetry in events with at least three electrons or muons, jets, and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV,” *JHEP* **02** (2018) 067. [arXiv:1710.09154](#) [\[hep-ex\]](#).

- [118] A. Avetisyan, T. Bose, M. Narain, R. Richardson, J. Rohlf, S. Sagir, and D. Zou, “Search for top quark partners with charge $5e/3$ in the same-sign dilepton final state with CMS detector (AN-2015/148),”. [\[link\]](#).
- [119] T. Müller, J. Ott, and J. Wagner-Kuhr, “theta - a framework for template-based modeling and inference,”. <http://www-ekp.physik.uni-karlsruhe.de/~ott/theta/theta-auto/>.
- [120] T. Müller, J. Ott, and J. Wagner-Kuhr, “theta - a framework for template-based modeling and inference,”. <http://www-ekp.physik.uni-karlsruhe.de/~ott/theta/theta-auto/>.
- [121] **Particle Data Group** Collaboration, K. Olive *et al.*, “Review of Particle Physics,” *Chin. Phys. C* **38** (2014 and 2015 update) 090001. Review 38: Statistics.
- [122] R. Barlow and C. Beeston, “Fitting using finite Monte Carlo samples,” *Comput. Phys. Commun.* (1993) 219.
- [123] J. Conway, “Nuisance Parameters in Likelihoods for Multisource Spectra,” *Proceedings of PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding* (2011) 115.
- [124] J. Thaler and T. F. Wilkason, “Resolving Boosted Jets with XCone,” *JHEP* **12** (2015) 051. [arXiv:1508.01518 \[hep-ph\]](#).
- [125] D. Bertolini, P. Harris, M. Low, and N. Tran, “Pileup Per Particle Identification,” *JHEP* **10** (2014) 059. [arXiv:1407.6013 \[hep-ph\]](#).
- [126] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, “Soft Drop,” *JHEP* **1405** (2014) 146. [arXiv:1402.2657](#).
- [127] **CMS** Collaboration, “New Developments for Jet Substructure Reconstruction in CMS,” <https://cds.cern.ch/record/2275226> (Jul, 2017) . [\[link\]](#).
- [128] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, “TMVA: Toolkit for Multivariate Data Analysis,” *PoS ACAT* (2007) 040. [arXiv:physics/0703039](#).

- [129] CMS Collaboration, V. Khachatryan *et al.*, “Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV,” *JHEP* **03** (2011) 136. [arXiv:1102.3194 \[hep-ex\]](#).
- [130] CMS Collaboration, “Performance of b tagging at $\sqrt{s}=8$ TeV in multijet, ttbar and boosted topology events,” Tech. Rep. CMS-PAS-BTV-13-001, CERN, Geneva, 2013. [\[link\]](#).
- [131] L. Randall and R. Sundrum, “A Large mass hierarchy from a small extra dimension,” *Phys. Rev. Lett.* **83** (1999) 3370–3373. [arXiv:hep-ph/9905221 \[hep-ph\]](#).
- [132] T. Sjostrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, “An Introduction to PYTHIA 8.2,” *Comput. Phys. Commun.* **191** (2015) 159–177. [arXiv:1410.3012 \[hep-ph\]](#).
- [133] P. Nason, “A New method for combining NLO QCD with shower Monte Carlo algorithms,” *JHEP* **11** (2004) 040. [arXiv:hep-ph/0409146 \[hep-ph\]](#).
- [134] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method,” *JHEP* **11** (2007) 070. [arXiv:0709.2092 \[hep-ph\]](#).
- [135] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX,” *JHEP* **06** (2010) 043. [arXiv:1002.2581 \[hep-ph\]](#).
- [136] S. Frixione, P. Nason, and G. Ridolfi, “A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction,” *JHEP* **09** (2007) 126. [arXiv:0707.3088 \[hep-ph\]](#).
- [137] E. Re, “Single-top Wt-channel production matched with parton showers using the POWHEG method,” *Eur. Phys. J. C* **71** (2011) 1547. [arXiv:1009.2450 \[hep-ph\]](#).
- [138] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, “The automated computation of tree-level and

