

การค้นหาอันตรายกิริยาตามแบบจำลองมาตรฐานที่ให้ที่ออปควาร์กสีตัวในเครื่องตรวจจับ
อนุภาคซีเอ็มเอส และการออกแบบโครงข่ายประสาทเทียมเพื่อการวัดค่าสัมประสิทธิ์
ยูกาวาของที่ออปควาร์กจากการค้นหาอันตรายกิริยาที่ให้ที่ออปควาร์กสีตัว

นายวิษณุพันธ์ วชิรภูษิตานันท์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต

สาขาวิชาฟิสิกส์ ภาควิชาฟิสิกส์

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2566

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย



SEARCH FOR STANDARD MODEL PRODUCTION OF FOUR TOP
QUARKS IN COMPACT MUON SOLENOID DETECTOR AND
DESIGNING NEURAL NETWORKS FOR TOP QUARK YUKAWA
COUPLING MEASUREMENT FROM FOUR TOP QUARK PRODUCTION

Mr. Vichayanun Wachirapusitanand

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Physics

Department of Physics

Faculty of Science

Chulalongkorn University

Academic Year 2023

Copyright of Chulalongkorn University

Dissertation Title SEARCH FOR STANDARD MODEL PRODUCTION
OF FOUR TOP QUARKS IN COMPACT MUON
SOLENOID DETECTOR AND DESIGNING NEURAL
NETWORKS FOR TOP QUARK YUKAWA COUPLING
MEASUREMENT FROM FOUR TOP QUARK
PRODUCTION

By Mr. Vichayanun Wachirapusitanand

Field of Study Physics

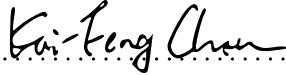
Dissertation Advisor Assistant Professor Norraphat Srimanobhas, Ph.D.

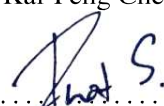
Dissertation Co-advisor Professor Freya Blekman, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment
of the Requirements for the Doctoral Degree

.....  Dean of the Faculty of Science
(Professor Pranut Potiyaraj, Ph.D.)

THESIS COMMITTEE

.....  Chair External Examiner
(Professor Kai-Feng Chen, Ph.D.)

.....  Dissertation Advisor
(Assistant Professor Norraphat Srimanobhas, Ph.D.)

.....  Dissertation Co-advisor
(Professor Freya Blekman, Ph.D.)

.....  Examiner
(Associate Professor Auttakit Chatrabhuti, Ph.D.)

.....  Examiner
(Thiparat Chotibut, Ph.D.)

.....  Examiner
(Assistant Professor Pawin Ittisamai, Ph.D.)

วิทยุณันท์ วชิรภูษิตานันท์: การ ค้นหา อันตรกิริยา ตาม แบบจำลอง มาตรฐาน ที่ให้
 ท็อป ควาร์ก สี่ ตัว ใน เครื่อง ตรวจจับ อนุภาค ซีเอ็มเอส และการ ออกแบบ โครง
 ข่าย ประสาท เทียม เพื่อ การ วัด ค่า สัมประสิทธิ์ ยูกาวา ของ ท็อป ควาร์ก จาก
 การ ค้นหา อันตรกิริยา ที่ให้ ท็อป ควาร์ก สี่ ตัว. (SEARCH FOR STANDARD
 MODEL PRODUCTION OF FOUR TOP QUARKS IN COMPACT MUON
 SOLENOID DETECTOR AND DESIGNING NEURAL NETWORKS FOR
 TOP QUARK YUKAWA COUPLING MEASUREMENT FROM FOUR
 TOP QUARK PRODUCTION) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : ผศ. ดร. นรพัทธ์
 ศรีมโนภาส, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม : ศ. ดร. เฟรยา เบลีคแมน 304 หน้า.

อันตรกิริยาการเกิดเหตุการณ์ท็อปควาร์กสี่ตัวเป็นอันตรกิริยาที่ทำนายได้ด้วยแบบ
 จำลองมาตรฐานของอนุภาค และเป็นอันตรกิริยาที่ใช้ทดสอบความแม่นยำของแบบจำลอง
 มาตรฐาน การวัดปริมาณต่าง ๆ เช่น ภาคตัดขวางของอันตรกิริยานี้สามารถนำไปเปรียบเทียบกับ
 ค่าที่ทำนายได้จากแบบจำลองมาตรฐาน ซึ่งสามารถแสดงให้เห็นว่าแบบจำลองมาตรฐาน
 เพียงพอที่จะอธิบายอันตรกิริยานี้หรือไม่ อันตรกิริยาการเกิดเหตุการณ์ท็อปควาร์กสี่ตัวยัง
 มีความไวต่อค่าสัมประสิทธิ์ที่บ่งบอกถึงขนาดของอันตรกิริยาระหว่างท็อปควาร์กกับฮิกส์
 โบซอนที่รู้จักกันในชื่อค่าสัมประสิทธิ์ยูกาวา จากการพิจารณาแผนภาพไฟน์แมนสำหรับ
 อันตรกิริยานี้แสดงให้เห็นว่าค่าสัมประสิทธิ์นี้มีผลต่อภาคตัดขวางและสมบัติทางจลนศาสตร์
 ของท็อปควาร์กที่เกิดจากอันตรกิริยานี้ ผลที่เกิดขึ้นจากการเปลี่ยนแปลงค่าสัมประสิทธิ์นี้
 สามารถให้ข้อมูลใหม่สำหรับวิธีการวัดค่าสัมประสิทธิ์วิธีใหม่ เช่น การใช้โครงข่ายประสาท
 เทียม วิทยานิพนธ์ฉบับนี้นำเสนอการรวมผลการค้นหาอันตรกิริยาที่ให้ท็อปควาร์กสี่ตัว
 จากสถานะสุดท้ายต่าง ๆ และการศึกษาการใช้โครงข่ายประสาทเทียมเพื่อการวัดค่า
 สัมประสิทธิ์ยูกาวาสำหรับท็อปควาร์ก

ภาควิชา ฟิสิกส์

สาขาวิชา ฟิสิกส์

ปีการศึกษา 2566

ลายมือชื่อนิสิต

ลายมือชื่อ อ.ที่ปรึกษาหลัก

ลายมือชื่อ อ.ที่ปรึกษาร่วม

วิทยุณันท์

 ผศ.ดร.

627 10333 23: MAJOR PHYSICS

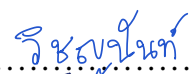
KEYWORDS: STANDARD MODEL / TOP QUARKS / FOUR TOP QUARK PRODUCTION / TOP YUKAWA COUPLING

VICHAYANUN WACHIRAPUSITANAND : SEARCH FOR STANDARD MODEL PRODUCTION OF FOUR TOP QUARKS IN COMPACT MUON SOLENOID DETECTOR AND DESIGNING NEURAL NETWORKS FOR TOP QUARK YUKAWA COUPLING MEASUREMENT FROM FOUR TOP QUARK PRODUCTION. ADVISOR : ASST. PROF. NORRAPHAT SRIMANOBHAS, Ph.D., DISSERTATION COADVISOR : PROF. FREYA BLEKMAN, Ph.D., 304 pp.

The production of four top quarks has been predicted by the Standard Model (SM) and has been considered as one of its important tests. The measurements obtainable from this production, such as its cross section, can be compared with predictions provided by SM. This comparison can determine whether SM is adequate to explain this phenomenon. The four-top quark production is also sensitive to the coupling strength between the top quark and the Higgs boson, also known as the top Yukawa coupling. By considering Feynman diagrams for this production, this coupling value affects both its cross section and the kinematics of the top quarks within the production. These changes caused by top Yukawa coupling offer potential new information for novel techniques, such as neural networks, to measure the coupling value. This dissertation describes the combination of four-top quark production searches in different final states, along with a study of the use of neural networks for the measurement of top quark Yukawa coupling.

Department: Physics

Student's Signature

.....


Field of Study: Physics

Advisor's Signature

.....


Academic Year: 2023

Co-advisor's signature

.....


“Never accept the world as it appears to be,
dare to see it for what it could be.”

– Harold Winston

Acknowledgements

After four years of being a PhD student, three research trips to CERN and in Europe, three Schengen visa applications, one lost luggage, countless nighttime work, sleepless nights, lonely days during COVID lockdown, mistakes, bugs, self-doubts, and heart attacks, I still can't believe I will finally receive my PhD, which I just realised a few days before writing this acknowledgement section that I have wished upon as long as since when I was 15 years old.

First and foremost, I can't thank you enough to my advisors, Assistant Prof Dr Norraphat Srimanobhas and Prof Dr Freya Blekman. Thank you for still believing in me after witnessing me working for two years during my MSc studies, even though I don't believe in myself sometimes, and for being fair and pointing outright if I am doing something wrong. I would also like to thank my BSc academic advisor Assistant Prof Dr Narumon Suwonjandee, who still advises me when I was around in the lab, and my BSc thesis advisor Assistant Prof Dr Burin Asavapibhop, who gave me a chance on working with High-Energy Physics in my first thesis ever.

The work in this thesis is partially supported by the computing infrastructure provided by the National Science and Technology Development Agency, National e-Science Infrastructure Consortium, Chulalongkorn University and the Chulalongkorn Academic Advancement into Its 2nd Century Project, and NSRF via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation [grant numbers B05F650021, B37G660013] (Thailand), and het Fonds Wetenschappelijk Onderzoek ODYSSEUSII programme (Belgium). This research has also received funding support from the NSRF via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation [grant number B05F650021] (Thailand) and Deutsches Elektronen-Synchrotron (Germany).

I would like to thank the four-top quark production search team, especially Sinan Sagir, Emanuele Usai, Wenyu Zhang, Nicolas Stylianou, Nikolas Pervan, Joel Goldstein, Robert Clare, Nicholas Manganelli, Melissa Quinnan, and Valentina Dutta. Thank you for working with me tirelessly for the past four years on this production search and giving me a chance to be right at the centre of the action. I would also like to dedicate this work to Stephen Wimpenny (1956-2022) and Meenakshi Narain (1964-2023) who unfortunately passed away during four years of this work. Thank you for all the advice you gave to me during the past four years over the Internet. I am saddened that I did not have a chance to meet you in person, and I wish I could pay a visit to your relatives and tell them all how brilliant you were.

I would also like to thank the ttX working group, such as Denise Muller, Matthias Komm, Gabriele Milella, and Shahzad Sanjrani. Special thanks to Matthias Komm for your invaluable comments during the final days of my working on my thesis. Thanks to your comments, I am now more confident that the content in my thesis is more accurate.

I would like to thank Martijn Mulders for being my advisor during my time as a CERN Summer Student in 2017 and showing me day-to-day work at CERN as a physicist, affirming that I am on the right path to becoming a physicist working with CERN. I would also like to thank Emanuele Simili and Gurpreet Singh Chahal for introducing me to this field of High Energy Physics. I have no idea where you are, but I wish we would meet again and share the journeys we have had as physicists during so many years we have not met.

I would also like to thank my family, especially my parents, for supporting me on my PhD path unconditionally. I know they would still not understand my work,

but they did not force me to get a job in other fields or to study medicine when I graduated from high school like a lot of families in Thailand out there.

I would like to thank the video game Hogwarts Legacy and all the characters for keeping me sane during the preparation of my thesis, such as Professor Eleazar Fig, Professor Matilda Weasley, Professor Abraham Ronen, Professor Dinah Hecat, Professor Mirabel Garlick, Professor Aesop Sharp, Professor Bai Howin, Madam Chiyo Kogawa, Natsai Onai, Sebastian and Anne Sallow, Garreth Weasley, Poppy Sweeting, Ominis Gaunt, Lucan Brattleby, Everett Clopton, Deek, Lodgok, Sirona Ryan, and Vethis Cormier. Thank you for allowing me to escape into the Wizarding World sometimes and embrace the inner hard-working Hufflepuff in me.

Also thanks to other video games I have been enjoying during four years of my PhD study, such as Animal Crossing: New Horizons (thanks Tom Nook!), Overwatch, Tetris Effect: Connected, Mario Kart 8, SUPERHOT, ABZU, Ancestors: The Humankind Odyssey, No Man's Sky, PC Building Simulator, House Flipper, Control, Warframe, and much, much more. Also, thanks to Ring Fit Adventure for keeping me sane and allowing me to exercise in an RPG world during the COVID lockdown, and all my friends from the Department of Physics in Chulalongkorn University for still talking to me over Discord and playing games together after all these years, especially during the lockdown.

Last but not least, I would like to thank Mahidol Wittayanusorn School and all my friends I met during my three years as a high school student. Thank you for changing my life from being a nerd with little to no friends and being bullied all the time into a happy nerd with a lot more friends and a lot more sanity. Thank you for being my friend regardless of how weird I am, and thank you for still being my friend after all these years. Also, thanks to the school for inspiring young brilliant minds like me in science and technology. Please keep on doing that.

CONTENTS

	Page
Abstract (Thai)	iv
Abstract (English)	v
Acknowledgements	vii
List of Tables	xvii
List of Figures	xxii
0 Individual contribution to the CMS Collaboration by the author . . .	1
1 Introduction and background theory	4
1.1 History of fundamental particle studies	4
1.2 Beginnings of the Standard Model	7
1.3 The Standard Model itself	10
1.3.1 Quarks	10
1.3.2 Leptons	14
1.3.3 Bosons	16
1.4 Mathematical formulation of the Standard Model	17
1.4.1 Symmetries and why we need them	20
1.4.2 Spontaneous Symmetry Breaking (SSB)	21
1.4.3 Brout-Englert-Higgs mechanism	23
1.4.4 Formulation of W and Z bosons	26
1.4.5 Yukawa Lagrangian	28
1.4.6 Electroweak theory	29
1.4.7 Quantum Chromodynamics primer	33

1.5	Problems with the Standard Model, and Beyond Standard Model theories	35
1.6	Top quarks	37
1.7	Yukawa couplings, especially the top quark Yukawa coupling	39
1.7.1	Yukawa coupling measurements	40
1.8	Four-top quark production	42
1.8.1	Recent searches of four-top quark production from the ATLAS and CMS Experiments	45
1.8.2	Challenges to the search for four top quarks	48
1.8.3	Yukawa coupling measurement based on four-top quark production	51
1.8.4	BSM interpretations based on four-top quark production . . .	51
2	The Compact Muon Solenoid detector	54
2.1	The Large Hadron Collider	54
2.2	The overall design of the CMS detector	56
2.3	Superconducting solenoid	57
2.4	Silicon trackers	58
2.5	Electromagnetic calorimeter (ECAL)	58
2.6	Hadronic calorimeter (HCAL)	60
2.7	Muon chamber	60
2.8	Trigger	61
2.9	Reconstruction of events using Particle Flow (PF) algorithm	61
2.10	Data Quality Monitoring (DQM) and Data Certification (DC)	64
2.10.1	Machine Learning applications for DQM	65
2.11	Monte Carlo simulation	66
3	Introduction to Machine Learning and Neural Networks	70
3.1	What is Machine Learning?	70

3.2	What are Boosted Decision Trees?	72
3.3	What are neural networks?	73
3.3.1	Perceptrons	73
3.3.2	Neurons	74
3.3.3	Constructing neural networks from neurons	76
3.3.4	Another interpretation of neural networks via linear models	78
3.4	How do neural networks learn?	80
3.4.1	Training and validating dataset	80
3.4.2	Overtraining versus overfitting	81
3.4.3	One epoch of network training	82
3.4.4	Loss function	83
3.4.4.1	Mean squared error (MSE)	83
3.4.4.2	Mean absolute error (MAE)	84
3.4.4.3	Binary cross-entropy	84
3.4.4.4	Categorical cross-entropy	84
3.4.5	Regularizers	84
3.4.6	Backpropagation	85
3.4.7	Optimization with optimizers	86
3.4.7.1	Gradient descent and Stochastic gradient descent (SGD) optimizer	87
3.4.7.2	Momentum SGD	88
3.4.7.3	RMSprop	88
3.4.7.4	Adam	89
3.4.8	Vanishing gradient problem	89
3.4.9	Dropout layers	90
3.5	How do we measure the performance of neural networks?	90
3.5.1	True positives, true negatives, false positives, and false negatives	91

3.5.2	Accuracy	92
3.5.3	Receiver operating characteristic (ROC) curve and area under ROC curve (AUC)	93

4 Search for four top quark production in Run 2 data of CMS

	in individual final states	95
4.1	Overview of final states and analysis strategy	96
4.2	Existing search results from CMS Experiment	98
4.2.1	2016 single-lepton and opposite-sign dilepton	99
4.2.1.1	Main background estimation	99
4.2.1.2	Baseline selection	100
4.2.1.3	Analysis technique used	101
4.2.1.4	Systematic uncertainties	102
4.2.1.5	Results and shortcomings of this analysis	106
4.2.2	Run 2 same-sign dilepton and multilepton	107
4.2.2.1	Main background estimation	108
4.2.2.2	Baseline selection	109
4.2.2.3	Analysis technique used	110
4.2.2.4	Systematic uncertainties	111
4.2.2.5	Results from the same-sign dilepton and multilepton final state	112
4.3	Run 2 single-lepton	113
4.3.1	Main background estimation	113
4.3.2	Baseline selection	116
4.3.3	Resolved top tagger	117
4.3.4	Analysis technique used	117
4.3.5	Systematic uncertainties	118
4.3.6	Results from the single-lepton final state	120

4.4	2017+2018 opposite-sign dilepton	121
4.4.1	Main background estimation	122
4.4.2	Baseline selection	122
4.4.3	Analysis technique used	123
4.4.4	Systematic uncertainties	124
4.4.5	Results from the opposite-sign dilepton final state	125
4.5	Run 2 all-hadronic	127
4.5.1	Main background estimation	127
4.5.2	Baseline selection	132
4.5.3	Boosted and resolved top tagging	132
4.5.4	Analysis technique used	133
4.5.5	Systematic uncertainties	134
4.5.6	Results from the all-hadronic final state	137
4.6	Summary of results from individual final states	137
5	Combination of search results for four top quark production	140
5.1	Why do we combine final states?	140
5.2	Profile likelihood fit	142
5.2.1	Signal strength measurement	145
5.2.2	Search significance	146
5.2.3	Asimov dataset	147
5.3	Systematic uncertainties	149
5.3.1	Types of systematic uncertainties, mathematically speaking	150
5.3.2	Correlating systematic uncertainties	151
5.4	Systematic uncertainties considered in combination	151
5.4.1	Theoretical uncertainties	152
5.4.1.1	$t\bar{t} + b\bar{b}$ normalisation	152
5.4.1.2	Monte Carlo cross section	155

5.4.1.3	Initial and final state radiation (ISR/FSR)	155
5.4.1.4	Parton Distribution Function (PDF)	156
5.4.1.5	Renormalisation and factorisation scale variations μ_R and μ_F	156
5.4.2	Experimental uncertainties	157
5.4.2.1	Luminosity uncertainty	157
5.4.2.2	Level-1 trigger prefiring	158
5.4.2.3	Pileup	159
5.4.3	Reconstruction uncertainties	161
5.4.3.1	Jet Energy Scale (JES)	161
5.4.3.2	Jet Energy Resolution (JER)	163
5.4.3.3	b tagging	163
5.4.3.4	Lepton scale factors	164
5.5	Applying theoretical uncertainties for signal strength and cross section measurement	164
5.6	Combination results	165
5.6.1	Expected signal strength, cross section, and search significance	165
5.6.2	Observed signal strength, cross section, and search significance	167
5.6.3	Negative log-likelihood scans	170
5.6.4	Nuisance parameter pulls and impacts	172
5.6.5	Channel compatibility tests	174
5.7	Comparison to recent results from CMS and ATLAS experiments . .	176
6	Top Yukawa coupling measurement	180
6.1	Overview of the study	180
6.2	Machine Learning techniques used	181
6.3	Simulated event samples	182
6.4	Event reconstruction and preselection criteria	184

6.5	NN-c training	184
6.6	NN- y_t training	191
6.6.1	$t\bar{t}t\bar{t}$ sample splitting	191
6.6.2	Input variables to NN- y_t	196
6.7	Direct measurement of y_t from $t\bar{t}t\bar{t}$ using NN- y_t	202
6.7.1	Output distribution construction	203
6.7.2	Systematic uncertainties included in the measurement	206
6.7.3	Direct measurement using the Asimov dataset	210
6.7.3.1	Adding $t\bar{t}H$ -enriched regions to the direct measurement	214
6.7.3.2	Two-dimensional fit of signal strength μ and y_t/y_t^{SM}	217
6.7.3.3	Two-dimensional fit of y_t/y_t^{SM} for $t\bar{t}H$ and $t\bar{t}t\bar{t}$	219
6.8	Indirect measurement of y_t via cross-section measurement	221
6.9	Projections of the measurement to higher integrated luminosity	233
6.10	Comparison of y_t measurements to existing measurements from other studies	238
7	Summary and Future Studies	241
7.1	Possible future studies on four-top quark production	241
7.2	Possible future studies for top Yukawa coupling measurement	243
7.3	Summary and final remarks	247
	References	249
	Biography	273

LIST OF TABLES

Table	Page
1.1 Particles in the Standard Model and corresponding fundamental forces of nature [24]. Gravitational force does not correspond to any particle in the Standard Model and is best explained by the curvature of the spacetime according to Einstein’s General Relativity.	10
1.2 Properties of the six quarks in the Standard Model. Mass measurements are determined using $\overline{\text{MS}}$ scheme, except top quark mass which is extracted from event kinematics [17].	12
1.3 Electric charge and lepton numbers L_e , L_μ , and L_τ for each lepton. Note that corresponding antipartners have the lepton number of the corresponding flavour as -1.	15
1.4 Cross section upper limit for four-top quark production searches by the CMS and ATLAS Collaborations, using data collected in 2012-2016. All results reported in this table do not have observed and expected cross section measurements reported.	46
1.5 Cross section measurement for four-top quark production searches by the CMS Collaboration, using data collected in 2016-2018. The SSDL+ML analysis in the last row uses an outdated Run 2 integrated luminosity of 137 fb^{-1} , compared to the latest calculation of Run 2 integrated luminosity of 138 fb^{-1} [76; 77; 78].	46
1.6 Cross section measurement for four-top quark production searches by the ATLAS Collaboration, using data collected in 2016-2018.	47
3.1 True positives, true negatives, false positives, and false negatives.	92
4.1 Summary of analyses involved in the latest search for four-top quark production.	96

4.2	Measured signal strength and cross section for single-lepton final state, broken by data-taking year.	121
4.3	Expected signal strength for opposite-sign dilepton final state, broken by data-taking year.	126
4.4	Observed signal strength and cross section for opposite-sign dilepton final state, broken by data-taking year.	127
4.5	12 event categories used in Run 2 all-hadronic analysis, separated by the number of resolved tops N_{RT} , the number of boosted tops N_{BT} , and H_T . In this analysis, the preselection requirement of $H_T > 700$ GeV is applied before categorisation.	134
4.6	Measured signal strength and cross section from individual analyses, including Run 2 SL (Section 4.3), 2017+2018 OSDL (Section 4.4), and Run 2 all-hadronic analysis (Section 4.5), as well as 2016 OSDL analysis (Section 4.2.1) and Run 2 SSDL&ML analysis (Section 4.2.2). The uncertainty values reported are statistical uncertainties and systematic uncertainties respectively, except for the original values of 2016 OSDL where only the total uncertainty is reported.	138
4.7	Expected and observed significance from individual analyses.	139
5.1	Summary table on nuisance parameter correlation assumptions, for theoretical uncertainties.	153
5.2	Summary table on nuisance parameter correlation assumptions, for experimental uncertainties.	153
5.3	Summary table on nuisance parameter correlation assumptions, for reconstruction uncertainties.	154
5.4	Luminosity uncertainty correlation scheme for Run 2 combination, as recommended by Luminosity Physics Object Group.	158

5.5	Integrated luminosity from the latest measurements performed in Ref. [76; 77; 78], compared to integrated luminosity values used in 2016 OSDL and Run 2 SSDL&ML analyses.	158
5.6	New luminosity uncertainty scheme applied to Run 2 SSDL&ML analysis, based on integrated luminosity value in Table 5.5.	159
5.7	b-tagging algorithms used in each analysis. Numbers in parenthesis denote corresponding section numbers in Chapter 4.	163
5.8	Expected signal strength and search significance for individual analysis and the combination. Results from 2016 OSDL [1] have been recalculated to allow negative signal strength values and adjust the four-top quark production cross section up to 12 fb from 9.2 fb (See Section 4.2.1.5).	166
5.9	Observed signal strength and search significance for individual analysis and the combination. Results from 2016 OSDL [1] have been recalculated to allow negative signal strength values and adjust four-top quark production cross section up to 12 fb from 9.2 fb (See Section 4.2.1.5).	167
5.10	Observed signal strength and cross section from individual analyses and combinations, with uncertainties broken down into statistical and systematic uncertainties. The uncertainty values reported are statistical uncertainties and systematic uncertainties respectively. The values for the 2016 OSDL analysis in this table have been recalculated to allow negative signal strength values and adjust four-top quark production cross section up to 12 fb from 9.2 fb (See Section 4.2.1.5).	168
5.11	Observed significance calculated from asymptotic limit method and generated toys, from 2017+2018 OSDL, Run 2 SL, and Run 2 all-hadronic analyses.	170

5.12	Test results for the compatibility of each final state, with their corresponding p -values [2].	176
6.1	Area under ROC curve (AUC) calculated from training and testing datasets and three output nodes of the NN-c.	190
6.2	Similarity index (Equation 6.1) for 10 variables with the least similarity index, calculated with different splitting schemes, sorted by the values from probability comparison splitting method. Note that the similarity index from the weighted randomisation method may fluctuate since the distributions are randomised.	193
6.3	Similarity index (Equation 6.1) for 10 variables with the least similarity index, calculated and sorted from the weighted randomisation method. Note that the similarity index from the weighted randomisation method may fluctuate since the distributions are randomised.	194
6.4	Similarity index (Equation 6.1) for 10 variables with the least similarity index, calculated and sorted from the direct weight comparison method.	194
6.5	Jet and b-tagged jet multiplicities used in the indirect measurement of y_t via cross section measurement.	222
6.6	Comparisons between the direct and indirect measurements of y_t/y_t^{SM} , projected to different integrated luminosity values. Negative values show the change in the y_t/y_t^{SM} range compared to the 2017 integrated luminosity.	237
6.7	Comparison of <i>expected</i> y_t/y_t^{SM} measurements between the direct measurements via $t\bar{t}t\bar{t}$ in this work and the measurement presented in Ref. [62].	239

7.1	Measured cross section, expected and observed search significance from individual analyses presented in this thesis, compared to new results reported by CMS [83] and ATLAS [84]. The reported uncertainty values for the cross section are statistical uncertainties and systematic uncertainties respectively.	242
7.2	Expected significance from the SSDL&ML analysis performed using 2016 data [79] and its projection to 300 fb^{-1} and 3000 fb^{-1} , as shown in Ref. [177]. “stat. only” refers to the scenario with only statistical uncertainties included. “Run 2” refers to the scenario where only statistical uncertainties are scaled by the factor of $1/\sqrt{L/L_{\text{ref}}}$. “YR18” refers to the scenario where several systematic uncertainties are scaled down according to Ref. [177] and [191].	244

LIST OF FIGURES

Figure	Page
1.1 Particles in the Standard Model. Mass measurements are shown as reported in Ref. [17]. Quark masses, except top quark mass, are determined using $\overline{\text{MS}}$ scheme. Neutrino masses are determined via tritium decay at a 90% confidence level [17].	11
1.2 The Standard Model shirt, often sold as a souvenir at CERN, depicts fundamental terms that build up SM, including Field strength tensor, Dirac Lagrangian, Yukawa Lagrangian, Higgs kinetic term, and Higgs potential term.	18
1.3 Visualization of the potential term $\frac{\lambda}{4} (\phi^* \phi - \nu^2)^2$, where ϕ is a complex scalar field ϕ . The symmetry point at $\phi = 0$, or the tip of the Mexican hat, does not have the lowest potential energy. The ground state for the scalar field, or the ridge of the Mexican hat, gives the lowest potential energy and breaks the symmetry. Diagram derived from Ref. [49]. . . .	22
1.4 Best fit values of six Yukawa couplings based on search results for $H \rightarrow \mu^+ \mu^-$ [61], combined with previous measurements performed in 2016 [60]. All measured values are consistent with the SM prediction depicted in the dashed line. Diagram taken from Ref. [61].	41
1.5 Negative log likelihood scan for Y_t , based on the electroweak corrections in kinematic distributions of $t\bar{t}$ production. The most probable value obtained from this fit is $Y_t = 1.16^{+0.24}_{-0.35}$ [62].	43
1.6 Leading order Feynman diagrams for four-top quark production. Notice that the third diagram has $t\bar{t}H$ vertices involved. Diagrams derived from Ref. [1] and Ref. [65].	44

1.7	BDT output distributions of four-top quark production compared with background processes ($t\bar{t}$, single-top production, electroweak, $t\bar{t}$ with Higgs or Z boson, $t\bar{t} + W$, and $t\bar{t}$ with two vector bosons (WW, WZ, or ZZ), from the single-lepton final state with muons (upper panel) and electrons (lower panel), and 10 or more jets [1]. BDT bins are arranged to have an equal amount of background events in each bin. Notice a large amount of postfit uncertainty in the 10-jet, 4+ b-jets category (rightmost section).	50
1.8	Observed four-top quark cross section and its upper limit, based on Ref. [65], compared to predicted values with top quark Yukawa coupling varied. The upper limit is obtained by comparing the observed upper limit (hashed line) against the predicted cross section (shaded band around the dashed line).	52
2.1	The CERN accelerator complex, depicting LINAC, Booster, PS, SPS, and LHC, along with other experiments at CERN [97].	55
2.2	A transverse slice of the CMS detector as of 2008, showing several components and energy residues caused by several types of particles passing through them [99].	57
2.3	CMS Pixel Tracker schematics comparison between the configuration used since 2009 (labelled “current” in this diagram) and the configuration since 2017 (labelled “upgrade” in this diagram) configurations [103]. The 2017 configuration of the tracker in this diagram is used in the Run 2 data-taking period.	59
2.4	CMS Silicon Strip Tracker Schematics on a longitudinal plane, showing a quarter of the whole Strip Tracker system [105].	59
2.5	Cumulative integrated luminosity for proton-proton collisions during 2016-2018, showing the amount of data delivered by LHC, recorded by CMS, and certified for physics [114].	65

2.6	Reconstruction of occupancy plots from pixel barrel layer 1 using residual networks [4], a class of neural networks. Notice the reconstruction failure in the middle row (run 297100, lumi section 230), which can be classified as an anomaly in the silicon pixel tracker. The top and bottom row represents lumi sections 229 and 233 with good reconstruction from the model, which can be classified as normal operations in the tracker [5; 115].	67
3.1	Sigmoid (Equation 3.4) and tanh (Equation 3.5) activation functions for $-5 < x < 5$	76
3.2	Rectified learning unit (ReLU, Equation 3.6), scaled exponential linear units (SELU, Equation 3.7), and leaky rectified linear units (LeakyReLU, Equation 3.8) activation functions for $-1 < x < 1$, with $\lambda = 1$, $\alpha = 0.3$ for SELU, and $\alpha = 0.1$ for LeakyReLU. Since $\lambda = 1$ for SELU, all three functions have the same output in the $x > 0$ region. . . .	77
3.3	Examples of receiver operating characteristic (ROC) curve. The green curve represents the ideal curve with perfect signal acceptance at zero background acceptance. The orange curve represents the typical shape of the curve for a model, and the red curve represents the curve obtained by random guessing.	94
4.1	BDT output distributions from 2016 single-lepton analysis, separated by the number of jets and the number of b-tagged jets [1].	103
4.2	BDT output distributions from three flavour combinations (from top to bottom, $\mu\mu$, $e\mu$, and ee) in 2016 opposite-sign dilepton analysis, separated by the number of jets and the number of b-tagged jets [1]. . . .	104
4.3	Event yield distribution of signal regions determined by BDT classifier from Run 2 SSDL&ML analysis [65].	111

4.4	Distribution ratio of background and signal events in Run 2 SL analyses per category, split by the number of resolved top quarks, the number of b-tagged jets, and the number of jets. Notice the dominance of $t\bar{t}$ background events, which are further split into ttbb ($t\bar{t} + \geq 1b$ in the plot) and ttnobb ($t\bar{t} + 0b$ in the plot). Figure derived from auxiliary materials of [2].	115
4.5	BDT output distributions of three jet multiplicity categories, defined by number of resolved top quarks (N_{RT}), number of b-tagged jets (N_b), and number of jets (N_j). Figure derived from [2].	121
4.6	Jet multiplicity distributions for events with 2, 3, and 4 b-tagged jets in 2017+2018 OSDL analyses. Figures derived from Ref. [2].	126
4.7	Example control and signal regions of two <i>uncorrelated</i> variables v_1 and v_2 for ABCD method. The control regions A, B, and C are used to estimate event yields in signal region D.	128
4.8	Example control and signal regions of two <i>uncorrelated</i> variables v_1 and v_2 for extended ABCD method. In this method, two additional control regions X and Y can be added to the estimation of the signal region D.	130
4.9	Signal region (SR), validation region (VR), signal control region (SCR), and validation control region (VCR) used in an estimation of data-driven backgrounds based on extABCD and ABCDnn methods [2].	131
4.10	BDT output distributions from two of the most sensitive categories: $N_{RT} = 1, N_{BT} \geq 1, H_T > 1400 \text{ GeV}$, and $N_{RT} \geq 2, H_T > 1100 \text{ GeV}$ [2].	135
4.11	BDT output distributions from two of the most sensitive categories: $N_{RT} = 1, N_{BT} \geq 1, H_T > 1400 \text{ GeV}$, and $N_{RT} \geq 2, H_T > 1100 \text{ GeV}$ in validation region [2].	138
4.12	Expected and observed significance from individual analyses, derived from Table 4.7.	139

5.1	Distributions of the mean number of interactions per crossing for proton-proton collisions during Run 2 data-taking period between 2015-2018 [184].	160
5.2	Observed signal strength measurement from individual analyses and the combination, with uncertainties broken down into statistical and systematic uncertainties [2].	168
5.3	Observed cross section measurement from individual analyses and the combination, with uncertainties broken down into statistical and systematic uncertainties [2].	169
5.4	Post-fit NLL scans for single-lepton, opposite-sign dilepton, and all-hadronic final state, separated per year and combined [2] Analysis from SSDL&ML final state cannot be separated per year, and thus no NLL scan can be performed per each year of data taking for that final state. Refer to Figure 5.5 for NLL scan from SSDL&ML final state. . . .	171
5.5	Post-fit NLL scans for all final states and the combination of all final states [2].	172
5.6	Nuisance parameter impacts plot for the final combination, listing nuisance parameters with the most impact.	175
5.7	Measurement of signal strength for $t\bar{t}H$ process from single-lepton and dilepton final states. Single-lepton final state offers the measurement with less statistical uncertainty [188].	179
6.1	Distributions for input variables of classification model from $t\bar{t}t\bar{t}$, $t\bar{t}$, and $t\bar{t}H$ samples, normalised.	187
6.2	Loss evolution during the training of NN-c, indicating that there is no overtraining since the testing loss is generally not greater than the training loss, as explained in Section 3.4.1.	189

6.3	Output distribution of NN-c from training and testing datasets, separated by three output nodes and three processes. The distributions from training and testing datasets are comparable.	189
6.4	ROC curve of NN-c from training and testing datasets, separated by three output nodes. The x -axis represents signal acceptance, while the y -axis represents background acceptance.	190
6.5	Distributions of example variables with different splitting methods. Note that the distributions from the weighted randomisation method (middle column) may fluctuate since the distributions are randomised. .	195
6.6	Correlation matrix for NN- y_t input variables for SM class.	197
6.7	Correlation matrix for NN- y_t input variables for non-SM class.	198
6.8	Evolution of area under ROC curve over 100 epochs of training, calculated with 20 training trials, each with a different random seed for the training dataset.	199
6.9	Output distribution (left) and ROC curve (right) of NN- y_t , calculated from training and testing datasets, <i>normalised</i>	200
6.10	Output distribution of NN- y_t , calculated on $t\bar{t}t\bar{t}$ dataset with $y_t/y_t^{\text{SM}} = 1.16, 1.74, 2.32$ (shown in the plot as $y_t = 200, 300, 400$, which are the value used in generator configurations) and SM value, <i>normalised</i>	201
6.11	Output distributions of NN- y_t calculated from $t\bar{t}$, $t\bar{t}H$, and $t\bar{t}t\bar{t}$ processes, normalised to 2017 integrated luminosity of 49.81 fb^{-1}	204
6.12	Output distributions of NN- y_t calculated from $t\bar{t}t\bar{t}$ process, where each bin in the histogram is normalised based on the event yield of SM y_t value. Note that the normalisation method applied in this figure is different from Figure 6.13.	205

6.13	Output distributions of NN- y_t calculated from $t\bar{t}t\bar{t}$ process, where each distribution from different y_t value is normalised to 100 events. Note that the normalisation method applied in this figure is different from Figure 6.12.	206
6.14	Fourth-order polynomial fitting to y_t for $t\bar{t}t\bar{t}$ yield per each histogram bin. The x -axis represents the ratio between y_t and SM y_t value, or y_t/y_t^{SM} . Notice that the fitting shows excellent agreement with $R^2 = 1.000$.	207
6.15	Renormalisation and factorisation uncertainty templates for $t\bar{t}t\bar{t}$ (upper row), $t\bar{t}$ (middle row), and $t\bar{t}H$ (lower row).	209
6.16	ISR uncertainty templates for $t\bar{t}t\bar{t}$ (left), $t\bar{t}$ (middle), and $t\bar{t}H$ (right). . . .	209
6.17	JES uncertainty templates for $t\bar{t}t\bar{t}$ (upper row), $t\bar{t}$ (middle row), and $t\bar{t}H$ (lower row), before (left) and after (right) applying LOWESS smoothing algorithm [169; 170].	211
6.18	JER uncertainty templates for $t\bar{t}t\bar{t}$ (upper row), $t\bar{t}$ (middle row), and $t\bar{t}H$ (lower row), before (left) and after (right) applying LOWESS smoothing algorithm [169; 170].	212
6.19	FSR uncertainty templates for $t\bar{t}t\bar{t}$ (upper row), $t\bar{t}$ (middle row), and $t\bar{t}H$ (lower row), before (left) and after (right) applying LOWESS smoothing algorithm [169; 170].	213
6.20	Negative log-likelihood scans over y_t/y_t^{SM} , showing the range of y_t/y_t^{SM} values under 68% confidence level.	215
6.21	Output distributions of NN- y_t calculated from $t\bar{t}$, $t\bar{t}H$, and $t\bar{t}t\bar{t}$ processes in $t\bar{t}H$ -enriched region, normalised to 2017 integrated luminosity of 49.81 fb^{-1}	216
6.22	Two-dimensional NLL scan of $\mu_{t\bar{t}t\bar{t}}$ versus y_t/y_t^{SM} , based on NN- y_t output distribution used for the direct measurement.	218

6.23 Two-dimensional NLL scan of two y_t parameters affecting $t\bar{t}t\bar{t}$ yield (x -axis) and $t\bar{t}H$ yield (y -axis), based on NN- y_t output distribution used for the direct measurement. The red line represents points where both parameters have the same value according to SM.	220
6.24 NN- y_t output distribution used in the indirect measurement of y_t via cross section measurement.	222
6.25 JES uncertainty template for $t\bar{t}t\bar{t}$ process, before (top) and after (bottom) applying LOWESS smoothing algorithm [169; 170].	223
6.26 JES uncertainty template for $t\bar{t}$ process, before (top) and after (bottom) applying LOWESS smoothing algorithm [169; 170].	223
6.27 JES uncertainty template for $t\bar{t}H$ process, before (top) and after (bottom) applying LOWESS smoothing algorithm [169; 170].	224
6.28 JER uncertainty template for $t\bar{t}t\bar{t}$ process, before (top) and after (bottom) applying LOWESS smoothing algorithm [169; 170].	224
6.29 JER uncertainty template for $t\bar{t}$ process, before (top) and after (bottom) applying LOWESS smoothing algorithm [169; 170].	225
6.30 JER uncertainty template for $t\bar{t}H$ process, before (top) and after (bottom) applying LOWESS smoothing algorithm [169; 170].	225
6.31 FSR uncertainty template for $t\bar{t}H$ process, before (top) and after (bottom) applying LOWESS smoothing algorithm [169; 170].	226
6.32 FSR uncertainty templates for $t\bar{t}t\bar{t}$ (top) and $t\bar{t}$ (bottom) processes.	226
6.33 ISR uncertainty templates for $t\bar{t}t\bar{t}$ (top), $t\bar{t}$ (middle), and $t\bar{t}H$ (bottom) processes.	227
6.34 Renormalisation and factorisation uncertainty templates (varying μ_R) for $t\bar{t}t\bar{t}$ (top), $t\bar{t}$ (middle), and $t\bar{t}H$ (bottom) processes.	228
6.35 Renormalisation and factorisation uncertainty templates (varying μ_F) for $t\bar{t}t\bar{t}$ (top), $t\bar{t}$ (middle), and $t\bar{t}H$ (bottom) processes.	229

6.36	Renormalisation and factorisation uncertainty templates (correlating μ_F and μ_R) for $t\bar{t}t\bar{t}$ (top), $t\bar{t}$ (middle), and $t\bar{t}H$ (bottom) processes.	230
6.37	$t\bar{t}t\bar{t}$ cross section comparison between theoretical predictions with varying y_t versus expected cross section measured with NN- y_t output distribution. Predicted cross section at SM is derived from Ref. [63]. . .	231
6.38	Negative log-likelihood scans over y_t/y_t^{SM} , projected with different integrated luminosity values.	234
6.39	Two-dimensional NLL scan of $\mu_{t\bar{t}t\bar{t}}$ versus y_t/y_t^{SM} , based on NN- y_t output distribution used for the direct measurement, <i>with projections to higher integrated luminosity values</i>	235
6.40	Two-dimensional NLL scan of two y_t parameters affecting $t\bar{t}t\bar{t}$ yield (x -axis) and $t\bar{t}H$ yield (y -axis), based on NN- y_t output distribution used for the direct measurement, <i>with projections to higher integrated luminosity values</i> . Red line represents points where both parameters have the same value according to SM.	235
6.41	$t\bar{t}t\bar{t}$ cross section comparison between theoretical predictions with varying y_t versus expected cross section measured with NN- y_t output distribution, <i>with projections to higher integrated luminosity values</i> . Predicted cross section at SM is derived from Ref. [63].	236

Chapter 0

INDIVIDUAL CONTRIBUTION TO THE CMS COLLABORATION BY THE AUTHOR

The CMS Collaboration is composed of a large group of physicists and engineers, working on different aspects of the Compact Muon Solenoid detector, from hardware design, maintenance, software commissioning, simulation, and beyond. Therefore, papers under the CMS Collaboration are traditionally published with an alphabetic author list, or collectively attributed to the CMS Collaboration. Contributions by the author are based on the following publications:

- Search for four-top quark production in single-lepton and opposite-sign dilepton final state in 2016 data within CMS detector [1]. In this analysis, event distributions before the fit to nuisance parameters (so-called prefit) are required as a cross-check before fitting to the actual data. The author's contributions to this study are prefit distribution plots for control regions of the single-lepton final state. The author also performed a feasibility study of neural networks as a discriminator in this final state but was not used in favour of Boosted Decision Trees (BDT) discriminators in the final result. This contribution has been published in the Journal of High Energy Physics [1].
- Search for four-top quark production in all final states in Run 2 data within the CMS detector [2]. The analysis is first performed individually by other collaborators, which is described in Chapter 4, and then combined using the maximum likelihood fit method [3], which will be further discussed in

Chapter 5. The combination part of this analysis, detailed in Chapter 5, is performed by the author and is included in the final result reported by the CMS Collaboration, published in Physics Letters B [2]. Furthermore, the author is also involved in designing main and supplementary figures for the publication and supplying the data from all main figures and tables to the HEPData platform.

- Studies of machine learning applications for data quality monitoring (DQM) workflow, which will be further discussed in Section 2.10.1. In addition to shifts as a DQM subsystem on-call person, the author has contributed to a study of applications of several machine learning techniques, such as residual networks [4], for data certification at lumi section granularity. A poster presentation summarising the studies is presented at the 20th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT2021). The corresponding proceeding article is also published in the Journal of Physics: Conference Series [5].

The author would also like to attribute individual new analyses on the search for the four-top quark production to the following individuals:

- **Run 2 single-lepton analysis:** Ulrich Heintz, Daniel Li, Meenakshi Narain, Nikolas Pervan, Sinan Sagir, Emanuele Usai, and Wenyu Zhang from Brown University, Freya Blekman¹ from Deutsches Elektronen-Synchrotron (DESY) and Universität Hamburg, Nicolas Stylianou² and Joel Goldstein from University of Bristol
- **2017+2018 opposite-sign dilepton analysis:** Steve Wimpenny, Robert Clare, and Nicholas Manganelli from the University of California Riverside

¹Formerly at Vrije Universiteit Brussel.

²Also affiliated with Vrije Universiteit Brussel.

- **Run 2 all-hadronic analysis:** Melissa Quinnan, Valentina Dutta, and Joseph Incandela from the University of California Santa Barbara, Suyong Choi, Chang Whan Jung, Hayong Oh, and Jae Hyeok Yoo of Korea University

Chapter 1

INTRODUCTION AND BACKGROUND THEORY

What is matter around us made of? This question has been repeatedly pondered by philosophers and scientists since Ancient Greece and further developed through subsequent scientific discoveries. Several models of fundamental particles have been made throughout history, from Democritus's indivisible atomic model, Thomson's plum pudding model, and Rutherford's atomic model, all the way to the Standard Model. In this chapter, we will go through the history of fundamental particles, as well as relevant underlying concepts to this thesis's work. Note that the discoveries presented in this chapter are not exhaustive.

1.1 History of fundamental particle studies

The idea of fundamental particles has been formulated since Ancient Greece by a philosopher named Democritus [6]. Along with several other Greek philosophers, he proposed an idea of indivisible particles, the atoms, that can be rearranged to form ever-changing appearances. According to the Atomist doctrine, these particles traverse through the void and can be collided or combined with tiny hooks or barbs (similar to chemical bonds in our current understanding of chemical compounds). Given that there was no adequate scientific apparatus to test this hypothesis from the Atomist doctrine, it is mainly philosophical and does not have any scientific evidence at all.

Fast forward to 1805, John Dalton, a British chemist, revived the idea of atoms, based on the observations of gas behaviours when mixed with water and

other liquids [7]. In 1897, J. J. Thompson discovered electrons via his study of cathode rays [8], which shows that the so-called indivisible atoms are not indivisible after all. Later, in his paper published in 1911, Ernest Rutherford assumed that “the diameter of the sphere of positive electricity is minute compared with the diameter of the sphere of influence of the atom” [9]. In modern terms, this refers to a small positively-charged nucleus compared to the effective diameter of the atom itself, including the electron orbital cloud. In his Bakerlean lecture in 1920 [10], Rutherford also suspected “the possible existence of an atom of mass 1 which has zero nucleus charge”, and in 1932, such particles, or neutrons, were discovered by James Chadwick by bombarding beryllium with protons and alpha particles [11].

And so the deluge of newly-discovered particles, such as mesons and baryons, arrived, and some physicists observed a pattern among those newly-discovered “fundamental” particles. In 1961, as a first attempt to explain such patterns, Gell-Mann [12] and Ne’eman [13] independently proposed a framework of strong interactions based on $SU(3)$ symmetry, later named The Eightfold Way in Gell-Mann’s technical report. Based on this framework, several mesons and baryons can be arranged in a geometric pattern according to their electrical charge and a new property called “strangeness” [14]. This early explanation would evolve into the quark model later on.

On the other hand, in 1936 Anderson and Neddermeyer observed particle showers occurring inside a cloud chamber apparatus located 4300 meters above sea level at Pike’s Peak in the Rocky Mountains in North America [15]. The observed showers significantly have more occurrence than the showers observed previously in Pasadena near sea level, and particle showers with electron energies up to 300 MeV are not attributed to protons. Based on the observation by Anderson and Neddermeyer, Street and Stevenson performed another experiment using a

cloud chamber with a lead filter on top to remove shower particles [16]. Evidence presented in the literature shows that there is a new particle with a rest mass of approximately 130 times the mass of an electron (and with an estimated error of about 25%). In modern literature, this particle is later known as a muon, and its mass has been measured by later measurements to be 105 MeV [17].

“Dear radioactive ladies and gentlemen”, Pauli stated in his letter addressed to Hans Geiger and Lise Meitner in 1930, “... I have hit upon a desperate remedy to save the ‘exchange theorem’ of statistics and the energy theorem” [18]. In this letter, he had proposed a particle “that I wish to call neutrons, which have spin $1/2$ and obey the exclusion principle¹... The mass of the neutron must be of the same order of magnitude as the electron mass...”

Several modifications were made to Pauli’s new particle model. Carlson and Oppenheimer modified the idea of Pauli’s neutron into a magnetic neutron, with their calculations published in 1932 [19]. However, after a lengthy calculation, they stated that “there is no experimental evidence [in the form of cloud chamber tracks] for the existence of a particle like the magnetic neutron”.

Later, in the Seventh Solvay Conference in 1933, Pauli reaffirmed his idea presented at a conference in Pasadena in 1931 that the beta particles must be emitted at the same time as “a very penetrating radiation of neutral particles, which has not been observed yet” [18]. Back in that time, neutrons which had been discovered by Chadwick in 1932 [11] were considered *heavy* neutrons, and Pauli announced that “to distinguish them from the heavy neutrons, E. Fermi proposed the name ‘neutrino’” [18].

¹Nowadays known as Fermi statistics [18].

The existence of neutrinos was later confirmed by Cowan et al. in 1956 with an experiment conducted at the Savannah River Plant, now known as the Savannah River Site in South Carolina [20]. The experiment apparatus made out of scintillators and a large tank, enclosed in lead and paraffin, are located underground, providing excellent shielding against reactor neutrons and cosmic rays. Still, neutrons from a reactor were detected via the beta decay reaction²:

$$\nu_{-} + p^{+} \rightarrow \beta^{+} + n^{0}. \quad (1.1)$$

1.2 Beginnings of the Standard Model

In 1964, Gell-Mann [21] and Zweig [22] independently proposed that mesons and baryons are not fundamental themselves, but are comprised of smaller particles called - according to Gell-Mann - quarks (or aces according to Zweig). In the beginning, three flavours of quarks were acknowledged - up, down, and strange.

Based on the model of three quarks and the Cabibbo mechanism, flavour-changing neutral current (FCNC) decay can occur, for instance, in the decay of a particle called long-lived neutral kaon K_L into two muons, $K_L \rightarrow \mu^{+}\mu^{-}$. However, the decay has not been observed. The branching ratio of this decay mode is measured today to be [17]

$$\text{BR}(K_L \rightarrow \mu^{+}\mu^{-}) = (6.84 \pm 0.11) \times 10^{-19} \quad (1.2)$$

which is considered nonexistent. Glashow, Iliopoulos, and Maiani, in response to this phenomenon, proposed a mechanism that requires the fourth quark, charm [23]. With the newly-introduced charm quark, the decay of kaon-long $K_L \rightarrow \mu^{+}\mu^{-}$ now has another diagram involving charm quarks in addition to diagrams involving strange quarks and matrix elements from both diagrams negate each other, resulting in a virtually nonexistent branching ratio [24].

²As written in Ref. [20]. In modern notation, this corresponds to $\bar{\nu}_e + p^{+} \rightarrow e^{+} + n^{0}$

In 1973, Kobayashi and Maskawa showed that the four-quark model could not explain CP-violation interactions, and proposed a six-quark model that gave rise to CP-violating interactions via a component later known as Cabibbo-Kobayashi-Maskawa (CKM) matrix [25]. In their six-quark model, an extra pair of quarks is introduced, resulting in the form of three left-handed doublets and six right-handed singlets. In a different six-quark model proposed by Harari [26], however, the six-quark setup was in the form of a three-quark triplet, up, down, and strange, and a three-quark antitriplet, right, top, and bottom. Ultimately Harari's model faltered in favour of Kobayashi and Maskawa's model, but the names of top and bottom quarks coined by Harari live on.

In 1974, two groups of physicists, one under the lead of Samuel C. C. Ting at Brookhaven National Laboratory in New York, and another under the lead of Burton Richter at Stanford Linear Accelerator Center in California (now known as SLAC National Accelerator Laboratory), discovered an observation of a new particle from an electron-positron annihilation. Both papers from Brookhaven [27] (naming it a heavy particle J) and SLAC [28] (naming it ψ [14]) reported that the new particle has a very narrow mass width, whereas SLAC reported that the "resonance" has a mass of 3.105 ± 0.003 GeV, with the upper limit of the mass with around 1.3 MeV. Modern measurements show that this particle, jointly called J/ψ , has a mass of 3096.900 ± 0.006 MeV, with a mass width of 92.6 ± 1.7 keV [17], suggesting that this particle has a lifetime of 7.1×10^{-21} seconds. However, it was well aware right after the discovery of this particle that its lifetime, due to its very narrow mass width, is much longer than the typical lifetime in the order of 10^{-23} seconds [14]. The paper from Brookhaven also suggested that this particle "may be one of the theoretically suggested charmed particles" [27], and now it is accepted that the J/ψ particle is a meson consisting of a charm quark and an anticharm quark ($c\bar{c}$).

Later, in 1977, invariant mass peaks of 9.4 GeV and 10.0 GeV were observed in dimuon productions generated from 400 GeV proton-nucleus collisions [29], named Υ . In the literature, the new Υ peaks were attributed to a new bound state of a new heavy quark (known later as bottom quark) and its antiquark counterpart (anti-bottom quark). In modern literature, however, these Υ peaks are associated with $\Upsilon(1S)$ and $\Upsilon(2S)$ particles, with masses of 9.5 and 10.0 GeV respectively [17].

The quest for the top quark in the Standard Model started in 1982. At that time it was clear that there must be a sixth quark as a counterpart to the newly discovered bottom quark. The SppS proton-antiproton collider at CERN discovered W and Z bosons in 1982 at masses of around 80-90 GeV [30]. W and Z bosons are gauge vector bosons proposed by Weinberg in his model of leptons in 1967 [31], and at that time the top quark is expected to be lighter than the W boson. Hence, the decay from W bosons into a pair of top quarks. Results from particle colliders at DESY, CERN, SLAC, and the Fermilab, however, show that the lower limit of the top quark mass was increasing, exceeding the mass of W bosons [32]. Finally, in 1995, the top quark was discovered jointly by CDF [33] and D0 [34] collaboration at the Tevatron via top-antitop ($t\bar{t}$) quark production in proton-antiproton collisions.

In 1964, three groups of theoretical physicists, Robert Brout and François Englert [35], Peter Higgs [36], and Gerald Guralnik, Carl Hagen, and Tom Kibble [37], proposed that vector bosons may gain masses by breaking symmetries of a scalar field. The mechanism also predicted that there is a new type of boson, later called Higgs boson, that arises from the perturbation of scalar fields around the vacuum state. (Mathematical formulation of the so-called Brout-Englert-Higgs mechanism will be discussed in Section 1.4.3.) The search for the Higgs boson became complete almost 50 years later, in 2012 when the ATLAS and

Table 1.1: Particles in the Standard Model and corresponding fundamental forces of nature [24]. Gravitational force does not correspond to any particle in the Standard Model and is best explained by the curvature of the spacetime according to Einstein’s General Relativity.

	Electromagnetic force	Weak force	Strong force
Fermions			
Quarks	✓	✓	✓
Leptons (e, μ, τ)	✓	✓	
Neutrinos (ν_e, ν_μ, ν_τ)		✓	
Bosons			
Gluons			✓
W and Z bosons	✓		
Photons	✓	✓	
Higgs boson		(none)	

CMS Collaboration announced the discovery of the new boson with a mass of 125 GeV [38; 39].

1.3 The Standard Model itself

The Standard Model contains three main types of particles: quarks, leptons, and bosons. Quarks and leptons are fermions since they obey the Fermi-Dirac statistics rule forbidding two particles to have the same quantum state at the same position. Bosons on the other hand obey the Bose-Einstein statistics rule allowing the same particles to reside in the same energy level. Fermions have half-integer spins ($1/2$, $3/2$, $5/2$, etc.) while bosons have integer spins. Figure 1.1 shows the comprehensive list of particles in the Standard Model.

1.3.1 Quarks

Currently, there are six quarks in the model: up, down, charm, strange, top, and bottom. Their properties are shown in Table 1.2. Quarks are categorised into three

	QUARKS			BOSONS	
mass	2.2 MeV	1.27 GeV	173 GeV	0	125 GeV
charge	+2/3	+2/3	+2/3	0	0
spin	1/2	1/2	1/2	1	1
	u up	c charm	t top	g gluon	H Higgs
	4.7 MeV	96 MeV	4.18 GeV	0	
	-1/3	-1/3	-1/3	0	
	1/2	1/2	1/2	1	
	d down	s strange	b bottom	γ photon	
	0.511 MeV	106 MeV	1777 MeV	80.4 GeV	
	-1	-1	-1	± 1	
	1/2	1/2	1/2	1	
	e electron	μ muon	τ tau	$W^{\pm}$ W boson	
	< 1.1 eV	< 1.1 eV	< 1.1 eV	91.2 GeV	
	0	0	0	0	
	1/2	1/2	1/2	1	
	ν_e electron neutrino	ν_{μ} muon neutrino	ν_{τ} tau neutrino	Z Z boson	
	LEPTONS				

As reported in Prog. Theor. Exp. Phys. 2022 (2022) 083C01
 Quark masses (except top) calculated using $\overline{\text{MS}}$ scheme
 Neutrino masses determined from tritium decay at 90% CL

Figure 1.1: Particles in the Standard Model. Mass measurements are shown as reported in Ref. [17]. Quark masses, except top quark mass, are determined using $\overline{\text{MS}}$ scheme. Neutrino masses are determined via tritium decay at a 90% confidence level [17].

generations since each generation of quarks forms a weak isospin pair. Up and down quarks are first-generation quarks, charm and strange are second-generation quarks and top and bottom are third-generation quarks.

Quarks experience electromagnetic, weak, and strong interactions. More precisely, these effects are caused by Electroweak theory and Quantum Chromodynamics (QCD) via gauge transformations of $\text{SU}(2) \times \text{U}(1)$ and $\text{SU}(3)$ respectively, which will be discussed later in Section 1.4. In addition to the electric charge of either +2/3 (for up, charm, and top) or -1/3 (for down, strange, and bottom), quarks possess an extra property called colour, which can have one of three values commonly referred to as red, green, or blue. Antiquarks are antiparticle partners of

Table 1.2: Properties of the six quarks in the Standard Model. Mass measurements are determined using $\overline{\text{MS}}$ scheme, except top quark mass which is extracted from event kinematics [17].

Quark	Mass		Charge (e)
up	$2.16^{+0.49}_{-0.26}$	MeV	+2/3
down	$4.67^{+0.48}_{-0.17}$	MeV	-1/3
charm	1.27 ± 0.02	GeV	+2/3
strange	$93.4^{+8.6}_{-3.4}$	MeV	-1/3
top	172.69 ± 0.30	GeV	+2/3
bottom	$4.18^{+0.03}_{-0.02}$	GeV	-1/3

quarks, possessing the same mass with opposite charges ($-2/3$ or $+1/3$) and opposite colour charges (antired, antigreen, or antiblue).

As dictated by QCD, quarks may not be free³ and must be coupled with other quarks, through the process called hadronization, to form larger particles called *hadrons*. Hadrons must be colour-neutral and can be classified into two types as follows:

- **Mesons** comprise a quark and an antiquark, $q\bar{q}$. Based on their constituents, a meson may have a neutral, positive, or negative electric charge. For example, a J/ψ meson consists of a charm quark and an anticharm quark, often denoted as $c\bar{c}$. Since a charm quark has a $+2/3$ electric charge and an anticharm quark has a $-2/3$ electric charge, the J/ψ meson is neutral and can be decayed into a pair of opposite-sign leptons such as $\mu^+\mu^-$. To make a meson colour-neutral, both the quark and the antiquark in a meson must have complementary colour charges, e.g. red with antired, green with antigreen, or blue with antiblue.
- **Baryons** comprise three quarks, qqq . For example, a proton consists of two up quarks and one down quark, denoted as uud , and thus has a positive charge

³Searches for the free quark since 1977 have been conducted. All searches do not show evidence of free quarks so far [17].

(since $2/3 + 2/3 - 1/3 = +1$). A neutron consists of one up quark and two down quarks, denoted as udd , and thus has a neutral charge (since $2/3 - 1/3 - 1/3 = 0$). To make a baryon colour-neutral, all three quarks in a baryon must have a different colour charge, e.g. one quark may have a red colour charge, another quark may have a blue colour charge, and the third quark may have a green colour charge. An antibaryon made up of three antiquarks $\bar{q}\bar{q}\bar{q}$ is also possible.

However, particles can also contain four quarks (tetraquark), five quarks (pentaquark), and beyond, as long as they are colour-neutral. Several discoveries of tetraquark and pentaquark have been made since the discovery of $Z_c(3900)$ by the BESIII Collaboration [40] and the Belle Collaboration [41]. Both collaborations speculated that this particle is charmonium-like, meaning that it is a bound state of a charm quark (c) and an anticharm quark ($\bar{c}$) [42]. It is later speculated [43] and shown [44] that this particle is a tetraquark containing $c\bar{d}\bar{c}u$. Furthermore, in 2015, the LHCb Collaboration also discovered two new resonances within the decay of $\Lambda_b^0 \rightarrow J/\psi K^- p$, which should contain at least *five* quarks $c\bar{c}uud$, and thus are called *pentaquarks* [45]. The discovery of such exotic hadrons is also performed by the CMS Collaboration, such as the observation of three new structures (or resonances) in the $J/\psi J/\psi$ mass spectrum [46]. The mass measurements for these three newly observed structures seem to be compatible with the mass spectrum of the tetraquark containing $c\bar{c}c\bar{c}$.

Since quarks must couple with each other to form hadrons via hadronization, and a pair of quark and antiquark can be created out of a vacuum, a single quark produced from a particle collision forms particle jets containing a large number of colourless particles⁴. Therefore, the signature for quarks in particle detectors is called a particle jet, usually called a jet. Jets can be detected by backtracking

⁴While there is a QCD-motivated model explaining hadronization, it is not definitive [24] and thus will not be discussed in this thesis.

detected particles inside the detector. More information will be discussed in Section 2.9. In physics analyses, one jet is usually considered as a signature for one quark (or one gluon). An exception is a top quark, which decays before hadronizing and cannot be associated with one single jet.

1.3.2 Leptons

There are three types of charged leptons: electrons (e), muons (μ), and tau leptons (τ), along with neutrinos of corresponding flavour: electron neutrinos (ν_e), muon neutrinos (ν_μ), and tau neutrinos (ν_τ). Electrons, muons, and tau leptons have masses and an electric charge of -1 , while neutrinos have a neutral charge and minuscule mass, and are usually considered massless. Antipartners of leptons have an electric charge of $+1$. Antipartners of electrons are also known as positrons, denoted as e^+ . Antipartners of neutrinos are written as $\bar{\nu}_e$, $\bar{\nu}_\mu$, and $\bar{\nu}_\tau$, and still have a neutral electric charge.

Unlike quarks, leptons do not experience strong interaction induced by QCD and thus can be moved freely. Massive leptons experience both electromagnetic and weak interactions, while neutrinos experience only electroweak interactions. This is due to the construction of the electroweak theory that does not include right-handed neutrinos.

In addition to the conservation of electric charges, interactions involving leptons also require the conservation of *lepton numbers*. Each flavour of lepton has its lepton number: that is, electrons and electron neutrinos have $L_e = +1$, muons and muon neutrinos have $L_\mu = +1$, and tau leptons and tau neutrinos have $L_\tau = +1$. Antipartners of these leptons, including antineutrinos without electric charges, always have a lepton number of -1 with a corresponding flavour. Table 1.3 lists the electric charge and lepton number for each lepton.

Table 1.3: Electric charge and lepton numbers L_e , L_μ , and L_τ for each lepton. Note that corresponding antipartners have the lepton number of the corresponding flavour as -1.

Lepton	Electric charge (e)	L_e	L_μ	L_τ
e^-	-1	+1	0	0
μ^-	-1	0	+1	0
τ^-	-1	0	0	+1
ν_e	0	+1	0	0
ν_μ	0	0	+1	0
ν_τ	0	0	0	+1

With this conservation rule, a corresponding flavour of neutrinos must be included in, for example, a W boson decay into leptons, even though one lepton can already satisfy the conservation of electric charges. The resulting decay pathways for both W bosons, following both conservation of electric charges and lepton numbers, are

$$W^+ \rightarrow b + \ell^+ + \nu_\ell \quad (1.3)$$

$$W^- \rightarrow b + \ell^- + \bar{\nu}_\ell \quad (1.4)$$

where ν_ℓ and $\bar{\nu}_\ell$ are the neutrino and the antineutrino with the same lepton flavour as ℓ .

Experimental results from the Super-Kamiokande experiment and the Sudbury Neutrino Observatory (SNO) experiment suggest that neutrinos may have their flavour transformed over long distances, such as from the Sun. This phenomenon is called *neutrino oscillation* and confirms that neutrinos have mass as this phenomenon requires the extension of the Standard Model allowing neutrino masses and flavour mixing like in quarks [47]. The *mass eigenstates* of neutrinos correspond to stationary states of the free particle Hamiltonian (since neutrinos are still free particles). Based on the current explanation, three neutrino flavours

corresponding to each lepton flavour (ν_e , ν_μ , and ν_τ), or *weak eigenstates*, can be written as a superposition of mass eigenstates with corresponding time evolution. As such, neutrino flavours, such as ν_e , may change, or *oscillate* while propagating through space into ν_μ , ν_τ , or remain as ν_e [24].

1.3.3 Bosons

Bosons (in the sense of fundamental particles according to SM, not particles obeying Bose-Einstein statistics) are considered force carriers analogous to three of the four fundamental forces⁵. Photons (γ) represent electromagnetic force, W (having either positive charge W^+ or negative charge W^-) and Z bosons represent weak nuclear force, and gluons (g) represent strong nuclear force. Currently, a particle representing gravity has not been discovered, and gravity is best explained by curvatures in four-dimensional spacetime, which is not in the scope of this thesis. Exceptionally, Higgs bosons do not represent any fundamental force but are byproducts of the Brout-Englert-Higgs mechanism which gives masses via interactions with Higgs bosons.

W and Z bosons are usually involved in weak decays. For example, in the beta decay reaction [14],

$$n \rightarrow p^+ + e^- + \bar{\nu}_e. \quad (1.5)$$

Since protons consist of two up quarks and one down quark, and neutrons consist of two down quarks and one up quark, the interaction above has the decay of the down quark into an up quark and a W^- boson (due to the conservation of electric charge). The W^- boson, in turn, decayed into an electron and an electron antineutrino (corresponding to the conservation of electric charge and lepton number). Since

⁵The four fundamental forces in Physics are gravitational force, electromagnetic force, weak nuclear force, and strong nuclear force.

W bosons have an electric charge, and Z bosons have a neutral electric charge, they are often called *charged current* and *neutral current* respectively.

Gluons are usually involved in QCD processes as a mediator for an interaction between two quarks [14]. While quarks may have one out of three colour charges (red, blue, or green), there are *eight* variants of gluons. This is because of the construction of QCD theory, which is based on SU(3) group symmetry, and the eight gluons come from the generators of the SU(3) group.

Usually, W and Z bosons are collectively called vector bosons. This is because both bosons can be written as a vector form in the electroweak theory. More information shall be discussed under the formulation of electroweak theory in Section 1.4.6. The Higgs boson, on the other hand, is often known as a boson that gives mass. In reality, the Higgs boson is the byproduct of the Brout-Englert-Higgs mechanism, which is the mechanism that gives mass to other particles through spontaneous symmetry breaking (SSB) on local gauge transformations. More information regarding this mechanism will be discussed in Section 1.4.3.

1.4 Mathematical formulation of the Standard Model

In this section, unless otherwise noted, two assumptions will be made.

- Natural units are used. That is, $c = \hbar = 1$, and all quantities are measured in energy units, preferably, MeV, GeV, or TeV. This assumption reduces the complexity in calculations involving c and $\hbar$ constants.

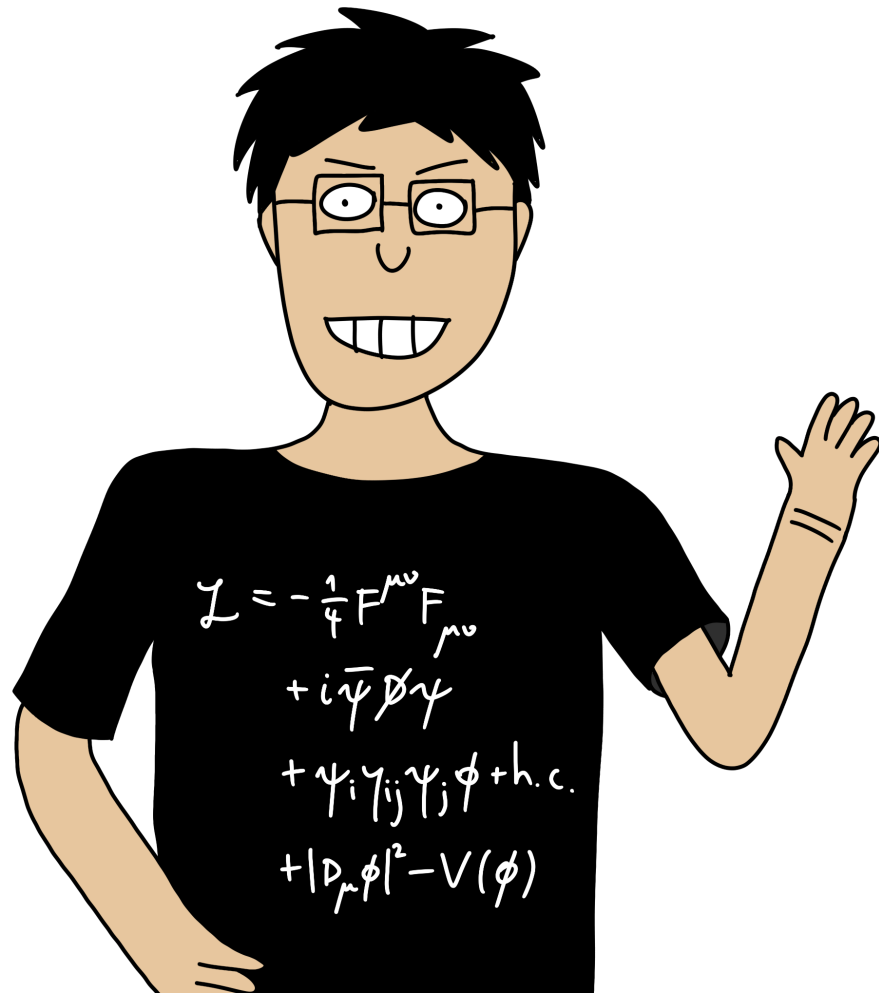


Figure 1.2: The Standard Model shirt, often sold as a souvenir at CERN, depicts fundamental terms that build up SM, including Field strength tensor, Dirac Lagrangian, Yukawa Lagrangian, Higgs kinetic term, and Higgs potential term.

- Einstein's summation convention is used [48]. Under this convention, if an index is repeated twice, it must be summed over the four dimensions. For example, $p_\mu p^\mu = p_0 p^0 + p_1 p^1 + p_2 p^2 + p_3 p^3$.

The famous Standard Model Lagrangian, often shown on souvenirs at CERN, is shown in Figure 1.2. The Lagrangian, although heavily abridged, consists of four components as follows:

- The field strength tensor term encoded the kinetic term for Electroweak bosons (photons, W and Z bosons) and gluons. The term is written as

$$\mathcal{L}_{\text{field strength tensor}} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} \quad (1.6)$$

where the field strength tensor $F^{\mu\nu}$ for *photons* is denoted as

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu \quad (1.7)$$

and A^μ denotes a photon. $F^{\mu\nu}$ is antisymmetric, i.e. $F^{\mu\nu} = -F^{\nu\mu}$ and diagonal components of this tensor are zero. Nonzero components can be written in the form of electric field and magnetic field components in Cartesian coordinates as

$$F^{\mu\nu} = \begin{bmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & -B_3 & B_2 \\ E_2 & B_3 & 0 & -B_1 \\ E_3 & -B_2 & B_1 & 0 \end{bmatrix} \quad (1.8)$$

Usually, field strength tensors are associated with photons, but W bosons and gluons can have their field strength tensors as well.

- The Dirac Lagrangian term explains the fermions, as well as their interactions with gauge bosons. It is in the form of

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}\not{D}\psi + \text{h.c.} \quad (1.9)$$

where h.c. refers to its Hermitian conjugate, and $\not{D} = \gamma^\mu D_\mu$. The most important part of this Dirac Lagrangian is the covariant derivative D_μ , which contains gauge fields representing photons, W and Z bosons, as well as gluons.

- The Yukawa coupling term explains the interaction between fermions and the Higgs field. It is written in the form of

$$\mathcal{L}_{\text{Yukawa}} = \psi_i y_{ij} \psi_j \phi + \text{h.c.} \quad (1.10)$$

where ψ represents fermions and ϕ represents Higgs boson.

- The Higgs kinetic term and Higgs potential term explain the interactions between Higgs bosons and gauge bosons, as well as the existence of the Higgs bosons themselves. It is written in the form of

$$\mathcal{L}_{\text{Higgs}} = |D_\mu \phi|^2 - V(\phi) \quad (1.11)$$

where $V(\phi)$ represents a Higgs potential. Since the term includes covariant derivative D_μ , this term contains the interactions between Higgs bosons and gauge bosons.

In the following sections, formulations of theories that are considered cornerstones of the SM shall be discussed. The theories are the **Brout-Englert-Higgs mechanism**, the **Electroweak theory**, proposed by Glashow, Salam, and Weinberg, and **Quantum Chromodynamics**. Other relevant concepts will also be discussed.

1.4.1 Symmetries and why we need them

In classical field theory, a field ϕ may undergo a transformation, which can be written in an infinitesimal form as

$$\phi(x) \rightarrow \phi'(x) = \phi(x) + \alpha \delta \phi(x). \quad (1.12)$$

If this transformation retains the equations of motion, derived by the Euler-Lagrange equation, this transformation can be called a *symmetry*.

This concept of symmetry is crucial to physics in general since the laws of physics should be applied in the same way even though the transformation is applied. For example, laws of classical mechanics should be applied in the same way at any point of space, which means they have translational symmetry. It should also be the case when the whole coordinate system is being rotated in the space, which can be described by rotational symmetry.

Some fields may exhibit *global* gauge symmetry with the transformation $\phi(x) \rightarrow \phi'(x) = \exp(i\theta)\phi(x)$, or *local* gauge symmetry with the transformation $\phi(x) \rightarrow \phi'(x) = \exp(i\alpha(x))\phi(x)$, where each point in space may be transformed differently as dictated by $\alpha(x)$. These types of symmetries in a Lagrangian, which must be invariant under such transformations, lead to different effects which will be illustrated in subsequent sections.

1.4.2 Spontaneous Symmetry Breaking (SSB)

Spontaneous symmetry breaking (SSB) is a mechanism in which a complex scalar field, under a certain configuration of potential Lagrangian, breaks symmetry by reaching the ground state of the Lagrangian.

Consider the following Lagrangian of a complex scalar field ϕ ,

$$\mathcal{L} = \partial_\mu \phi^* \partial^\mu \phi - \frac{\lambda}{4} (\phi^* \phi - \nu^2)^2. \quad (1.13)$$

The Lagrangian $\mathcal{L}$ has symmetry under global U(1) gauge transformation, where $\phi \rightarrow \phi' = e^{i\alpha}\phi$ and α is a global constant. The potential term

$$V(\phi, \phi^*) = \frac{\lambda}{4} (\phi^* \phi - \nu^2)^2 \quad (1.14)$$

clearly has the minimum of zero at $\phi^* \phi = \nu^2$. This potential term has the shape of a famous Mexican hat, as shown in Figure 1.3. If we define

$$\phi = \phi_1 + i\phi_2 \quad (1.15)$$

$$\phi^* = \phi_1 - i\phi_2 \quad (1.16)$$

$$\phi_1^2 + \phi_2^2 = \nu^2 \quad (1.17)$$

we now have two scalar fields ϕ_1 and ϕ_2 , where we can choose the ground state to be $\bar{\phi} = \nu$. This state is not invariant under U(1) transformation anymore, because

$$\bar{\phi} \rightarrow \bar{\phi}' = e^{i\alpha}\bar{\phi} \neq \bar{\phi}. \quad (1.18)$$

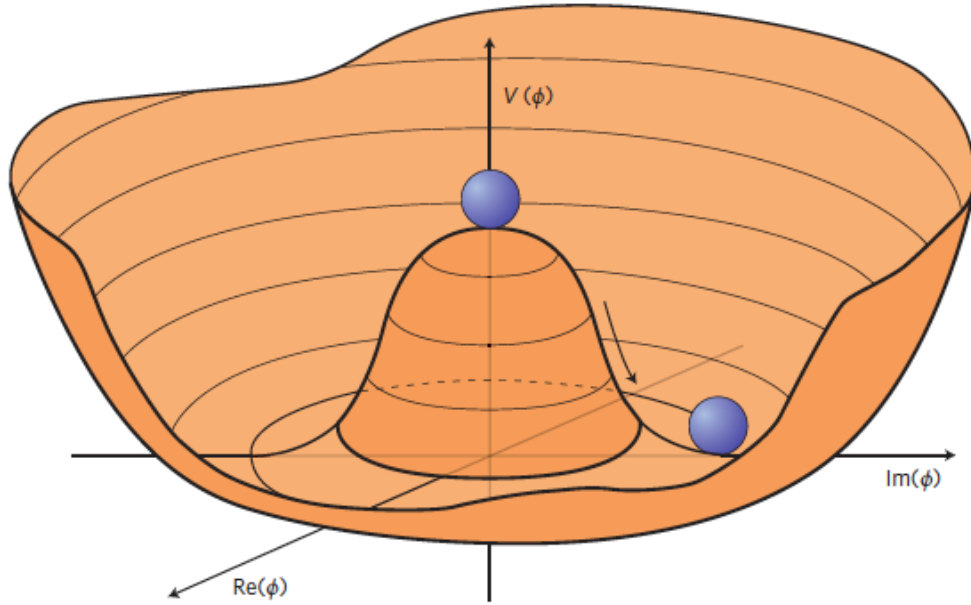


Figure 1.3: Visualization of the potential term $\frac{\lambda}{4} (\phi^* \phi - \nu^2)^2$, where ϕ is a complex scalar field ϕ . The symmetry point at $\phi = 0$, or the tip of the Mexican hat, does not have the lowest potential energy. The ground state for the scalar field, or the ridge of the Mexican hat, gives the lowest potential energy and breaks the symmetry. Diagram derived from Ref. [49].

This phenomenon is called *spontaneous symmetry breaking*. Now consider the perturbation around the ground state, where

$$\phi = \frac{1}{\sqrt{2}} (\eta + i\theta) + \nu. \quad (1.19)$$

Then

$$\partial_\mu \phi = \frac{1}{\sqrt{2}} (\partial_\mu \eta + i \partial_\mu \theta) \quad (1.20)$$

$$\partial_\mu \phi^* = \frac{1}{\sqrt{2}} (\partial_\mu \eta - i \partial_\mu \theta) \quad (1.21)$$

$$\partial_\mu \phi^* \partial^\mu \phi = \frac{1}{2} \partial_\mu \eta \partial^\mu \eta + \frac{1}{2} \partial_\mu \theta \partial^\mu \theta. \quad (1.22)$$

With this perturbation, the potential now becomes

$$\frac{\lambda}{4} (\phi^* \phi - \nu^2)^2 = \frac{\lambda}{4} \left(\left(\frac{1}{\sqrt{2}} \eta + \nu \right)^2 + \frac{\theta^2}{2} - \nu^2 \right)^2 \quad (1.23)$$

$$= \frac{\lambda}{4} \left(\frac{\eta^2}{2} + \sqrt{2} \eta \nu + \frac{\theta^2}{2} \right)^2 \quad (1.24)$$

$$= \frac{\lambda}{4} \left(\frac{1}{4} \eta^4 + 2\nu^2 \eta^2 + \sqrt{2} \nu \eta^3 + \frac{1}{4} \theta^4 + \frac{1}{2} \eta^2 \theta^2 + \sqrt{2} \nu \eta \theta^2 \right) \quad (1.25)$$

and the mass term for the field η , $\frac{\lambda \nu^2}{2} \eta^2$, arises from the perturbation. Now the Lagrangian, given in Equation 1.13, becomes

$$\mathcal{L} = \partial_\mu \phi^* \partial^\mu \phi - \frac{\lambda}{4} (\phi^* \phi - \nu^2)^2 \quad (1.26)$$

$$= \frac{1}{2} \partial_\mu \eta \partial^\mu \eta + \frac{1}{2} \partial_\mu \theta \partial^\mu \theta + \frac{\lambda \nu^2}{2} \eta^2 + \mathcal{O}(\eta^3, \eta^4, \theta^4, \eta \theta^2, \eta^2 \theta^2) \quad (1.27)$$

where $\mathcal{O}(\eta^3, \eta^4, \theta^4, \eta \theta^2, \eta^2 \theta^2)$ represents the interaction terms of fields η and θ . The resulting Lagrangian now contains a massive scalar field η and a massless field θ (since it does not have mass terms remaining in this Lagrangian). This massless field is called a Goldstone boson, and it can be perturbed without inertia.

In summary, spontaneous symmetry breaking of a scalar field can give rise to a massive boson by perturbing the scalar field around the ground state that breaks the symmetry.

1.4.3 Brout-Englert-Higgs mechanism

The mechanism proposed by Robert Brout, François Englert [35], and Peter Higgs [36] is simply the spontaneous symmetry breaking (SSB) of local gauge transformation. Consider the Lagrangian

$$\mathcal{L} = (D_\mu \phi)^* D^\mu \phi - \frac{\lambda}{4} (\phi^* \phi - \nu^2)^2 - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \quad (1.28)$$

where D_μ is a covariant derivative including a gauge boson A_μ .

$$D_\mu = \partial_\mu + ig A_\mu. \quad (1.29)$$

Here we expect to see the mass term of A_μ as $m_A^2 A_\mu A^\mu$. However, the transformation under U(1) where

$$\phi \rightarrow \phi' = e^{i\Lambda(x)} \phi \quad (1.30)$$

where $\Lambda(x)$ is a scalar function representing local transformation, gives

$$A_\mu \rightarrow A'_\mu = A_\mu - \frac{1}{g} \partial_\mu \Lambda(x). \quad (1.31)$$

As seen in the Equation 1.28, The potential term for a scalar field ϕ is in the form of $-\frac{\lambda}{4}(\phi^* \phi - \nu^2)^2$, and the ground state of this scalar field can be chosen, for simplicity, as $\bar{\phi} = \nu$. If we perturb the field around the ground state into the form of

$$\phi = \frac{1}{\sqrt{2}} \eta(x) + \frac{1}{\sqrt{2}} i\xi(x) + \nu. \quad (1.32)$$

This perturbation from the ground state is called spontaneous symmetry breaking (SSB). With this perturbation, we would simply get

$$\partial_\mu \phi = \frac{1}{\sqrt{2}} (\partial_\mu \eta + i\partial_\mu \xi) \quad (1.33)$$

$$D_\mu \phi = (\partial_\mu + igA_\mu) \left(\frac{1}{\sqrt{2}} \eta(x) + \frac{1}{\sqrt{2}} i\xi(x) + \nu \right) \quad (1.34)$$

$$= \frac{1}{\sqrt{2}} (\partial_\mu \eta + i\partial_\mu \xi) + \frac{1}{\sqrt{2}} (-\xi + i\eta) gA_\mu + ig\nu A_\mu \quad (1.35)$$

$$(D_\mu \phi)^* (D^\mu \phi) = \left[\frac{1}{\sqrt{2}} (\partial_\mu \eta - i\partial_\mu \xi) - \frac{1}{\sqrt{2}} (\xi + i\eta) gA_\mu - ig\nu A_\mu \right] \left[\frac{1}{\sqrt{2}} (\partial^\mu \eta + i\partial^\mu \xi) + \frac{1}{\sqrt{2}} (-\xi + i\eta) gA^\mu + ig\nu A^\mu \right] \quad (1.36)$$

$$\approx \frac{1}{2} \partial_\mu \eta \partial^\mu \eta + \frac{1}{2} \partial_\mu \xi \partial^\mu \xi + \sqrt{2} g\nu A^\mu \partial_\mu \xi + g^2 \nu^2 A_\mu A^\mu + \dots \quad (1.37)$$

with the perturbation introduced in ϕ , the mass term for gauge field $g^2 \nu^2 A_\mu A^\mu$ arises.

Now if we apply the perturbation to the scalar field potential term we obtain

$$-\frac{\lambda}{4}(\phi^* \phi - \nu^2)^2 = \frac{\lambda}{4} \left[\left(\frac{1}{\sqrt{2}} \eta + \nu + \frac{i\xi}{\sqrt{2}} \right) \left(\frac{1}{\sqrt{2}} \eta + \nu - \frac{i\xi}{\sqrt{2}} \right) - \nu^2 \right]^2 \quad (1.38)$$

$$= -\frac{\lambda}{4} \left[\frac{\eta^2}{2} + \sqrt{2} \eta \nu + \frac{\xi^2}{2} \right]^2 = -\frac{\lambda}{2} \nu^2 \eta^2 + \dots \quad (1.39)$$

The perturbation applied to ϕ causes the system to have 5 degrees of freedom, 2 from two scalar fields η and ξ , and 3 from a massive vector field A_μ . Before the

perturbation, we only had 2 degrees of freedom from the complex scalar field ϕ and 2 degrees of freedom from the *massless* vector field A_μ ⁶. Hence, an excess degree of freedom must be removed.

If we rearrange terms in $(D_\mu\phi)^*(D^\mu\phi)$ as

$$\frac{1}{2}\partial_\mu\xi\partial^\mu\xi + \sqrt{2}g\nu A^\mu\partial_\mu\xi + \nu^2 A_\mu A^\mu = g^2\nu^2 \left(A_\mu + \frac{1}{g\nu\sqrt{2}}\partial_\mu\xi \right) \left(A^\mu + \frac{1}{g\nu\sqrt{2}}\partial^\mu\xi \right). \quad (1.40)$$

This corresponds to the gauge transformation pattern

$$A_\mu \rightarrow A'_\mu = A_\mu - \frac{1}{g}\partial_\mu\Lambda(x) \quad (1.41)$$

where

$$\Lambda(x) = -\frac{\xi(x)}{\nu\sqrt{2}}. \quad (1.42)$$

Then the transformation on the scalar field should be

$$\phi \rightarrow \phi' = e^{i\Lambda(x)}\phi = \exp\left(-\frac{i\xi(x)}{\nu\sqrt{2}}\right)\phi. \quad (1.43)$$

Given $\phi = \frac{1}{\sqrt{2}}\eta + \nu + \frac{1}{\sqrt{2}}i\xi$ and η and ξ are perturbation terms,

$$\phi' = \exp\left(-\frac{i\xi}{\nu\sqrt{2}}\right) \left(\frac{1}{\sqrt{2}}\eta + \nu + \frac{1}{\sqrt{2}}i\xi \right) \quad (1.44)$$

$$= \exp\left(-\frac{i\xi}{\nu\sqrt{2}}\right) \left(\frac{\eta}{\sqrt{2}} + \nu \right) \left(1 + \frac{i\xi}{\nu\sqrt{2}} \right) \quad (1.45)$$

$$= \exp\left(-\frac{i\xi}{\nu\sqrt{2}}\right) \left(\frac{\eta}{\sqrt{2}} + \nu \right) \exp\left(+\frac{i\xi}{\nu\sqrt{2}}\right) \quad (1.46)$$

$$= \frac{\eta}{\sqrt{2}} + \nu \equiv \frac{h}{\sqrt{2}}. \quad (1.47)$$

The new field h that has just been defined is the Higgs field, and now the Lagrangian $\mathcal{L}$ from Equation 1.28 becomes

$$\mathcal{L} = (D_\mu\phi)^*D^\mu\phi - \frac{\lambda}{4}(\phi^*\phi - \nu^2)^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (1.48)$$

$$= \frac{1}{2}\partial_\mu h\partial^\mu h - \frac{\lambda}{2}\nu^2 h^2 + g^2\nu^2 A_\mu A^\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \quad (1.49)$$

⁶The vector field initially has 4 degrees of freedom, but 2 degrees of freedom have been removed due to Maxwell's equations and symmetry.

With this transformation, we have treated $\xi(x)$ as part of the local gauge transformation, and now only 4 degrees of freedom remain, 3 from a *massive* vector field A_μ and 1 from a massive scalar field h .

In summary, the Brout-Englert-Higgs mechanism imposes spontaneous symmetry breaking upon local gauge transformations to obtain mass terms of gauge bosons. The remaining scalar boson is then called the Higgs boson and is the byproduct of this mechanism.

1.4.4 Formulation of W and Z bosons

W and Z bosons, along with photons, arise from a symmetry under $SU(2)_W \times U(1)_Y$ transformation. Given the following Lagrangian

$$\mathcal{L} = (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi^\dagger \phi) \text{ where } \phi = \frac{1}{\sqrt{2}} \begin{bmatrix} \phi^+ \\ \phi^0 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{bmatrix}. \quad (1.50)$$

To make this Lagrangian invariant under $SU(2)_W \times U(1)_Y$ symmetry, the covariant derivative must take the form of

$$D_\mu = \partial_\mu + i\frac{g_W}{2}T^i \cdot W_\mu^i + i\frac{g_Y}{2}B_\mu Y \quad (1.51)$$

where W_μ^i is a collection of 3 $SU(2)$ gauge fields, and T^i is a collection of $SU(2)$ weak generators, which are Pauli matrices as follows:

$$T^1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, T^2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, T^3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (1.52)$$

On the other hand, B_μ is a gauge field for $U(1)_Y$ symmetry, and Y is a hypercharge operator. g_W and g_Y are the couplings for $SU(2)_W$ and $U(1)_Y$ symmetry respectively. If the potential term $V(\phi^\dagger \phi)$ permits SSB, such as in the form of

$$V(\phi^\dagger \phi) = \lambda (\phi^\dagger \phi)^2 - \mu^2 \phi^\dagger \phi. \quad (1.53)$$

The vacuum state $\bar{\phi}$ is at $\phi^\dagger\phi = \mu^2/\lambda \equiv \nu^2/2$ as usual, so we can choose⁷

$$\bar{\phi} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ \nu \end{bmatrix}. \quad (1.54)$$

With this choice, we have now broken the symmetry from $SU(2)_W \times U(1)_Y$ into $U(1)_Y$. Based on this vacuum state choice, we now have

$$\begin{aligned} & \left[i \frac{g_W}{2} T^i \cdot W_\mu^i + i \frac{g_Y}{2} Y B_\mu \right] \phi \\ &= \frac{i}{\sqrt{2}} \left[\frac{g_W}{2} T^i \cdot W_\mu^i + \frac{g_Y}{2} Y B_\mu \right] \begin{bmatrix} 0 \\ \nu \end{bmatrix} \end{aligned} \quad (1.55)$$

$$\begin{aligned} &= \frac{i}{\sqrt{2}} \left(\frac{g_W}{2} W_\mu^1 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \frac{g_W}{2} W_\mu^2 \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \right. \\ &\quad \left. + \frac{g_W}{2} W_\mu^3 \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + \frac{g_Y}{2} Y B_\mu \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} 0 \\ \nu \end{bmatrix} \end{aligned} \quad (1.56)$$

$$= \frac{i}{2\sqrt{2}} \begin{bmatrix} g_W W_\mu^3 + g_Y Y B_\mu & g_W W_\mu^1 - i g_W W_\mu^2 \\ g_W W_\mu^1 + i g_W W_\mu^2 & g_Y Y B_\mu - g_W W_\mu^3 \end{bmatrix} \begin{bmatrix} 0 \\ \nu \end{bmatrix} \quad (1.57)$$

$$= \frac{\nu}{2\sqrt{2}} \begin{bmatrix} g_W W_\mu^2 + i g_W W_\mu^1 \\ i g_Y Y B_\mu - i g_W W_\mu^3 \end{bmatrix}. \quad (1.58)$$

Here, we can choose $Y = 1$ for our vacuum state. Based on this calculation, the term $(D^\mu \phi)^\dagger (D_\mu \phi)$ should contain

$$\begin{aligned} & \frac{\nu^2}{8} \begin{bmatrix} g_W (W^2)^\mu - i g_W (W^1)^\mu & -i g_Y B^\mu + i g_W (W^3)^\mu \end{bmatrix} \begin{bmatrix} g_W W_\mu^2 + i g_W W_\mu^1 \\ i g_Y B_\mu - i g_W W_\mu^3 \end{bmatrix} \\ &= \frac{\nu^2}{8} \left[g_W^2 (W_\mu^2)^2 + g_W (W_\mu^1)^2 + (g_Y B_\mu - g_W W_\mu^3)^2 \right]. \end{aligned} \quad (1.59)$$

⁷We can also choose to add perturbative Higgs field h , such that $\phi = [0 \quad \nu + h]^T / \sqrt{2}$, which gives interaction terms between vector bosons and the Higgs field.