- next-to-leading order differential cross sections, and their matching to parton shower simulations,” *JHEP* **07** (2014) 079. [arXiv:1405.0301 \[hep-ph\]](#).
- [139] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, “MadGraph 5 : Going Beyond,” *JHEP* **06** (2011) 128. [arXiv:1106.0522 \[hep-ph\]](#).
- [140] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, “Matching Matrix Elements and Shower Evolution for Top-Quark Production in Hadronic Collisions,” *JHEP* **01** (2007) 013. [arXiv:hep-ph/0611129 \[hep-ph\]](#).
- [141] NNPDF Collaboration, R. D. Ball *et al.*, “Parton distributions for the LHC Run II,” *JHEP* **04** (2015) 040. [arXiv:1410.8849 \[hep-ph\]](#).
- [142] CMS Collaboration, V. Khachatryan *et al.*, “Event generator tunes obtained from underlying event and multiparton scattering measurements,” *Eur. Phys. J. C* **76** (2016) no. 3, 155. [arXiv:1512.00815 \[hep-ex\]](#).
- [143] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET,” *CMS Physics Analysis Summary CMS-PAS-PFT-09-001* (2009) . [\[link\]](#).
- [144] CMS Collaboration, “Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV,” *CMS Physics Analysis Summary CMS-PAS-PFT-10-002* (2010) . [\[link\]](#).
- [145] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet User Manual,” *Eur. Phys. J. C* **72** (2012) 1896. [arXiv:1111.6097 \[hep-ph\]](#).
- [146] M. Cacciari and G. P. Salam, “Dispelling the N^3 myth for the k_t jet-finder,” *Phys. Lett. B* **641** (2006) 57–61. [arXiv:hep-ph/0512210 \[hep-ph\]](#).
- [147] CMS Collaboration, “Determination of jet energy calibration and transverse momentum resolution in CMS,” *JINST* **6** (Nov., 2011) 11002. [arXiv:1107.4277 \[physics.ins-det\]](#).

- [148] CMS Collaboration, “Pileup Jet Identification,” *CMS Physics Analysis Summary CMS-PAS-JME-13-005* (2013) .
- [149] CMS Collaboration, “Identification of b quark jets at the CMS Experiment in the LHC Run 2,” *CMS Physics Analysis Summary CMS-PAS-BTV-15-001* (2016) .
- [150] D. E. Soper and M. Spannowsky, “Finding physics signals with shower deconstruction,” [arXiv:1102.3480v2](https://arxiv.org/abs/1102.3480v2) [[hep-ph](#)].
- [151] J. Thaler and K. Van Tilburg, “Identifying Boosted Objects with N-subjettiness,” *JHEP* **03** (2011) 015. [arXiv:1011.2268](https://arxiv.org/abs/1011.2268) [[hep-ph](#)].
- [152] S. D. Ellis, A. Hornig, T. S. Roy, D. Krohn, and M. D. Schwartz, “Qjets: A Non-Deterministic Approach to Tree-Based Jet Substructure,” *Phys. Rev. Lett.* **108** (2012) 182003. [arXiv:1201.1914](https://arxiv.org/abs/1201.1914) [[hep-ph](#)].
- [153] T. Sjostrand, P. Eden, C. Friberg, L. Lonnblad, G. Miu, S. Mrenna, and E. Norrbin, “High-energy physics event generation with PYTHIA 6.1,” *Comput. Phys. Commun.* **135** (2001) 238–259. [arXiv:hep-ph/0010017](https://arxiv.org/abs/hep-ph/0010017) [[hep-ph](#)].
- [154] G. Corcella, I. G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M. H. Seymour, and B. R. Webber, “HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes),” *JHEP* **01** (2001) 010. [arXiv:hep-ph/0011363](https://arxiv.org/abs/hep-ph/0011363) [[hep-ph](#)].
- [155] S. D. Ellis and D. E. Soper, “Successive combination jet algorithm for hadron collisions,” *Phys. Rev. D* **48** (1993) 3160–3166. [arXiv:hep-ph/9305266](https://arxiv.org/abs/hep-ph/9305266) [[hep-ph](#)].
- [156] CMS Collaboration, “Inclusive Vertex Finder (IVF) and Double b-Tag Prototype.” <https://twiki.cern.ch/twiki/bin/view/CMS/InclusiveVertexFinderRecipes>.
- [157] D. Bloch, K.-F. Chen, J. P. Chou, D. Ferencek, P. V. Hove, S. Kaplan, D. Majumder, I. Marchesini, A. Schmidt, J.-F. Tsai, E. Usai, and M. Zientek, “Identification of b-quark jets in boosted topologies,” CMS Note 2013/185, 2013.

- [158] CMS Collaboration, “Jet flavour identification (MC truth).”
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideBTagMCTools>.
- [159] Y. L. Dokshitzer, G. Leder, S. Moretti, and B. Webber, “Better jet clustering algorithms,”
JHEP **08** (1997) 001. [arXiv:hep-ph/9707323](https://arxiv.org/abs/hep-ph/9707323) [[hep-ph](#)].