Define the following fields

$$A_\mu = \frac{1}{\sqrt{g_W^2 + g_Y^2}} (g_Y W_\mu^3 + g_W B_\mu) \quad (1.60)$$

$$Z_\mu = \frac{1}{\sqrt{g_W^2 + g_Y^2}} (g_W W_\mu^3 - g_Y B_\mu) \quad (1.61)$$

$$W_\mu^+ = \frac{1}{\sqrt{2}} (W_\mu^1 - iW_\mu^2) \quad (1.62)$$

$$W_\mu^- = \frac{1}{\sqrt{2}} (W_\mu^1 + iW_\mu^2) . \quad (1.63)$$

With this definition, the above terms can be rewritten as

$$\begin{aligned} & \frac{\nu^2}{8} \left[g_W^2 (W_\mu^2)^2 + g_W (W_\mu^1)^2 + (g_Y B_\mu - g_W W_\mu^3)^2 \right] \\ &= \frac{\nu^2}{8} \left[g_W^2 (W_\mu^+)^2 + g_W^2 (W_\mu^-)^2 + (g_W^2 + g_Y^2) (Z_\mu)^2 + 0(A_\mu)^2 \right] . \end{aligned} \quad (1.64)$$

These terms now become mass terms for W and Z bosons, with no mass terms for photons A_μ . If we define a Weinberg mixing angle, sometimes called a weak mixing angle, θ_W as $\tan \theta_W = g_Y/g_W$, we will get

$$A_\mu = \frac{1}{\sqrt{g_W^2 + g_Y^2}} (g_Y W_\mu^3 + g_W B_\mu) \quad (1.65)$$

$$= W_\mu^3 \sin \theta_W + B_\mu \cos \theta_W \quad (1.66)$$

$$Z_\mu = \frac{1}{\sqrt{g_W^2 + g_Y^2}} (g_W W_\mu^3 - g_Y B_\mu) \quad (1.67)$$

$$= W_\mu^3 \cos \theta_W - B_\mu \sin \theta_W . \quad (1.68)$$

In summary, W and Z bosons are the byproducts of applying $SU(2)_W \times U(1)_Y$ symmetry. After SSB, W and Z bosons regain masses while photons remain massless.

1.4.5 Yukawa Lagrangian

The Yukawa Lagrangian term, first used by Hideki Yukawa in 1935 [50], was originally used to describe the interactions between nucleons and pions. As our understanding of fundamental particles becomes deeper, the term is now used to

describe interactions between fermions and a Higgs boson. The typical form of the Yukawa Lagrangian term is [51]

$$\mathcal{L}_{\text{Yukawa}} = -g\bar{\psi}\psi\phi \quad (1.69)$$

where ψ represents fermions written in the doublet form, and ϕ represents a scalar boson (which is a Higgs boson for the SM case). The coupling g is called Yukawa coupling and is different for each fermion. In the standard model, the coupling for fermions can be written as [17]

$$g = \frac{m_f}{\nu} \quad (1.70)$$

where m_f is the fermion mass and ν is the vacuum expectation value, which is $\nu = (\sqrt{2}G_F)^{-1/2} \approx 246$ GeV, where G_F is Fermi coupling constant [17].

1.4.6 Electroweak theory

The interactions between fermions (quarks and leptons) and W/Z bosons and photons can be described by the electroweak theory. In this theory, fermions, such as quarks and leptons, can be separated into left-handed and right-handed components. Left-handed and right-handed components can be written in the form of a doublet and a singlet respectively as

$$\psi_L \equiv \frac{1}{2}(1 + \gamma_5)\psi \quad (1.71)$$

$$\psi_R \equiv \frac{1}{2}(1 - \gamma_5)\psi \quad (1.72)$$

where γ_5 is a matrix given by

$$\gamma_5 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (1.73)$$

Notice that $(\gamma_5)^2 = I$ and $\psi_L + \psi_R = \psi$. Given the split of fermions into the left-handed and right-handed components, we can rewrite Lagrangian terms involving fermions into left-handed and right-handed components.

The Electroweak theory, as introduced by Weinberg in 1967 [31], is based on the Lagrangian containing a fermion field ψ representing *electrons and electron neutrinos*⁸, a scalar doublet Higgs boson field ϕ , three SU(2) gauge bosons W_μ^i , and one U(1) gauge boson B_μ . The left-handed field ψ_L can be written as a doublet [52],

$$\psi_L = \begin{bmatrix} \nu_e \\ e_L \end{bmatrix} \quad (1.74)$$

containing both left-handed neutrino and electron components, while the right-handed field ψ_R is simply a singlet,

$$\psi_R = e_R \quad (1.75)$$

containing only right-handed electrons, without right-handed neutrinos. Quarks, on the other hand, have separate right-handed fields for all quark flavours, which are u_R , d_R , and so on, as well as left-handed doublets such as $[u_L \ d_L]^T$.

W_μ^i gauge bosons interact only with the left-handed component of the fermion, while B_μ gauge boson interacts with both left and right-handed components of the fermion. Both types of gauge bosons have their field strength tensor, defined as

$$f_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \quad (1.76)$$

$$F_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i - g_W \epsilon^{ijk} W_\mu^j W_\nu^k. \quad (1.77)$$

Both bosons are also added to the covariant derivative. However, since the right-handed component is a singlet, but the left-handed component is a doublet, the covariant derivatives are applied differently to left-handed and right-handed components as follows:

$$D_\mu \psi_L = \left[\partial_\mu + i \frac{g_W}{2} \vec{\tau} \cdot \vec{W}_\mu + i \frac{g_Y}{2} B_\mu \right] \psi_L \quad (1.78)$$

$$D_\mu \psi_R = \left[\partial_\mu + i \frac{g_Y}{2} B_\mu \right] \psi_R. \quad (1.79)$$

⁸The model proposed by Weinberg in Ref. [31] made clear that it should focus on “electron-type leptons only”. However, this fermion field can also represent muon-type and tau-type leptons.

The base Lagrangian used in Electroweak theory, as introduced in Ref. [31] contains several components as follows:⁹

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{4}f_{\mu\nu}f^{\mu\nu} \quad (\text{boson kinetic}) \quad (1.80a)$$

$$-i\bar{\psi}\gamma^\mu D_\mu\psi \quad (\text{Dirac}) \quad (1.80b)$$

$$-\frac{1}{2}|D_\mu\phi|^2 \quad (\text{Higgs kinetic}) \quad (1.80c)$$

$$-g\left(\bar{\psi}_L\phi\psi_R + \bar{\psi}_R\phi^\dagger\psi_L\right) \quad (\text{Yukawa}) \quad (1.80d)$$

$$-\mu^2\phi^\dagger\phi + \lambda\left(\phi^\dagger\phi\right)^2 \quad (\text{Higgs potential, prone to SSB}) \quad (1.80e)$$

Notice that almost all terms in the Lagrangian contain ψ , which is not yet split into left-handed and right-handed components. The Yukawa Lagrangian term 1.80d, however, must be rewritten since ϕ is a doublet field. The only Lagrangian component containing ψ is the Dirac term (Equation 1.80b). With the formulations written in Equations 1.71 and 1.72, we can rewrite the term 1.80b as

$$-i\bar{\psi}\gamma^\mu D_\mu\psi = -i(\bar{\psi}_L + \bar{\psi}_R)\gamma^\mu D_\mu(\psi_L + \psi_R) \quad (1.81)$$

$$= -i(\bar{\psi}_L\gamma^\mu D_\mu\psi_L + \bar{\psi}_L\gamma^\mu D_\mu\psi_R + \bar{\psi}_R\gamma^\mu D_\mu\psi_L + \bar{\psi}_R\gamma^\mu D_\mu\psi_R) . \quad (1.82)$$

⁹The notations used in this section are different from the ones used in Ref. [31]. For example, W_μ is used here instead of $\vec{A}_\mu$ in the literature. Equation 1.80 is also written here with a shortened form of the boson kinetic term.

Since $\gamma_5 \gamma^\mu = -\gamma^\mu \gamma_5$, the first and last term vanish because

$$\bar{\psi}_L \gamma^\mu D_\mu \psi_L = \bar{\psi} \left(\frac{1 + \gamma_5}{2} \right) \gamma^\mu D_\mu \left(\frac{1 + \gamma_5}{2} \right) \psi \quad (1.83)$$

$$= \frac{1}{4} \bar{\psi} (1 + \gamma_5) \gamma^\mu D_\mu (1 + \gamma_5) \psi \quad (1.84)$$

$$= \frac{1}{4} \bar{\psi} (\gamma^\mu D_\mu + \gamma_5 \gamma^\mu D_\mu + \gamma^\mu D_\mu \gamma_5 + \gamma_5 \gamma^\mu D_\mu \gamma_5) \psi \quad (1.85)$$

$$= \frac{1}{4} \bar{\psi} (\gamma^\mu D_\mu + \gamma_5 \gamma^\mu D_\mu - \gamma_5 \gamma^\mu D_\mu - \gamma_5 \gamma_5 \gamma^\mu D_\mu) \psi \quad (1.86)$$

$$= \frac{1}{4} \bar{\psi} (\gamma^\mu D_\mu + \gamma_5 \gamma^\mu D_\mu - \gamma_5 \gamma^\mu D_\mu - \gamma^\mu D_\mu) \psi = 0 \quad (1.87)$$

$$\bar{\psi}_R \gamma^\mu D_\mu \psi_R = \bar{\psi} \left(\frac{1 - \gamma_5}{2} \right) \gamma^\mu D_\mu \left(\frac{1 - \gamma_5}{2} \right) \psi \quad (1.88)$$

$$= \frac{1}{4} \bar{\psi} (1 - \gamma_5) \gamma^\mu D_\mu (1 - \gamma_5) \psi \quad (1.89)$$

$$= \frac{1}{4} \bar{\psi} (\gamma^\mu D_\mu - \gamma_5 \gamma^\mu D_\mu + \gamma_5 \gamma^\mu D_\mu - \gamma^\mu D_\mu) \psi = 0. \quad (1.90)$$

Hence, the term in Equation 1.80b is simply

$$-i \bar{\psi} \gamma^\mu D_\mu \psi = -i \bar{\psi}_L \gamma^\mu D_\mu \psi_R - i \bar{\psi}_R \gamma^\mu D_\mu \psi_L. \quad (1.91)$$

Once we have the Lagrangian with fermion components split into left-handed and right-handed components, we can perform spontaneous symmetry breaking on ϕ by setting a vacuum state with perturbation field h [52]

$$\bar{\phi} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ \nu + h \end{bmatrix}. \quad (1.92)$$

This is similar to the choice of vacuum we made in Equation 1.54. With this vacuum choice, the only terms affected are the Higgs kinetic term 1.80c, Yukawa term 1.80d, and Higgs potential term 1.80e. Since we choose ϕ to be in a vacuum, the term in 1.80e becomes zero. The Higgs kinetic term 1.80c will be transformed in the same manner as discussed in Section 1.4.4 and gives out mass terms for W and Z bosons.

The Yukawa term, on the other hand, can be transformed as [52]

$$-g\bar{\psi}_L\phi\psi_R = \frac{g}{\sqrt{2}} \begin{bmatrix} \bar{\nu}_e & \bar{e}_L \end{bmatrix} \begin{bmatrix} 0 \\ \nu + h \end{bmatrix} e_R \quad (1.93)$$

$$= \frac{g}{\sqrt{2}} \bar{e}_L(\nu + h)e_R \quad (1.94)$$

$$-g\bar{\psi}_R\phi^\dagger\psi_L = \frac{g}{\sqrt{2}} \bar{e}_R \begin{bmatrix} 0 & \nu + h \end{bmatrix} \begin{bmatrix} \bar{\nu}_e \\ e_L \end{bmatrix} \quad (1.95)$$

$$= \frac{g}{\sqrt{2}} \bar{e}_R(\nu + h)e_L \quad (1.96)$$

and we obtain mass terms $g\nu\bar{e}_Le_R/\sqrt{2}$ and $g\nu\bar{e}_Re_L/\sqrt{2}$. The mass of the electron m_e is $g\nu/\sqrt{2}$. We also obtain the interaction term between the scalar Higgs field and both left and right-handed electrons. However, the interaction terms involving neutrinos and Higgs bosons do not appear as a consequence of the vacuum choice we made in Equation 1.92.

Since neutrinos do not have a right-handed component, the mass terms do not appear in SM, and thus they are treated as massless particles. However, neutrino masses can be added in several extensions to SM [17].

1.4.7 Quantum Chromodynamics primer

In Quantum Chromodynamics (QCD), the strong interactions between quarks with colour charges are described by the SU(3) component of $SU(3) \times SU(2) \times U(1)$ gauge symmetry. The Lagrangian for QCD is given by [17]

$$\mathcal{L} = \sum_q \bar{\psi}_{q,a}(i\gamma^\mu\partial_\mu\delta_{ab} - g_s\gamma^\mu t_{ab}^C A_\mu^C - m_q\delta_{ab})\psi_{q,b} - \frac{1}{4}F_{\mu\nu}^A F^{A,\mu\nu} \quad (1.97)$$

where q represents quark flavour, $a, b \in \{1, 2, 3\}$ represent colour index, commonly known as red, green, and blue, and $A, C \in \{1, 2, \dots, 8\}$ represent SU(3) generator index. To better understand what happens in this Lagrangian, we will simplify Equation 1.97 by considering only up-quark flavour $q = u$. With this assumption,

we can rewrite the Lagrangian above as

$$\mathcal{L} = \bar{u}_a(i\gamma^\mu\partial_\mu\delta_{ab} - g_s\gamma^\mu t_{ab}^C\mathcal{A}_\mu^C - m_u\delta_{ab})u_b - \frac{1}{4}F_{\mu\nu}^A F^{A,\mu\nu} \quad (1.98)$$

$$= \bar{u}_a(i\gamma^\mu\partial_\mu - m_u)u_a - g_s\bar{u}_a\gamma^\mu t_{ab}^C\mathcal{A}_\mu^C u_b - \frac{1}{4}F_{\mu\nu}^A F^{A,\mu\nu}. \quad (1.99)$$

The first term is simply the Dirac Lagrangian describing the kinematics and mass of an up-type quark of all colour charges. The second term contains the strong coupling constant g_s , SU(3) generator t_{ab}^C , written in the form of 3×3 matrices (similar to SU(2) generators we have during W and Z boson formation), and $\mathcal{A}_\mu^C$ is a *gluon field*. Since there are 8 generators in the SU(3) special unitary group, there exist exactly 8 gluon fields. This term explains the interactions between two quarks (up quarks in this case) exchanging colour charges (a and b) and a gluon.

The final term contains field strength tensor for gluon fields, which can be written as [17]

$$F_{\mu\nu}^A = \partial_\mu\mathcal{A}_\nu^A - \partial_\nu\mathcal{A}_\mu^A - g_s f_{ABC}\mathcal{A}_\mu^B\mathcal{A}_\nu^C \quad (1.100)$$

where f_{ABC} is a structure constants of SU(3) group, defined by its Lie algebra

$$[t^A, t^B] = if_{ABC}t^C. \quad (1.101)$$

With the introduction of colour charges and gluon fields by QCD, the best explanation of why quarks may not be free is a hypothesis called *colour confinement*. This hypothesis states that objects with coloured charges (such as quarks) must be confined to colour singlet states, which is a combination of complementing colour states, and objects with non-zero colour charges can propagate. For mesons with a quark and an antiquark, and baryons with three quarks, the colour wavefunction would be

$$\psi^c(q\bar{q}) = \frac{1}{\sqrt{3}}(r\bar{r} + g\bar{g} + b\bar{b}) \quad (1.102)$$

$$\psi^c(qqq) = \frac{1}{\sqrt{6}}(rgb - rbg + gbr - grb + brg - bgr). \quad (1.103)$$

However, this hypothesis has not been proven analytically [24].

Another problematic aspect of QCD is that its coupling constant α_s is large at low energy scales. This coupling is also dependent, or *running*, to the energy scale q^2 of the gluon mediator.

To illustrate how problematic this is, let us look at what happens in Quantum Electrodynamics, or QED, where it has the same design of the coupling constant α which is also running to the energy scale q^2 of the photon mediator. In QED, a photon exchange between two electrons (with two $e^- \gamma e^-$ vertices) can have up to an infinite number of diagrams with an infinite number of loops. These diagrams form an infinite series which can be summed into one simple diagram of photon exchange, with a photon mediator q and an *effective* coupling value α . Fortunately for QED, at low energy scales of q^2 [17],

$$\alpha(q^2 \approx 0) = \frac{1}{137.035\,999\,084\,(21)} . \quad (1.104)$$

However, the same cannot be said for QCD with a gluon exchange between two quarks. As with QED, QCD has the *effective* coupling value α_s which depends on the energy scale q^2 of the mediator. However, at low levels of energy scale at $|q| \sim 1$ GeV, $\mathcal{O}(\alpha_s) \sim 1$, and perturbation theory cannot be used. Even at higher energy scales such as $|q| \sim 100$ GeV, $\alpha_s \sim 0.1$, and still requires higher-order corrections in QCD calculations [24].

1.5 Problems with the Standard Model, and Beyond Standard Model theories

Even though the Standard Model has been successfully predicting numerous phenomena in particle physics, there are still issues that it cannot perfectly describe

and need solutions in the form of extensions to the Standard Model. This section will introduce a few important issues along with some extensions to the model.

One of the problems with SM itself is called the Hierarchy problem. In the current SM, quantum loops of the Higgs boson propagator contribute to the Higgs mass, and if SM is valid up to a very high energy scale of 10^{19} GeV, corrections from these loops can be very large since they are quadratic to the scale itself. Still, the measured Higgs mass is considered surprisingly low at merely 125 GeV. The Hierarchy problem can be solved by fine-tuning new contributions to Higgs mass corrections so that these corrections are cancelled at higher precision [24].

However, another extension to SM, called Supersymmetry (SUSY), is hypothesized to be a more natural solution to this problem [24]. This extension proposes *superpartners* to particles in SM, which, according to Minimally Supersymmetric Standard Model (MSSM), are squarks, sleptons, Higgsinos, gluinos, winos, and binos. Superpartners for fermions, such as quarks and leptons, are called sfermions and are spin-0 bosons, while superpartners for bosons, such as Higgs, gluons, and W bosons, are spin-1/2 fermions [53]. With Supersymmetry, quantum loops that provide corrections to Higgs mass will have corresponding loops containing superpartner particles with opposite signs, and cancel out those corrections. However, supersymmetric particles have not been found as of this writing, and it is speculated that Supersymmetry is broken (or else superpartners would have been found in the first place) [24].

Several astronomical observations, such as the measurement of galactic rotational curves [54], suggest that there must be non-luminous and non-absorbing matter, known as *dark matter*, that does not interact with other particles but exhibits the gravitational force towards other matter in the observable universe. In 2004, a weak gravitational lensing effect was detected based on astronomical observation

data, providing evidence for the existence of dark matter [55]. Several candidates for dark matter are hypothesised, such as weakly interacting massive particles (WIMPs) that have a neutral electrical charge and do not interact with other particles [56].

One other problem with the current design of SM is that it does not give masses to neutrinos, which are required to explain the neutrino oscillation phenomenon [47]. One of the possible ways to construct neutrino mass terms is to introduce sterile neutrinos. With this construction, gauge symmetry is maintained and several scenarios from this extension with sterile neutrinos are possible, such as Dirac neutrinos and the See-saw mechanism [17].

Several extensions to SM are collectively called Beyond Standard Model (BSM) theories, and there are currently searches performed in particle detectors, such as the CMS detector, involving these BSM theories. There are two main approaches for searches in the BSM sector: model-dependent and model-independent searches. Model-dependent searches involve searching for a new particle based on a particular BSM model, such as the search for top squarks (or *stops*) in the CMS detector [57]. On the other hand, model-independent searches do not focus on a particular new particle but instead look for deviations from SM, such as the measurement of Yukawa coupling of top quarks and Higgs boson, or deviations from SM in the form of Effective Field Theory.

1.6 Top quarks

As of the current understanding of SM, the top quark is the largest out of six quarks, with a mass of 172.69 ± 0.30 GeV [17]. Even after almost 30 years of its discovery in Tevatron [33; 34], the top quark remains an active research topic in High Energy Physics. Studies performed involving top quarks inside the CMS Collaboration include, but are not limited to, the following:

- Cross section measurements of top-antitop quark production in final states involving zero, one, and two leptons
- Cross section measurements of single-top quark production
- Top quark mass precision measurements
- Top quark polarization measurements
- Observation of $t\bar{t}H$ production [58]
- Flavour-Changing Neutral Current (FCNC) and CP-violation searches
- Four-top quark production

Top quarks will almost always decay into a W boson and a bottom quark, with a branching ratio of 99.7% [17]. Since a W boson can either decay into a pair of quarks or a lepton and a neutrino, the signature for productions involving top quarks can be categorised by the number of leptons present in each event. The almost exclusive decay pathway $t \rightarrow Wb$ is explained by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [25], written in the form of

$$V_{\text{CKM}} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix}. \quad (1.105)$$

Each element in the CKM matrix indicates the probability of a quark decaying from one flavour to another. For the case of top quarks, the probability of top quarks decaying into bottom quarks corresponds to $|V_{tb}|^2$. The most recent global fit of CKM matrix elements is [17]

$$|V_{\text{CKM}}| \approx \begin{bmatrix} 0.974 & 0.225 & 0.004 \\ 0.225 & 0.973 & 0.042 \\ 0.009 & 0.041 & 0.999 \end{bmatrix} \quad (1.106)$$

with $|V_{tb}| = 0.999\,118^{+0.000\,031}_{-0.000\,036}$ [17]. This means that the top quark will almost always decay into a bottom quark and a W boson.

Another interesting property of the top quark is that it has a very short lifetime, so short that it decays before hadronization. Normally, quarks will undergo a process called hadronization, as dictated by QCD. This process causes a pair of quarks to generate other quarks as they are separated with high velocities, causing a large group of quarks with energies low enough to form hadrons. [8] As a consequence, the signatures for quarks and gluons detectable in particle detectors are a particle jet consisting of hadrons. The timescale of hadronization can be estimated by the uncertainty principle to be $\Delta\tau \leq 0.7 \text{ fm}/c \approx 2 \times 10^{-24} \text{ s}$ [59]. However, the top quark mass width is $\Gamma = 1.42 \text{ GeV}$ [17], corresponding to the decay time of $\tau = \hbar/\Gamma \approx 5 \times 10^{-25} \text{ s}$, approximately one order of magnitude less than the hadronization timescale. Therefore, top quarks do not have enough time to hadronize into a jet of particles and must decay.

Compared to the mass of the bottom quark, which is the second largest quark, at $4.18^{+0.03}_{-0.02} \text{ GeV}$ [17], and even the mass of Higgs boson at $125.25 \pm 0.17 \text{ GeV}$ [17], this prompts the question as to why the top quark has its enormous mass. A consequence of its large mass is that the Yukawa coupling value between top quarks and a Higgs boson is the largest among all fermions in SM. More details regarding the Yukawa coupling will be discussed in Section 1.7.

1.7 Yukawa couplings, especially the top quark

Yukawa coupling

As discussed in Section 1.4.5, the Yukawa Lagrangian term is used to describe the interaction between any fermion and a Higgs boson scalar field. From the SM value

of the Yukawa coupling given in Equation 1.70, along with the top quark mass of 173 GeV, the top quark Yukawa coupling is almost at the order of unity, giving the largest connection between Higgs bosons and top quarks.

Based on Equation 1.69, the Lagrangian term for the interaction between top quarks and a Higgs boson can be written as

$$\mathcal{L}_{t\bar{t}H} = -y_t \bar{t}tH \quad (1.107)$$

which involves the top Yukawa coupling y_t . This means that, for every top-higgs interaction vertex, the matrix element $\mathcal{M}$ of that Feynman diagram must be multiplied by y_t , and as the cross-section for that diagram is proportional to $|\mathcal{M}|^2$, its cross-section may be proportional to y_t^2 , assuming that there are no interference effects. Therefore, particle productions involving both top quarks and Higgs bosons are susceptible to this coupling value and are good candidates for the measurement of this coupling.

1.7.1 Yukawa coupling measurements

The measurements of the Yukawa coupling for fermions or vector bosons can be performed by numerous particle productions. One result from the CMS Collaboration uses the latest search result for $H \rightarrow \mu^+\mu^-$ particle production, combined with the previous measurement performed in 2016 [60], to perform a simultaneous fit for six coupling modifiers, κ_W , κ_Z , κ_t , κ_τ , κ_b , and κ_μ , which modifies SM Yukawa coupling values for W boson, Z boson, top quark, tau lepton, bottom quark, and muon respectively [61]. These six coupling modifiers affect directly the cross section and branching ratio of each production category. The measurement results are shown in Figure 1.4. As seen from the diagram, the measured values are in agreement with the predicted values by SM.

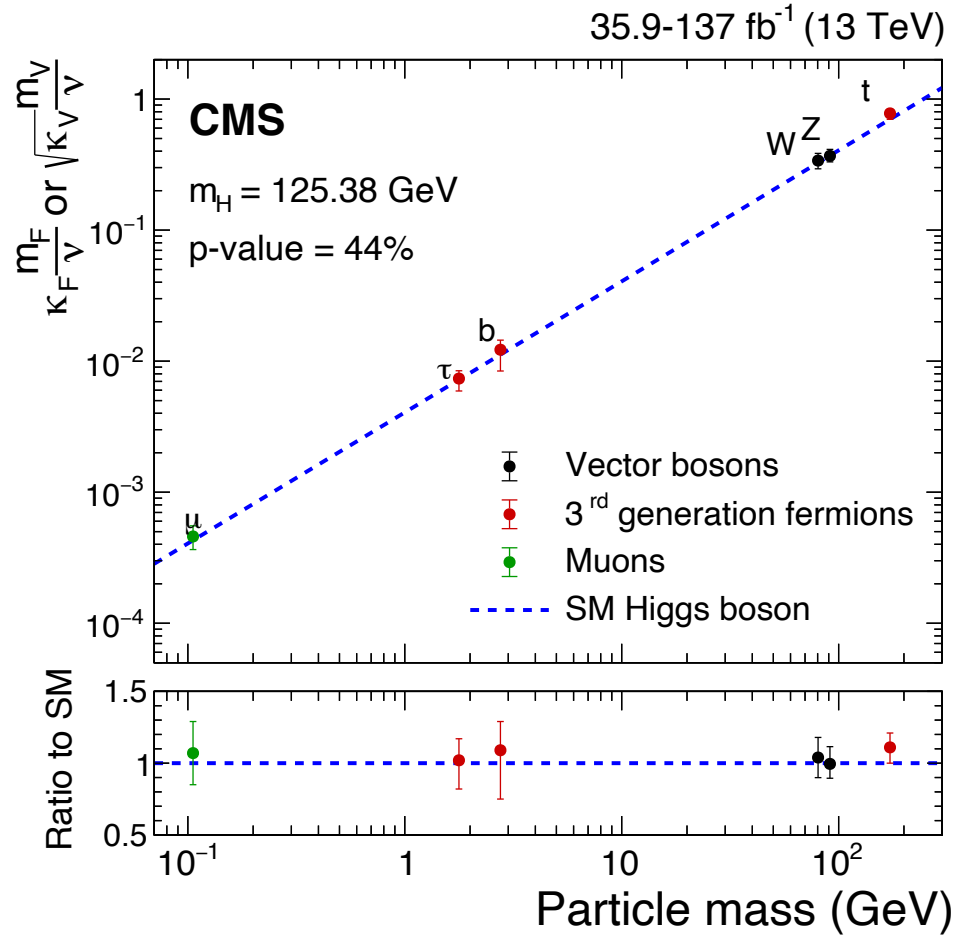


Figure 1.4: Best fit values of six Yukawa couplings based on search results for $H \rightarrow \mu^+ \mu^-$ [61], combined with previous measurements performed in 2016 [60]. All measured values are consistent with the SM prediction depicted in the dashed line. Diagram taken from Ref. [61].

Another example of the measurement of top quark Yukawa coupling is the measurement based on kinematic distributions of the $t\bar{t}$ production [62]. It has been shown in Ref. [62] that corrections from electroweak terms in sub-leading order corrections introduce diagrams with loops including neutral vector and scalar bosons (including Higgs bosons). Therefore, a large Yukawa coupling value can alter the kinematic distributions of the invariant mass of the top quark pair, $M_{t\bar{t}}$, and the difference in rapidity of the two top quarks, $\Delta y_{t\bar{t}}$.

The measurement is performed on the distribution of two proxy variables: $M_{bb\ell\ell}$, which is an invariant mass of two bottom quarks and two leptons, and $|\Delta y_{b\ell b\ell}|$. An electroweak correction rate modifier is introduced to modify the yields of each bin of the distribution for these two proxy variables. For each region in the kinematics distribution, the modifiers are observed to have quadratic behaviour to the top Yukawa coupling value. With these modifiers, the best fit of the Yukawa coupling value ratio, as shown in Figure 1.5, is derived to be

$$Y_t = 1.16^{+0.24}_{-0.35} \quad (1.108)$$

where $Y_t = y_t/y_t^{\text{SM}}$, which is the ratio between the measured value of top Yukawa coupling versus the SM value. If SM is correct, the ratio must be exactly 1. This measurement of the Yukawa coupling differs from the one presented in Ref. [61] since this measurement takes into account changes in kinematic distributions, not only the cross section and branching ratio of the decay.

1.8 Four-top quark production

Four-top quark ($t\bar{t}t\bar{t}$) production is, as its name suggests, the production mechanism giving four top quarks, or to be more precise, two pairs of top-antitop quarks. Figure 1.6 shows leading-order diagrams of this production. The most recent calculation for the four-top quark production cross section,

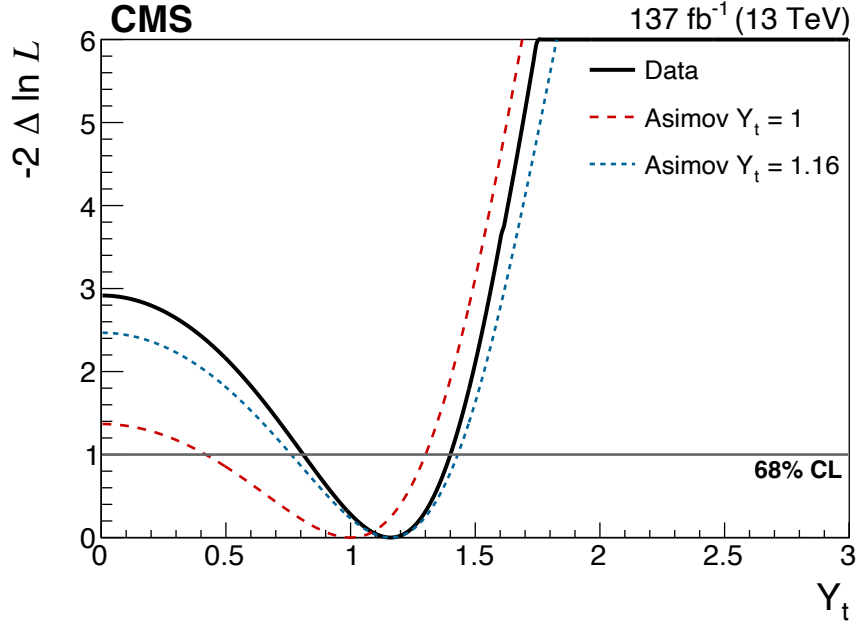


Figure 1.5: Negative log likelihood scan for Y_t , based on the electroweak corrections in kinematic distributions of $t\bar{t}$ production. The most probable value obtained from this fit is $Y_t = 1.16^{+0.24}_{-0.35}$ [62].

at next-to-leading-logarithmic (NLL') accuracy, is $13.4^{+1.0}_{-1.8}$ fb at 13 TeV centre-of-mass energy [63]. In this thesis, however, the cross section value calculated at next-to-leading-order (NLO) with electroweak corrections, which is $12.0^{+2.2}_{-2.5}$ fb at 13 TeV centre-of-mass energy [64], will be used as it is the main cross section value used in all analyses mentioned in this thesis. However, the difference in theoretical cross section values assumed for four-top quark production would only be important when comparing signal strength measurements to other reported results. The cross section on the other hand is measured and reported in fb, thus should not be affected by the change in cross section assumption.

As seen from Figure 1.6, the four-top quark production already has $t\bar{t}H$ vertex involved at the leading order. The cross section of the four top quark production can be affected by the production induced with Higgs bosons and interaction terms

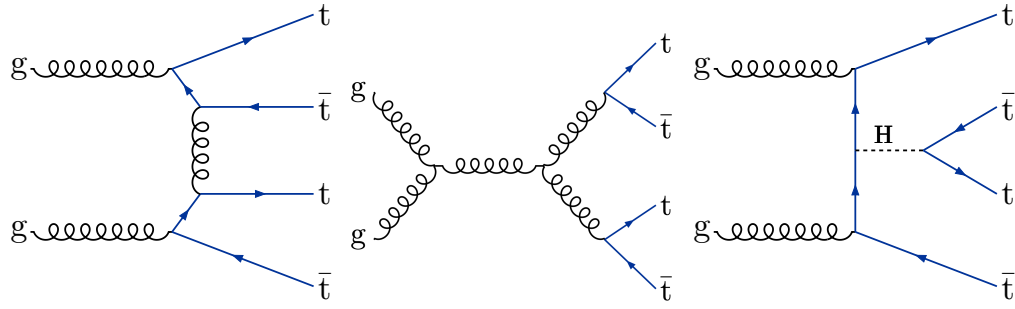


Figure 1.6: Leading order Feynman diagrams for four-top quark production. Notice that the third diagram has $t\bar{t}H$ vertices involved. Diagrams derived from Ref. [1] and Ref. [65].

between Higgs-induced production versus gluon/Z/photon-induced production. It is also proportional to the fourth power of the top quark Yukawa coupling where [66]

$$\sigma(t\bar{t}t\bar{t}) = \sigma^{\text{SM}}(t\bar{t}t\bar{t})_{g+Z/\gamma} + \kappa_t^2 \sigma^{\text{SM}}(t\bar{t}t\bar{t})_{\text{int}} + \kappa_t^4 \sigma^{\text{SM}}(t\bar{t}t\bar{t})_H \quad (1.109)$$

where $\sigma^{\text{SM}}(t\bar{t}t\bar{t})_{g+Z/\gamma}$, $\sigma^{\text{SM}}(t\bar{t}t\bar{t})_H$, and $\sigma^{\text{SM}}(t\bar{t}t\bar{t})_{\text{int}}$ are cross sections calculated from $t\bar{t}t\bar{t}$ diagrams with gluon/Z/photon-induced production, with Higgs production, and interaction terms between two types of productions, and $\kappa_t \equiv y_t/y_t^{\text{SM}}$. For 13 TeV centre-of-mass energy, the calculation at the leading order cross section yields the following [66]:

$$\sigma^{\text{SM}}(t\bar{t}t\bar{t})_{g+Z/\gamma} = 9.997 \text{ fb} \quad (1.110)$$

$$\sigma^{\text{SM}}(t\bar{t}t\bar{t})_H = 1.168 \text{ fb} \quad (1.111)$$

$$\sigma^{\text{SM}}(t\bar{t}t\bar{t})_{\text{int}} = -1.547 \text{ fb}. \quad (1.112)$$

Several final states can be categorised based on the leptons present in each event. The common categorisation of these final states is as follows:

- Single-lepton (SL) final state, containing either one electron or one muon

- Opposite-sign dilepton (OSDL) final state, containing two leptons (two electrons, two muons, or an electron and a muon), both having *opposite* electric charges
- Same-sign dilepton (SSDL) final state, containing two leptons (two electrons, two muons, or an electron and a muon), both having the *same* electric charges
- Multilepton (ML) final state, containing at least three leptons (electrons or muons)
- All-hadronic final state, containing no leptons

The main goal of searching this particle production is twofold. First, it can serve as a test for a prediction of the cross-section made based on SM [64]. Second, it can also serve as a test for several Beyond Standard Model (BSM) theories, which propose changes to the four-top quark production cross section by adding diagrams with extra interactions [67]. Examples of these BSM theories are top quark compositeness [68], extra mediator particle X in $t\bar{t} + X$, which subsequently decays into another pair of top quarks [69], and scalar gluon pair production [70].

1.8.1 Recent searches of four-top quark production from the ATLAS and CMS Experiments

Numerous searches for four-top quark production have been performed both in the ATLAS and CMS Experiments. The summary of search results by the ATLAS and CMS experiments since 2012 is shown in Tables 1.4, 1.5 and 1.6. As seen from all four tables, the search result for four-top quark production has improved over time, largely due to larger integrated luminosity $\mathcal{L}$ accumulated, the combination of several final states together, better reconstruction techniques, and the use of Machine Learning. These improvements are made to tackle the challenge posed by the search

Table 1.4: Cross section upper limit for four-top quark production searches by the CMS and ATLAS Collaborations, using data collected in 2012-2016. All results reported in this table do not have observed and expected cross section measurements reported.

$\sqrt{s}$	$\mathcal{L}$	Final state	Expected upper limit	Observed upper limit	
CMS Collaboration					
8 TeV	19.6 fb ⁻¹	SL	32 ± 17 fb	32 fb	[71]
13 TeV	2.3 fb ⁻¹	SSDL	102 ⁺⁵⁷ ₋₃₅ fb	119 fb	[72]
13 TeV	2.6 fb ⁻¹	SL+OSDL	118 ⁺⁷⁶ ₋₄₁ fb	94 fb	[73]
13 TeV	2.6 fb ⁻¹ + 2.3 fb ⁻¹	SL+OSDL + SSDL	71 ⁺³⁸ ₋₂₄ fb	69 fb	[73]
ATLAS Collaboration					
13 TeV	36.1 fb ⁻¹	SL	84 fb	60 fb	[74]
13 TeV	36.1 fb ⁻¹	SSDL+ML	29 fb	69 fb	[75]

Table 1.5: Cross section measurement for four-top quark production searches by the CMS Collaboration, using data collected in 2016-2018. The SSDL+ML analysis in the last row uses an outdated Run 2 integrated luminosity of 137 fb⁻¹, compared to the latest calculation of Run 2 integrated luminosity of 138 fb⁻¹ [76; 77; 78].

$\sqrt{s}$	$\mathcal{L}$	Final state	Observed significance	Cross section (fb)	
CMS Collaboration					
13 TeV	35.9 fb ⁻¹	SSDL+ML	1.6	16.9 ^{+13.8} _{-11.4}	[79]
13 TeV	35.8 fb ⁻¹	SL+OSDL	0	0 ⁺²⁰	[1]
13 TeV	35.8 fb ⁻¹ + 35.9 fb ⁻¹	SL+OSDL + SSDL+ML	1.4	13 ⁺¹¹ ₋₉	[1]
13 TeV	137 fb ⁻¹	SSDL+ML	2.6	12.6 ^{+5.8} _{-5.2}	[65]

for production, which will be further discussed in Section 1.8.2. Tables 1.4 and 1.5 also contain several results from Ref. [1] and Ref. [65].

The first search results for four-top quark production led by the CMS and ATLAS Collaborations, based on early Run 2 data, only yield cross-section upper limits. This is due to low statistics given by the limited amount of data collected. As more data have been collected from the detector, along with more sophisticated

Table 1.6: Cross section measurement for four-top quark production searches by the ATLAS Collaboration, using data collected in 2016-2018.

$\sqrt{s}$	$\mathcal{L}$	Final state	Observed significance	Cross section ($\sigma_{\text{SM}}^{t\bar{t}t\bar{t}}$)	
ATLAS Collaboration					
13 TeV	36.1 fb ⁻¹	SL+OSDL	1.6	1.7 ^{+1.9} _{-1.7}	[80]
13 TeV	36.1 fb ⁻¹	SL+OSDL			
	36.1 fb ⁻¹	+ SSDL+ML	2.8	3.1 ^{+1.3} _{-1.2}	[80]
13 TeV	139 fb ⁻¹	SSDL+ML	4.3	2.0 ^{+0.8} _{-0.6}	[81]
13 TeV	139 fb ⁻¹	SL+OSDL	1.9	2.2 ^{+1.6} _{-1.2}	[82]
13 TeV	139 fb ⁻¹	SL+OSDL			
	139 fb ⁻¹	+ SSDL+ML	4.3	2.0 ^{+0.8} _{-0.6}	[82]

analysis techniques and a combination of results from different final states, the search significance for this production greatly improves over time, approaching closer to 3 standard deviations. In the High Energy Physics community, the search significance of 3 standard deviations is considered as the *evidence*, corresponding to a probability¹⁰ of the result being a statistical fluctuation of 1 in 740. If the search result achieves a search significance of 5 standard deviations, it is considered an *observation*, with a probability of the result being a statistical fluctuation of 1 in 3.5 million. More details on search significance can be found in Section 5.2.2.

To measure the effectiveness of the analysis technique employed, an *expected significance* is calculated, based on the distribution dictated by Monte-Carlo samples with an assumption that the distribution contains an amount of signal events predicted by the SM. This can be done by using Asimov datasets [3], which is essentially the distribution populated by Monte-Carlo samples. The *observed significance*, however, is calculated from the data collected from the detector and may be vastly different from the expected significance. Further details shall be discussed in Chapter 5.

¹⁰Calculated from an upper-tail p -value [3]. See Section 5.2.2 for more details.

Since the new search result for four-top quark production, which will be presented in this thesis, contains 2016 analysis on opposite-sign dilepton (OSDL) final state and Run 2 analysis on same-sign dilepton (SSDL) and multilepton (ML) final states, an in-depth review of both analyses will be discussed in Sections 4.2.1 and 4.2.2 respectively. Section 4.2.1 also discusses the 2016 analysis on the single-lepton (SL) final state, to which the author of this thesis has made contributions.

During the preparation of this thesis, two new search results from both the CMS and ATLAS Collaborations on same-sign dilepton and multilepton have been announced [83; 84], with the observed search significance of 5.6 and 6.1 standard deviations respectively. These search significance values equate to the probability where the results are statistical fluke at *1 in 93 million* and *1 in 1.9 billion* respectively.

1.8.2 Challenges to the search for four top quarks

There are two main challenges to the search for four-top quark production. Due to a very low cross-section of 12 fb at centre-of-mass energy ($\sqrt{s}$) of 13 TeV, calculated at NLO with EW corrections [64], the four-top quark production suffers from a dominant background by top-antitop quark ($t\bar{t}$) production, where its cross section is 834 pb [85], roughly 70 000 times larger than the cross-section of four-top quark production. Furthermore, given a low cross-section of 12 fb, the total expected number of four-top quark production events (on average) with an integrated luminosity of 138 fb^{-1} (equivalent to data collected by the CMS detector during 2016-2018) is only $12 \times 138 = 1656$ events. This low number of events is not the absolute number of events that will be included for analysis, since collision events are subject to selection efficiencies, branching ratios, and acceptance for final states included in the analysis. These factors, especially the selection efficiencies

which are at a few per cent at best, cause the final number of events that is included in the analysis to be merely in the order of tens of events only. Therefore, large statistical uncertainties will affect the search sensitivity.

Large systematic uncertainties may also cause difficulties in the search. Recent searches performed by the CMS Collaboration [1; 65] reported that the uncertainty in $t\bar{t}$ + heavy flavour normalisation has the largest impact on the cross-section measurement. This uncertainty type is caused by the normalisation of events containing $t\bar{t}$ production associated with bottom quarks, or $t\bar{t} + b\bar{b}$, which is measured with uncertainty in Ref. [86]. From the theoretical standpoint, the $t\bar{t} + b\bar{b}$ production is difficult to model, since the b jet can be produced at very different energy scales in association with $t\bar{t}$ pair production, and can have its accuracy improved via perturbative QCD calculations [87].

Furthermore, a search in SSDL+ML final state also reported [65] that jet energy scale (JES) uncertainty has the second largest impact on the cross-section measurement. This type of uncertainty is caused by the uncertainties during jet energy corrections in the detector [88].

Figure 1.7 shows the distribution of BDT output from four-top quark production against several background processes. As seen from the figure, uncertainties compared to the number of predicted events are high.

On the other hand, recent searches performed by the ATLAS Collaboration [81; 82] reported that the uncertainty in the four-top quark production cross section and modelling caused the largest impact on its cross section measurement itself. This is followed by the modelling of events with $t\bar{t} + W$ and extra jets in the SSDL+ML final state [81], and the modelling of the $t\bar{t} + b\bar{b}$ process and $t\bar{t}$ with an extra charm quark in the SL+OSDL final state [82].

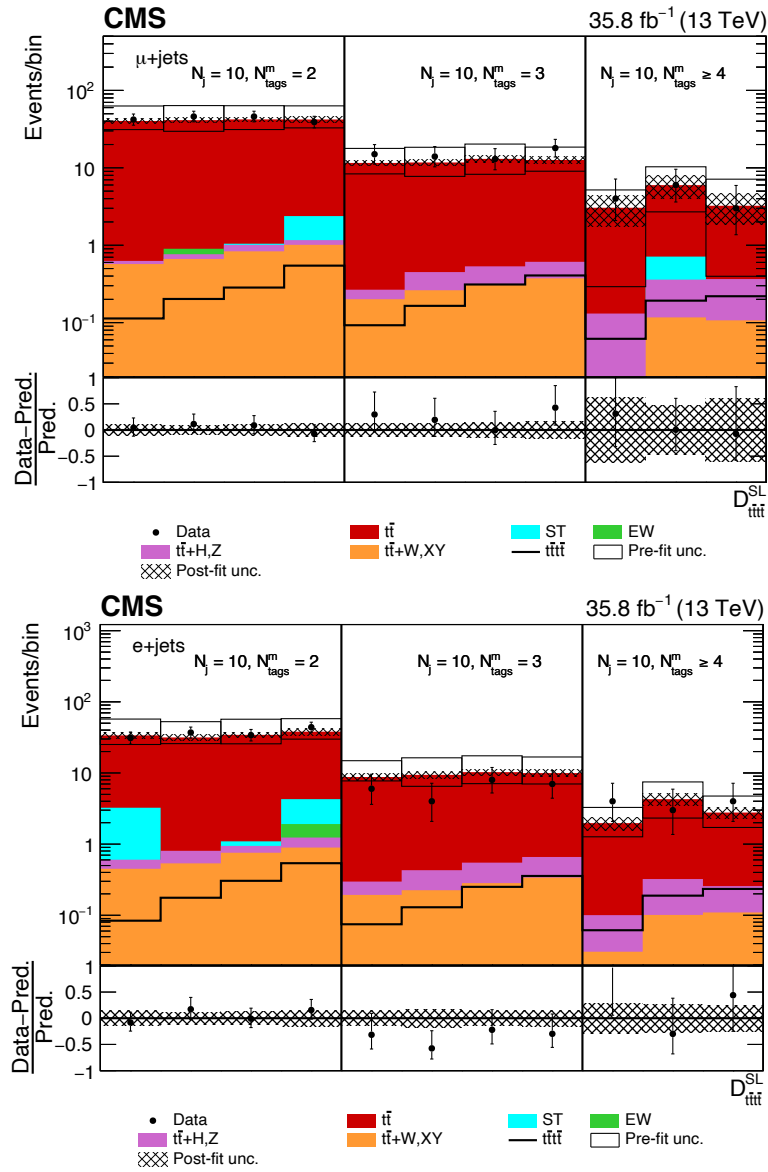


Figure 1.7: BDT output distributions of four-top quark production compared with background processes ($t\bar{t}$, single-top production, electroweak, $t\bar{t}$ with Higgs or Z boson, $t\bar{t} + W$, and $t\bar{t}$ with two vector bosons (WW, WZ, or ZZ), from the single-lepton final state with muons (upper panel) and electrons (lower panel), and 10 or more jets [1]. BDT bins are arranged to have an equal amount of background events in each bin. Notice a large amount of postfit uncertainty in the 10-jet, 4+ b-jets category (rightmost section).

1.8.3 Yukawa coupling measurement based on four-top quark production

The measured four-top quark cross section can be further used in the top quark Yukawa coupling (y_t) measurement. Based on the search result of the production performed in same-sign dilepton and multilepton (SSDL&ML) final state over Run 2 data of 138 fb^{-1} , reported in Ref. [65], the value of y_t is constrained. Figure 1.8 illustrates the measurement made by the CMS Collaboration. In this measurement, the observed four-top quark cross section and its upper limit are calculated with different values of y_t and then compared with the predicted values. Based on the comparison between the cross section upper limit and the predicted cross section values, the upper limit of y_t is, at a 95% confidence level,

$$\left| \frac{y_t}{y_t^{\text{SM}}} \right| < 1.7 \pm 0.3 \quad (1.113)$$

1.8.4 BSM interpretations based on four-top quark production

In addition to searches for the four-top quark production, the search results can also be used for interpretations based on Beyond Standard Model (BSM) theories. Several interpretations have been made in results from both CMS and ATLAS. This section, however, will only discuss measurements of Wilson coefficients in Effective Field Theory (EFT).

The EFT framework was first introduced in 1986 by Buchmüller and Wyler as “an extension to Standard Model Lagrangian, assuming extra terms in the order of intermediate-mass scale Λ ” [89]. The extension is in the form of extra terms for SM Lagrangian in a similar fashion to the Taylor series expansion.

$$\mathcal{L}_{\text{effective}} = \mathcal{L}_{\text{SM}} + \frac{1}{\Lambda} \sum_k C_k^{(5)} \mathcal{O}_k^{(5)} + \frac{1}{\Lambda^2} \sum_k C_k^{(6)} \mathcal{O}_k^{(6)} + \dots \quad (1.114)$$

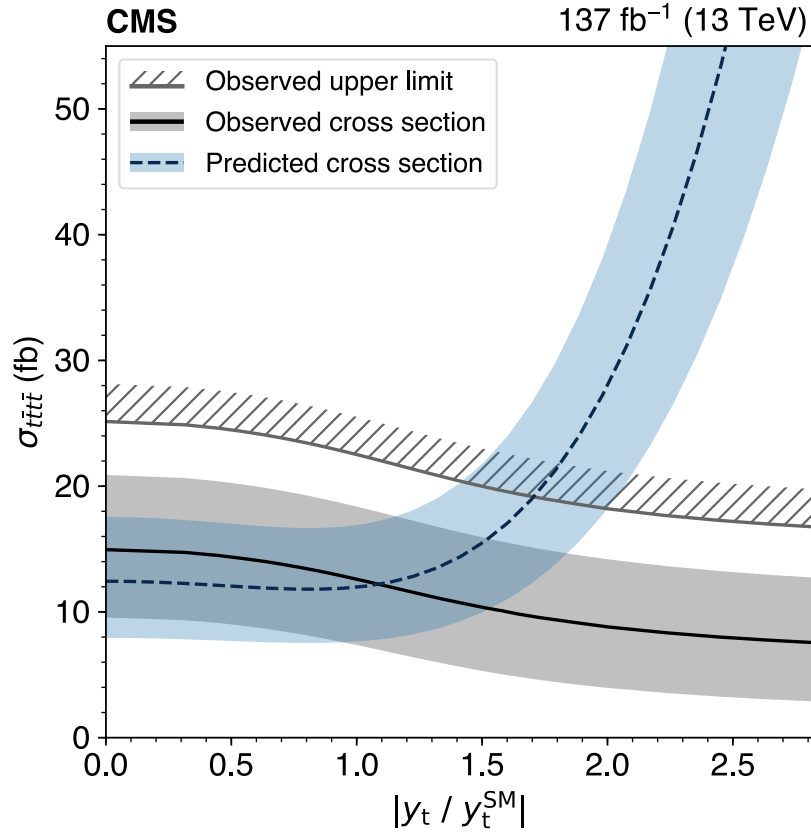


Figure 1.8: Observed four-top quark cross section and its upper limit, based on Ref. [65], compared to predicted values with top quark Yukawa coupling varied. The upper limit is obtained by comparing the observed upper limit (hashed line) against the predicted cross section (shaded band around the dashed line).

However, the expansion shown in the equation above is valid at low energy up to the order of Λ [89]. The extension consists of two parts, the *operator* $\mathcal{O}_k^{(n)}$ and its *Wilson coefficients* $C_k^{(n)}$. Based on this framework, if SM is perfect, all Wilson coefficients will be zero, and the Lagrangian will simply collapse into SM Lagrangian.

Based on the four-top quark cross-section upper limit in single-lepton, opposite-sign dilepton, same-sign dilepton, and multilepton final states combined, the CMS Collaboration measured the Wilson coefficients of only *four* operators as

follows [1]:

$$\mathcal{O}_{tt}^1 = (\bar{t}_R \gamma^\mu t_R) (\bar{t}_R \gamma_\mu t_R) \quad (1.115)$$

$$\mathcal{O}_{QQ}^1 = (\bar{Q}_L \gamma^\mu Q_L) (\bar{Q}_L \gamma_\mu Q_L) \quad (1.116)$$

$$\mathcal{O}_{Qt}^1 = (\bar{Q}_L \gamma^\mu Q_L) (\bar{t}_R \gamma_\mu t_R) \quad (1.117)$$

$$\mathcal{O}_{Qt}^8 = (\bar{Q}_L \gamma^\mu T^A Q_L) (\bar{t}_R \gamma_\mu T^A t_R) . \quad (1.118)$$

Since the operators are added directly to SM Lagrangian, the cross-section for four-top quark production can be altered by Wilson coefficients, and this analysis uses the altered cross section to constrain the values of each Wilson coefficient.

On the other hand, the ATLAS Collaboration has set an upper limit of *one* Wilson coefficient from the search for four-top quark production [80], but the procedure is different from the CMS Collaboration. In this analysis, event samples for four-top quark production with a modified Wilson coefficient for the four-fermion interaction operator, $|C_{4t}|/\Lambda^2 = 4\pi \text{ TeV}^{-2}$, are generated and considered as an extra signal process. The observed cross-section upper limit is then calculated with the extra signal process. Based on the combination of SL+OSDL+SSDL+ML final states, the observed upper limit of the Wilson coefficient $|C_{4t}|$ is calculated to be $|C_{4t}|/\Lambda^2 < 1.9 \text{ TeV}^{-2}$.

Chapter 2

THE COMPACT MUON SOLENOID DETECTOR

The Compact Muon Solenoid (CMS) detector is a general-purpose particle detector situated along the Large Hadron Collider, along with three other main particle detectors: ATLAS [90], ALICE [91], and LHCb [92], as well as other smaller particle detectors installed along the beamline. Its purpose is to study the widest range of physics possible, including the discovery of the Higgs boson in 2012 [39] along with ATLAS [38]. Although the purposes of the CMS detector are mostly identical to ATLAS, the designs used in both detectors are different, and physics results between CMS and ATLAS Collaborations are sometimes combined for better precision.

The CMS detector has undergone several upgrades since its first technical proposal was released in 1994 [93]. In this thesis, the design and specifications of the detector used during the data-taking of Run 2 in CMS, spanning from 2015-2018, shall be discussed.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a circular particle accelerator with a circumference of 27 km. It usually accelerates two beams of protons¹, each with 1.2×10^{11} and collide with each other at the centre-of-mass energy of 13 TeV. During

¹LHC also accelerates beams of heavy lead ions, which are used for heavy-ion studies in the ALICE detector [91], the ATLAS detector, and the CMS detector. The LHC also accelerates ions of other elements such as Xenon [94]. Such studies will not be discussed here.

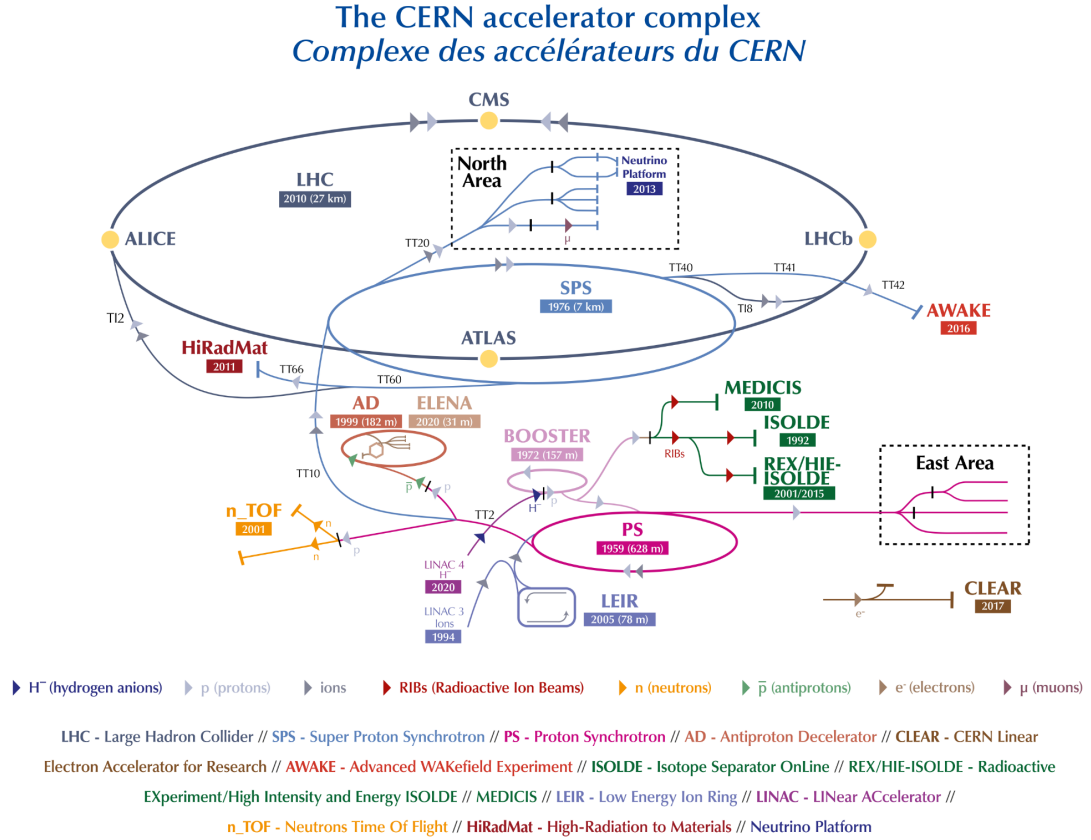


Figure 2.1: The CERN accelerator complex, depicting LINAC, Booster, PS, SPS, and LHC, along with other experiments at CERN [97].

the Run 2 data-taking campaign, the peak instantaneous luminosity achieved by the LHC of $2.0 \times 10^{-34} \text{ cm}^{-2} \text{ s}^{-1}$ [95], which is twice the initial goal set in 2004 [96].

The main component of the LHC accelerator is the superconducting dipole magnets, providing a magnetic field of 8 T to steer the proton beams along the circular path. The magnets operate at an extremely low temperature of 1.9 K to maintain the superconducting properties of Niobium Titanium (Nb-Ti) cables [95; 96]. However, the LHC will not accelerate particles from scratch, but rather receive particles from an existing accelerator complex, starting from Linear Accelerator (LINAC), Booster, Proton Synchrotron (PS), and Super Proton Synchrotron (SPS) [96], as shown in Figure 2.1.

2.2 The overall design of the CMS detector

The CMS detector is a general-purpose particle detector that is designed to detect as many types of particles as possible, unlike more specialised detectors, such as the LHCb detector, that focus on highly-forwarded particles, or particles that travel near the beamline. As such, the detector consists of several layers of particle detectors, covering the pseudorapidity area of $|\eta| \leq 2.4$, and a superconducting magnet, necessary for charged particle momentum measurements.

As documented in the original CMS technical proposal in 1994 [93], the design goal for CMS is to have “a very good and redundant muon system”, “the best possible electromagnetic calorimeter” consistent with the muon system, “a high quality central tracking” for the muon system and electromagnetic calorimeter, and “a financially affordable detector”. These design ethos are focused on the muon system, which is different from ATLAS’s design goal where, as stated in its technical design report in 1994 [98], they are focused on “very good electromagnetic calorimetry” and “stand-alone, precision, muon-momentum measurements up to the highest luminosity”. Over time, the design of the CMS detector changed, with improvements in detector components to accommodate higher luminosities delivered by the LHC.

Figure 2.2 shows a transverse slice of the CMS detector with several components. Starting from the collision point at the centre of the detector (or the leftmost point of the figure), the components are **trackers**, **electromagnetic calorimeter (ECAL)**, **hadronic calorimeter (HCAL)**, **superconducting solenoid**, and **Muon chamber**. Different types of particles, which are muons, electrons, photons, and hadrons (both charged and neutral), leave energy residues in different layers of the detector, which can be reconstructed into different types of particles during physics analyses. Neutrinos and many exotic particles do not leave

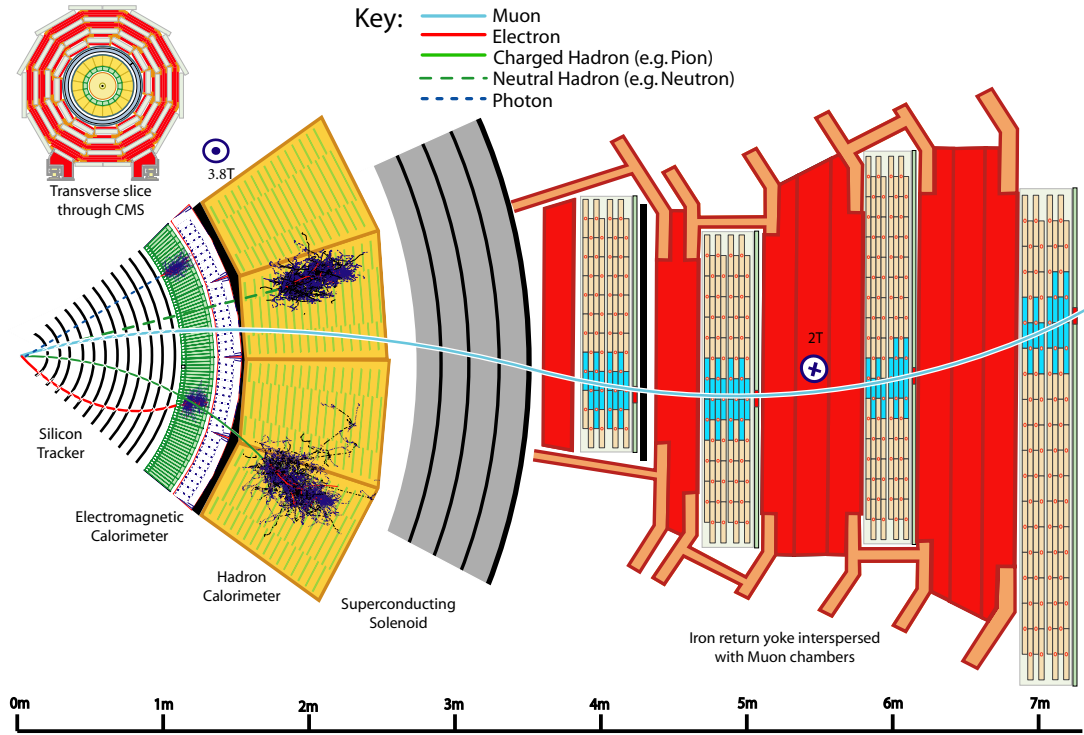


Figure 2.2: A transverse slice of the CMS detector as of 2008, showing several components and energy residues caused by several types of particles passing through them [99].

traces in the detector² and result in missing transverse energy for each collision event [101].

2.3 Superconducting solenoid

The superconducting solenoid provides a magnetic field of 3.8 T (4 T at ideal conditions) within the cavity with a diameter of 6 m and a length of 12.5 m. To maintain a strong amount of magnetic field, niobium-titanium (NbTi) superconducting wires are used to conduct 19.14 kA of electrical current. To

²One exception for exotic particles is heavy stable charged particles, which leave a signature in the form of a particle with high momentum and a large rate of energy loss through ionization. The CMS Collaboration has performed a search for this type of exotic particle by using the information from trackers and muon chambers [100].

maintain low temperatures within the superconducting wires, the vacuum system and cryogenic plant are also installed to maintain vacuum and temperature in the solenoid [101].

2.4 Silicon trackers

Silicon trackers are the innermost component of most LHC detectors, which are designed to track charged particles right after the collision. The CMS tracker has two main components, which are a pixel detector and a silicon strip tracker [101]. The pixel detector contains four barrel layers and three forward disks per side, with the innermost barrel layer 2.9 cm away from the interaction point [102; 103]. As it is closest to the interaction region, they have a small pixel size of $100 \times 150 \mu\text{m}^2$ [101].

Silicon strip trackers are made of microstrip sensors and located further away from the collision point [101], with ten barrel layers (four inner layers plus six outer layers) and twelve forward disks (three inner disks plus nine outer disks) on each side of the detector [104].

2.5 Electromagnetic calorimeter (ECAL)

Electromagnetic calorimeters (ECAL) are the next component outside of the Silicon Trackers, consisting of 61 200 lead tungstate (PbWO_4) crystals arranged in the barrel part, and 7324 crystals arranged at each endcap [101]. They are designed to detect high-energy photons and electrons by detecting scintillating light caused by the particles within the crystal³, using avalanche photodiodes (APDs) and vacuum phototriodes (VPTs). To help reject neutral pions, a special component named the

³A scintillator, such as PbWO_4 crystals, converts high-energy photons and charged particles into ultraviolet or visible light [106].

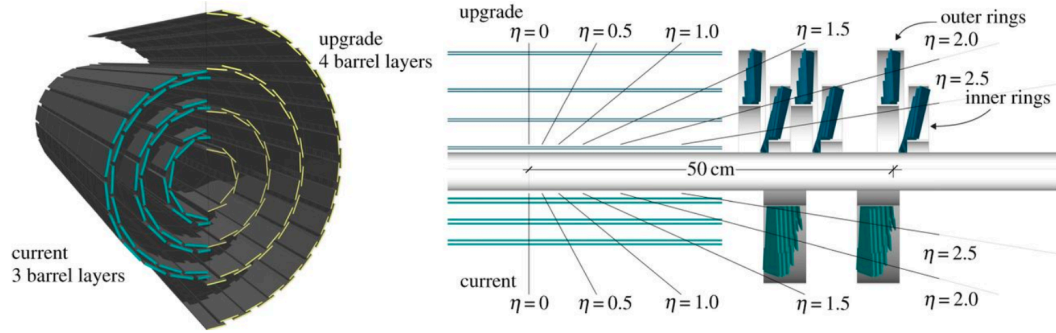


Figure 2.3: CMS Pixel Tracker schematics comparison between the configuration used since 2009 (labelled “current” in this diagram) and the configuration since 2017 (labelled “upgrade” in this diagram) configurations [103]. The 2017 configuration of the tracker in this diagram is used in the Run 2 data-taking period.

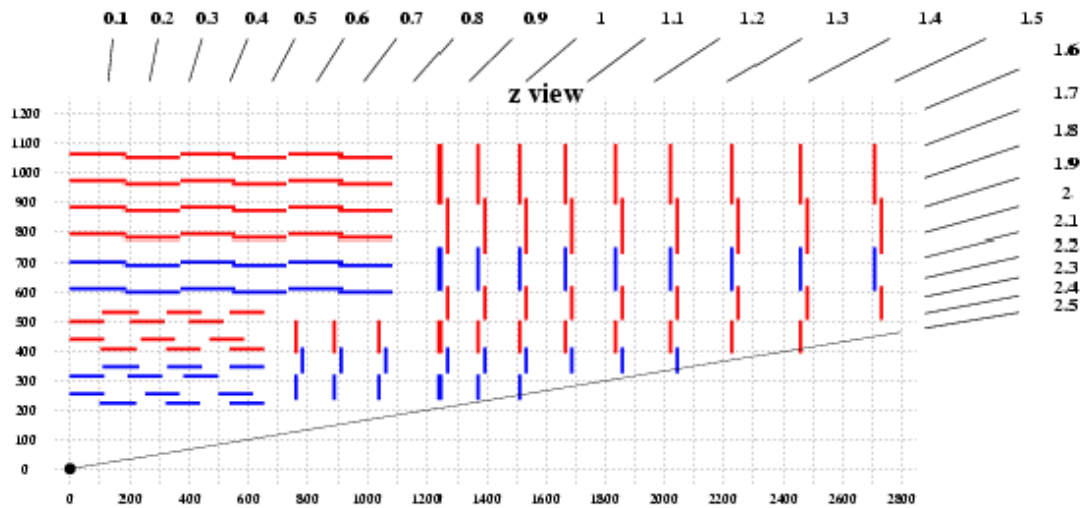


Figure 2.4: CMS Silicon Strip Tracker Schematics on a longitudinal plane, showing a quarter of the whole Strip Tracker system [105].

preshower detector is added before the endcap component of ECAL. ECAL has a coverage of up to $|\eta| < 3.0$ [101].

2.6 Hadronic calorimeter (HCAL)

The hadronic calorimeter (HCAL) immediately follows the ECAL, detecting both charged and neutral hadrons, with a coverage of $|\eta| < 3$. The calorimeter contains a combination of brass absorber plates and scintillators. Unlike silicon trackers and ECAL, which are located entirely inside the superconducting solenoid, HCAL also has an *outer* component outside of the solenoid to ensure that all hadron showers are contained within HCAL. Furthermore, an additional set of forward HCAL detectors is placed 11.2 m from the interaction point, increasing the coverage up to $|\eta| < 5.2$ [101].

2.7 Muon chamber

Muon measurements in the CMS detector have been the most important theme since the earliest designs of the detector, since they can lead to high discovery potential of an array of physics in the Standard Model, such as Higgs boson decay into two Z bosons, SUSY models, and so on. The current muon system present in the detector identifies muons, measures momentum and provides triggers. The muon system consists of three components, which are drift tubes (DT) located in the barrel region, cathode strip chambers (CSC) in the endcap region, and resistive plate chambers (RPC) located in both barrel and endcap regions [101].

To complement the CSC in the endcap region, gas electron multipliers (GEMs) were first installed in 2017. This upgrade can maintain, under higher luminosity, or even improve the forward muon triggering and reconstruction at regions of $1.6 <$

$|\eta| < 2.2$. They will be gradually implemented within the CMS detector, with the plan to finalise the integration by the end of Long Shutdown 3 (LS3) in 2026 [107].

2.8 Trigger

The LHC provides proton bunches which are 25 ns apart. This means new collision will take place within the CMS detector every *25 nanoseconds*, resulting in a maximum collision rate of 40 MHz. The detector cannot store a large amount of collision data at this rate. Hence, the detector must decide which events are interesting enough to save for physics analyses. To narrow down the amount of data, two trigger systems are used in the detector. The Level-1 trigger trims down the event rate to 100 kHz by utilising custom hardware [108], and the High Level Trigger (HLT) trims the data rate further to 1000 Hz by performing particle tracking and reconstruction using a normal computer CPU. Events passing the HLT requirement will be further sent to the computing farm for further reconstruction (with more accuracy) and analysis [109].

2.9 Reconstruction of events using Particle Flow (PF) algorithm

Since physics analyses require information on *physics objects* in the collision event, such as leptons, photons, and particle jets coming from quarks or gluons, they must be *reconstructed* from raw data from all components of the detector. This can be done using the Particle Flow algorithm [99], which has been first used since the era of the Large Electron-Positron (LEP) collider [110]. Comparisons between simulation and data recorded by the detector at $\sqrt{s} = 8$ TeV show excellent agreement of the reconstruction of these physics objects. Subsequent upgrades to

the detector and the algorithm itself should allow good performance for up to 200 pile-up interactions delivered by the LHC [99].

The reconstruction starts from the raw data from the detector, especially from silicon trackers. The high-resolution tracker already available in CMS provides momentum information of charged hadrons, which is essential to jet reconstruction [99].

Tracks for charged hadrons are reconstructed using an iterative tracking strategy, where particle tracks are reconstructed in the first iteration with very tight criteria, and then reconstructed with loosened criteria in subsequent iterations. The fake rate for charged particles with low hit count in the tracker, low p_T and an origin vertex far from the collision point is in the order of 1% [111].

Energy deposits detected in calorimeters are grouped into clusters. With clustering, we can detect photons, neutral hadrons, electrons, and charged hadrons separately, and improve energy measurement of charged hadrons in addition to track reconstruction.

Charged-particle tracks, clusters in calorimeters, and muon tracks in muon chambers need to be connected by a link algorithm to reconstruct particles. The algorithm produces “blocks” of these elements (tracks and clusters) linked based on the extrapolation of the track position.

The Particle Flow algorithm then takes this information to reconstruct physics objects, starting from muons, electrons, photons, particle jets, and finally missing transverse energy. First, tracks corresponding to particle-flow muons are removed from the block of elements, and then a particle track and ECAL clusters corresponding to particle-flow electrons are removed. After tracks associated with electrons and muons are removed, the remaining tracks are identified as

particle-flow charged hadrons. If the energy deposit in the calorimeter cluster is larger than the momentum associated with the charged particle, particle-flow photons and neutral hadrons are reconstructed based on the excess energy in those clusters. The remaining calorimeter clusters not associated with any remaining tracks are treated as particle-flow photons and neutral hadrons. Finally, as collisions in the CMS detector are head-on, the vector sum of the transverse momentum of all particles must be zero. Thus, transverse missing energy is defined as the complement of that vector sum [111].

From these particle-flow candidate particles, jets can be reconstructed using anti- k_T algorithm [112]. This jet clustering algorithm takes into account the transverse momentum, rapidity, and azimuth angle of each particle, and clusters them together based on the distance between each particle until no particles are left. With this jet clustering algorithm, reconstructed particle jets can have a cone shape, or partly conical if they are close together. The algorithm also requires the distance parameter (radius parameter as in Ref. [112]) R , which determines how large the jet cone can become. Usually, the value of $R = 0.4$ is used, and jets reconstructed with this distance parameter are referred to as AK4 jets. In some cases, however, $R = 0.8$ can also be used to reconstruct larger jets to accommodate larger decay, such as top quark decay at low p_T , and are usually called AK8 jets or fat jets.

Jet energies are calibrated by correcting for energy caused by other proton-proton collisions (also called pileup), the detector response to hadrons, and the difference between data and MC simulations. Jet p_T resolution, on the other hand, is extracted from data and MC simulations [88].

As many analyses, especially in the top physics sector, rely on the information on which jet originated from bottom quarks, a procedure named b tagging is crucial to physics analyses. Many b-tagging algorithms are commonly used in several

analyses. DeepJet [113], for example, is a b-tagging algorithm based on neural networks, leveraging low-level features from as many jet constituents as possible.

2.10 Data Quality Monitoring (DQM) and Data Certification (DC)

As with all scientific apparatus, components in the CMS detector may fail and will sometimes be unable to detect particles. Failures in these components may cause the recorded data to be unsuitable for physics analyses. Hence, monitoring the detector during the data taking (also called online operations) is critical to keeping the recorded data to be of high quality. Furthermore, certification of the recorded data after the data taking (also called offline operations) is also crucial to determining which sets of recorded data are suitable for physics analyses.

For this purpose, the Data Quality Monitoring (DQM) and Data Certification (DC) workflows within the CMS Collaboration monitor the quality of raw data during online operations, as well as certify the data during offline operations. During a typical online operation, a small portion of the data is sent to the online DQM backend computers, which is then passed onto the DQM GUI⁴, which is a web application displaying important histograms derived from that small portion of the data. Human DQM shifters will inspect the histograms, and notify detector experts if they detect anomalies within the histograms.

Due to the large amount of data recorded from the detector, DQM shifters and subsystem experts must certify the data on a run-by-run basis, which means they must certify a set of data known as *a run* recorded during stable conditions as a whole. A run can be as short as 30 minutes or as long as 8 hours or more,

⁴Graphical User Interface

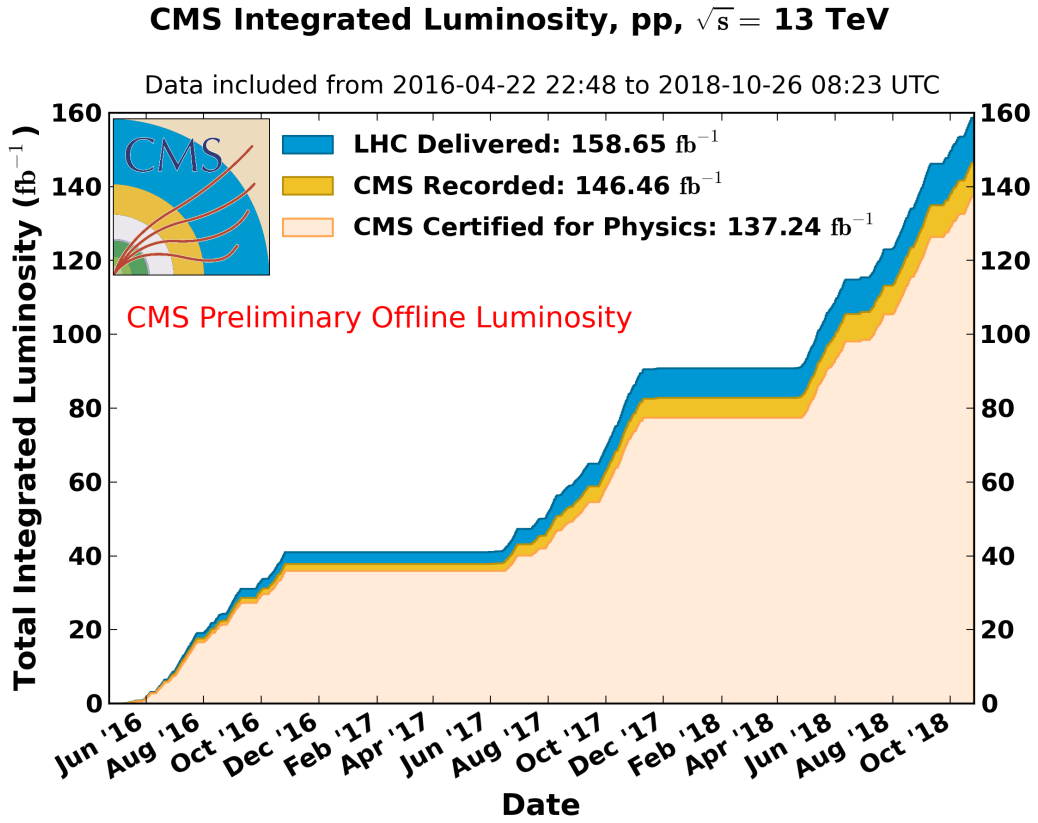


Figure 2.5: Cumulative integrated luminosity for proton-proton collisions during 2016-2018, showing the amount of data delivered by LHC, recorded by CMS, and certified for physics [114].

which is usually made by turning the detector on during the whole night. Figure 2.5 illustrates the amount of data delivered by the LHC, recorded by CMS, and certified for physics analyses, which shows high efficiency in retaining quality data by the whole collaboration. However, the Run 3 data-taking campaign, which started in 2022, and High-Luminosity LHC (HL-LHC) upgrades are challenging to both workflows in terms of a larger amount of data to monitor and certify [5].

2.10.1 Machine Learning applications for DQM

The current workflow certifying data on a run-by-run basis imposes two problems. Since one run of the data can be split into smaller portions called *lumi sections*,

some detector problems may surface in a small portion of the run, which may not be detectable if it is certified run-wide. Furthermore, with a large amount of data that needs to be certified *by humans*, errors may arise due to human fatigue [5].

To alleviate the issues caused by the current human-based workflow on a run-by-run basis, machine learning techniques have been explored to assist DQM shifters and subsystem experts in certifying the data. With these improvements, the data can be certified with higher granularity, down to a single lumi section, and the repetitive part of workloads to human shifters and experts can be offloaded to machines.

Figure 2.6 shows an example implementation using residual networks [4] to detect anomalies in the silicon pixel tracker. The figure presents a reconstruction of occupancy plots in pixel barrel layer 1 using residual networks, which is a class of neural networks, from three lumi sections of run 297100. The first and third lumi sections show good reconstruction of the occupancy plots, which means the pixel tracker was operating normally. However, the second lumi section depicted in the figure shows reconstruction failure, as shown in the mean squared error loss function in the third column, signifying an anomaly during the lumi section.

Furthermore, there is currently a development effort for a framework that consolidates detector data for machine learning training, as well as automates machine learning models and provides reports on the model's performance after the training [5].

2.11 Monte Carlo simulation

So far, the complete chain of the data from colliding particles inside the detector to physics analyses start from the particle accelerator (LHC), particle detector

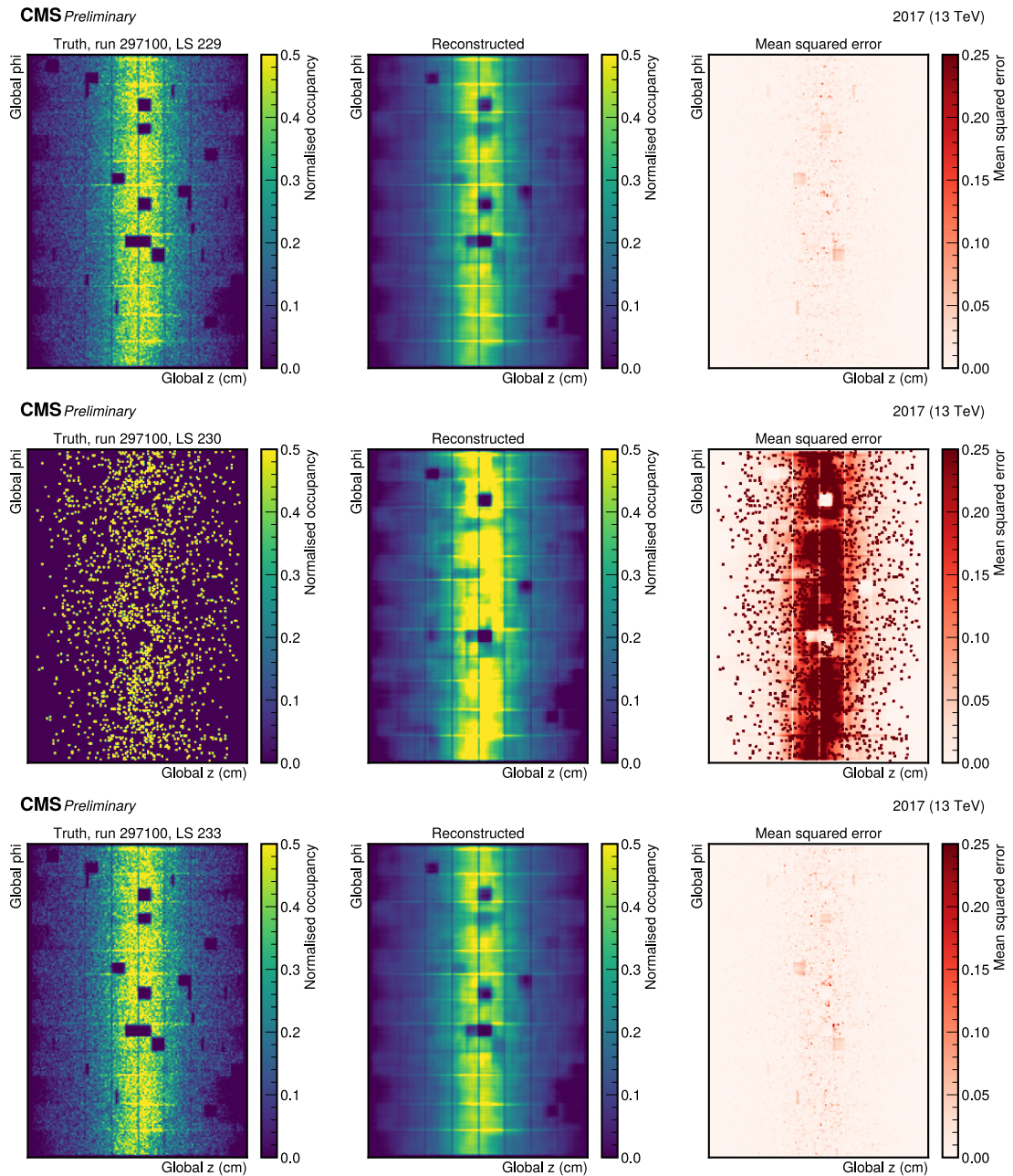


Figure 2.6: Reconstruction of occupancy plots from pixel barrel layer 1 using residual networks [4], a class of neural networks. Notice the reconstruction failure in the middle row (run 297100, lumi section 230), which can be classified as an anomaly in the silicon pixel tracker. The top and bottom row represents lumi sections 229 and 233 with good reconstruction from the model, which can be classified as normal operations in the tracker [5; 115].

(CMS), event reconstruction (using Particle Flow algorithm), and physics analyses (using data analysis framework such as ROOT [116]). However, according to quantum mechanics, identical particles coming out of a collision cannot be distinguished [117]. Therefore, we may not be able to understand what kind of physics interactions happen inside the particle detector due to this limitation.

To design new analysis strategies and estimate performance for certain physics scenarios, we perform Monte Carlo-based simulations to study the signals of certain physics processes, such as $t\bar{t}t\bar{t}$ versus $t\bar{t}$ production, by using a set of programs working on different parts of the data chain. They are [118]:

- Event generators, designed to generate hard scattering processes based on the given production physics. Famous general-purpose event generators are Python-based MADGRAPH5_AMC@NLO [119], PYTHIA 8 [120], and HERWIG [121]. However, there are many more programs for specific scenarios available as extensions to general-purpose event generators [118].
- Detector simulator, designed to simulate the detector response once particles leave the collision into different parts of detector materials. The gold standard used in several experiments is GEANT4 [122], which is widely used in different particle detectors, such as CMS, ATLAS, or even Jiangmen Underground Neutrino Observatory (JUNO), located just outside of Jiangmen, China [123], and many more.

The event reconstruction algorithm and data analysis framework are designed to be the same as in the data chain. Furthermore, fast simulation programs, such as Delphes [124] and CMS Collaboration's FASTSIM [125] can perform fast detector simulation and physics object reconstruction with less accuracy.

Since event generators are built based on limited modelling principles to reduce complexity and increase the speed of event generation, they may provide inaccuracies compared to actual data from the detector. Several comparisons between different event generators and data are usually made, and small correction factors must be included to match predictions provided by event generators to data at sub-percent agreement. For example, b-tagging algorithms have different efficiency measured between Monte Carlo samples and data. Scale factors for b-tagging algorithms are calculated to calibrate the efficiency of the simulation to match the efficiency from data [126].

Chapter 3

INTRODUCTION TO MACHINE LEARNING AND NEURAL NETWORKS

Machine learning has been widely used in many aspects to identify patterns among huge amounts of data. It has a very diverse set of applications, ranging from simple discrimination and value prediction to more complex use cases such as data generation and anomaly detection. As the work in this thesis has connections to Machine Learning techniques, especially on the measurement of top Yukawa coupling based on four-top quark production, we need to understand the basic concepts first. In this chapter, we will discuss briefly about Machine Learning and neural networks, which are a very popular class of Machine Learning techniques.

3.1 What is Machine Learning?

In essence, Machine Learning is a technique that tries to find a *model* that gives a target value based on its input data, much like a function that accepts inputs and outputs either a single number, a vector, a matrix, or a tensor. The data can be in any form, such as numbers in tables, text, and images. Machine Learning models that give outputs as numbers are a subclass of ML models.

Mathematically speaking, given a collection of data $\{\vec{x}_i\}$ and a target value $\{y_i\}$, we need to find a model $f(\vec{x})$ that satisfies

$$f(\vec{x}_i) = y_i. \quad (3.1)$$

The input data can have more than one variable, or feature, such as kinematic information of leptons in each collision event. Thus, the input data that is inputted

into the model is usually written in the form of vectors. The function $f(\vec{x})$ can be in any form as long as it satisfies Equation 3.1, that is, it can explain the relationship between input data $\vec{x}_i$ and target output y_i .

With this definition, we can apply it to solve various types of problems. An example case is a classification of flower species, first proposed by Fisher in 1936 [127]. The input data, $\{\vec{x}_i\}$, is a collection of four flower features, while the output data is the flower species, namely *Iris setosa*, *Iris versicolour*, and *Iris virginica*. The input data can be interpreted as a four-dimensional vector, whereas the output target can be interpreted as a number or a three-dimensional vector, where $[1\ 0\ 0]^T$ represents *I. setosa*, $[0\ 1\ 0]^T$ represents *I. versicolour*, and $[0\ 0\ 1]^T$ represents *I. virginica*. Therefore, the goal of our Machine Learning model is to find a model that translates a four-dimensional vector encoding sepal and petal length and width into a three-dimensional vector encoding one of the three species of *Iris* flowers.

The original paper by Fisher also proposed a model which is very similar to a linear model, where a vector of four flower features is multiplied by a 1×4 matrix, resulting in one discriminator value. The discriminator value is then evaluated further using statistics. Note that, since the model we are trying to find can be in any form, the model proposed by Fisher which uses a linear model and outputs a single discriminator value can be called a Machine Learning model. On the other hand, we can develop a neural network that accepts a vector of four flower features and outputs a three-dimensional vector encoding the probability that a flower is one of the three species, and the neural network can still be considered a Machine Learning discriminator as well.

Two basic types of Machine Learning problems are usually considered:

- Classification problems require the model to predict *certain classes of entries*, such as the classification of flower species as described above.
- Regression problems require the model to predict the *value* of a variable, given the information for each entry. An easy example of this type of problem is linear regression, where the model is a straight line $y = mx + c$, and the input data is the data points to be fitted with the straight line using linear regression.

3.2 What are Boosted Decision Trees?

Boosted Decision Trees (BDTs) are very simple Machine Learning models based on the idea of decision trees. Normally, one decision tree can be constructed based on the training data to classify several classes of data. A decision tree contains decision nodes based on a feature value that splits the dataset into smaller sets, or a terminal node that decides which class the remaining dataset belongs, in the same fashion as a Dichotomous key in Biology.

One decision tree can be considered as a *weak* classifier since it relies on a series of linear cuts based on the input features. To increase the classification power, an ensemble of decision trees can be combined into *Boosted* decision trees. For boosted decision trees, the first tree is trained on a training dataset, and then subsequent trees are trained based on the training data that is misclassified by prior trees (the second tree trains on the data misclassified by the first tree, the third tree trains on the data misclassified by the first two trees, and so on) [128].

With its simplicity, BDTs are used in several physics analyses involving the classification of collision events from signal and background processes, such as the discovery of the Higgs boson in CMS detector [39] and several searches for four top quark production, which will be introduced in Chapter 4.

3.3 What are neural networks?

Neural networks are a class of Machine Learning that is inspired by the inner workings of human brains. They have been developed and successfully applied to several types of problems, ranging from classification and regression to data generation, anomaly detection, and beyond. There are now several classes of neural networks being developed, such as shallow neural networks, deep neural networks, convolutional neural networks [129], graph neural networks [130], residual networks [4], and recurrent neural networks [131], and so on. Furthermore, new classes of neural networks are under active development, and numerous scientists are trying to understand how and why neural networks can explain the results with such a good performance.

Neural networks have the flexibility for a myriad of applications, such as the estimation of BDT output distribution in a data-driven approach using neural autoregressive flows [132], which will be discussed in Section 4.5.1, or autoencoders based on residual networks used in Data Quality Monitoring workflows, as discussed in Section 2.10.1 [5]. It can also be constructed to distinguish between collision events originating from SM versus non-SM cases, which will be discussed further in Chapter 6.

In this thesis, we shall discuss the most basic type of neural network, which sometimes can be called feed-forward neural network in some old literature. Complex networks, such as convolutional neural networks, graph neural networks, residual networks, and recurrent neural networks are beyond the scope of this thesis.

3.3.1 Perceptrons

The most fundamental principle of neural networks is a *neuron*. In Biology, a neuron contains two unique structures, dendrites and axons. Dendrites receive

electrical signals from other neurons, and axons send electrical signals towards other neurons [133]. If an input towards a neuron is strong enough, the neuron will emit its signal towards other neurons. The neuron will emit a full output signal if and only if an input passes its threshold. This relationship can also be called “all-or-nothing” [134].

In 1958, Rosenblatt proposed a mathematical model that mimics the neurons called a *perceptron*. In essence, it is a function that accepts inputs $x_1, x_2, \dots, x_n$ and emits output in a piecewise fashion as follows:

$$f(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{if } \sum_i w_i x_i + b > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3.2)$$

This model employs the idea of a neuron accepting multiple input channels and the “all or nothing” relationship of the neuron as described earlier. It also contains two types of parameters: *weights* w_i and *bias* b . These two sets of parameters dictate how a perceptron gives its output given a set of inputs. One can connect these perceptrons and form layers upon layers of perceptrons connecting, resulting in a larger network of perceptrons.

3.3.2 Neurons

Due to the piecewise nature of perceptrons, it is possible that tuning a parameter of one perceptron, be it a weight or a bias, can result in a drastic change of the perceptron output from 1 to 0 and vice versa. To avoid this, the perceptron model can be changed into something more differentiable. Hence, we arrive at a *revised* version of the perceptron model as

$$f(x_1, x_2, \dots, x_n) = F\left(\sum_i w_i x_i + b\right) \quad (3.3)$$

where $F(x)$ is called an activation function. Activation functions can be in any form, such as sigmoid, hyperbolic tangent (tanh), Rectified Learning Unit (ReLU), and its

variations as follows:

$$\text{sigmoid}(x) = \frac{1}{1 + \exp(-x)} \quad (3.4)$$

$$\tanh x = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} \quad (3.5)$$

$$\text{ReLU}(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (3.6)$$

$$\text{SELU}(x) = \begin{cases} \lambda x & \text{if } x \geq 0 \\ \lambda \alpha (e^x - 1) & \text{otherwise} \end{cases} \quad (3.7)$$

$$\text{LeakyReLU}(x) = \begin{cases} x & \text{if } x \geq 0 \\ \alpha x & \text{otherwise.} \end{cases} \quad (3.8)$$

Figure 3.1 shows the shape of sigmoid (Equation 3.4) and tanh (Equation 3.5), while Figure 3.2 shows the shape of ReLU (Equation 3.6), scaled exponential linear units (SELU, Equation 3.7), and leaky rectified linear units (LeakyReLU, Equation 3.8).

The designs of scaled exponential linear units (SELU, Equation 3.7) [135] and leaky rectified linear units (LeakyReLU, Equation 3.8) can fix the vanishing gradient problem.

A perceptron model, as stated in Equation 3.2, can be rewritten as an activation function as follows:

$$F(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3.9)$$

However, due to the piecewise nature of a perceptron function, it is not usually used as an activation function. In a modern sense, our “revised perceptron model” in Equation 3.3 is usually called a neuron. Like a perceptron model, neurons can be connected in the form of layers upon layers, forming a neural network.

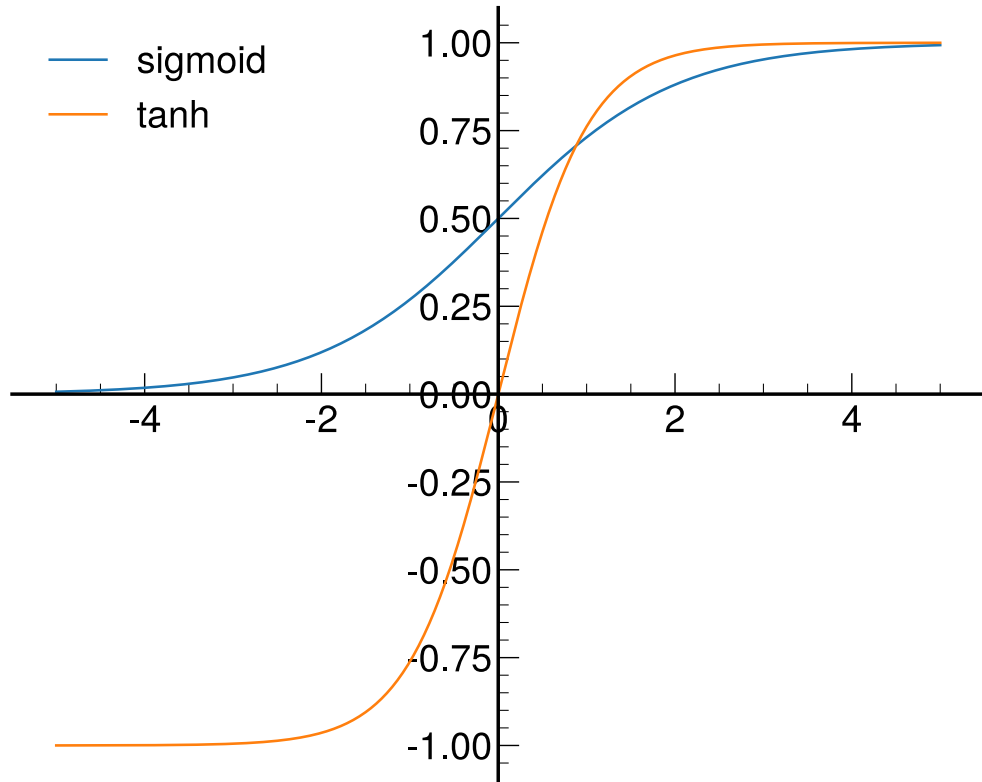


Figure 3.1: Sigmoid (Equation 3.4) and tanh (Equation 3.5) activation functions for $-5 < x < 5$.

3.3.3 Constructing neural networks from neurons

As explained in a previous section, we can construct several neurons in the form of layers upon layers. Since each neuron accepts a set of inputs and gives an output, we can combine a set of neurons, each containing its own set of weights and biases, into a layer of neurons as follows:

$$\begin{aligned}
 f_i(x_1, x_2, \dots, x_n) &= F(w_{i1}x_1 + w_{i2}x_2 + \dots + w_{in}x_n + b_i) \\
 &= F\left(\sum_j w_{ij}x_j + b_i\right)
 \end{aligned}$$

where i denotes the i th neuron in the layer. Since one neuron can output one number, an output from a set of m neurons can be arranged as an m -dimensional vector. Furthermore, given n inputs to the neuron layer, the input of neurons can be interpreted as a vector of n dimensions. Hence, the weights of all neurons in the layer

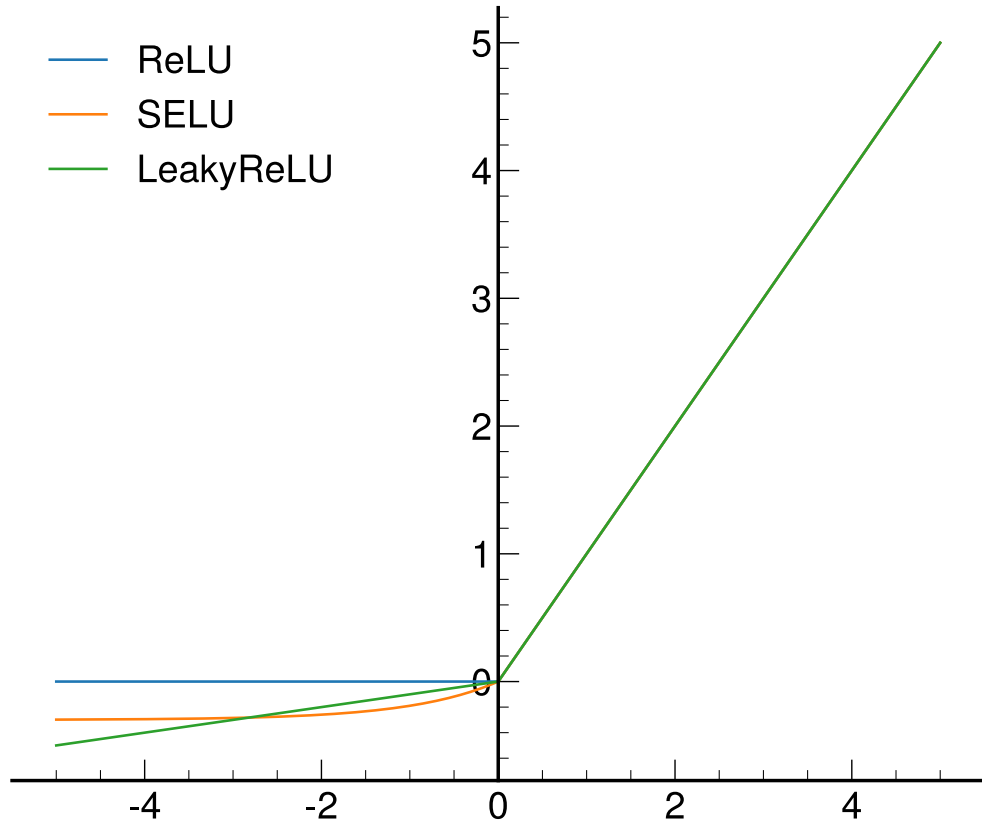


Figure 3.2: Rectified learning unit (ReLU, Equation 3.6), scaled exponential linear units (SELU, Equation 3.7), and leaky rectified linear units (LeakyReLU, Equation 3.8) activation functions for $-1 < x < 1$, with $\lambda = 1$, $\alpha = 0.3$ for SELU, and $\alpha = 0.1$ for LeakyReLU. Since $\lambda = 1$ for SELU, all three functions have the same output in the $x > 0$ region.

can be interpreted as a matrix of $m \times n$ dimensions, and the biases of all neurons can be interpreted as a vector of m dimensions as well. With this matrix-based interpretation, we can interpret it as follows:

$$f(\vec{x}) = F\left(W\vec{x} + \vec{b}\right). \quad (3.10)$$

With this construction of neuron layers, we can stack together several neuron layers into one neural network by passing the output of one neuron layer as the input of the next neuron layer. By stacking several neuron layers, the output from each

layer will be in the form of vectors of the same dimension as the number of neurons in that layer. The result of the neural network constructed this way will be a vector with its dimension equal to the number of neurons in the final neuron layer.

Based on this construction, neural networks can either output a scalar (with one neuron in the final layer) or a vector (with two or more neurons in the final layer). Each neuron accepts one feature and passes on to the next layer, and a final neuron layer that gives outputs is called an *output layer*, which may have one or more neurons depending on its design. All intermediate layers between an input layer and an output layer are called a *hidden layer*.

Since a layer of neurons can have more than one neuron at a time, an activation function that normalises its outputs can also be used, such as the softmax activation function,

$$\text{softmax}(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \quad (3.11)$$

where x_i represents the value in i th neuron per layer. This activation function guarantees that the sum of all outputs in the network will sum to one. It can be used to convert the output into the probability for an entry to be assigned to one class.

3.3.4 Another interpretation of neural networks via linear models

A linear model is a model based on simple matrix manipulation. Given an input in the form of n -dimensional vector $\vec{x}$, a linear model takes the form of [136]

$$\vec{y} = W\vec{x} + \vec{b}$$

where W is an $m \times n$ matrix and $\vec{b}$ is a vector of m dimensions. The output of this linear model, $\vec{y}$, is a vector of m dimensions. An easy example is one-dimensional

regression, where both $\vec{x}$ and $\vec{y}$ have one dimension. The equation above can be simply reduced into $y = mx + c$.

We can string together several linear models as follows,

$$\begin{aligned}\vec{y}_1 &= W_1\vec{x} + \vec{b}_1 \\ \vec{y}_2 &= W_2\vec{y}_1 + \vec{b}_2 \\ &\vdots \\ \vec{y}_n &= W_n\vec{y}_{n-1} + \vec{b}_n\end{aligned}$$

where $\vec{y}_i$ represents the output of i th linear model. With this construction, the output of a linear model is passed on as an input of the next linear model. We can rewrite $\vec{y}_2, \vec{y}_3, \dots, \vec{y}_n$ as

$$\begin{aligned}\vec{y}_2 &= W_2(W_1\vec{x} + \vec{b}_1) + \vec{b}_2 \\ \vec{y}_3 &= W_3(W_2(W_1\vec{x} + \vec{b}_1) + \vec{b}_2) + \vec{b}_3 \\ &\vdots \\ \vec{y}_n &= W_n(W_{n-1}(W_{n-2}(\dots W_2(W_1\vec{x} + \vec{b}_1) + \vec{b}_2) + \dots + \vec{b}_{n-2}) + \vec{b}_{n-1}) + \vec{b}_n\end{aligned}$$

By introducing two new matrices, W' and $\vec{b}'$, we can rewrite $\vec{y}_n$ as

$$\vec{y}_n = W'\vec{x} + \vec{b}'$$

where

$$\begin{aligned}W' &= W_n W_{n-1} \dots W_2 W_1 \\ \vec{b}' &= W_n W_{n-1} \dots W_2 \vec{b}_1 + W_n W_{n-1} \dots W_3 \vec{b}_2 + \dots + W_n \vec{b}_{n-1} + \vec{b}_n \\ &= \sum_{i=1}^n W_n W_{n-1} \dots W_{i+1} \vec{b}_i\end{aligned}$$

This means we can collapse a string of linear models into one linear model, and a string of linear models preserves linearity. This also means combining several

linear models becomes useless since we can always rewrite a combination of linear models into one.

To break this linearity, we can introduce a nonlinear function to be applied to an output of every linear model, thereby breaking the linearity between linear models, and our model cannot be collapsed into a single linear model [136]. This means the input of the next linear model is the output of that nonlinear function applied to the output of the previous linear model, not the output of the previous linear model directly.

The introduction of a nonlinear function to an output of a linear model is analogous to Equation 3.10. This construction of linear models with a nonlinear function added is therefore analogous to a neural network, which has been introduced earlier in this chapter.

3.4 How do neural networks learn?

Neural networks “learn” by adjusting their parameters, which are weights and biases, to better predict outputs that match the goal value. The procedure where the network adjusts itself to provide better predictions is called network training.

3.4.1 Training and validating dataset

As stated in Section 3.1, neural networks are designed to find a target value *based on its input data*. This means it must be trained with the given dataset in the first place. The network can be trained on the entire dataset given for the problem, but this risks overtraining, where the network tends to “memorise” the target answer given the input information, and not try to grasp the essential information appearing in the dataset. This causes the network to be unable to predict accurately when new input is given.

An analogy to overtraining in neural networks is fitting data points with a function containing too many parameters. Such functions would be able to perfectly fit all the data points available but may fail at predicting the value of new data points.

To avoid overtraining, the dataset is usually split into training and testing datasets. The training dataset is used to train the network and adjust weights and biases in the network, while the testing dataset is solely used to evaluate the performance of the network during the training. One may also split the dataset further into the third dataset called the validating dataset, which can be used to validate the network's performance after the training. During the training, several metrics can also be monitored using the training and testing datasets. If the metrics from the testing dataset show poorer performance than the same metrics from the training dataset, this means the network is overtrained.

3.4.2 Overtraining versus overfitting

Overtraining and overfitting are two different terms referring to the design and training of neural networks. According to Tetko, Livingstone, and Luik [137], *overfitting* refers to “exceeding some optimal ANN size”, while *overtraining* refers to “the time of ANN training that may finally result in the worse predictive ability of a network”, similar to the definition discussed in Section 3.4.1.

In 1995, Tetko, Livingstone, and Luik showed with several dataset examples [137] that *overfitting* does not have “any influence on ANN predictive ability”. This is shown via the indifference of the network performance when the number of neurons in a hidden layer increases.

The fact that *overfitting* does not affect the predicting power of neural networks is later affirmed by the work of Hornik, Stinchcombe, and White [138], which claimed that neural networks can approximate “any measurable function to any

desired degree of accuracy”, and thus neural networks are universal approximators which can be used to model any function [136]. Therefore, suspicion of *overfitting*, where the size of neural networks may be too large, is not a concern anymore. Sadly, *overtraining* remains a problem in network training that needs to be dealt with, which can be fixed in several ways, such as introducing dropouts and regularizers, and avoiding vanishing gradients during the training.

3.4.3 One epoch of network training

During network training, the data in both training and testing datasets are used repeatedly. An epoch of the network training means one loop of network training using an entire set of training and testing datasets, and the procedure is as follows:

1. Break down the entire dataset into smaller *batches* of entries. Entries that go into each batch are usually randomised in every epoch to make sure that the data is drawn randomly from the dataset.
2. For each batch:
 - a) Calculate the output of all entries in the batch.
 - b) Calculate the loss mean based on the output of the network (predicted value) against the target value corresponding to entries in the batch (truth value). See Section 3.4.4 for more details on loss functions. Also, calculate all necessary metrics based on predicted values against truth values.
 - c) Calculate the gradient of the loss function value (obtained in the previous step) against *every weight and bias value in the network*. This procedure is called backpropagation. See Section 3.4.6.
 - d) Update the value of weights and biases with optimizers, based on the gradient values obtained via backpropagation. See Section 3.4.7.

3. After all batches of data are used, calculate all necessary metrics again, this time using data from the testing dataset.

By separating the data into smaller *batches* during the training, the network gets updated more frequently instead of being updated once when the whole dataset is trained at once. If the batch size is large enough, the adjustments derived from the batch can be a *stochastic*¹ *approximation* for the whole dataset. This effect will be later demonstrated later.

3.4.4 Loss function

An output from the loss function can be used as a metric to determine how predictions of entries correspond to their true values. They are designed to be differentiable and give lower output as the prediction and truth values get closer.

There are several types of loss functions, which can be used depending on the task used to solve with neural networks. Here are some examples of loss functions commonly used in neural networks. In these loss functions, p represents the predicted value and q represents the truth value.

3.4.4.1 Mean squared error (MSE)

$$\text{MSE}(p, q) = (p - q)^2 \quad (3.12)$$

This loss function will output zero if both p and q are equal. This can be used in both classification problems (which usually have outputs in the range of 0 to 1) and regression problems (which may have arbitrary output values). However, the gradient of this loss function becomes smaller when p approaches q , leading to a vanishing gradient problem.

¹According to the Oxford English Dictionary, “stochastic” means “Randomly determined; that follows some random probability distribution or pattern, so that its behaviour may be analysed statistically but not predicted precisely” [139].

3.4.4.2 Mean absolute error (MAE)

$$\text{MAE}(p, q) = |p - q| \quad (3.13)$$

This loss function has a constant gradient, except at $p - q = 0$, which has no gradient, unlike the mean squared error loss function. As this loss function is differentiable at all other points of p , this is not a problem until p is exactly equal to q . One disadvantage of this loss function is that it has a constant gradient, which may cause the descent to the optimal point with the least loss to be slower than MSE.

3.4.4.3 Binary cross-entropy

$$\text{binary cross-entropy}(p, q) = -q \log(p) - (1 - q) \log(1 - p) \quad (3.14)$$

This loss function can only be used in classification problems since the function would not otherwise work when p and q have a value beyond the range of 0 to 1. Also, since it uses logarithms, many neural network libraries will have a cutoff for values of p and $1 - p$ to a certain value to ensure numerical stability.

3.4.4.4 Categorical cross-entropy

$$\text{categorical cross-entropy}(p_i, q_i) = - \sum_i q_i \log(p_i) \quad (3.15)$$

This loss function is designed for classification problems with multiclass output (more than one class of entries). It can also be reduced to a binary cross-entropy loss function when there are only two classes of entries.

3.4.5 Regularizers

So far in our setup of neural networks, there are no restrictions as to how large weights and biases in all neurons can be. This may result in networks having large weights or biases, making the network susceptible to small changes in output, which may be caused by statistical fluctuations of the input variables. Ultimately, it causes the network to overtrain.

To avoid the complexity caused by large weights and biases, extra terms to the loss function can be added, called *regularizers*. Commonly used regularizers are called L1 and L2 regularizers, defined as:

$$\text{L1 regularizer} = \ell_1 \sum_i |w_i| \quad (3.16)$$

$$\text{L2 regularizer} = \ell_2 \sum_i w_i^2. \quad (3.17)$$

As seen from both formulations, this extra term to the loss function will penalize the network when the sum of weights (or biases) gets too large. Parameters ℓ_1 and ℓ_2 can be adjusted. By adding regularizers, additional constraints are added for model fitting, therefore avoiding model overfitting. L1 and L2 regularizers utilise the L^p -norm of a vector, which can be written as [140]

$$|\vec{x}|_p = \left(\sum_i |x_i|^p \right)^{1/p}. \quad (3.18)$$

3.4.6 Backpropagation

As described in Section 3.3.3, each layer of the neural network receives input and gives output towards the next layer. Each layer has weights and biases that are used to calculate the output, and during the training, they must be adjusted so that the loss of the network decreases over each training epoch. To do this, the derivative of loss per each weight or bias element must be calculated.

Suppose that the network has n layers. The total derivative of the loss function L per weight elements w_i and bias elements b_i is

$$\delta L = \sum_i \frac{\partial L}{\partial w_i} \delta w_i + \sum_i \frac{\partial L}{\partial b_i} \delta b_i. \quad (3.19)$$

This can be written as a dot product between a *gradient* of L per all parameters and a vector containing all infinitesimal amounts of w_i and b_i

$$\delta L = \nabla L \cdot \delta \vec{\theta}_i \quad (3.20)$$

where θ_i collectively represents both weights and biases (or parameters) of the network.

To calculate the derivative of loss function L per weight element w_i belonging to layer i , the chain rule must be applied. This can be roughly written as

$$\frac{\partial L}{\partial w_i} = \frac{\partial L(\vec{y}_n)}{\partial w_i} \quad (3.21)$$

$$= \frac{\partial L(\vec{y}_n)}{\partial \vec{y}_n} \frac{\partial \vec{y}_n}{\partial \vec{y}_{n-1}} \frac{\partial \vec{y}_{n-1}}{\partial \vec{y}_{n-2}} \dots \frac{\partial \vec{y}_{i+1}}{\partial \vec{y}_i} \frac{\partial \vec{y}_i}{\partial W_i} \quad (3.22)$$

where $\vec{y}_n$ is an output of the n th layer. The derivative per bias element b_i can also be calculated in the same way.

Modern software libraries such as TensorFlow [141], which the Keras neural network library [142] uses as a backend, can automatically perform gradient calculations for all outputs in each hidden layer of the network. Once the loss is calculated per each entry, the network will have the information for the derivative of loss per *all* weights and biases in the network, which sometimes are called the *gradient of the loss function*.

3.4.7 Optimization with optimizers

After the gradient has been calculated, all weights and biases in the network should be adjusted to decrease the network loss. This can be done via *optimizers*. However, not all optimizers have the same adjustment mechanism. This section discusses a few optimizers that are most commonly used.

Note that the implementation for optimizers presented here may differ between each documentation or neural network library. The examples shown here are derived from the Keras neural network library [142].

3.4.7.1 Gradient descent and Stochastic gradient descent (SGD) optimizer

Gradient descent [143] is the simplest optimizer to understand and implement. The analogy for this optimizer is a ball rolling down a hill. Let $U(x)$ be the height of the hill at point x where the ball is located. The ball rolls down a hill in the direction of the gradient of the height U , which can be written as

$$F = -\nabla U. \quad (3.23)$$

Gradient descent optimizer employs the very same idea. After the gradient of the loss is calculated in each batch, each parameter θ (including weights and biases) is adjusted as

$$\theta \rightarrow \theta' = \theta - \eta \nabla_{\theta} L \quad (3.24)$$

where η is the *learning rate* of the optimizer, and $\nabla_{\theta} L$ is the gradient of loss per the parameter θ . This parameter controls the amount of change in the network's parameters. If the learning rate is too large, the network may skip over the minimum within the parameter space. If the learning rate is too small, converging to the minimum would require a lot of iterations.

One problem with gradient descent is that it requires the network to calculate loss and gradients from all training samples. If the network is trained with a large dataset, the training will be slowed down since the parameter update is done only once per epoch. Another problem is that it may result in parameters providing local minima of the loss instead of global minima. To solve the latter problem, it has been shown that a high learning rate η can allow gradient descent optimizers to escape local minima into global minima [144].

To speed up the process, stochastic gradient descent (SGD) is used [143] As its name suggests, instead of calculating the gradient from all inputs, SGD calculates only gradients from a randomised sample of training events from the dataset, called

mini-batches, and update the network parameters based on the calculated gradient. If the mini-batch size is large enough, it can be used as a good approximation for the gradient from the dataset.

3.4.7.2 Momentum SGD

Momentum SGD [145; 146] is an extension of SGD, adding a concept of momentum. Instead of adjusting the parameter θ directly with gradients, an extra term called momentum m is introduced. The updating mechanism is now as follows:

$$m_k \rightarrow m_{k+1} = \beta m_k - \eta \nabla_{\theta} L \quad (3.25)$$

$$\theta_k \rightarrow \theta_{k+1} = \theta_k + m_k \quad (3.26)$$

This type of optimizer introduces another adjustable parameter β that controls the momentum of the network parameter's progression. If $\beta = 0$, this optimizer reduces to a simple SGD.

3.4.7.3 RMSprop

RMSprop [147] is another class of optimizers extending SGD, where the learning rate can be automatically adjusted based on the size of the gradient. Instead of using a fixed learning rate, this optimizer has its updating mechanism as follows:

$$v_k \rightarrow v_{k+1} = \rho v_k + (1 - \rho) [\nabla_{\theta} L]^2 \quad (3.27)$$

$$\theta_k \rightarrow \theta_{k+1} = \theta_k - \frac{\alpha}{\sqrt{v_k + \epsilon}} \nabla_{\theta} L \quad (3.28)$$

where α and ρ are adjustable parameters for this optimizer. As seen from the equations above, the learning rate α is divided by the square root of velocity v_k . This means if the gradient $\nabla_{\theta} L$ becomes smaller, the learning rate for θ can be increased. The cutoff value ϵ is introduced in the denominator for numerical stability, which can be set at a very small value such as 1×10^{-7} .

3.4.7.4 Adam

Adam² [148] is yet another class of optimizers extending SGD, this time employing both the ideas of momentum from Momentum SGD optimizers and adjustable learning rates from RMSprop. With both ideas extending SGD, the updating mechanism is now as follows:

$$m_k \rightarrow m_{k+1} = \beta_1 m_k + (1 - \beta_1) \nabla_{\theta} L \quad (3.29)$$

$$v_k \rightarrow v_{k+1} = \beta_2 v_k + (1 - \beta_2) [\nabla_{\theta} L]^2 \quad (3.30)$$

$$\theta_k \rightarrow \theta_{k+1} = \theta_k - \frac{\alpha m_k}{\sqrt{v_k} + \epsilon} \quad (3.31)$$

where α , β_1 , and β_2 are adjustable parameters. The idea from Momentum SGD is present in the form of m_k , and the idea of an adjustable learning rate from RMSprop is present in v_k and an extra denominator. As with RMSprop, ϵ is added to ensure numerical stability during the calculation.

3.4.8 Vanishing gradient problem

During the training of neural networks, gradients of the loss per each network parameter must be calculated by the use of the chain rule, as explained in Section 3.4.6. However, if one term of the gradient (such as $\partial \vec{y}_n / \partial \vec{y}_{n-1}$) gets closer to zero, the whole gradient per the parameter *vanishes* to zero. This can be caused by an activation function containing regions with near-zero gradients, such as tanh, sigmoid, or softmax.

To avoid this problem, we can do the following:

- Use activation functions that do not have regions with near-zero gradients, such as SELU and LeakyReLU.

²Adam is not an acronym. It is introduced with this name in Ref. [148].

- Add batch normalisation to inputs. Batch normalisation is an extra hidden layer that will learn the mean and variance of the input data, and normalise the input accordingly so that the output has zero mean and standard deviation close to 1. This removes the possibility that a large region of the input causes the gradient to be near zero.

3.4.9 Dropout layers

Dropout layers are hidden layers that temporarily turn off each input node at random during training with a predetermined probability. The probability of turning off (or dropping out) input nodes is called the *dropout rate*.

Dropout layers can be used to avoid overfitting by reducing the connections available in the network at random. With dropout, we effectively create weaker networks with fewer connections, so, during the training of the network with this type of layer, it is equivalent to training a large number of weaker networks together. During network testing and usage, however, dropout is deactivated, allowing all input nodes to be activated, equivalent to combining several weak networks [149].

3.5 How do we measure the performance of neural networks?

After the training of neural networks, it is important to measure their performance in predicting the class or the regression value of the input data. Several metrics can be used to evaluate its performance, such as accuracy and area under the receiver operating characteristic (ROC) curve.

3.5.1 True positives, true negatives, false positives, and false negatives

In binary classification problems, a model can predict outputs which may or may not be correct when compared to the true values of the input data. Four quantities can be used to describe the number of entries based on the output's prediction and truth values. The summary of descriptions is shown in Table 3.1.

- True positive (TP) represents the number of entries with *positive* prediction and *positive* truth value. For example, a pregnant female patient is diagnosed as pregnant.
- True negative (TN) represents the number of entries with *negative* prediction and *negative* truth value. For example, a male patient is diagnosed as not pregnant.
- False positive (FP) represents the number of entries with *positive* prediction and *negative* truth value. In other words, it represents the number of *false predictions stating that the input entry represents the positive class*. For example, a male patient who cannot get pregnant biologically is diagnosed as pregnant.
- False negative (FN) represents the number of entries with *negative* prediction and *positive* truth value. In other words, it represents the number of *false predictions stating that the input entry represents the negative class*. For example, a pregnant female patient is diagnosed as not pregnant.

Table 3.1: True positives, true negatives, false positives, and false negatives.

		Truth value	
		Positive	Negative
Prediction	Positive	TP	FP
	Negative	FN	TN

3.5.2 Accuracy

This is the simplest metric to calculate the performance of a neural network designed for classification tasks. Accuracy is defined by:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3.32)$$

In other words, accuracy measures the *accuracy* of the model to predict the output class correctly.

This metric, however, has its shortcomings due to the definition of the output classification. Since the predicted class must be binary, but the output directly from the model can be of any value, usually ranging from 0 to 1, a cutoff must be decided. Usually, the cutoff for this metric is 0.5, which means all entries with output larger than 0.5 will be classified into class 1 (or positive class), while entries with output smaller than 0.5 will be classified into class 0 (or negative class). If the cutoff value changes from 0.5 to another value, such as 0.75, the value of accuracy changes.

Another shortcoming of this metric is that the denominator is the number of entries included. However, if the dataset used to determine the accuracy does not have an equal number of entries in positive and negative classes, the value may be influenced by the class with a larger number of entries.

3.5.3 Receiver operating characteristic (ROC) curve and area under ROC curve (AUC)

The receiver operating characteristic curve, commonly known as the ROC curve, is a curve used to determine how well the model accepts *signal* (or positive) entries as signal entries compared to how well the model accepts *background* (or negative) entries as signal entries. To calculate the curve, *signal acceptance* and *background acceptance* are calculated *at different values of cutoff* as follows:

$$\text{signal acceptance} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3.33)$$

$$\text{background acceptance} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (3.34)$$

Note that, if the cutoff value is low enough to allow all entries to be predicted as the signal, both FN and TN will be zero since no events are predicted as background, giving both signal acceptance and background acceptance to be exactly one. On the other hand, if the cutoff value is high enough to deny all entries to be predicted as background, both TP and FP will be zero instead, leading to both signal and background acceptance being exactly zero.

This type of metric can be used to calculate the performance of the model, independent of the class imbalance of the dataset used and the cutoff values used to determine the classification, since both signal acceptance and background acceptance are calculated using only signal and background entries respectively, and several cutoff values are used to determine different values of both acceptance metrics.

The ideal shape of this type of curve should show that the model has a perfect signal acceptance without any background acceptance. This means that the model can perfectly classify entries without confusion. On the other hand, the worst possible shape of this curve is the curve determined by random guesses. By random

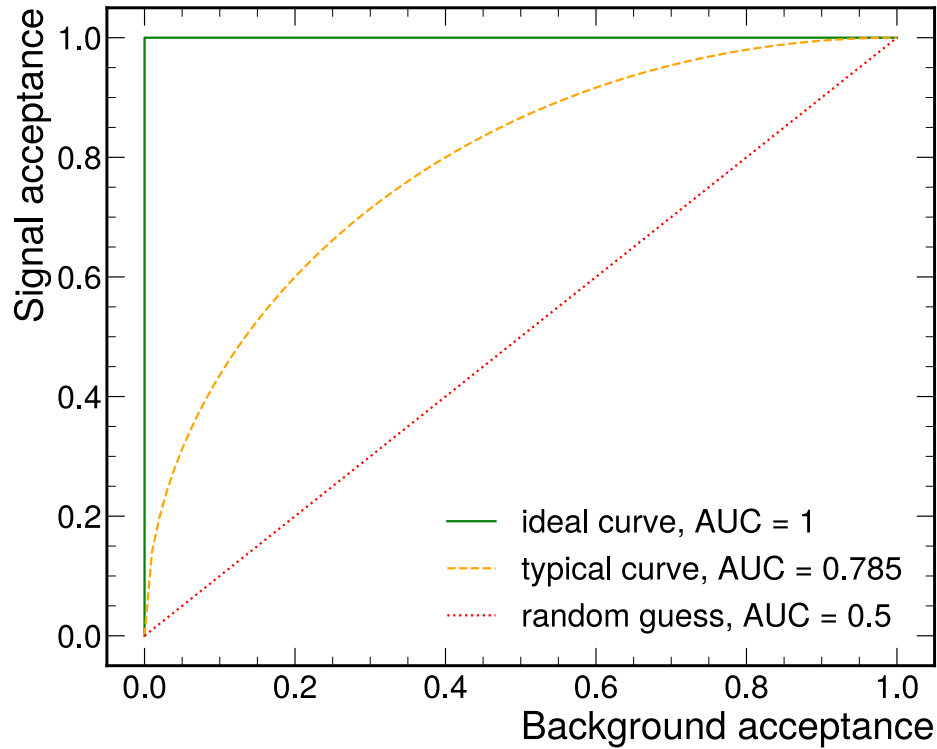


Figure 3.3: Examples of receiver operating characteristic (ROC) curve. The green curve represents the ideal curve with perfect signal acceptance at zero background acceptance. The orange curve represents the typical shape of the curve for a model, and the red curve represents the curve obtained by random guessing.

guessing classes of entries, the probability of an input entry being classified as signal or background is equal, rendering an equal amount of signal and background acceptance at every value of cutoff. Example ROC curves are shown in Figure 3.3.

The area under the ROC curve, commonly abbreviated to AUC, can be used to determine the performance of the model based on the ROC curve. If the ROC curve has an ideal shape, the AUC is 1. On the other hand, if the ROC curve resembles random guesses, the AUC will become 0.5. Since ROC curves can be determined independent of class imbalance and cutoff values, AUC is immune to both properties as well.

Chapter 4

SEARCH FOR FOUR TOP QUARK PRODUCTION IN RUN 2 DATA OF CMS IN INDIVIDUAL FINAL STATES

Numerous searches for four-top quark production have been conducted by ATLAS and CMS experiments since 2012, especially searches by CMS experiment which are performed with data collected from the Run 2 period, spanning from 2016-2018. The results cover four final states, which are the single-lepton (SL), the opposite-sign dilepton (OSDL), the same-sign dilepton and multilepton (SSDL&ML), and for the first time in CMS Collaboration, the all-hadronic final state. It contains both existing search results from the OSDL analysis in 2016 and the SSDL&ML from Run 2, as well as brand new analyses performed on SL and all-hadronic final states using the Run 2 dataset and OSDL final state using the 2017-2018 dataset. More information will be discussed in Section 4.1.

Details regarding analysis from each final state shall be explained in this chapter, while the combination between all final states shall be discussed in Chapter 5. As stated in Section 1.8.1, two new search results from ATLAS and CMS in same-sign dilepton and multilepton final state have achieved the observational level with the observed significance of more than 5 standard deviations. However, this thesis will focus on the search results based on single-lepton, opposite-sign dilepton, and all-hadronic final states. Therefore, these two new results will not be discussed in this chapter.

Table 4.1: Summary of analyses involved in the latest search for four-top quark production.

Analysis	Ref.	Analysis strategy
2016 SL+OSDL	[1]	BDT
Run 2 SSDL&ML	[65]	BDT
Run 2 SL	Section 4.3	BDT
2017+2018 OSDL	Section 4.4	HT
Run 2 all-hadronic	Section 4.5	BDT with extABCD [150] and ABCDnn [151] background estimation methods

4.1 Overview of final states and analysis strategy

The latest search for four-top quark production [2] involves four final states spanning data collected over three years in the Run 2 campaign, equivalent to an integrated luminosity of 138 fb^{-1} . Table 4.1 summarises the analyses involved in this latest search.

This search for four-top quark production can be broken down into three main parts:

- Run 2 SL analysis (Section 4.3) - This analysis implements a Boosted Decision Tree (BDT) to discriminate between four-top quark production and $t\bar{t}$ production. The discriminator is then used to construct event distributions, separated into 60 categories based on year, lepton flavour, jet multiplicity, and b-tagged jet multiplicity, for maximum search significance.
- 2017 + 2018 OSDL analysis (Section 4.4) - This analysis constructs a distribution of events based on the sum of jet transverse momentum H_T as a discriminating variable, separated by jet multiplicity, b-tagged jet multiplicity, lepton flavour, and data-taking year, for maximum search

significance. It has several improvements in Monte Carlo simulation and b-tagging algorithms compared to OSDL analysis performed on 2016 data.

- Run 2 all-hadronic analysis (Section 4.5) - For the first time in CMS Collaboration, this analysis performs a search on the final state with no leptons. It employs BDT discriminators to discriminate between $t\bar{t}t\bar{t}$ against simulated $t\bar{t}$ and QCD multijet background events. To alleviate poor Monte-Carlo simulation quality of QCD multijet and $t\bar{t}$ production, two methods inspired by the ABCD method are used to estimate the background yield and shape distribution. The search is then performed on 12 event categories based on the number of resolved tops, the number of boosted tops, and the sum of jet transverse momentum H_T .

In addition to the new analyses described above, two existing analysis results are also included:

- 2016 OSDL analysis (Section 4.2.1), part of the search for four-top quark production performed on 2016 data using SL and OSDL final states [1]. This analysis employs two BDTs to detect a set of three jets hadronically decayed from a top quark and to discriminate between $t\bar{t}t\bar{t}$ and $t\bar{t}$ production. The second BDT is then used to populate the distribution of events separated by the number of jets, the number of b-tagged jets, and the lepton flavour combination.
- Run 2 SSDL&ML analysis (Section 4.2.2), which has the highest search significance for particle production [65]. This analysis employs a BDT discriminator to separate between $t\bar{t}t\bar{t}$ events from all other SM backgrounds, and separate events into 17 signal regions, with a special control region to constrain $t\bar{t}Z$ normalisation.

The techniques used in these analyses mentioned above are vastly different from searches performed by the ATLAS Collaboration. In CMS analyses, events are categorised with a large number of event categories based on the number of physics objects, such as the number of jets or the number of b-tagged jets, as signal and background processes have different characteristics that result in different numbers of jets and b-tagged jets. ATLAS analyses, on the other hand, employ a small number of event categories, aiming to separate between main backgrounds as much as possible [74; 75; 80; 81; 82]. Nonetheless, these two different techniques tend to deliver relatively similar performance, which provides additional confidence when both techniques provide the same conclusion.

For each final state, analysis techniques, background processes, and systematic uncertainty treatments are performed differently, which will be discussed in subsequent sections. Search results from all final states can also be combined to obtain more search sensitivity. The details of the combination of Run 2 SL analysis, 2017+2018 OSDL analysis, Run 2 all-hadronic analysis, 2016 OSDL analysis, and Run 2 SSDL&ML analysis will be discussed in Chapter 5.

4.2 Existing search results from CMS Experiment

This section will discuss two existing search results for four-top quark production from CMS Experiment before the results presented in this thesis. They are **2016 single-lepton and opposite-sign dilepton analyses** [1] (Section 4.2.1) and **Run 2 same-sign dilepton and multilepton (SSDL&ML) analysis** [65] (Section 4.2.2).

4.2.1 2016 single-lepton and opposite-sign dilepton

CMS Collaboration has performed the search of four-top quark production using 2016 data on single-lepton and opposite-sign dilepton final states [1]. In this section, analyses in both final states will be discussed, but only the result from opposite-sign dilepton analysis will be used in the combination. The 2016 single-lepton analysis result in this section is superseded by Run 2 single-lepton analysis, which will be discussed in Section 4.3.

4.2.1.1 Main background estimation

The main background process for both final states is $t\bar{t}$ with additional jets, which is simulated at the NLO level generator. However, the cross section is normalised to the predicted value at the NNLO level [85]. Furthermore, the predictions at the NNLO level suggest that the distribution of transverse momentum p_T of top quarks from data may differ from the predictions provided at the NLO level. To solve this discrepancy, a scale factor derived from Ref. [152; 153] is applied to $t\bar{t}$ events.

Due to the low statistics of events with a high number of jets, dedicated samples with a generator-level requirement of a high number of jets are generated to provide better statistics in the region. The use of dedicated samples with the same generator-level requirement is also used in the newer Run 2 single-lepton analysis (see Section 4.3).

However, the modelling of $t\bar{t}$ with additional jets background does not match the data in regions with high jet multiplicity. To solve this, several scale factors are calculated from fits to the data from the single-lepton final state in regions with 8 and 9 AK4 jets with 2 and 3 b-tagged jets. These scale factors are later propagated to signal-rich regions in the single-lepton final state and to the opposite-sign dilepton

final state (where the difference in jet multiplicity between each final state is taken into account).

4.2.1.2 Baseline selection

Baseline requirements for events to be included in both single-lepton and opposite-sign dilepton final states are as follows:

- Trigger requirements
 - Single-lepton events require single-lepton triggers, requiring at least one isolated muon with $p_T > 24$ GeV and $|\eta| < 2.4$, or one isolated electron with $p_T > 32$ GeV and $|\eta| < 2.1$ [1].
 - Opposite-sign dilepton events require single-lepton and dilepton triggers. Dilepton triggers used require either two muons with p_T greater than 17 and 8 GeV, two electrons with p_T greater than 23 and 12 GeV, or a muon and an electron with p_T greater than 23 and 8 GeV (regardless of lepton flavour) [1]. Even though this final state looks for two opposite-sign leptons in an event, including single-lepton triggers can increase the acceptance for this final state by roughly 20% [154]. However, the requirements for this analysis are designed in coordination between single-lepton and opposite-sign dilepton final states to prevent double counting.
- For single-lepton events, exactly one isolated muon with $p_T > 26$ GeV or one isolated electron with $p_T > 35$ GeV within the range of $|\eta| < 2.1$. For dilepton events, 2 leptons with opposite electric charge, each with $p_T > 25$ and 20 GeV respectively, both within $|\eta| < 2.4$.

- The invariant mass of both leptons in dilepton events must be $m_{\text{OSDL}} > 20 \text{ GeV}$ and $|m_{\text{OSDL}} - m_Z| > 15 \text{ GeV}$. This requirement is added to remove events containing two opposite-sign leptons from Z boson decays.
- Requiring different numbers of jets per different lepton final states.
 - 7 or more jets for single-muon events, and 8 or more jets for single-electron events, each with $p_T > 30 \text{ GeV}$ and $|\eta| < 2.5$.
 - 4 or more jets for dilepton events, with $|\eta| < 2.4$ and different p_T requirements based on b tagging (see below).
- Among the selected jets in an event, 2 or more jets must be classified as b-tagged using the Combined Secondary Vertex version 2 (CSVv2) algorithm with a medium working point [155]. The medium working point of this algorithm offers the b-tagging efficiency of 68% with a mistagging rate of 1%. For dilepton events, jets that are classified as b-tagged jets may have $p_T > 25 \text{ GeV}$, while jets that are *not* classified as b-tagged jets must have $p_T > 30 \text{ GeV}$. The threshold for p_T of b-tagged jets in dilepton events is lowered to increase acceptance for events with more b-tagged jets.
- Sum of jet transverse momentum $H_T > 500 \text{ GeV}$. This requirement is imposed to suppress $t\bar{t}$ background while retaining the acceptance of signal events.
- Missing transverse energy $p_T^{\text{miss}} > 50 \text{ GeV}$ for single-lepton events only. This requirement suppresses a small residual QCD background.

4.2.1.3 Analysis technique used

The analysis is performed using two sets of Boosted Decision Trees (BDTs) [156] trained with the AdaBoost algorithm [157], implemented in the TMVA package [158]. The first set of BDT is designed to determine how likely a combination of three jets (trijet) originated from a hadronic top quark decay. The

BDT is trained with simulation data from $t\bar{t}$ production with additional jets ($t\bar{t}$ +jets) to discriminate between a trijet combination decayed hadronically from a top quark versus any other trijet combination. It is then used to calculate the most likely trijet combination consistent with a top quark from all possible combinations of three jets. The discriminant from the most probable combination of three jets is used as an input to the second BDT.

The second set of BDT classifies events into signal and background. Two BDTs are trained to output discriminants for single-lepton and dilepton events separately. Both BDTs use input variables based on hadronic activity in each event, such as the number of jets, the scalar sum of p_T , etc. Event topology variables, such as event sphericity¹ [73] and hadronic centrality², are also used as inputs to these BDTs. The output from the corresponding BDT classifier is then used to populate histogram templates shown in Figures 4.1 and 4.2 for single-lepton and opposite-sign dilepton final states respectively.

4.2.1.4 Systematic uncertainties

In this analysis, systematic uncertainties are included as follows:

- Theoretical uncertainties
 - **Heavy-flavour reweighting uncertainty** (See Section 5.4.1.1)
 - **Normalisation uncertainties** from rare processes of $t\bar{t}$ production with one or two Z, W, or Higgs bosons, and three-top production (See Section 5.4.1.2)

¹Given a matrix $S_{ij} = \sum_k p_{k,i} p_{k,j} / \sum_k p_k^2$, where i and j are Cartesian coordinates x , y , and z , and k is an index running over all objects included in a calculation. Sphericity is calculated as $3(\lambda_2 + \lambda_3)/2$ where λ_2 and λ_3 are the two smallest eigenvalues of S_{ij} .

² H_T divided by the sum of all jet energies. If all jets are located on a transverse plane, or $\eta = 0$ for all jets, this quantity will be 1. If all jets are located along the beamline, or $\eta = \infty$, this quantity will be 0.

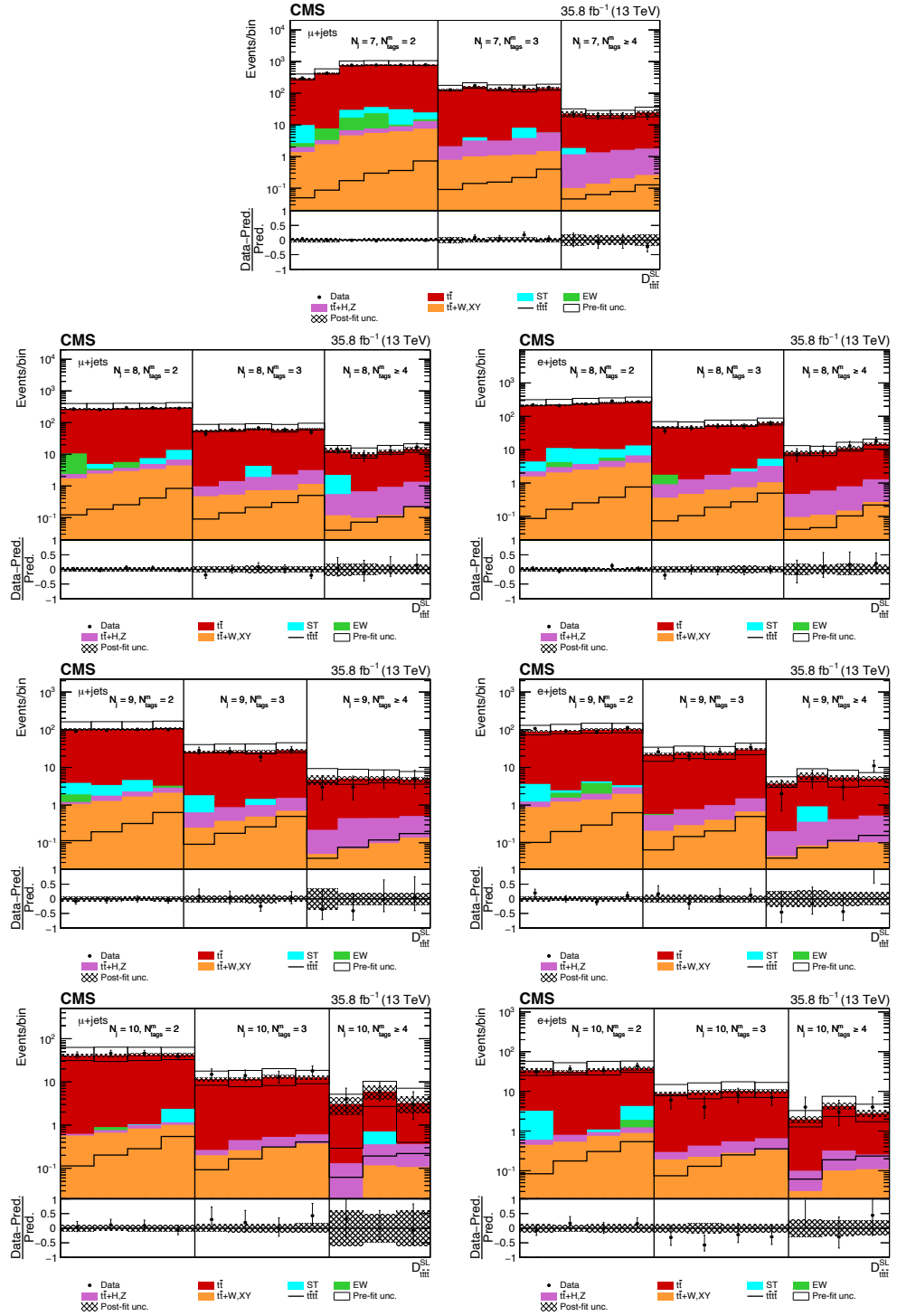


Figure 4.1: BDT output distributions from 2016 single-lepton analysis, separated by the number of jets and the number of b-tagged jets [1].

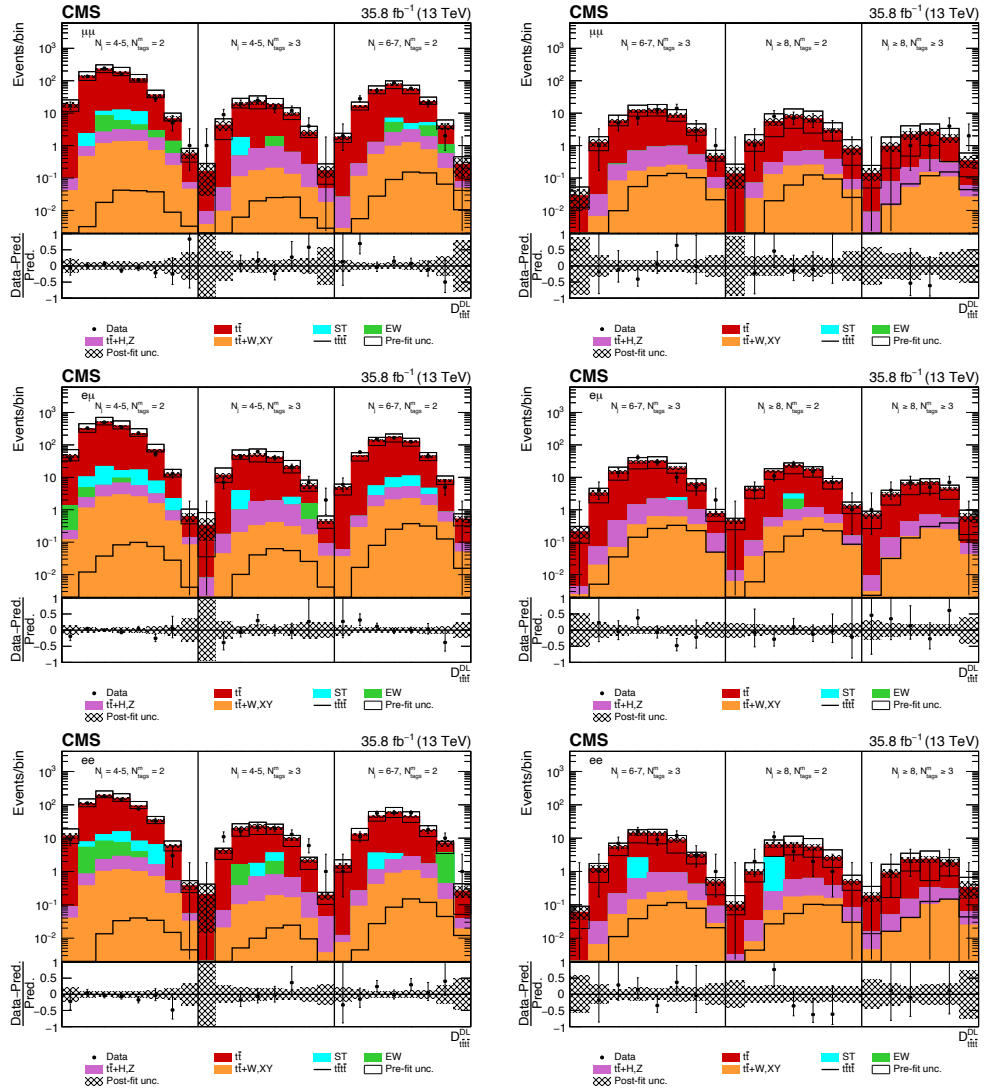


Figure 4.2: BDT output distributions from three flavour combinations (from top to bottom, $\mu\mu$, $e\mu$, and ee) in 2016 opposite-sign dilepton analysis, separated by the number of jets and the number of b-tagged jets [1].

- **Parton shower scales uncertainty**, also known as initial and final state radiation uncertainty (See Section 5.4.1.3)
- **Parton distribution function (PDF) uncertainty** (See Section 5.4.1.4)
- **Uncertainties from renormalisation and factorisation scale variations μ_R and μ_F** (See Section 5.4.1.5)
- **Matrix element to parton shower matching (ME-PS matching) uncertainty** estimated by varying a parameter that controls ME-PS matching by ± 1 standard deviation of the reported value in Ref. [159].
- **Underlying event (UE) uncertainty**, evaluated by varying parameters related to CUETP8M2T4 tuning function [159] used in generating MC simulations of the $t\bar{t}$ process.
- **Jet multiplicity correction uncertainty** due to a scale factor introduced to solve the imperfect modelling of MC simulations at high jet multiplicity regions.
- **Top quark p_T reweighting uncertainty** due to corrections made to solve discrepancy in top quark p_T distribution, as stated in Section 4.2.1.1
- Experimental uncertainties
 - **Luminosity** uncertainty of 2.5% for 2016 data (See Section 5.4.2.1)
 - **Pileup** (See Section 5.4.2.3)
- Reconstruction uncertainties
 - **Jet energy corrections (JEC)**, also known as Jet energy scale (JES), and **jet energy resolution (JER)** uncertainties (See Sections 5.4.3.1 and 5.4.3.2)
 - **b-tagging** uncertainties from Combined Secondary Vertex version 2 (CSVv2) b tagger (See Section 5.4.3.3)

- **Lepton reconstruction and identification**, known as the lepton scale factor in more recent analyses (See Section 5.4.3.4)

4.2.1.5 Results and shortcomings of this analysis

Analysis results from both single-lepton and opposite-sign dilepton final states indicate that the best fit value of signal strength³ are $1.6^{+4.6}_{-1.6}$ for single-lepton final state and $0.0^{+2.7}$ for opposite-sign dilepton final state. These results translate to 15^{+42}_{-15} fb and 0^{+25} fb of the measured cross section respectively. The reason why there is no negative uncertainty reported in the opposite-sign dilepton result is that the fit imposes a constraint where the signal strength must be positive. This assumption may cause difficulty when comparing with other results where it is not imposed, which is the signal strength may be allowed to have a negative value. Cross section, on the other hand, must always be positive since it represents the probability of a process to occur in the particle detector. By combining results from both final states, the best fit values of signal strength and cross section are $0.0^{+2.2}$ and 0^{+20} fb respectively.

The expected search significance⁴ obtained from single-lepton and opposite-sign final state are 0.21 and 0.36 respectively. However, the observed search significance becomes 0.36 and 0.0 for SL and OSDL final state respectively.

Due to the positivity constraint imposed on this analysis, a recalculation of signal strength and cross section best fit values have been performed to be compatible with other analyses during the combination, where both quantities are allowed to have negative values, along with updates to the four-top quark production cross section from 9.2 fb to 12 fb and luminosity uncertainties schemes (see

³Signal strength μ indicates how many signal events are observed compared to SM prediction. If SM is correct, $\mu = 1$.

⁴See Section 5.2.2 for the definition of search significance, and Section 5.2.3 for the difference between expected and observed search significance

Section 5.4.2.1). The recalculated best fit values for signal strength μ and cross section for opposite-sign dilepton final state are $-0.2_{-1.5}^{+1.7}$ (stat.) ± 1.5 (syst.) and -2_{-18}^{+20} (stat.) ± 18 (syst.) fb respectively [2]. The expected and observed search significance, after scaling the four-top quark production cross section, are 0.45 and 0.0 respectively.

Another important point of this analysis is that the four-top quark production cross section assumed is merely 9.2 fb [160; 119], which is calculated at NLO with QCD processes only, or in other words, *without electroweak corrections* (later added in Ref. [64]). This decision was made for the results to be directly comparable with search results from ATLAS [1]. However, to make the analysis compatible with newer results from other final states in the combination stage, the cross section of four-top quark production has to be modified at the analysis level from 9.2 fb to 12 fb [64]. As seen from the recalculated result, the measured value is consistent with the previous result, since the best fit value of the previous result is within the uncertainty of the recalculated result.

4.2.2 Run 2 same-sign dilepton and multilepton

This analysis [65] is performed on Run 2 data in the same-sign dilepton final state and multilepton final state. More precisely, this analysis requires events with two leptons with the same electric charge and events with at least three leptons. Since the branching ratio of a W boson (which can be decayed from a top quark along with a bottom quark) into a lepton and a neutrino is approximately 1/3 compared to 2/3 for a pair of quarks [17], this signature of four-top quark production has a branching ratio of only 12% [65]. However, since the background processes in this final state, mainly $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}H$, and $t\bar{t}$ with nonprompt leptons, have relatively low yields compared to other final states, this final state offers the cleanest search results of four-top quark production.

CMS Collaboration performed the search on same-sign dilepton and multilepton final state using data recorded in 2016, with an integrated luminosity of 35.9 fb^{-1} [79]. With this analysis, CMS obtained an observed search significance of 1.6, corresponding to a probability of 1 in 18 where the results obtained are statistical fluctuation. Since this analysis uses the entire Run 2 dataset, corresponding to an integrated luminosity of 137 fb^{-1} , the search significance is expected to increase due to the increased number of events and fewer statistical uncertainties.

4.2.2.1 Main background estimation

In the SSDL&ML final state, the main backgrounds are top-antitop quark ($t\bar{t}$) production associated with W, Z, and Higgs boson ($t\bar{t}W$, $t\bar{t}Z$, and $t\bar{t}H$), as well as $t\bar{t}$ with nonprompt leptons. Nonprompt leptons refer to leptons produced in hadron decays or conversions of photons in jets, as well as hadrons misidentified as leptons. This type of lepton is defined to distinguish from *prompt leptons* defined as leptons produced from W and Z boson decays. Distinguishing between nonprompt and prompt leptons is important since events with prompt leptons can later be classified into a dedicated control region to constrain the normalisations of $t\bar{t}W$ and $t\bar{t}Z$ processes [65].

Background estimation for these processes is different as follows:

- $t\bar{t}W$, $t\bar{t}Z$, **and** $t\bar{t}H$ samples are generated with Monte Carlo generators at NLO with additional jet flavour correction. $t\bar{t}W$ and $t\bar{t}Z$ samples also require initial state radiation and final state radiation (ISR/FSR) corrections, while $t\bar{t}H$ samples do not have this correction since they already have additional jets from the decay of $H \rightarrow WW$, and thus have more expected number of light jets compared to $t\bar{t}W$ and $t\bar{t}Z$.

- **Events with nonprompt leptons** are estimated using the “tight-to-loose” ratio method, previously used in a CMS search for new physics in this final state [72], which is shown to give consistent nonprompt lepton background prediction compared to another estimation method used in ATLAS search [161]. In this method, the ratio of leptons passing both *loose* and *tight* requirements against leptons passing *loose* requirement is measured for lepton flavour, transverse momentum p_T and pseudorapidity $|\eta|$. Backgrounds of this type are estimated based on this ratio.
- **Events containing charge-misidentified leptons** are estimated using charge misidentification probability measured in simulation. The probability for charge-misidentified electrons is approximately 10^{-5} to 10^{-3} , while the same probability for muons is lower by an order of magnitude [65]. Hence, backgrounds from charge-misidentified muons are negligible, while backgrounds from charge-misidentified electrons are estimated by applying a correction factor from the sample with $Z \rightarrow e^+e^-$ events to a control region of opposite-sign dilepton events defined for each same-sign dilepton signal region [65].

4.2.2.2 Baseline selection

In this analysis, the baseline selection criteria are as follows:

- Requiring dilepton + H_T trigger for 2016 data, or three dilepton triggers for 2017 and 2018 data.
- 2 leptons with the same electric charge, or 3 or more leptons. The lepton with the highest p_T must have $p_T > 25$ GeV, and the lepton with lower p_T must have $p_T > 20$ GeV.

- For events with two same-sign leptons, the invariant mass of both leptons must be $m_{\text{SSDL}} > 20$ GeV. For three or more leptons, the invariant mass of a pair of leptons with opposite charges must be $m_{\text{OSDL}} > 12$ GeV, and $|m_{\text{OSDL}} - m_Z| > 15$ GeV. These requirements are added to reject events containing Z boson decays. If an event has the invariant mass in the Z boson mass window, it will be included in a special control region which will be used to constrain the normalisation of $t\bar{t}Z$ process.
- 2 or more jets with $p_T > 40$ GeV.
- 2 or more b-tagged jets with $p_T > 25$ GeV using the DeepCSV algorithm [162], with a working point offering 55-70% efficiency for b jets with p_T between 20-400 GeV. With this working point, the misidentification rate is 1-2% for light flavour and gluon jets, and 10-15% for charm jets [65].
- Sum of jet transverse momentum $H_T > 300$ GeV.
- Missing transverse energy $p_T^{\text{miss}} > 50$ GeV.

4.2.2.3 Analysis technique used

The analysis is performed in two ways, using cut-based analysis and BDT analysis. Events passing the baseline selection described above are categorised into different signal and control regions. For cut-based analysis, the categories are based on the number of leptons, number of b-tagged jets, and number of jets, resulting in 14 signal regions and two control regions CRW and CRZ designed to contain mostly $t\bar{t}W$ and $t\bar{t}Z$ background respectively. For BDT analysis, a BDT classifier is trained to discriminate between $t\bar{t}t\bar{t}$ and all other SM backgrounds. It is then used to separate events into 17 signal regions in addition to a special control region CRZ. Event yields from these regions determined by the BDT classifier are shown in Figure 4.3.

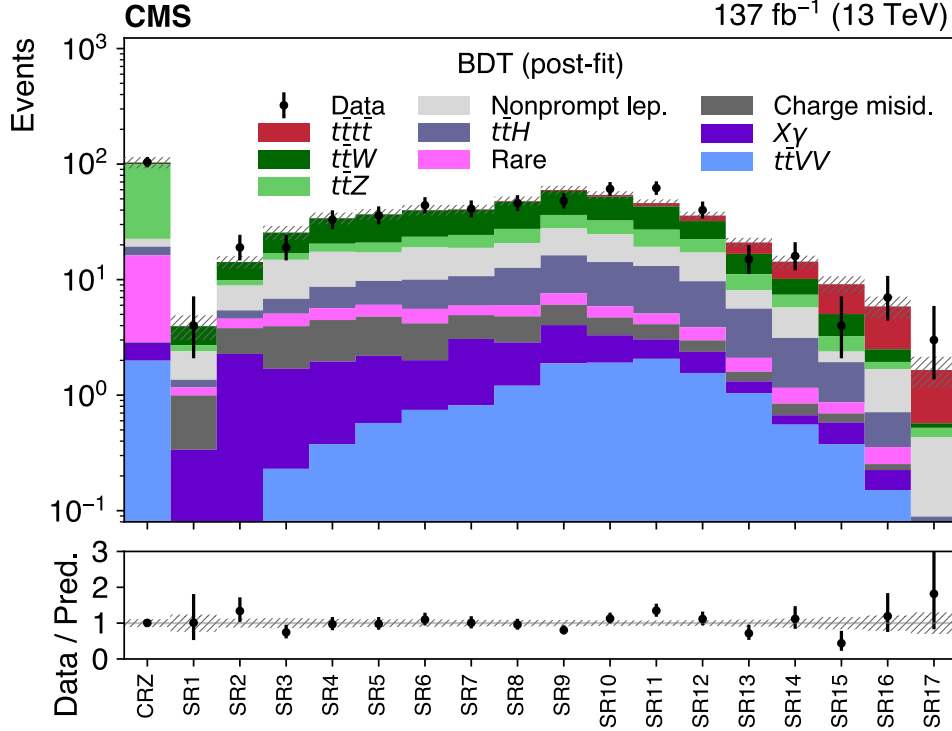


Figure 4.3: Event yield distribution of signal regions determined by BDT classifier from Run 2 SSDL&ML analysis [65].

As shown in Ref. [65], the results from BDT analysis give better expected search results based on SM expectations and are thus chosen as the main result.

4.2.2.4 Systematic uncertainties

In this analysis, systematic uncertainties are included as follows:

- Theoretical uncertainties
 - $t\bar{t} + b\bar{b}$ **normalisation uncertainty** (See Section 5.4.1.1)
 - **Normalisation uncertainties** from $t\bar{t}H$, $t\bar{t}W$, $t\bar{t}Z$, and other minor backgrounds (See Section 5.4.1.2)
 - **Initial and final state radiation (ISR and FSR)** (See Section 5.4.1.3)
 - **Parton distribution function (PDF)** (See Section 5.4.1.4)

- **Renormalisation (μ_R) and factorisation scale (μ_F) uncertainties** (See Section 5.4.1.5)
- Experimental uncertainties
 - **Luminosity uncertainties** - In this analysis, the integrated luminosity across three years of data taking is reported as 137 fb^{-1} , which is outdated compared to the latest integrated luminosity calculation of 138 fb^{-1} . (See Section 5.4.2.1)
 - **Prefire** (See Section 5.4.2.2)
 - **Pileup** (See Section 5.4.2.3)
 - **Trigger efficiency uncertainty**, assigned as 2%
 - **Uncertainty from charge misidentification background**, assigned as 20%
 - **Uncertainty from nonprompt lepton background**, assigned as 30% and 60% for nonprompt electrons with $p_T > 50 \text{ GeV}$
- Reconstruction uncertainties
 - **Jet energy scale (JES) and jet energy resolution (JER)** (See Sections 5.4.3.1 and 5.4.3.2)
 - **b-tagging uncertainty** using DeepCSV [162] b-tagging algorithm (See Section 5.4.3.3)
 - **Lepton scale factor** (See Section 5.4.3.4)

4.2.2.5 Results from the same-sign dilepton and multilepton final state

The observed search significance of this final state is 2.6 standard deviations, equivalent to a probability of 1 in 215 that the result is a statistical fluctuation. The observed best fit of signal strength μ is $1.0^{+0.5}_{-0.4}$. These results are improvements over

Ref. [79] and are considered the search result with the highest search significance before the latest combination. In addition to the measurement of signal strength and search significance, the results from this analysis are used to measure top quark Yukawa coupling, which has already been discussed in Section 1.8.3.

4.3 Run 2 single-lepton

This analysis, performed on data collected over the Run 2 phase in 2016-2018, is performed on the single-lepton (SL) final state. This analysis supersedes the 2016 SL analysis performed back in Ref. [1] with updated analysis techniques (see Section 4.2.1 for the previous 2016 SL analysis). As explained in Section 4.2.1, the superseded analysis has the expected and observed significance of 0.21 and 0.36 respectively, which are measly low amounts since, based on these values, the probability of the search results being a statistical fluke is 0.4168 and 0.3594 respectively. A full detail of this analysis is available in Ref. [163].

4.3.1 Main background estimation

As with the opposite-sign dilepton final state, the main background for this final state is $t\bar{t}$ with additional jets, which is simulated at the NLO level generator. To simplify the uncertainty scheme on this background process, $t\bar{t}$ events are further separated into smaller categories based on the number of b-jets at the generator level, using the Ghost-Hadron Clustering algorithm [164]. Two main categories that are used in this analysis are

- **ttbb**, containing
 - $t\bar{t} + b$, containing one additional b-jet with one b hadron
 - $t\bar{t} + 2b$, containing one additional b-jet with at least two b hadrons

- $t\bar{t} + b\bar{b}$, containing at least two additional b-jets, regardless of the number of b hadrons in each b-jet
- **ttnobbb**, containing
 - $t\bar{t} + c\bar{c}$, containing at least one additional c-jet, regardless of the number of c hadrons in each c-jet
 - $t\bar{t} + \text{LF}$, containing no additional b-jets or c-jets

The reason for the splitting into ttbb and ttnobb is because the $t\bar{t} + b\bar{b}$ process itself is hard to model since there are both b jets produced from top quark decays and b jets not from the decay, but created in association with the $t\bar{t}$ pair. These two types of b jets have different energy scales, which can lead to sensitivity to the renormalisation and factorisation scale parameters μ_R and μ_F [87; 165]. Furthermore, the uncertainty for $t\bar{t} + b\bar{b}$ cross section measured by CMS Collaboration is 30% [87], and it has been determined from previous analyses as the uncertainty with the most impact [1; 65]. By splitting into ttbb and ttnobb, we can directly apply $t\bar{t} + b\bar{b}$ uncertainties to the ttbb background instead of constructing the uncertainty inside the $t\bar{t}$ background.

Figure 4.4 shows the ratio of background and signal events per category in Run 2 SL analysis, split by the number of resolved top quarks, the number of b-tagged jets, and the number of jets. (See Section 4.3.3 for more information on resolved top tagger.) According to the plot, ttbb and ttnobb are dominant backgrounds across all jet multiplicity categories.

In the measurement of the $t\bar{t}$ production with additional b jets performed previously by the CMS Collaboration [166], it is shown that the $t\bar{t} + b\bar{b}$ event yield is underestimated in simulation. Therefore, a scale factor is applied to $t\bar{t} + b\bar{b}$ events based on the measured value reported in Ref. [166]. This background process also

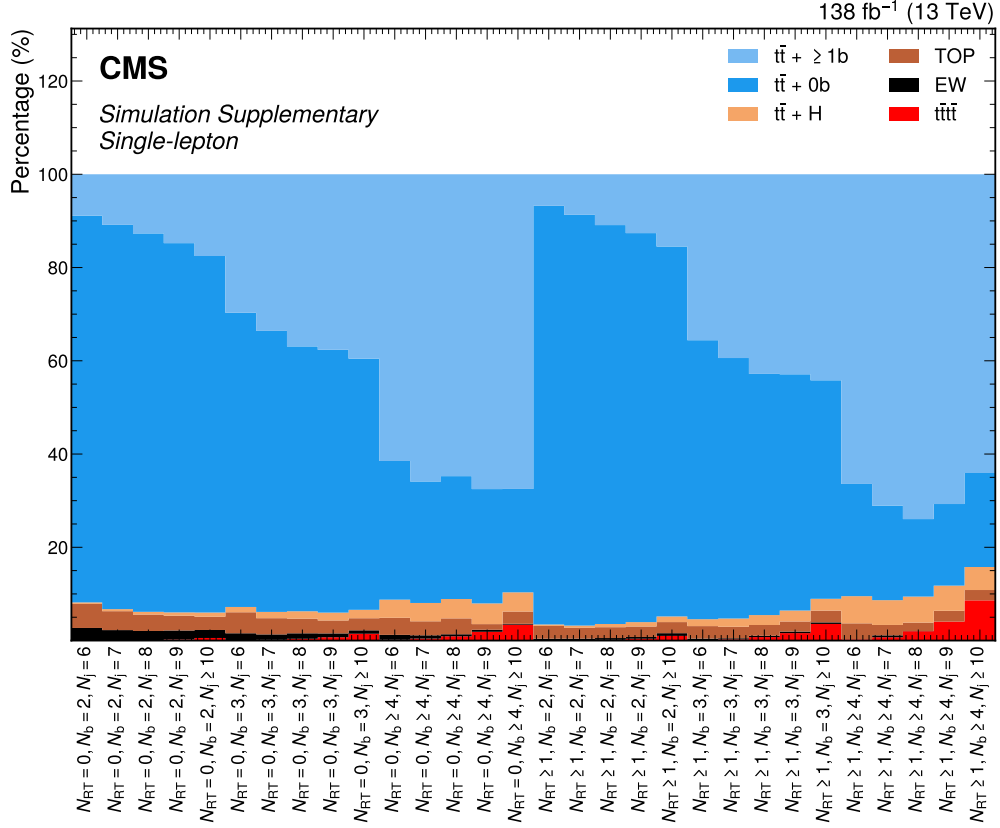


Figure 4.4: Distribution ratio of background and signal events in Run 2 SL analyses per category, split by the number of resolved top quarks, the number of b-tagged jets, and the number of jets. Notice the dominance of $t\bar{t}$ background events, which are further split into $t\bar{t}b\bar{b}$ ($t\bar{t} + \geq 1b$ in the plot) and $t\bar{t}nobb$ ($t\bar{t} + 0b$ in the plot). Figure derived from auxiliary materials of [2].

suffers from low statistics in the region with high jet multiplicity and the high sum of transverse momentum H_T . As with the 2016 OSDL analysis (Section 4.2.1), a dedicated $t\bar{t}$ sample with nine or more jets, reconstructed with anti- k_t algorithm using $R = 0.4$ (also called AK4 jets), with $p_T > 30$ GeV and $H_T > 500$ GeV filter at the generator level is generated and used to reduce statistical uncertainties at the region.

4.3.2 Baseline selection

The baseline selection criteria for this analysis are as follows:

- Trigger requirements
 - 2016: requiring the same set of triggers as used in the 2016 single-lepton analysis, which requires at least one isolated muon with $p_T > 24$ GeV and $|\eta| < 2.4$, or one isolated electron with $p_T > 32$ GeV and $|\eta| < 2.1$ [1].
 - 2017 and 2018: requiring either
 - * isolated electron with $p_T > 35$ GeV
 - * isolated muon with $p_T > 50$ GeV
 - * very loosely isolated electron or muon with $p_T > 15$ GeV and missing transverse energy of $\cancel{E}_T > 450$ GeV

This set of triggers offers efficiencies close to 1 for events with lepton $p_T > 50$ GeV. For events with lower lepton p_T , a scale factor of 0.98 is introduced to account for the efficiency difference between MC and data [163].

- Lepton requirements
 - 2016: 1 tight, isolated electron with $p_T > 35$ GeV and $|\eta| < 2.1$, or 1 tight, isolated muon with $p_T > 26$ GeV and $|\eta| < 2.1$
 - 2017+2018: 1 tight, isolated electron or muon with $p_T > 20$ GeV and $|\eta| < 2.1$
 - no additional loose leptons to preserve orthogonality with other final states
- 4 or more jets, classified with anti- k_t algorithm [112] with a distance parameter of 0.4 (AK4), with $p_T > 30$ GeV.

- Among the jets in an event, there must be 2 or more b-tagged jets with $p_T > 25$ GeV using the DeepCSV algorithm [162] at a medium working point, which provides average b-tagging efficiency of 60% and mistagging rate of 1% [163].
- Sum of jet transverse momentum $H_T > 500$ GeV
- Missing transverse energy $p_T^{\text{miss}} > 60$ GeV
- Transverse mass of lepton + neutrino $M_T > 60$ GeV
- ΔR between the lepton and the closest jet > 0.4 , where $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$

4.3.3 Resolved top tagger

Since this final state contains three top quarks each decaying hadronically into three jets with low p_T , the CMS-specific method of top-tagging with AK8 jets⁵ is not sufficient to detect them [163]. Instead, to detect jets with low p_T originating from a top quark hadronic decay, a neural network-based top quark tagger named DeepResolved [167; 168] is used. For each event, all possible combinations of three jets are made, and the tagger excludes combinations that do not originate from a top quark or overlap with other combinations. The remaining combinations are considered top quark candidates.

4.3.4 Analysis technique used

This analysis employs BDT discriminators to discriminate between $t\bar{t}t\bar{t}$ events as signal and $t\bar{t}$ as background. Since $t\bar{t}$ is the dominant background in this final state, training the discriminator against two classes of events is sufficient for this analysis. Several sets of input parameters, such as the kinematics of jets and

⁵AK8 jets refer to jets classified with anti- k_t algorithm [112], but with a distance parameter of 0.8 instead of 0.4 used in standard AK4 jets. See Section 2.9 for more information.

leptons, b-tagging discriminators, pair of b-tagged jets, the kinematics of top quarks tagged with DeepResolved top tagger and event shapes are considered as inputs. Ultimately, the BDT used in this analysis contains 40 input variables and is trained with events containing at least 6 jets, as it provides the best performance in terms of expected significance and cross section limit.

After the BDT discriminator is trained, the events are separated into 60 categories based on the lepton flavour (electron or muon), number of jets (6, 7, 8, 9, and 10+), number of b-tagged jets (2, 3, and 4+), and number of resolved top-tagged jets (0 and 1+). This categorisation scheme has been optimised to achieve the best expected search significance considering the statistics of MC samples. Within each jet multiplicity, b-jet multiplicity, and resolved top category, the distributions are populated in variable-sized bins (with the minimum size requirement for each b-tagged jet category) to have the statistical uncertainty from the total background be less than 30%. The binning scheme is optimised simultaneously in the electron and muon categories so that both categories will have the same binning scheme. The adjustment of the binning scheme is done to optimise the search significance to be as high as possible. However, the distributions of BDT output differ between simulation and data, and several scale factors are calculated, separated by year and b-jet multiplicity, to correct the shape discrepancy.

4.3.5 Systematic uncertainties

In this analysis, several uncertainties are applied in the list below. Detailed discussions on the origin of each type of uncertainty will be discussed in Section 5.3.

- Theoretical uncertainties
 - $t\bar{t} + b\bar{b}$ **normalisation** (See Section 5.4.1.1.)

- **Theoretical normalisation uncertainties** for Electroweak, single-top, ttbb, ttnobb, and $t\bar{t}H$ backgrounds (See Section 5.4.1.2.)
- **Initial and final state radiations (ISR and FSR) uncertainties**, calculated for $t\bar{t}t\bar{t}$, ttbb, ttnobb, electroweak, single-top, $t\bar{t}H$, and QCD backgrounds. Due to low statistics in 2016, this type of uncertainty for QCD is not applied in 2016 data. (See Section 5.4.1.3.)
- **Parton distribution function (PDF) uncertainty** (See Section 5.4.1.4.)
- **Uncertainties from renormalisation and factorisation scale variations μ_R and μ_F** (See Section 5.4.1.5.)
- **Matrix element to parton shower matching (ME-PS matching) uncertainty**, commonly known as h_{damp}
- Experimental uncertainties
 - **Luminosity uncertainty** applied differently across years. Also contains uncertainties correlated between 2017+2018 and Run 2. (See Section 5.4.2.1.)
 - **Prefire uncertainty** (See Section 5.4.2.2.)
 - **Pileup uncertainty** (See Section 5.4.2.3.)
 - **BDT shape correction uncertainty**, as discussed in Section 4.3.4
- Reconstruction uncertainties
 - **Jet energy scale (JES) and jet energy resolution (JER) uncertainties.** For this final state, JES uncertainties are broken down into five sources. (See Sections 5.4.3.1 and 5.4.3.2.)
 - **b-tagging** uncertainties from the DeepCSV algorithm [162], including both statistical and non-statistical parts in b tagging, c tagging, and light-flavoured tagging. (See Section 5.4.3.3.)

- **Lepton scale factors uncertainty**, both electrons and muons (See Section 5.4.3.4.)
- **DeepResolved top tagger uncertainties** [167; 168], including
 - * closure uncertainty, calculated from deviations from unity in the scale factor provided by the tagger
 - * control sample (CS) purity, calculated from the dependence of the tagging/mistagging rate on the purity of real/fake top quarks of the control sample
 - * statistical uncertainty in the scale factors

Due to a fine splitting of the binning of the BDT discriminator and some uncertainties, such as jet energy scale, causing some events to migrate from one jet multiplicity category to another category, some uncertainties gain high fluctuations within each jet category. To alleviate this, this analysis uses the LOWESS algorithm (LOcally WEighted Scatterplot Smoothing) [169; 170] to smooth the uncertainties as used in the search for heavy Higgs bosons decaying into a pair of top quarks performed by CMS Collaboration [171]. The resulting systematic uncertainty is symmetrized around nominal value and then used in this analysis.

4.3.6 Results from the single-lepton final state

Based on the BDT distribution, partly shown in Figure 4.5, the expected and observed significance is 1.2 and 1.4 respectively. The measured signal strength and cross section from this final state are $1.2^{+0.7}_{-0.6}$ (stat.) ± 0.6 (syst.) and 15 ± 8 (stat.) $^{+10}_{-7}$ (syst.) fb respectively. These results are a significant improvement from the previous analysis presented in Section 4.2.1 [1], where the expected and observed significance was 0.21 and 0.36 respectively. Improvements come mainly from the categorisation and binning scheme to maximise the expected search significance,

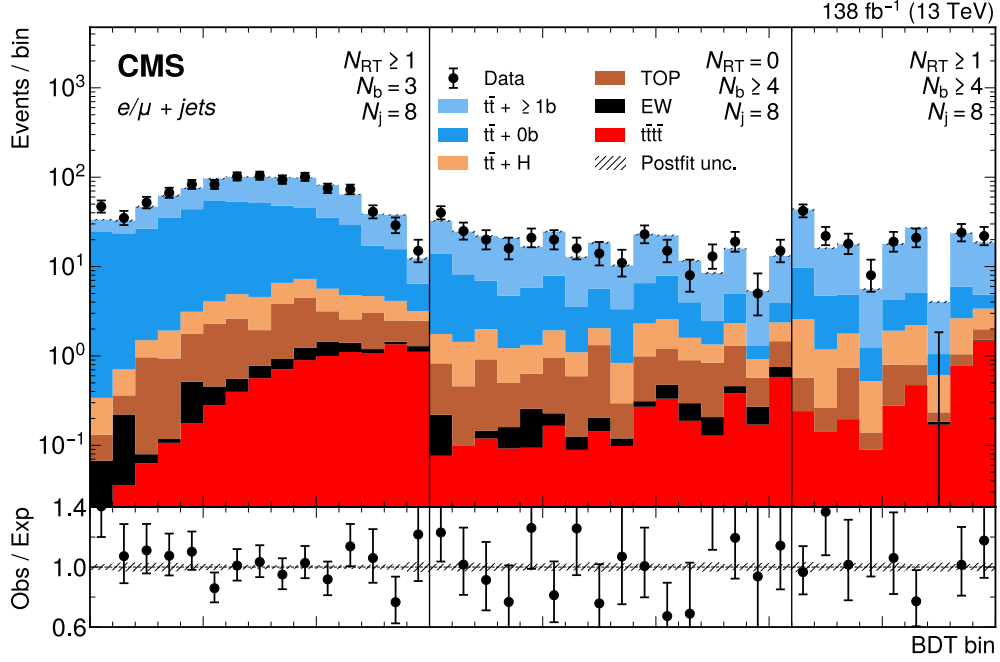


Figure 4.5: BDT output distributions of three jet multiplicity categories, defined by number of resolved top quarks (N_{RT}), number of b-tagged jets (N_b), and number of jets (N_j). Figure derived from [2].

Table 4.2: Measured signal strength and cross section for single-lepton final state, broken by data-taking year.

Year	Signal strength ($\sigma_{t\bar{t}t\bar{t}}^{\text{SM}}$)			Cross section (fb)		
	(stat.)	(syst.)		(stat.)	(syst.)	
2016	-2.8	$+1.1$ -0.8	$+1.2$ -0.9	-31	$+13$ -8	$+13$ -11
2017	0.5	$+1.1$ -1.0	$+1.0$ -0.9	6	$+14$ -12	$+14$ -11
2018	3.0	± 1.0	$+1.0$ -0.9	36	± 12	$+18$ -12
Combined	1.2	$+0.7$ -0.6	± 0.6	15	± 8	$+10$ -7

since more constraints are introduced from a larger number of histogram bins included in the fit template.

4.4 2017+2018 opposite-sign dilepton

This analysis is performed on the opposite-sign dilepton final state using data collected during 2017-2018, which amounts to an integrated luminosity of

101.5 fb^{-1} . The main goal of this analysis is to employ an H_T -based strategy on the sum of jet transverse momentum H_T . This is different to the 2016 OSDL analysis (presented in Section 4.2.1), which is a multivariate analysis (MVA) utilising BDT discriminators (see Section 4.2.1). This analysis has improved Monte Carlo simulation and better b-tagging algorithms compared to the 2016 OSDL. A full detail of this analysis is available in Ref. [172].

4.4.1 Main background estimation

The main background for this final state is $t\bar{t}$ production, and similar to Run 2 SL analysis explained in Section 4.3, they are split into $t\bar{t} + b\bar{b}$ and $t\bar{t} + \text{not } b\bar{b}$. The same splitting scheme as in the Run 2 SL analysis is applied here as well. (See Section 4.3.1 for more details.)

4.4.2 Baseline selection

The baseline requirements for this analysis are as follows:

- Requiring dilepton triggers as follows, with a priority cascade rule where events passing $e\mu$ triggers are chosen first, followed by events passing $\mu\mu$ triggers but not $e\mu$ triggers, and finally events passing ee triggers but not both $e\mu$ and $\mu\mu$ triggers [172]

– 2017 data

- * $e\mu$ channel: HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ
and HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ
for data eras B, C, D, E, and F
- * $\mu\mu$ channel: HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ for data
era B, and HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8
for data eras C, D, E, and F

- * ee channel: HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL for data eras B, C, D, E, and F
- 2018 data with data eras A, B, C, and D
 - * $e\mu$ channel: HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ and HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ
 - * $\mu\mu$ channel: HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8
 - * ee channel: HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL
- Two isolated leptons (electrons or muons) with opposite charges, where leading lepton must have $p_T > 25$ GeV, and sub-leading lepton must have $p_T > 15$ GeV. Muons must be within $|\eta| < 2.4$, while electrons must be within $|\eta| < 2.5$.
- For events with two electrons or two muons, the invariant mass $m_{\ell\ell}$ of both leptons must satisfy $m_{\ell\ell} \geq 20$ GeV and $|m_{\ell\ell} - m_Z| > 15$ GeV. These requirements are put in place to remove low-mass resonances and events with two leptons decaying from Z bosons [172].
- 4 or more jets, classified with anti- k_t algorithm [112] with a distance parameter of 0.4 (AK4), with $p_T > 30$ GeV and $|\eta| < 2.5$.
- Among the jets in an event, there must be 2 or more b-tagged jets, tagged using the DeepJet algorithm [113] at a medium working point, offering a misidentification rate of 1% [151].
- Sum of jet transverse momentum $H_T > 500$ GeV to suppress $t\bar{t}$ backgrounds.

4.4.3 Analysis technique used

This analysis uses only distributions of the sum of scalar jet transverse momentum (H_T) to determine the best fit of signal strength and search significance. Events

are categorised based on the number of jets (4, 5, 6, 7, and 8 or more) and the number of b-tagged jets (2, 3, and 4 or more), resulting in 15 jet categories in total. Distributions of H_T are then generated based on 15 jet multiplicity categories, further separated by year (2017 and 2018) and each of the three lepton flavour combinations (ee , $e\mu$, and $\mu\mu$), resulting in $5 \times 3 \times 2 \times 3 = 90$ sets of histograms.

The binning scheme for H_T distribution is designed so that statistical uncertainties based on unweighted $t\bar{t}$ events (including both $t\bar{t}b\bar{b}$ and $t\bar{t}n\bar{b}b$) in each histogram bin becomes below 30% in all bins, and the width of each bin is close to the H_T resolution determined from $t\bar{t}t\bar{t}$ simulation. The binning scheme is applied to all 90 histogram templates used to calculate the best fit of signal strength value [172]. In practice, having more histogram bins mean more constraints to the parameter of interest, which in this case is signal strength or four top quark production cross section.

4.4.4 Systematic uncertainties

Several systematic uncertainties are applied in this analysis as follows:

- Theoretical uncertainties
 - $t\bar{t} + b\bar{b}$ **normalisation** (See Section 5.4.1.1)
 - **Normalisation uncertainties** (See Section 5.4.1.2.)
 - **Initial and final state radiation** (See Section 5.4.1.3.)
 - **Parton distribution function (PDF)** (See Section 5.4.1.4.)
 - **Renormalisation and factorisation scale variations μ_R and μ_F** (See Section 5.4.1.5.)
 - **Matrix element to parton shower matching (ME-PS matching)**, commonly known as h_{damp}

- Experimental uncertainties
 - **Luminosity uncertainty** (See Section 5.4.2.1.)
 - **Prefire** (See Section 5.4.2.2.)
 - **Pileup** (See Section 5.4.2.3.)
- Reconstruction uncertainties
 - **Jet energy scale (JES) and jet energy resolution (JER)** (See Sections 5.4.3.1 and 5.4.3.2.)
 - **b-tagging** uncertainties from DeepJet [113] algorithm (See Section 5.4.3.3.)
 - **Lepton scale factor** (See Section 5.4.3.4.)

4.4.5 Results from the opposite-sign dilepton final state

Figure 4.6 shows the distribution of events per jet multiplicity and b-tagged jet multiplicity. As shown in the figure, both ttbb (shown as $t\bar{t} + \geq 1b$ in the plots) and ttnobb (shown as $t\bar{t} + 0b$ in the plots) become dominant backgrounds in all jet multiplicity categories.

Based on a fit of H_T distribution over 90 histogram templates across 2 years, 3 lepton flavour combinations, 5 jet multiplicities, and 3 b-jet multiplicities, the expected and observed search significance for this analysis is 0.6 and 1.9 standard deviations. The measured signal strength and cross section from this final state are 2.8 ± 1.0 (stat.) $^{+1.9}_{-1.2}$ (syst.) and 33 ± 12 (stat.) $^{+15}_{-14}$ (syst.) fb respectively. Table 4.4 shows the measured values broken down per year.

Compared to the results in the 2016 OSDL analysis, which gives the expected and observed significance of only 0.36 and 0.0 standard deviations (or 0.45 and 0.0 with four-top quark production cross section scaled, see Section 4.2.1.5), this result

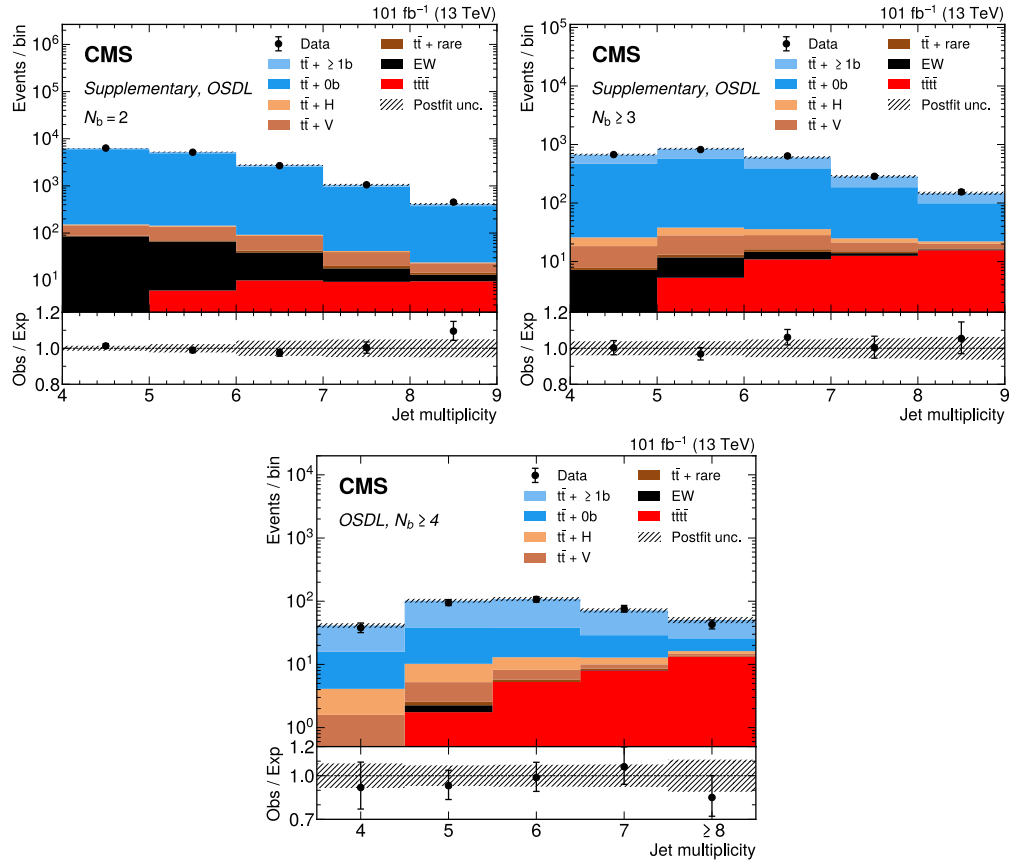


Figure 4.6: Jet multiplicity distributions for events with 2, 3, and 4 b-tagged jets in 2017+2018 OSDL analyses. Figures derived from Ref. [2].

Table 4.3: Expected signal strength for opposite-sign dilepton final state, broken by data-taking year.

Year	Signal strength ($\sigma_{t\bar{t}t\bar{t}}^{\text{SM}}$)
2017	$1^{+2.6}_{-2.3}$
2018	$1^{+2.2}_{-2.0}$
2017+2018	$1^{+1.8}_{-1.6}$

is once again a substantial improvement, due to a larger dataset and a more sensitive histogram template binning. The improvements seen in this analysis has the same characteristic as in the Run 2 single-lepton analysis introduced in Section 4.3.

Table 4.4: Observed signal strength and cross section for opposite-sign dilepton final state, broken by data-taking year.

Year	Signal strength ($\sigma_{t\bar{t}t\bar{t}}^{\text{SM}}$)			Cross section (fb)		
	(stat.)	(syst.)		(stat.)	(syst.)	
2017	3.6	+1.7 -1.5	+2.6 -1.6	44	+20 -19	+20 -18
2018	2.2	+1.4 -1.3	+2.3 -1.5	27	+17 -16	+20 -18
2017+2018	2.8	± 1.0	+1.9 -1.2	33	± 12	+15 -14

4.5 Run 2 all-hadronic

For the first time in CMS Collaboration, the search for four-top quark production is performed in the all-hadronic final state, where all four W bosons from four top quarks decay hadronically. This leaves the final product of collision events in this final state to have a large number of jets and no significant missing transverse energy. Furthermore, this final state is dominated by QCD process backgrounds which are notoriously difficult to simulate with Monte Carlo methods due to their high uncertainty. It also has a signal-to-background ratio of approximately 10^{-5} , assuming that the cross section of the signal process is equal to the SM prediction [2].

This analysis is performed on the full Run 2 dataset, spanning from 2016-2018 with an integrated luminosity of 138 fb^{-1} . The main feature of this analysis is, that instead of relying on Monte Carlo simulations for its dominant background, it utilises a data-driven approach to estimate both the yield and shape of BDT output distribution, which will be further discussed in Section 4.5.1. A full detail of this analysis is available in Ref. [151].

4.5.1 Main background estimation

For this final state, the dominant background is QCD multijet and $t\bar{t}$ processes. Monte Carlo simulations for both dominant backgrounds, however, are problematic

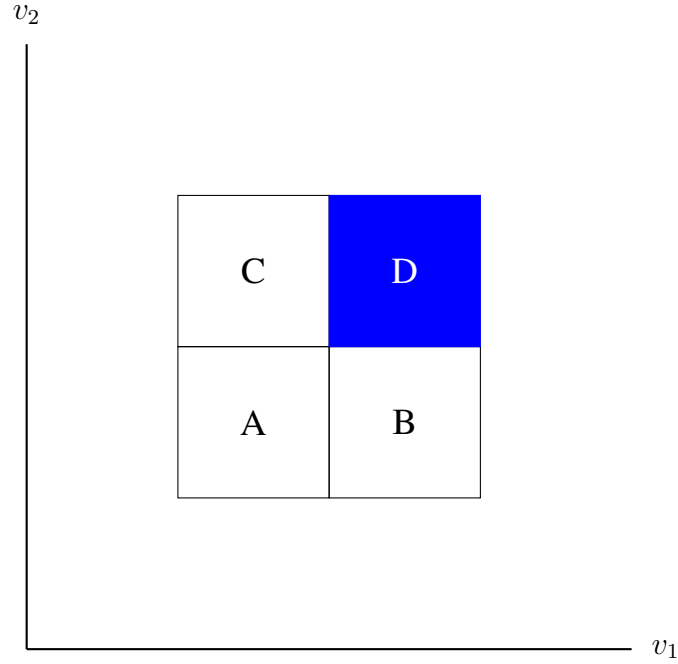


Figure 4.7: Example control and signal regions of two *uncorrelated* variables v_1 and v_2 for ABCD method. The control regions A, B, and C are used to estimate event yields in signal region D.

due to large uncertainties even at the NLO level generators and limited statistics [151]. To avoid the inaccuracies from Monte Carlo simulations, these dominant backgrounds are estimated using data-driven methods instead. However, Monte Carlo simulations of $t\bar{t}$ production are still used to train a neural network predicting BDT discriminant distributions, which will be explained later in Section 4.5.4.

In this analysis, two data-driven estimation techniques are used to estimate the event yields of a jet category, as well as the output distribution of the BDT discriminator. Both techniques are inspired by the standard ABCD method, where the number of events from three control regions is used to estimate the number of events in the signal region [151].

The idea of the ABCD method is simple. Given a phase space distribution of two *uncorrelated* variables v_1 and v_2 shown in Figure 4.7. To estimate the event yield in the signal region D, we can use the event yields in adjacent regions A, B, and C. By assuming that the rate of change of event yields from one region of a variable v_1 or v_2 remains the same, we can construct an equation

$$\frac{A}{B} = \frac{C}{D} \quad (4.1)$$

which we can solve for D as

$$D = \frac{BC}{A} \quad (4.2)$$

To predict the event yield of each analysis category, this analysis employs an extended ABCD (extABCD) method [150]. Instead of relying only on three control regions to predict the signal region, this analysis uses two additional regions⁶, X and Y, for a more precise prediction, depicted in Figure 4.8. With the additional two control regions, the estimation of D now becomes [151]

$$D = \left(\frac{BC}{A} \right) \left(\frac{CX}{AY} \right) \quad (4.3)$$

This new estimation technique has been proven by toy examples and the distribution of $t\bar{t}$ events associated with two additional jets to be more accurate than the vanilla ABCD method, as shown in Ref. [150].

The distribution of BDT output is predicted by the second data-driven estimation method, ABCDnn [173] based on the idea of neural autoregressive flows [132], in the signal region based on the same type of distributions as inputs. In this analysis, a deep neural network⁷ is trained to transform BDT output distributions from simulated $t\bar{t}$ events to match data-driven BDT output distributions in each control region.

⁶The original paper for the extABCD method [150] also offers a variation where eight control regions are used to estimate the number of events in the signal region.

⁷Not to be confused with a BDT discriminator, used to classify between $t\bar{t}t\bar{t}$ events and $t\bar{t}$ and QCD events. This will be discussed in Section 4.5.4.

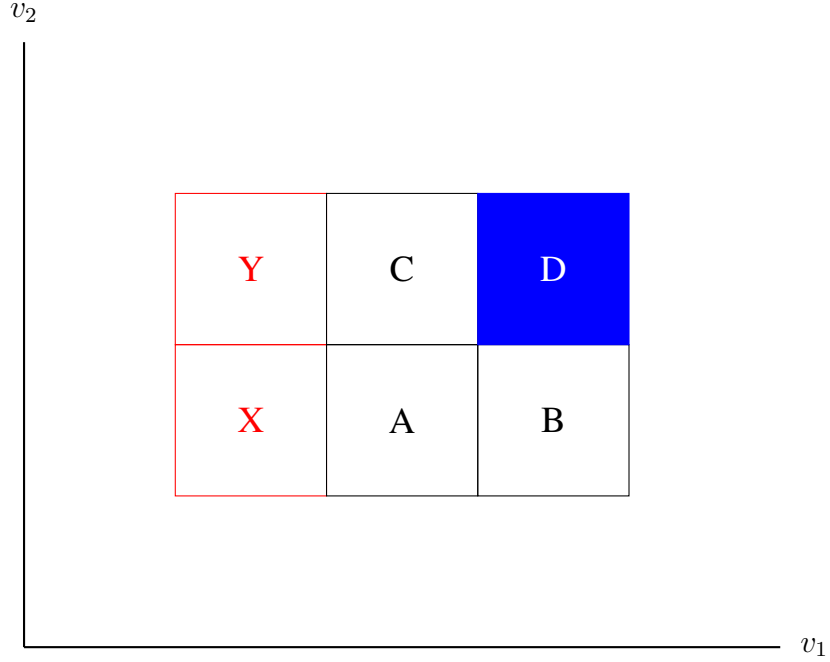


Figure 4.8: Example control and signal regions of two *uncorrelated* variables v_1 and v_2 for extended ABCD method. In this method, two additional control regions X and Y can be added to the estimation of the signal region D.

The idea of normalizing flow, which is the basis for neural autoregressive flows, is that it is an invertible function that transforms the distribution of random variables [132]. As such, the neural network does this only by generating an invertible function which translates, stretches, or squeezes the phase space under different conditions defined by control variables. After the training, the neural network is then used to transform $t\bar{t}$ distribution into data-driven prediction in the signal region [151].

In this analysis, the data-driven backgrounds from $t\bar{t}$ and QCD multijet processes are estimated using both extABCD and ABCDnn methods. Background estimation in the signal region with 9 or more jets and 3 or more b-tagged jets is performed first by the ABCDnn method to estimate the BDT discriminator distribution shape. Then, the normalisation of the BDT distribution is calculated

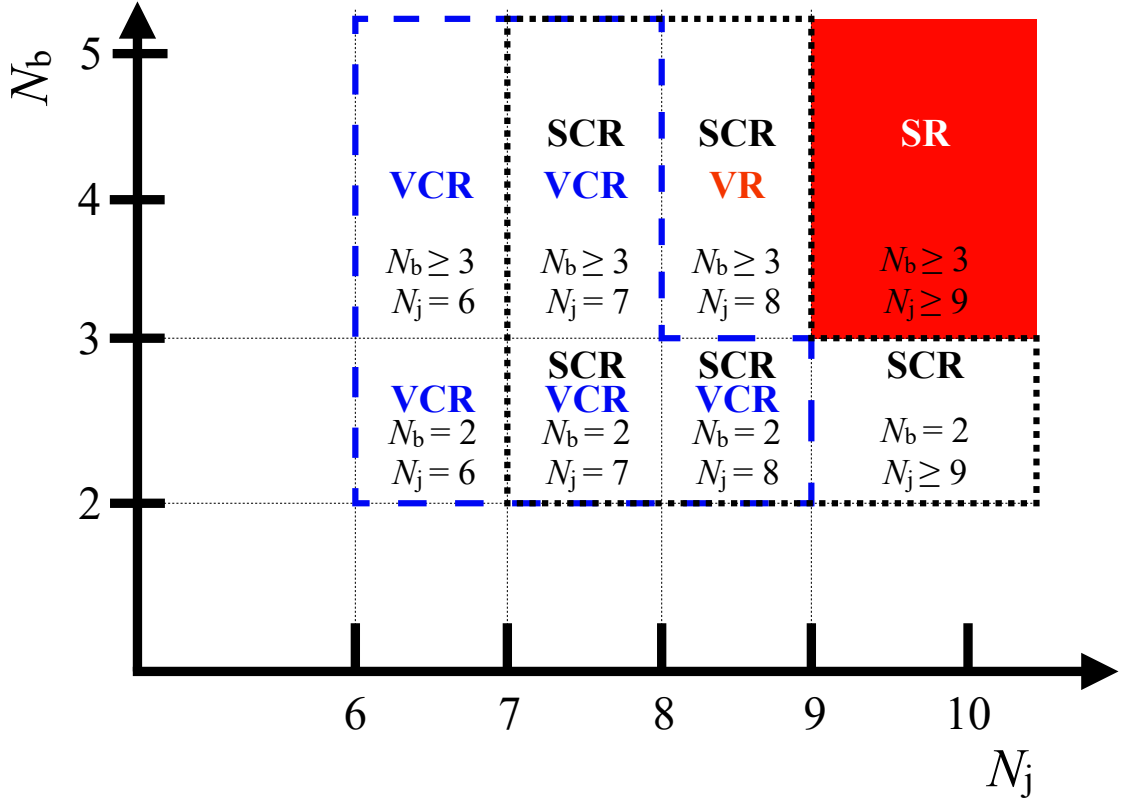


Figure 4.9: Signal region (SR), validation region (VR), signal control region (SCR), and validation control region (VCR) used in an estimation of data-driven backgrounds based on extABCD and ABCDnn methods [2].

by the extABCD method. The estimation for the signal region uses the information from five control regions as illustrated in Figure 4.9.

To prove that this method works, validation checks have also been performed in a validation region with exactly 8 jets and 3 or more b-tagged jets, using a different set of five control regions as shown in Figure 4.9 [151].

Since this analysis employs data-driven estimation methods to estimate its dominant backgrounds, several checks in validation regions have been performed to ensure that the predictions are valid in that region where we do not expect to see signal events. With these techniques, unique systematic uncertainties are also introduced, which are calculated from statistical uncertainties from the

estimated yields from the extABCD method, from the $t\bar{t}$ events used to simulate data-driven background shape distributions, from trigger efficiency corrections, and uncertainties from disagreements in data compared to prediction.

4.5.2 Baseline selection

The baseline requirements for this analysis are as follows:

- Using triggers requiring 6 or more jets, with additional b-jet and H_T requirements as follows:
 - 2016: 2 or more b-tagged jets and $H_T > 400$ GeV *or* one or more b-tagged jets and $H_T > 450$ GeV
 - 2017 and 2018: 2 or more b-tagged jets and $H_T > 380$ GeV *or* one or more b-tagged jets and $H_T > 430$ GeV.
- Zero leptons, required to preserve orthogonality with other final states
- 9 or more jets, with at least 3 of them being b-tagged jets, tagged using the DeepJet algorithm [113] at a medium working point. This requirement is used to reduce $t\bar{t}$ backgrounds [151] which mostly have fewer b-tagged jets. However, events with a lower number of jets are also used as control regions to estimate data-driven backgrounds, as explained in Section 4.5.1.
- Sum of jet transverse momentum $H_T > 700$ GeV to reduce QCD-induced multijet backgrounds [151].

4.5.3 Boosted and resolved top tagging

This analysis also applies boosted and resolved top tagging and uses the number of tagged top quarks for further classification of events, which will be discussed in Section 4.5.4. Resolved top tagging is performed at lower jet energies, where decay

products can be separated by standard jet reconstruction. Boosted top tagging, on the other hand, is performed at higher jet energies, where the top quark is boosted with high energy so that the decay products become closer together, thus requiring them to be reconstructed as a fat jet or a jet with a larger acceptance cone size.

To detect boosted top quarks, this analysis uses reconstructed large-radius jets via anti- k_T algorithm with the distance parameter $R = 0.8$ (also known as AK8 jets) and tags them using the DeepAK8 algorithm [174] using a medium working point, offering a mistagging rate of 1% [151]. On the other hand, resolved top quarks (from moderately boosted top quarks with resolvable decay products) are detected using a dedicated BDT, based on a top tagger used in CMS search for supersymmetric partners of top quarks [175]. The top-tagging BDT classifies resolved top candidates, each comprising one b-tagged jet based on the DeepJet b tagger and two light jets probably decayed from a W boson [151].

With both boosted and resolved top taggers in place, systematic uncertainties from both top taggers are also applied exclusively in this final state, which will be discussed in detail in Section 4.5.5.

4.5.4 Analysis technique used

Contrary to analyses in signatures with leptons, events in this final state are classified into different categories based on the number of *resolved tops* N_{RT} , number of *boosted tops* N_{BT} , and H_T , as shown in Table 4.5.

A BDT discriminator⁸ is trained per each category to discriminate between $t\bar{t}t\bar{t}$ signal events against $t\bar{t}$ and QCD multijet background events. The training incorporating both dominant backgrounds has been shown to have better

⁸Not to be confused with a deep neural network used to predict the distribution of data-driven backgrounds, based on the ABCDnn method [173], as explained in Section 4.5.1.

Table 4.5: 12 event categories used in Run 2 all-hadronic analysis, separated by the number of resolved tops N_{RT} , the number of boosted tops N_{BT} , and H_T . In this analysis, the preselection requirement of $H_T > 700$ GeV is applied before categorisation.

N_{RT}	N_{BT}	H_T
1	0	700-800
1	0	800-900
1	0	900-1000
1	0	1000-1100
1	0	1100-1200
1	0	1200-1300
1	0	1300-1500
1	0	> 1500
1	≥ 1	700-1400
1	≥ 1	> 1400
2	≥ 0	700-1100
2	≥ 0	> 1100

performance than training against $t\bar{t}$ background alone [151]. After the training, the BDT output distributions of data-driven backgrounds are predicted using the ABCDnn method, as explained in Section 4.5.1.

Figure 4.10 shows the distribution of BDT outputs in two of the most sensitive categories: $N_{\text{RT}} = 1$, $N_{\text{BT}} \geq 1$, $H_T > 1400$ GeV, and $N_{\text{RT}} \geq 2$, $H_T > 1100$ GeV. As seen from both figures, the excess of $t\bar{t}t\bar{t}$ signal is visible in the histogram bin representing a high BDT output value.

4.5.5 Systematic uncertainties

As with all other analyses, this analysis also employs several systematic uncertainties present in other analyses, along with some uncertainties unique to this analysis as follows [151]:

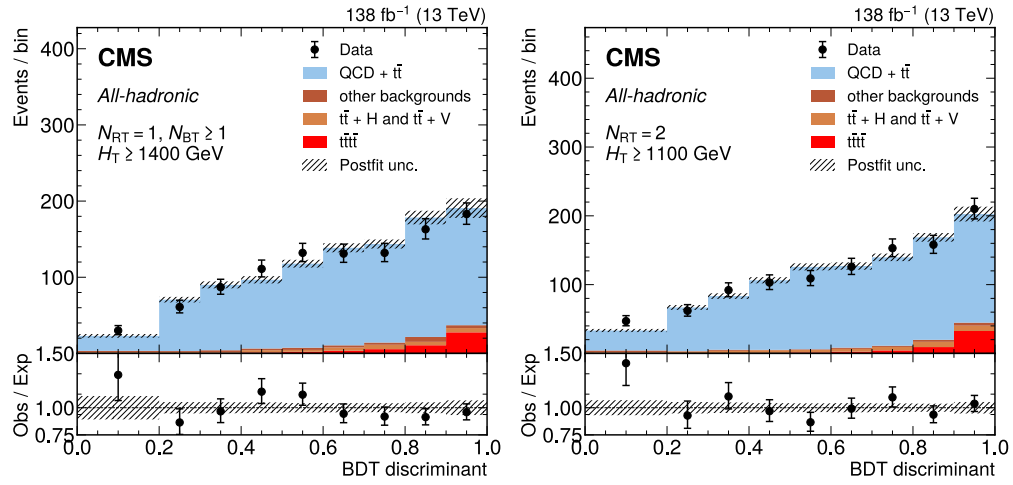


Figure 4.10: BDT output distributions from two of the most sensitive categories: $N_{RT} = 1$, $N_{BT} \geq 1$, $H_T > 1400$ GeV, and $N_{RT} \geq 2$, $H_T > 1100$ GeV [2].

- **Uncertainties caused by data-driven estimation**, unique to this analysis, which are

- Statistical uncertainties of estimated *yields* for QCD multijet and $t\bar{t}$ in each category propagated via the extABCD method [150].
- Statistical uncertainty of $t\bar{t}$ events used to transform into data-driven BDT output distribution using the ABCDnn method [173].
- Statistical uncertainties from the trigger efficiency corrections, dependent on the number of b-tagged jets N_b and number of jets N_j
- Data-prediction disagreement uncertainties, both in normalisation and shape form, derived from the validation region containing 8 jets.

By design, these uncertainties will be constrained during the fitting of corresponding nuisance parameters.

- Theoretical uncertainties

- **Theoretical normalisation uncertainties** for $t\bar{t}W + t\bar{t}Z + t\bar{t}H$ backgrounds. This uncertainty is entirely enveloped by $t\bar{t}H$. (See Section 5.4.1.2.)
- **Initial and final state radiation uncertainties** (See Section 5.4.1.3.)
- **Parton distribution function (PDF) uncertainty** (See Section 5.4.1.4.)
- **Uncertainties from renormalisation and factorisation scale variations μ_R and μ_F** (See Section 5.4.1.5.)
- Experimental uncertainties
 - **Luminosity uncertainty** (See Section 5.4.2.1.)
 - **Prefire uncertainty** (See Section 5.4.2.2.)
 - **Pileup uncertainty** (See Section 5.4.2.3.)
 - **Trigger efficiency uncertainty**, assigned as 0-5% [151]
- Reconstruction uncertainties
 - **Jet energy scale (JES) and jet energy resolution (JER) uncertainties** (See Sections 5.4.3.1 and 5.4.3.2.)
 - **b-tagging uncertainty**, using DeepJet [113] (See Section 5.4.3.3.)
 - **Lepton scale factors uncertainty** (See Section 5.4.3.4.)
 - **DeepAK8 uncertainties** [174], calculated from correction factors to match simulation to data. This type of uncertainty has separate nuisance parameters for the boosted top and boosted W bosons.
 - **Resolved top tagger uncertainties**, derived from the dedicated resolved top tagger explained in Section 4.5.3

4.5.6 Results from the all-hadronic final state

Based on a fit of BDT distributions, the expected and observed search significance for this analysis is 0.4 and 2.5 standard deviations. The measured signal strength and cross section from this final state are 5.8 ± 1.4 (stat.) ± 2.0 (syst.) and 70 ± 17 (stat.) $^{+25}_{-23}$ (syst.) fb respectively. A significant increase in observed search significance compared to expected significance is caused by a large excess of signal events found in this final state, as shown in Figure 4.10. This large excess causes a high value of signal strength best fit and pushes the observed significance higher than the expected value.

Several cross-checks have already been performed in this final state, especially in the predicted data-driven background. Validation of the data-driven background estimation method is performed by estimating the background in the *validation region* with 8 jets and 3 or more b-tagged jets and comparing it with data in the region, with the method explained in Section 4.5.1. Estimated backgrounds from two of the most sensitive categories in the validation region are shown in Figure 4.11.

4.6 Summary of results from individual final states

Tables 4.6 and 4.7 show the best fit values of signal strength and cross section, as well as the expected and observed significance of each analysis. As seen from both tables, none of the analyses performed in the individual final states achieved the expected level with the search significance of 3 standard deviations, even though many of the analyses have upgraded their techniques compared to their previous iterations.

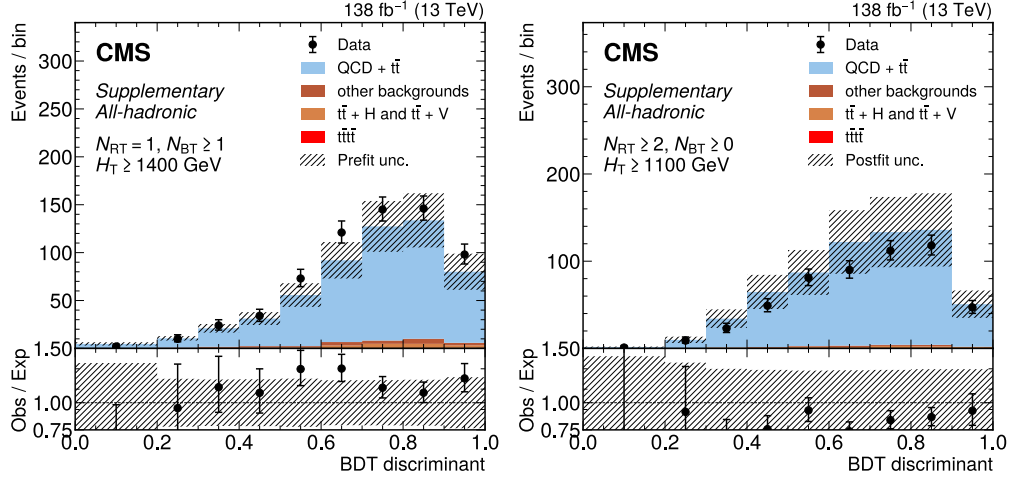


Figure 4.11: BDT output distributions from two of the most sensitive categories: $N_{RT} = 1$, $N_{BT} \geq 1$, $H_T > 1400$ GeV, and $N_{RT} \geq 2$, $H_T > 1100$ GeV in validation region [2].

Table 4.6: Measured signal strength and cross section from individual analyses, including Run 2 SL (Section 4.3), 2017+2018 OSDL (Section 4.4), and Run 2 all-hadronic analysis (Section 4.5), as well as 2016 OSDL analysis (Section 4.2.1) and Run 2 SSDL&ML analysis (Section 4.2.2). The uncertainty values reported are statistical uncertainties and systematic uncertainties respectively, except for the original values of 2016 OSDL where only the total uncertainty is reported.

Analysis	Signal strength (μ)			Cross section (fb)		
		(stat.)	(syst.)		(stat.)	(syst.)
Run 2 SL	1.2	$^{+0.7}_{-0.6}$	± 0.6	15	± 8	$^{+10}_{-7}$
2017+2018 OSDL	2.8	± 1.0	$^{+1.9}_{-1.2}$	33	± 12	$^{+15}_{-14}$
Run 2 All-hadronic	5.8	± 1.4	± 2.0	70	± 17	$^{+25}_{-23}$
Run 2 SSDL&ML [65]	1.0	± 0.4	$^{+0.3}_{-0.2}$	13	$^{+5}_{-4}$	± 3
2016 OSDL [1]	0.0	(combined)		0	(combined)	
<i>recalculated</i> [2]	-0.2	$^{+1.7}_{-1.5}$	± 1.5	-2	$^{+20}_{-18}$	± 18

To attain the level of evidence with a search significance of 3 standard deviations or more, several analyses can be combined to reach more statistical sensitivity, provided that they are dominated by statistical uncertainties, by using the entire dataset. Details on the combination shall be discussed in Chapter 5.

Table 4.7: Expected and observed significance from individual analyses.

Analysis	Significance (s.d.)	
	Expected	Observed
Run 2 SL	1.2	1.4
2017+2018 OSDL	0.6	1.8
Run 2 All-hadronic	0.4	2.5
Run 2 SSDL&ML [65]	2.7	2.6
2016 OSDL [1]	0.4	0.0

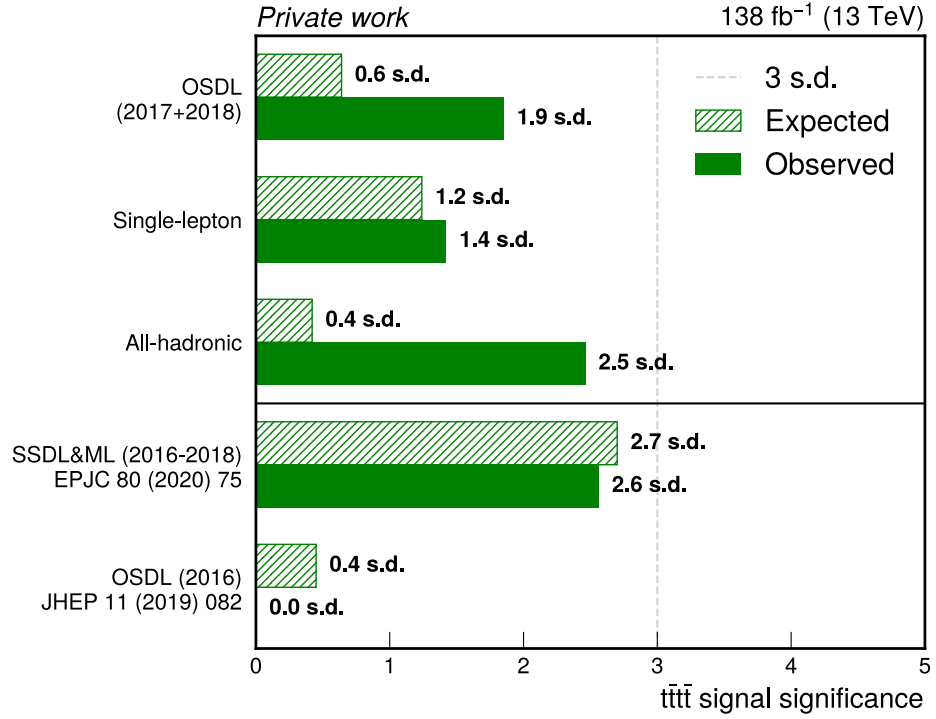


Figure 4.12: Expected and observed significance from individual analyses, derived from Table 4.7.

Chapter 5

COMBINATION OF SEARCH RESULTS FOR FOUR TOP QUARK PRODUCTION

As seen in Chapter 4, every analysis has several differences, such as the choice of the main background, analysis techniques, baseline selection criteria, and systematic uncertainties applied. Still, none of the results has reached the point of evidence or attained a search significance of 3 standard deviations. Even Run 2 same-sign dilepton and multilepton (SSDL&ML) analysis [65] obtained the expected and observed search significance of only 2.7 and 2.6 respectively. By combining search results from different final states together, we may be able to achieve better results in the form of better search significance, hopefully reaching the evidence level. Several benefits like this due to the combination will be discussed in Section 5.1.

In this chapter, we shall also discuss the basic ideas of the statistical tools used in interpreting the results, as well as systematic correlations applied in combining the results of several final states in each analysis as detailed in Chapter 4.

5.1 Why do we combine final states?

There are several ways to increase the search significance. The easiest way is to obtain more data from particle detectors, which works when the systematic uncertainties of the analysis are small compared to statistical uncertainties.

For a comparison, in the Run 2 campaign spanning three years of data taking (2016-2018), data collected by CMS detector amounts to an integrated luminosity¹ of 138 fb^{-1} . The High-Luminosity LHC (HL-LHC) upgrade project, established in 2010, has a target of obtaining 250 fb^{-1} per year, amounting to 3000 fb^{-1} through 12 years of operation [176]. CMS Collaboration has projected, based on the four-top quark production search result in the same-sign dilepton final state on 2016 data (with 35.9 fb^{-1} of data) [79], that we may be able to achieve evidence level with 300 fb^{-1} of data [177]. Even with this projection and HL-LHC upgrade in place, there is no guarantee at all that we will be able to achieve evidence level by waiting for two years' worth of data from the HL-LHC since not all the data delivered by HL-LHC can always be recorded by CMS detector or has enough quality for physics analyses. The projection, however, assumed that the techniques used in reconstructing physics objects (such as b tagging) and in analysis (such as signal and background discrimination) do not improve over larger integrated luminosities. Therefore, improvements beyond this projection are still possible.

Another way to increase search significance is to combine the data from several final states together. This increases the statistics of the search, since by combining two or more final states, the number of events included in the analysis increases, provided that there is orthogonality in each analysis. With this approach, we do not have to wait for more data from the particle detectors to increase the search significance. This approach has also been used several times for the search of four-top quark production within CMS, such as the combination between 2015 SL+OSDL and 2015 SSDL analyses [72; 73], and the combination of 2016 SL+OSDL and 2016 SSDL&ML analyses [79; 1]. Another famous example is the combination of search results for the Higgs boson from the CMS and ATLAS Collaborations, leading to the discovery of the Higgs boson itself [39; 38].

¹This amount of integrated luminosity is based on the latest calculation from the Run 2 campaign [76; 77; 78].

In addition to increased search significance, combining data from several final states together allows us to cross-check between individual results from each final state. In the combination, we expect to measure one final value of signal strength μ , which should apply to all final states at the same time. This can suppress the outliers among all final states, and give us a more stable result compared to results reported in individual final states. However, checks for compatibility among all final states are also required, which will be discussed further in Section 5.6.5.

Another benefit of combination is that it can better constrain systematic uncertainties. Some types of uncertainties can be applied in multiple final states and should bend the distribution in the same direction across all of the final states with them applied. By assigning correlations to systematic uncertainties, we can better constrain the changes applied to the distribution, since we have more constraints due to the same types of uncertainties being applied to more histogram bins.

However, the combination of data from different final states together should be handled carefully, since there are several reasons why the combination can go wrong. First, all the final states used in the combination must be orthogonal to prevent double counting issues where one event can be counted in two or more final states at the same time. Furthermore, several checks should be performed such as negative log-likelihood scans and compatibility checks to make sure that all nuisance parameters have an expected behaviour in the combination and the final measurements from the combination are compatible to each final state.

5.2 Profile likelihood fit

The profile likelihood fit [3] is a statistical formulation used as a gold standard in CMS analyses to interpret and measure signal strength and cross section of individual particle productions. This formulation is used in a statistical program

named `HiggsCombine` [178], which is used within the CMS Collaboration. Note that the formulation presented in this section is abridged from the original work, and a complete mathematical formulation is available in Ref. [3].

The starting idea is simple: given a set of the expectation value of events in each histogram bin $E[n_i]$, and corresponding background events b_i and signal events s_i for each bin, the signal strength μ must satisfy

$$E[n_i] = b_i + \mu s_i \quad (5.1)$$

as best as possible. s_i and b_i can be written as an integral of the probability density functions (PDFs) f_s and f_b as [3]

$$s_i = s_{\text{tot}} \int_{\text{bin } i} dx f_s(x; \theta_s) \quad (5.2)$$

$$b_i = b_{\text{tot}} \int_{\text{bin } i} dx f_b(x; \theta_b) \quad (5.3)$$

where θ_s and θ_b are nuisance parameters that can adjust the shape of PDFs, and s_{tot} and b_{tot} denote the total number of signal and background events respectively. From here, θ_s , θ_b , s_{tot} , and b_{tot} can adjust the shape and normalisation of both signal and background distributions, but in many analyses, we are mainly interested in finding the most probable amount of s_{tot} , which can be dependent to θ_s , θ_b , and b_{tot} . Hence, the latter three parameters can be collectively labelled as $\theta = (\theta_s, \theta_b, b_{\text{tot}})$. s_{tot} , on the other hand, can be set to a nominal value predicted by SM, which is usually done by normalising the signal distribution to SM predicted cross section.

In addition to histogram bins containing background and signal events, we can also add information from a histogram constructed from a control region where no signal events are expected. The expectation value $E[m_i]$ of each bin in that histogram would equal to

$$E[m_i] = u_i(\theta) \quad (5.4)$$

where u_i is a calculable amount of expected events in each bin of the control region, subject to parameters θ .

As seen from the formulation above, the distributions of s_i , b_i , and u_i should match the distributions n_i from regions with signal and m_i from control regions as much as possible. s_i , b_i , and u_i are subject to nuisance parameters θ , and the expectation values of n_i and m_i are also subject to μ in addition to θ . Based on Bayesian statistics, the likelihood function of μ and θ is the product of Poisson probabilities [3]

$$L(\mu, \theta) = \prod_i \frac{(b_i + \mu s_i)^{n_i}}{n_i!} e^{-(b_i + \mu s_i)} \prod_i \frac{u_i^{m_i}}{m_i!} e^{-u_i}. \quad (5.5)$$

Since L depends on μ and θ , we can find the optimal value of L per both μ and θ , which we can call $\hat{\mu}$ and $\hat{\theta}$. Furthermore, for each value of μ , an optimal set of θ can be calculated for that μ value, which we can call $\hat{\theta}$. With the values of L from these two sets of parameters, we can consider the profile likelihood ratio $\lambda(\mu)$ and a test statistics t_μ as [3]

$$\lambda(\mu) = \frac{L(\mu, \hat{\theta})}{L(\hat{\mu}, \hat{\theta})} \quad (5.6)$$

$$t_\mu = -2 \ln \lambda(\mu). \quad (5.7)$$

The quantity t_μ , introduced in Equation 5.7, is also called Negative Log Likelihood (NLL). Since the denominator of $\lambda(\mu)$ is the largest likelihood value derived from the most probable values of μ and θ , $\lambda(\mu)$ will always be less than or equal to 1, leading to $\ln(\mu) \leq 0$ and t_μ being positive. If $\mu = \hat{\mu}$, $\lambda(\mu) = 1$ and $t_\mu = 0$, which means the NLL value will be exactly zero at the most optimal point $\mu = \hat{\mu}$.

To measure signal strength, search significance, and upper limits of the signal strength, we must construct the probability density functions for test statistics, such as t_μ and others that are derived based on t_μ , under specific conditions to obtain the p -value specific to each measurement. These p -values can then be used to rule

out conditions where they fall under the determined threshold, such as confidence level. The construction of these probability density functions may be performed using Monte Carlo simulations, which will not be discussed in this thesis.

Based on a result by Wald [179], given that the number of events included in the histogram is large enough, the test statistic t_μ can be approximated as [3]

$$t_\mu = -2 \ln \lambda(\mu) = \frac{(\mu - \hat{\mu})^2}{\sigma^2} + \mathcal{O}\left(1/\sqrt{N}\right) \quad (5.8)$$

where N is the data sample size, which means the term $\mathcal{O}\left(1/\sqrt{N}\right)$ can be negligible if the sample size, or the number of events in the histogram, is large enough.

As shown in Ref. [3], the actual distributions of test statistics, such as t_μ and all other test statistics derived from t_μ that are used to calculate signal strength measurement, search significance, and upper limits of signal strength, may not follow the distribution derived from Wald's approximation when the sample size in the histogram is not large enough. In this case, the calculation of all quantities can also be calculated from Monte Carlo simulations, without the use of Wald's approximation.

5.2.1 Signal strength measurement

To measure the signal strength μ , we can calculate the p -value of t_μ as [3]

$$p_\mu = 2(1 - \Phi(\sqrt{t_\mu})) \quad (5.9)$$

where Φ represents the cumulative distribution function of a Gaussian distribution, with zero mean and variance of exactly 1. Since t_μ , introduced in Equation 5.7, is always positive, $\Phi(\sqrt{t_\mu})$ is equivalent to calculating a right-tailed p -value of the Gaussian distribution. From here, we can set the confidence interval of μ at the confidence level (CL) of $1 - \alpha$, by setting $p_\mu = \alpha$, to cover the desired range of μ ,

such as one standard deviation range. We can solve Equation 5.9 and obtain

$$t_\mu = [\Phi^{-1}(1 - \alpha/2)]^2. \quad (5.10)$$

This means the NLL value (t_μ) of 1 will give the μ value at the confidence interval of 1 standard deviation, since for 1 standard deviation, $\alpha = 0.3173$, then $t_\mu = [\Phi^{-1}(0.8413)]^2 = 1$. This is usually used in reporting the best fit value of μ . If we want the confidence interval of 95% instead, α will be 0.05, hence $t_\mu = [\Phi^{-1}(0.975)]^2 = 3.84$. Based on this conclusion, we can numerically calculate the NLL value for each value of μ and obtain its best fit interval based on the scanning of NLL values over the range of μ .

To prove that this assumption is valid, we can use Equation 5.8, based on the approximation by Wald [179], and report the signal strength value as [3]

$$\mu = \hat{\mu} \pm \sigma \Phi^{-1}(1 - \alpha/2) \quad (5.11)$$

$$= \hat{\mu} \pm \sigma \sqrt{t_\mu} \quad (5.12)$$

which means $t_\mu = 1$ will correspond to the confidence interval of $\sqrt{t_\mu} = 1$ standard deviation.

5.2.2 Search significance

The search significance Z is defined as [3]

$$Z = \Phi^{-1}(1 - p) \quad (5.13)$$

where p represents the upper-tail p -value, and Φ^{-1} is the inverse of the cumulative distribution function of a Gaussian distribution. Based on this definition, the search significance of 3 standard deviations (*evidence*) and 5 standard deviations (*observation*) are equivalent to the p -values of

$$p(3\sigma) = 1 - \Phi(Z = 3) = 0.00135 = 1 \text{ in } 740 \quad (5.14)$$

$$p(5\sigma) = 1 - \Phi(Z = 5) = 2.867 \times 10^{-7} = 1 \text{ in } 3.5 \text{ million.} \quad (5.15)$$

The search significance can be calculated based on a special case of test statistics t_μ [3].

$$q_0 = \begin{cases} -2 \ln \lambda(0) & \text{if } \hat{\mu} \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (5.16)$$

These test statistics can be used to reject the hypothesis where no signal is present in the data ($\mu = 0$). Based on Wald's approximation in Equation 5.8, we obtain [3]

$$q_0 = \begin{cases} \hat{\mu}^2 / \sigma^2 & \text{if } \hat{\mu} \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (5.17)$$

The cumulative distribution of this test statistics, with the hypothesis where no signal is present, is [3]

$$F(q_0|0) = \Phi(\sqrt{q_0}). \quad (5.18)$$

The p -value for the hypothesis is [3]

$$p_0 = 1 - F(q_0|0) \quad (5.19)$$

which leads to the significance of [3]

$$Z_0 = \Phi^{-1}(1 - p_0) = \sqrt{q_0} = \mu / \sigma \quad (5.20)$$

Since the final result only contains μ and σ , which can be obtained from the best fit of signal strength value along with its confidence interval, we can make a quick approximation of the search significance based on the best fit of signal strength and vice versa.

5.2.3 Asimov dataset

In some calculations of expected quantities, such as search significance, we can calculate it using a large set of randomised distributions based on simulated samples called *random toys*. However, generating random toys to calculate the quantity can

be time-consuming, especially with a very large combination of the search. An alternative is to use Asimov datasets [3], which is simply a dataset generated by copying distributions from simulated background samples and adding a distribution from the simulated signal sample (usually with $\mu = 1$ to mimic the number of signal events predicted in SM).

The name of the Asimov dataset is inspired by Isaac Asimov’s short science fiction *Franchise*, first published in 1955 [180; 3]. In the story, the protagonist Norman Muller is chosen from 200 million Americans by a computer to represent the whole population of Americans in a presidential election. Below is an excerpt from the story:

“All right,” said Handley, “let’s get that straight to begin with. Multivac weighs all sorts of known factors, billions of them. One factor isn’t known, though, and won’t be known for a long time. That’s the reaction pattern of the human mind. All Americans are subjected to molding pressure of what other Americans do and say, to the things that are done to him and the things he does to others. Any American can be brought to Multivac to have the bent of his mind surveyed. From that the bent of all other minds in the country can be estimated. Some Americans are better for the purpose than others at some given time, depending upon the happenings of that year. **Multivac picked you as most representative this year. Not the smartest, or the strongest, or the luckiest, but just the most representative. Now we don’t question Multivac, do we?**” [180]

Asimov datasets can be used to calculate expected search significance, upper limits of signal strength, and best fit value of signal strength. As the Asimov dataset

is formulated with the fixed μ value, the best fit value of signal strength will always be at the fixed value. The search significance, however, can be impacted by its variance, which is affected by nuisance parameters present in the analysis. The calculation of signal strength best fit and search significance, often called *expected signal strength best fit* and *expected search significance* can be a good indication of how powerful the search is. Based on Equation 5.20, if the expected signal strength best fit can be measured with less uncertainty, the expected search significance increases.

On the other hand, the *observed signal strength best fit* and *observed search significance* are calculated based on data, which is compared directly against simulation and can be vastly different from results obtained with the Asimov dataset, since the data used to fit the signal strength may be completely different than Asimov dataset. For example, the search for four-top quark production performed by ATLAS Collaboration [82] on single-lepton (SL), opposite-sign dilepton (OSDL), same-sign dilepton (SSDL), and multilepton (ML) final states, using Run 2 data, has obtained the expected significance of 2.6 but the observed search significance is 4.3.

5.3 Systematic uncertainties

In every physics experiment, there are two types of systematic uncertainties: statistical uncertainties and systematic uncertainties. Statistical uncertainties are random fluctuations limited by the amount of data collected during experiments. This type of uncertainty can be reduced by taking more data and reducing the fluctuations. On the other hand, systematic uncertainties arise from several components within the analysis, such as from theory and simulation, experiment, and reconstruction. Understanding the source of these uncertainties, such as in the

reconstruction step, will help us in reducing them. In this section, several types of systematic uncertainties applied in the combination shall be discussed.

5.3.1 Types of systematic uncertainties, mathematically speaking

Two main types of systematic uncertainties are often used in statistical inferences in the CMS Collaboration.

- **Log-normal uncertainties** adjust the number of events with the same factor in all histogram bins applied. It is usually defined in the form of $(1 + \sigma)^\theta$, where σ is the size of uncertainty at 1 standard deviation and θ is the nuisance parameter associated with this type of uncertainty. The reason why this formulation of uncertainty is used is to keep the number of events in a histogram bin positive, even if θ becomes highly negative.
- **Shape uncertainties** adjust the number of events with *different* factors for each histogram bin. This type of uncertainty must be accompanied by two histogram templates denoting the $+1\sigma$ and -1σ variant of the uncertainty. The fractional change $f(\theta)$ of each histogram bin can be interpolated with the nuisance parameter value θ for $-1 \leq \theta \leq 1$ as [178]

$$f(\theta) = \frac{1}{2} \left((\delta^+ - \delta^-)\theta + \frac{1}{8}(\delta^+ + \delta^-)(3\theta^6 - 10\theta^4 + 15\theta^2) \right) \quad (5.21)$$

where δ^+ and δ^- are the fractional changes of each histogram bin for $+1$ and -1 sigma respectively, and for $|\theta| > 1$,

$$f(\theta) = \begin{cases} \theta\delta^+ & ; \theta > 1 \\ \theta\delta^- & ; \theta < -1. \end{cases} \quad (5.22)$$

The formulations above are designed so that $f(\theta)$ is continuous in its value and derivatives at $\theta = \pm 1$.

5.3.2 Correlating systematic uncertainties

Usually, systematic uncertainties can change the event yield for each process with nuisance parameters θ . One nuisance parameter represents the effect of one systematic uncertainty. Correlating two or more uncertainties (or correlating two or more nuisance parameters) means forcing these nuisance parameters to have the same up/down effect. By correlating uncertainties together, we are making assumptions that the same type of uncertainties should have the same effect across different processes and final states.

One limitation of the tool used in this work for statistical inferences is that we cannot assume partial correlations to the nuisance parameters. The usual procedure is to break down the nuisance parameter into two or more smaller nuisance parameters, which correspond to the correlated and uncorrelated parts. Correlations to the first part are still treated as full correlations.

In this combination, the correlations applied are full correlations between different nuisance parameters. For this type of correlation, several nuisance parameters from different analyses are replaced with one nuisance parameter. With this modification, one nuisance parameter will cause the same effect across different analyses with the same set of nuisance parameters as intended.

5.4 Systematic uncertainties considered in combination

As stated in Section 4.1, the combination of searches for four-top quark production includes

- Run 2 single-lepton (Section 4.3)

- 2016 opposite-sign dilepton (Section 4.2.1)
- 2017+2018 opposite-sign dilepton (Section 4.4)
- Run 2 same-sign dilepton (Section 4.2.2)
- Run 2 all-hadronic (Section 4.5)

In this combination, several types of uncertainties from different analyses must be correlated properly, some correlated between processes, some correlated between years, and some correlated between both processes and years. In this subsection, each type of systematic uncertainty will be explained, along with the correlation scheme applied in this combination. Tables 5.1 and 5.2 give a summary of uncertainty correlations applied. This section will discuss details on theoretical, experimental, and reconstruction uncertainties that are common across different analyses in Sections 5.4.1, 5.4.2, and 5.4.3.

5.4.1 Theoretical uncertainties

5.4.1.1 $t\bar{t} + b\bar{b}$ normalisation

$t\bar{t} + b\bar{b}$ normalisation is correlated across years in the OSDL, SL, and SSDL&ML final states. This uncertainty does not apply to the all-hadronic final state.

The $t\bar{t} + b\bar{b}$ normalisation uncertainty is the uncertainty obtained from the *measurement* of $t\bar{t} + b\bar{b}$ fraction in $t\bar{t}$ production, performed by CMS experiment [166]. Since this type of uncertainty is reported by previous analyses [1; 65] to be the most impactful in four-top quark cross section measurement, this uncertainty is included in 2017+2018 OSDL and Run 2 SL analyses, and the correlation of this type of uncertainty is strongly considered. The correlation is applied across years and in OSDL, SL, and SSDL&ML final states.

Table 5.1: Summary table on nuisance parameter correlation assumptions, for theoretical uncertainties.

Uncertainty	2016 SL	2016 OSDL	2017 SL	2017 OSDL	2018 SL	2018 OSDL	SSDL&ML	all-hadronic
$t\bar{t} + b\bar{b}$ normalisation, Section 5.4.1.1								
	✓	✓	✓	✓	✓	✓	✓	
Monte Carlo cross section, Section 5.4.1.2								
Electroweak	✓	✓	✓	✓	✓	✓		
Single-top	✓	✓	✓		✓			
$t\bar{t}$		✓		✓		✓		
$t\bar{t}H$	✓		✓	✓	✓	✓	✓	✓
$t\bar{t} + V$				✓		✓		
$t\bar{t} + \text{ultra rare}$				✓		✓		
ISR and FSR, Section 5.4.1.3								
$t\bar{t} + b\bar{b}$	✓		✓	✓	✓	✓		
$t\bar{t} + \text{light jets}$	✓		✓	✓	✓	✓		
$t\bar{t}H$	✓		✓	✓	✓	✓		
$t\bar{t}t\bar{t}$	✓	✓	✓	✓	✓	✓		✓
PDF, Section 5.4.1.4								
	✓	✓	✓	✓	✓	✓	✓	✓

Table 5.2: Summary table on nuisance parameter correlation assumptions, for experimental uncertainties.

Uncertainty	2016 SL	2016 OSDL	2017 SL	2017 OSDL	2018 SL	2018 OSDL	SSDL&ML	all-hadronic
Luminosity, Section 5.4.2.1								
2016	✓	✓					✓	✓
2017			✓	✓			✓	✓
2018					✓	✓	✓	✓
2017+2018			✓	✓	✓	✓	✓	✓
2016+2017+2018	✓	✓	✓	✓	✓	✓	✓	✓
Prefire, Section 5.4.2.2								
2016	✓						✓	✓
2017			✓	✓			✓	✓
Pileup, Section 5.4.2.3								
	✓	✓	✓	✓	✓	✓	✓	✓

Table 5.3: Summary table on nuisance parameter correlation assumptions, for reconstruction uncertainties.

Uncertainty	2016 SL	2016 OSDL	2017 SL	2017 OSDL	2018 SL	2018 OSDL	SSDL&ML	all-hadronic
Jet energy scale (JES), Section 5.4.3.1								
Absolute	✓		✓	✓	✓	✓		
BB/EC1	✓		✓	✓	✓	✓		
EC2	✓		✓	✓	✓	✓		
FlavorQCD	✓		✓	✓	✓	✓		
HF	✓		✓	✓	✓	✓		
RelativeBal	✓		✓	✓	✓	✓		
Absolute '17			✓	✓				
BB/EC1 '17			✓	✓				
EC2 '17			✓	✓				
HF '17			✓	✓				
RelativeSample '17			✓	✓				
Absolute '18					✓	✓		
BB/EC1 '18					✓	✓		
EC2 '18					✓	✓		
HF '18					✓	✓		
RelativeSample '18					✓	✓		
Jet energy resolution (JER), Section 5.4.3.2								
2016	✓	✓					✓	✓
2017			✓	✓			✓	✓
2018					✓	✓	✓	✓
b tagging (DeepJet), Section 5.4.3.3								
purity				✓		✓		✓
charm-flavoured				✓		✓		✓
heavy flavour				✓		✓		✓
light flavour				✓		✓		✓
HF stat. 2017				✓				✓
LF stat. 2017				✓				✓
HF stat. 2018						✓		✓
LF stat. 2018						✓		✓
Lepton scale factor, Section 5.4.3.4								
2016 e	✓	✓						
2017 e			✓	✓				
2018 e					✓	✓		
2016 μ	✓	✓						
2017 μ			✓	✓				
2018 μ					✓	✓		

However, the Run 2 all-hadronic analysis derives most of its backgrounds from a data-driven approach (see Section 4.5), this type of uncertainty is not applied in the all-hadronic final state and thus not included in the correlation.

5.4.1.2 Monte Carlo cross section

Monte Carlo cross section uncertainties are correlated across years and in different analyses. Since it is applied separately for each background process, it is not correlated between processes.

Monte Carlo cross section uncertainties are obtained *based on theoretical calculations*, as opposed to $t\bar{t}b\bar{b}$ normalisation uncertainty explained in Section 5.4.1.1. In this combination, four correlations of this type of uncertainty are applied across different analyses as follows:

- **Electroweak:** Run 2 OSDL + Run 2 SL
- **Single-top production:** 2016 OSDL + Run 2 SL
- $t\bar{t}$: Run 2 OSDL
- $t\bar{t}H$: Run 2 SL + 2017+2018 ODSL + Run 2 SSDL&ML + Run 2 all-hadronic

5.4.1.3 Initial and final state radiation (ISR/FSR)

This type of uncertainty is correlated across years (where applicable), separated between $t\bar{t} + b\bar{b}$, $t\bar{t} + \text{light jets}$, $t\bar{t} + H$, and $t\bar{t}t\bar{t}$. They are applied in Run 2 SL, 2017+2018 OSDL, and Run 2 all-hadronic as detailed below.

The initial and final state radiations (ISR and FSR) refer to the radiations of partons giving extra particles, such as gluons, *before* or *after* the hard process. This can result in higher jet multiplicity and less energy from final state particles detected in an event.

5.4.1.4 Parton Distribution Function (PDF)

All nuisance parameters for this type of uncertainty are correlated together across years, processes, and final states.

A parton distribution function, or PDF, is a set of distribution functions determining the density of *partons*, or quarks and gluons, for momentum fraction x and the virtuality scale Q^2 for each parton. This is used to calculate the cross section for a process theoretically [181], and they are determined via several analyses from experimental data [17]. This type of uncertainty is calculated from different variations of the PDF set provided by the PDF4LHC working group [182; 151].

5.4.1.5 Renormalisation and factorisation scale variations μ_R and μ_F

Due to different treatment schemes of this uncertainty type in each analysis, nuisance parameters belonging to this type of uncertainty are incompatible across different analyses. Thus, correlations cannot be performed on this type of uncertainty.

The renormalisation (μ_R) and factorisation (μ_F) scale parameters, sometimes also called matrix element scale variations, are used in QCD calculations of heavy quark (top and bottom quarks, and sometimes charm is also included) productions. The renormalisation parameter μ_R affects directly the strong coupling α_s , which directly affects the interactions of gluons. On the other hand, the factorisation parameter μ_F affects the distribution functions for quarks and gluons [183]. They can be used to calculate the generator weights of each simulated event. To determine the uncertainty caused by these two scale variations, each factor can be scaled up by a factor of 2 or down by a factor of 0.5. Six recommended combinations can be done as follows:

- Fixed μ_R , vary μ_F by 2 or 0.5
- Fixed μ_F , vary μ_R by 2 or 0.5
- Simultaneously vary μ_R and μ_F by 2 or 0.5

In the Run 2 SL analysis, this type of uncertainty is calculated from the maximum variation out of the six variations described above *for each histogram bin*. This results in one nuisance parameter per process, which is a combination of all six variations of μ_R and μ_F [163].

In the 2017+2018 OSDL analysis, however, this type of uncertainty is calculated separately for each up/down variation as described above. This results in *three* nuisance parameters per process, where each one represents variations in μ_R (fixed μ_F), μ_F (fixed μ_R), and both of them simultaneously [172].

Given the different treatment schemes of this uncertainty type, correlating nuisance parameters for this uncertainty type is unfeasible. Therefore, this type of uncertainty is not correlated.

5.4.2 Experimental uncertainties

5.4.2.1 Luminosity uncertainty

Luminosity uncertainties are correlated across all processes and applied differently across years as detailed below. This uncertainty applies to all final states.

The integrated luminosity recorded by the CMS detector is measured by year in 2016 [76], 2017 [77], and 2018 [78]. In addition to uncertainties measured per year, correlated uncertainties must also be taken into account in combination with data from 2016-2018. Luminosity uncertainties for 2016-2018, both uncorrelated

Table 5.4: Luminosity uncertainty correlation scheme for Run 2 combination, as recommended by Luminosity Physics Object Group.

Uncertainty	Analysis year		
	2016	2017	2018
Uncorrelated 2016	1.0		
Uncorrelated 2017		2.0	
Uncorrelated 2018			1.5
Correlated 2017+2018		0.6	0.2
Correlated Run 2	0.6	0.9	2.0

Table 5.5: Integrated luminosity from the latest measurements performed in Ref. [76; 77; 78], compared to integrated luminosity values used in 2016 OSDL and Run 2 SSDL&ML analyses.

	Integrated luminosity (fb^{-1})			
	2016	2017	2018	2016-2018
Latest measured luminosity	36.31	41.53	59.74	137.58
2016 OSDL	35.83			
Run 2 SSDL&ML	35.922	41.53	59.71	

and correlated, are shown in Table 5.4. The new luminosity uncertainty scheme is applied to new analyses, namely Run 2 SL and 2017+2018 OSDL.

In addition to the new uncertainty scheme applied to new analyses, integrated luminosity and luminosity uncertainties must also be updated in 2016 OSDL and Run 2 SSDL&ML, shown in Table 5.5. Since the Run 2 SSDL&ML analysis does not separate the histogram template into years, a new luminosity uncertainty scheme must be applied according to Table 5.6, where five new nuisance parameters representing uncorrelated and correlated parts are introduced.

5.4.2.2 Level-1 trigger prefiring

This type of uncertainty is correlated across different processes, separated by year.

Table 5.6: New luminosity uncertainty scheme applied to Run 2 SSDL&ML analysis, based on integrated luminosity value in Table 5.5.

Uncertainty	Size
Uncorrelated 2016	$1\% \times 36.31 \div 137.58 = 0.26\%$
Uncorrelated 2017	$2\% \times 41.53 \div 137.58 = 0.60\%$
Uncorrelated 2018	$1.5\% \times 59.74 \div 137.58 = 0.65\%$
Correlated 2017+2018	$(0.6\% \times 41.53 + 0.2\% \times 59.74) \div 137.58 = 0.27\%$
Correlated Run 2	$(0.6\% \times 36.31 + 0.9\% \times 41.53 + 2\% \times 59.74) \div 137.58 = 0.60\%$

During the collection of data from the CMS detector in 2016-2017, a fault in the electromagnetic calorimeter in the region of $|\eta| > 2.5$ caused an energy deposit to be incorrectly assigned to a bunch crossing that happened earlier than the bunch crossing of interest [108]. This issue affects data collected in 2016 and 2017, so analyses using data from both years are affected by this type of systematic uncertainty, which are 2016+2017 SL, 2017 OSDL, 2016+2017 all-hadronic, and 2016+2017 SSDL&ML. Unfortunately, the 2016 OSDL analysis does not have this uncertainty included [1], so prefiring uncertainty is not correlated in that analysis.

5.4.2.3 Pileup

All nuisance parameters for this type of uncertainty are correlated together across years, processes, and final states.

To increase the number of particle collisions to be as high as possible, the LHC has been designed to deliver a very high amount of instantaneous luminosity in the order of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ [96] and has achieved such a high luminosity with $O(10^{10})$ protons in each proton bunch during Run 2 [76; 77; 78]. However, one side effect of this high luminosity is that more than one proton-proton collision may occur in one collision event. This effect is called pileup, which is mainly caused by colliding two bunches of protons together, each with a very large number of protons

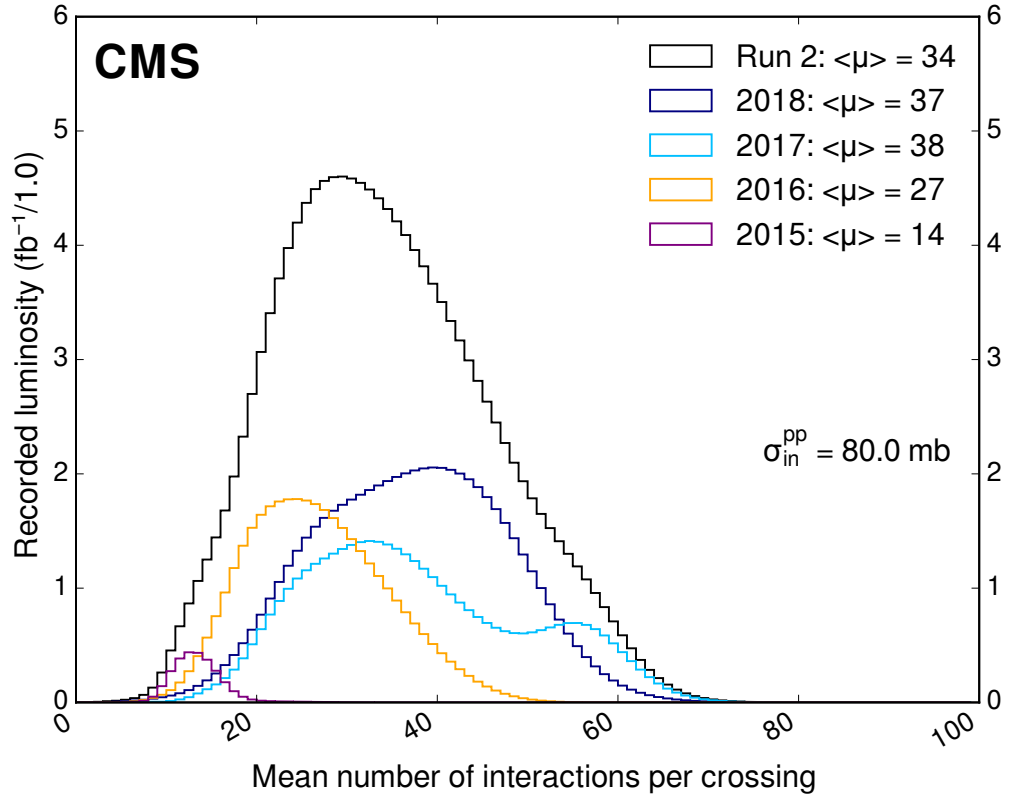


Figure 5.1: Distributions of the mean number of interactions per crossing for proton-proton collisions during Run 2 data-taking period between 2015-2018 [184].

to increase the possibility of the collision. Figure 5.1 shows the distribution of the mean number of interactions per bunch crossing during Run 2.

According to the recommendation by the Top Physics Analysis Group in CMS, this uncertainty should be correlated across years, processes, and final states. Analyses affected by this type of uncertainty are Run 2 SL, 2016 OSDL, 2017+2018 OSDL, Run 2 SSDL&ML, and Run 2 all-hadronic.

5.4.3 Reconstruction uncertainties

5.4.3.1 Jet Energy Scale (JES)

This type of uncertainty is correlated across SL and OSDL analyses, except 2016 OSDL, and separated by sources detailed below.

Jet Energy Scale uncertainty is an uncertainty calculated from the calibrations made to jet energies [88]. This type of uncertainty may be called Jet Energy Corrections (JEC) in some literature. Since this correction may alter the jet energy, there may be more jets being included or excluded from the preselection criteria, thus causing collision events to be assigned into different jet multiplicity categories when JES uncertainty is applied. This effect is sometimes called event migration. Since events may migrate to different jet multiplicity categories when JES uncertainty is applied, evaluating the effect of this type of uncertainty requires the simulation of this effect at the reconstruction level.

This type of uncertainty is applied differently in several analyses. In Run 2 SL and 2017+2018 OSDL analyses, the following sources of JES uncertainties are applied:

- Flat absolute scale p_T dependence uncertainties extracted from the lepton momentum scale for muons in $Z \rightarrow \mu\mu + \text{jet}$ and single-pion response in HCAL [88]. This uncertainty has both uncorrelated parts per year and a correlated part across three years.
- BB/EC1 uncertainties obtained from detector regions $|\eta| < 2.5$. BB refers to regions of $|\eta| < 1.3$, and EC1 refers to regions of $1.3 < |\eta| < 2.5$ [88]. This uncertainty has both uncorrelated parts per year and a correlated part across three years.

- EC2 uncertainties obtained from EC2 region, defined as $2.5 < |\eta| < 3.0$ [88]. This uncertainty has both uncorrelated parts per year and a correlated part across three years.
- HF uncertainties obtained from HF region, defined as $|\eta| > 3.0$ [88]. This uncertainty has both uncorrelated parts per year and a correlated part across three years.
- Uncertainties due to differences between log-linear fits of MPF and p_T balance methods.
- Flavour response differences uncertainties calculated from differences in simulations from PYTHIA 6.4 [185] and HERWIG++ [121] MC generators. This type of uncertainty only has correlated part across three years.

However, in the 2016 OSDL analysis, this type of uncertainty is applied differently, where separate sources of uncertainties are combined into a shorter list of uncertainties as follows [88]:

- Pileup offset uncertainties
- Relative η correction uncertainties
- Flat absolute scale p_T dependence uncertainties
- Scale uncertainties independent of p_T , with statistical uncertainty included
- Jet energy corrections time dependence uncertainty between different epochs of data-taking

Furthermore, in Run 2 SSDL&ML and Run 2 all-hadronic analyses, this type of uncertainty is combined as one nuisance parameter per year, which could not be correlated to nuisance parameters of the same type in other final states.

Table 5.7: b-tagging algorithms used in each analysis. Numbers in parenthesis denote corresponding section numbers in Chapter 4.

Analysis	CSV [155]	DeepCSV [162]	DeepJet [113]
Run 2 SL (4.3)		✓	
2016 OSDL (4.2.1)	✓		
2017+2018 OSDL (4.4)			✓
Run 2 SSDL&ML (4.2.2)		✓	
Run 2 all-hadronic (4.5)			✓

5.4.3.2 Jet Energy Resolution (JER)

This type of uncertainty is correlated across different processes but in different years. They are applied in all final states throughout Run 2.

The calibration of jet energy resolution is performed at the same time as the jet energy scale, and can also give systematic uncertainties to the analysis. This type of uncertainty is applied to all final states differently across three years of Run 2 and should be correlated per year across different processes.

5.4.3.3 b tagging

This type of uncertainty is correlated between analyses using DeepJet b tagger [113], namely 2017+2018 OSDL and Run 2 all-hadronic. The correlation is applied differently based on the source as described below.

Every b-tagging algorithm, such as Combined Secondary Vertices (CSV) [155], DeepCSV [162], and DeepJet [113], has its uncertainties in classifying jet flavours. Table 5.7 show the list of b-tagging algorithms used in Run 2 SL (Section 4.3), 2016 OSDL (Section 4.2.1), 2017+2018 OSDL (Section 4.4), Run 2 SSDL&ML (Section 4.2.2), and Run 2 all-hadronic (Section 4.5).

As seen from the table, the DeepCSV b-tagging algorithm is used in both Run 2 SL and Run 2 SSDL&ML analyses, and the DeepJet b-tagging algorithm is used in both 2017+2018 OSDL and Run 2 all-hadronic analyses. However, due to a different treatment of this type of systematic uncertainty between Run 2 SL and Run 2 SSDL&ML analyses, nuisance parameters of DeepCSV from both analyses are incompatible and thus could not be correlated. DeepJet nuisance parameters from the other two analyses can still be correlated with the following components:

- Purity uncertainties, correlated across years with the same type of uncertainty
- Charm-flavoured uncertainties, correlated across years with the same type of uncertainty
- Heavy-flavoured and light-flavoured statistical uncertainties, correlated separately by year within the same type of uncertainty

5.4.3.4 Lepton scale factors

This type of uncertainty is correlated across different processes, separated by year and lepton flavour.

This type of uncertainty takes into account lepton reconstruction, including identification and isolation, and trigger efficiencies. They are different for muons and electrons and thus are separated for both lepton flavours.

5.5 Applying theoretical uncertainties for signal strength and cross section measurement

As seen in Section 5.3.2, several theoretical uncertainties are applied to background and signal processes. Since the histogram templates have the distribution of the

signal process fixed at 12 fb [64] but we want to measure the cross section of the signal process, we should not add theoretical uncertainties that will change the normalisation of the signal process (as known as *rate-changing uncertainties*). With this conclusion, theoretical uncertainty treatments for signal strength and cross section measurements are different as follows:

- **Signal strength:** all theoretical uncertainties applied, including rate-changing uncertainties for the signal process
- **Cross section:** all theoretical uncertainties *except rate-changing uncertainties for signal process* applied

As a result of this difference, the reported value of signal strength and cross section may differ due to the different uncertainty treatment schemes, especially in the final combination.

5.6 Combination results

After applying the correlations to nuisance parameters as detailed in Section 5.4, the best fit values of signal strength and cross section are calculated, both in expected (calculated from Asimov datasets, see Section 5.2.3) and observed values. To make sure that the results are stable and nuisance parameters have expected behaviours, additional statistical analyses are performed, which will be discussed in this section.

5.6.1 Expected signal strength, cross section, and search significance

In this section, the *expected* signal strength and search significance will be presented for reduced sets of the combination first, leading up to the final combination with

Table 5.8: Expected signal strength and search significance for individual analysis and the combination. Results from 2016 OSDL [1] have been recalculated to allow negative signal strength values and adjust the four-top quark production cross section up to 12 fb from 9.2 fb (See Section 4.2.1.5).

Analysis	Signal strength (μ)	Significance (s.d.)
Run 2 SL	$1^{+0.9}_{-0.8}$	1.2
2017+2018 OSDL	$1^{+1.8}_{-1.6}$	0.6
Run 2 all-hadronic	$1^{+2.5}_{-2.4}$	0.4
Combination of above	1 ± 0.7	1.5
Run 2 SSDL&ML [65]	1 ± 0.4	2.7
2016 OSDL [1; 2]	$1^{+2.6}_{-2.2}$	0.4
Full combination	$1^{+0.4}_{-0.3}$	3.2

all analyses included. As these quantities are calculated using the Asimov dataset (constructed based on the expected number of events with $\mu = 1$), signal strength and cross section values will centre around 1 and 12 fb respectively, since the Asimov dataset is based on SM prediction. Table 5.8 shows the expected signal strength and search significance from each analysis, as well as the combination of new results and full combination.

As seen from Table 5.8, combining the results between three years of data taking greatly, and combining all search results available from new analyses increases the expected search significance up to 1.5. This indicates that by improving analysis techniques and including orthogonal final states together, we can gain higher search significance than what we have achieved in previous analyses which are limited by old analysis methods and a small amount of data. As the Run 2 SSDL&ML analysis [65] has the expected significance of 2.7, this analysis boosts the combined significance (calculated from all final states) up to 3.2 standard deviations. However, the result from Run 2 SSDL&ML analysis alone does not reach the evidence level while the combination with other final states can. This affirms the rationale for performing the combination.

Table 5.9: Observed signal strength and search significance for individual analysis and the combination. Results from 2016 OSDL [1] have been recalculated to allow negative signal strength values and adjust four-top quark production cross section up to 12 fb from 9.2 fb (See Section 4.2.1.5).

Analysis	Signal strength (μ)	Significance (s.d.)
Run 2 SL	1.2 ± 0.9	1.4
2017+2018 OSDL	$2.7^{+2.2}_{-1.5}$	1.8
Run 2 all-hadronic	$5.8^{+2.5}_{-2.4}$	2.5
Combination of above	2.5 ± 0.7	3.9
Run 2 SSDL&ML [65]	$1.0^{+0.5}_{-0.4}$	2.6
2016 OSDL [1; 2]	$-0.2^{+2.2}_{-2.1}$	0.0
Full combination	1.4 ± 0.4	4.0

Although this result does not reflect the actual search results performed on data (which will be discussed in Section 5.6.2), the expected search significance achieved by the full combination is already very impressive. This is because it already broke the evidence level of 3 standard deviations, with the probability of 1 in 1455 that the results are statistical fluke, which is predominantly contributed by Run 2 SSDL&ML analysis with an expected search significance of 2.7.

5.6.2 Observed signal strength, cross section, and search significance

Unlike the expected results reported in Section 5.6.1, observed signal strength and search significance values are obtained from the fit to actual data, thus it can be different from the expected results obtained from the Asimov dataset. Tables 5.9 and 5.10 show observed search significance, signal strength, and cross section from each analysis, as well as the combination of new results and full combination.

Since the observed search significance from the all-hadronic channel is very high compared to the expected search significance of 0.4, the analysis boosts the observed search significance of new results up to 3.9, and the significance of the full

Table 5.10: Observed signal strength and cross section from individual analyses and combinations, with uncertainties broken down into statistical and systematic uncertainties. The uncertainty values reported are statistical uncertainties and systematic uncertainties respectively. The values for the 2016 OSDL analysis in this table have been recalculated to allow negative signal strength values and adjust four-top quark production cross section up to 12 fb from 9.2 fb (See Section 4.2.1.5).

Analysis	Signal strength (μ)			Cross section (fb)		
	(stat.)	(syst.)		(stat.)	(syst.)	
Run 2 SL	1.2	$^{+0.7}_{-0.6}$	± 0.6	15	± 8	$^{+10}_{-7}$
2017+2018 OSDL	2.8	± 1.0	$^{+1.9}_{-1.2}$	33	± 12	$^{+15}_{-14}$
Run 2 All-hadronic	5.8	± 1.4	± 2.0	70	± 17	$^{+25}_{-23}$
Combination of above	2.5	± 0.5	± 0.5	36	± 7	$^{+10}_{-8}$
Run 2 SSDL&ML [65]	1.0	± 0.4	$^{+0.3}_{-0.2}$	13	$^{+5}_{-4}$	± 3
2016 OSDL [1; 2]	-0.2	$^{+1.7}_{-1.5}$	± 1.5	-2	$^{+20}_{-18}$	± 18
Full combination	1.4	± 0.3	± 0.2	17	± 4	± 3

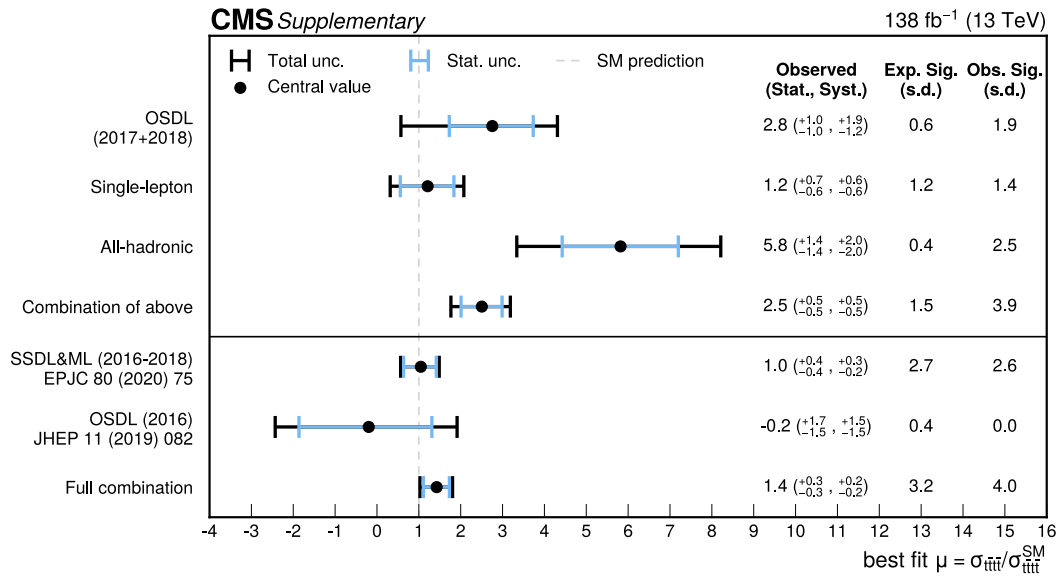


Figure 5.2: Observed signal strength measurement from individual analyses and the combination, with uncertainties broken down into statistical and systematic uncertainties [2].

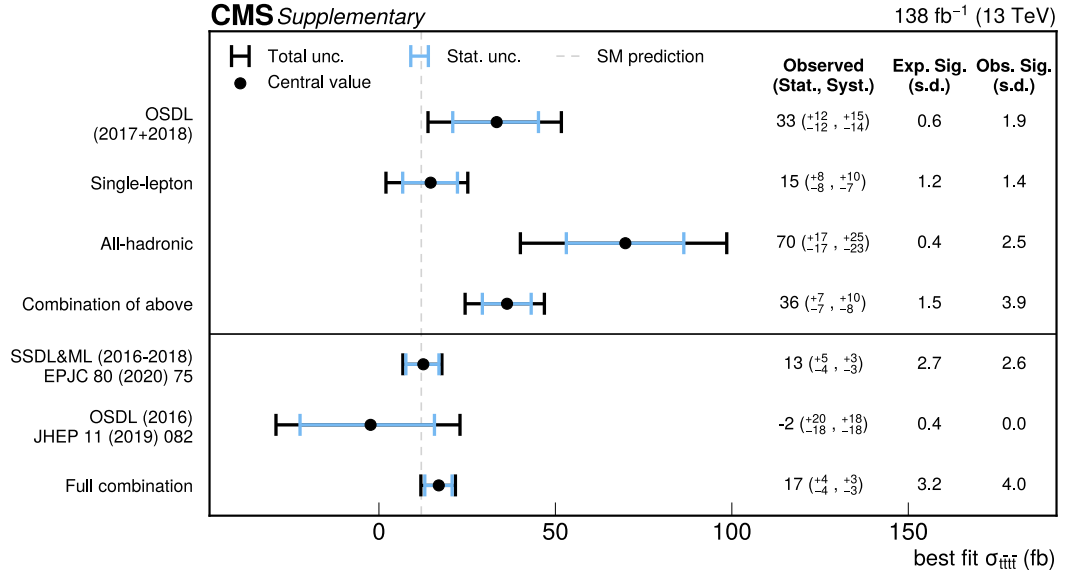


Figure 5.3: Observed cross section measurement from individual analyses and the combination, with uncertainties broken down into statistical and systematic uncertainties [2].

combination up to 4.0. This level of observed significance equates to the probability of 1 in 35 855 that the results are a statistical fluke.

Furthermore, the observed cross section measurement obtained from the full combination in Table 5.10 is 17 ± 4 (stat.) ± 3 (syst.) fb. This result is consistent, within uncertainties, to SM predictions of four-top quark production cross section, which is $12.0^{+2.2}_{-2.5}$ fb with NLO precision and Electroweak corrections [64] at the time of analysis preparation, and more recently $13.4^{+1.0}_{-1.8}$ fb with next-to-leading-logarithmic (NLL') accuracy [63].

Since the observed significance reported in Table 5.9 is calculated based on Wald's approximation in Equation 5.8 [179], one can argue that these results should be based on a high number of events included. To prove that these results, assuming asymptotic approximation, are still valid, the observed significance for Run 2

Table 5.11: Observed significance calculated from asymptotic limit method and generated toys, from 2017+2018 OSDL, Run 2 SL, and Run 2 all-hadronic analyses.

	Asymptotic limit	Generated toys
2017+2018 OSDL	1.8	1.77 ± 0.05
Run 2 SL	1.4	1.39 ± 0.04
Run 2 all-hadronic	2.5	2.38 ± 0.02

single-lepton, 2017+2018 opposite-sign dilepton, and Run 2 all-hadronic analyses are also calculated by generating *toy distributions* using Monte Carlo. Unlike Asimov datasets, briefly discussed in Section 5.2.3, toy distributions are generated with random values around the nominal values for all nuisance parameters [178].

As shown in Table 5.11, the observed significance results reported in Table 5.9, obtained using asymptotic limits, agree with the results obtained using generated toys. Since the observed significance from these analyses is consistent, the observed significance for the full combination, obtained using asymptotic limits, should be valid as well.

5.6.3 Negative log-likelihood scans

As explained in Section 5.2.1, the value of negative log-likelihood, or NLL, can be used to measure the signal strength value to cover a certain confidence level, such as 68% or one standard deviation. Since the quantity can be calculated per each value of signal strength μ , we can repeatedly calculate NLL per each value of μ and plot them as NLL scans. The scan can be performed to inspect the behaviour of NLL values in the fit and see if the measurement of signal strength is as expected.

The NLL scan plot can reveal the best fit value (with the lowest value of NLL) and the uncertainty within the desired confidence level. As shown in Section 5.2.1, the NLL value of 1 covers 68% CL, while the value of 3.84 covers 95% CL. Figure 5.4 shows NLL scans derived after nuisance parameter fitting (post-fit) for

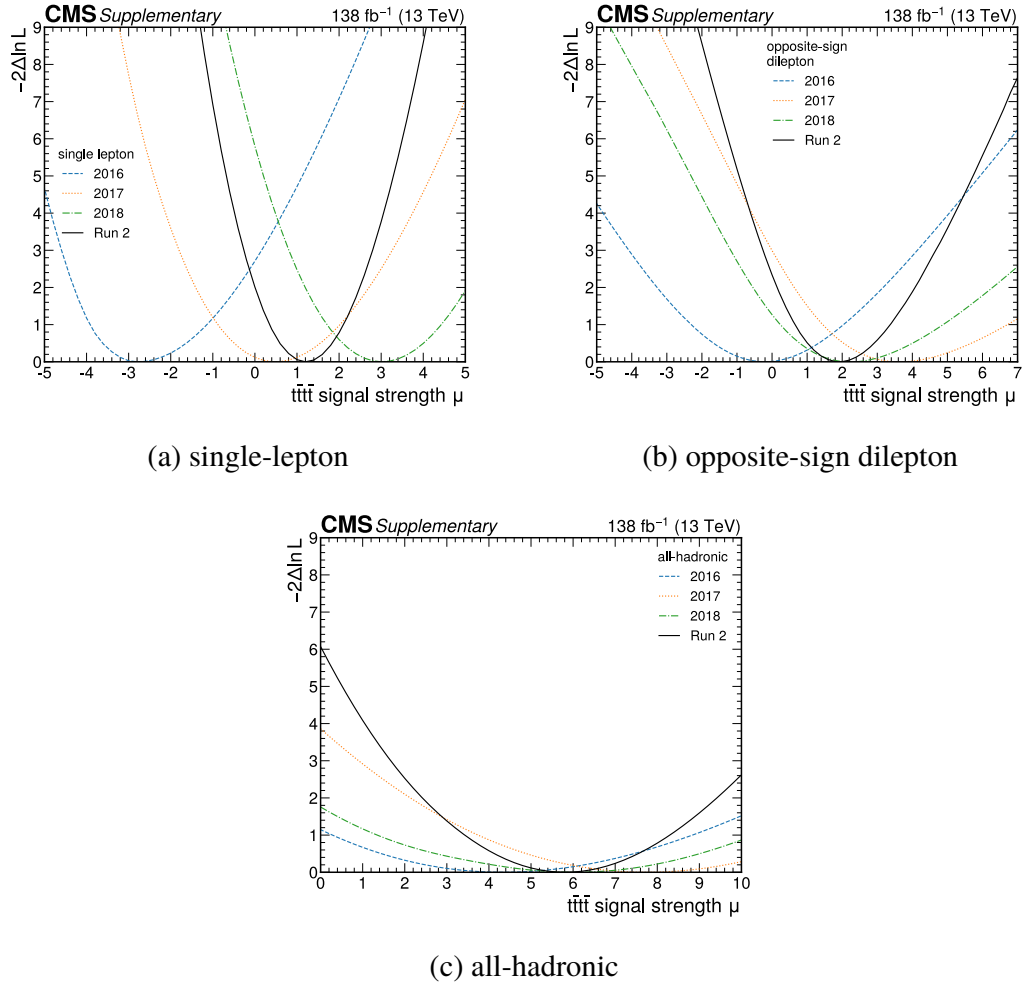


Figure 5.4: Post-fit NLL scans for single-lepton, opposite-sign dilepton, and all-hadronic final state, separated per year and combined [2] Analysis from SSDL&ML final state cannot be separated per year, and thus no NLL scan can be performed per each year of data taking for that final state. Refer to Figure 5.5 for NLL scan from SSDL&ML final state.

single-lepton, opposite-sign dilepton, and all-hadronic final state, separated per year and combined across three years (denoted in the plots as Run 2).

As seen from both Figures 5.4 and 5.5, NLL values have a smooth curve with the minimum point at the observed signal strength value. These NLL scan plots suggest that the observed signal strength measurement behaves as expected.

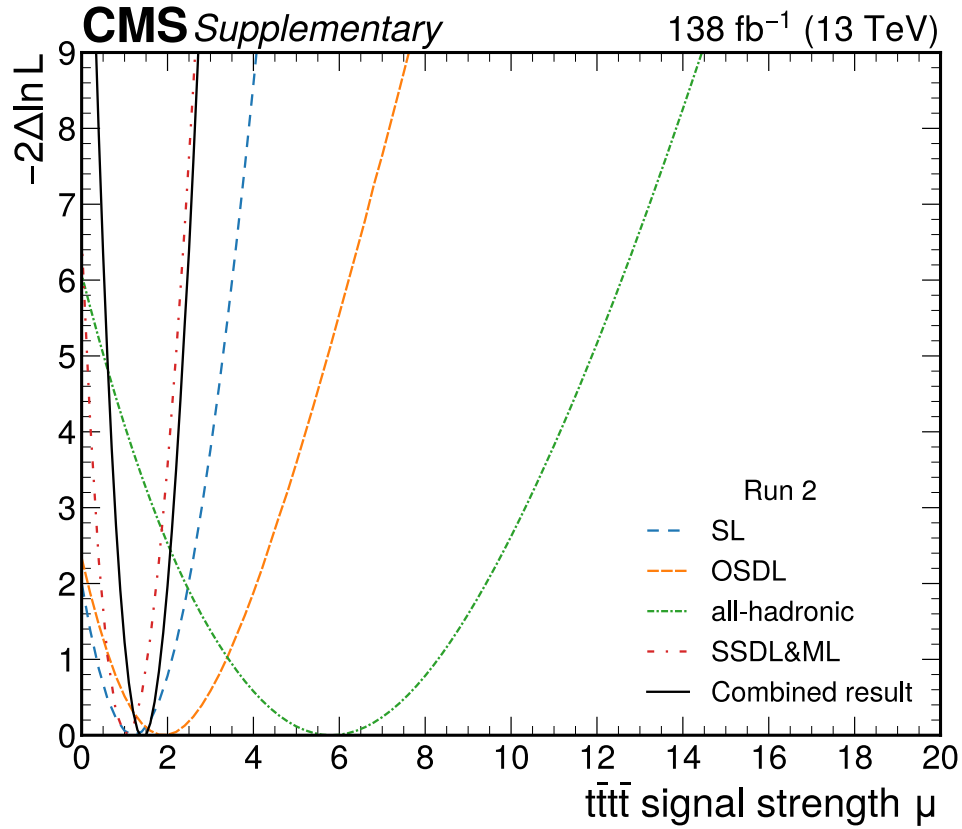


Figure 5.5: Post-fit NLL scans for all final states and the combination of all final states [2].

5.6.4 Nuisance parameter pulls and impacts

During the modelling of nuisance parameters before the actual fit, some nuisance parameters may have higher or lower deviations from the estimated value after the fit. This may be caused by design where nuisance parameters are designed to be constrained based on data fit or may be caused by an incorrect modelling of the uncertainty as well. The deviation of each nuisance parameter is called a *pull* and is calculated separately with background-only assumption or signal+background assumption.

Each nuisance parameter can also affect the measured value of signal strength differently. The *impact* of each nuisance parameter towards the measured signal

strength is calculated by setting the nuisance parameter to $+1\sigma$ or -1σ variation from the post-fit value and measuring the signal strength as usual, with all other nuisance parameters floating [178; 186]. In other words, the impacts for each nuisance parameter are calculated with the following procedure [178]:

1. Perform the initial fit for signal strength and all nuisance parameters simultaneously. The result from this step is the central value of signal strength and the best fit values of all nuisance parameters, with one standard deviation.
2. Set the value of the nuisance parameter we want to measure the impacts to $+1\sigma$ or -1σ value and perform the second fit allowing all other nuisance parameters to float freely.
3. Calculate the impacts from the shift of signal strength over its central value, or $\Delta\mu/\mu$.

Nuisance parameters with higher impacts can be considered as the most important uncertainties to the analysis and should be carefully considered in subsequent analyses. For example, previous analyses [1; 65] reported that $t\bar{t} + b\bar{b}$ normalisation uncertainty has the most impact on signal strength, and is included first when new analyses are being prepared.

The `HiggsCombine` package [178], along with `CombineHarvester` package [187] provided by CMS Collaboration already has tools to automate impacts calculation and ranking for all nuisance parameters in physics analyses, including this combination. The plot generated from these packages will show the impacts from each nuisance parameter, ranked from the ones with the highest impacts first, along with the nuisance parameter pulls calculated after the fit to data.

Figure 5.6 shows the nuisance parameter impacts plot for the full combination, with the ranking of nuisance parameters with the highest impacts on the signal strength. The rightmost coloured part of the plot shows the impact of nuisance parameters on the signal strength, where the red and blue bars represent the change induced when the parameter is shifted away $+1\sigma$ and -1σ from its central value respectively. The middle part with error bars shows pulls of each nuisance parameter, which can denote how well the nuisance parameter is modelled before the fit. If the parameter is well-designed, the error bar of the pull should have its range as -1 to 1 around the central value, since this means the estimated uncertainty imposed by the nuisance parameter is correct.

As seen from the figure, the $t\bar{t}H$ normalisation uncertainty (denoted as `ttH_norm_correlated`) has the highest impact on the signal strength, followed by $t\bar{t}+b\bar{b}$ normalisation uncertainty (denoted as `ttbb_correlated`). Several nuisance parameters representing data-prediction disagreement uncertainties from Run 2 all-hadronic analysis also have considerable impacts on the analysis, and have a tighter constraint as seen by short error bars. This behaviour, however, is expected, as these uncertainties are designed to be automatically constrained during the data fit.

From the impacts ranking shown in this figure, it is clear that, in addition to $t\bar{t}+b\bar{b}$ background, $t\bar{t}H$ background should also be considered as another main background for subsequent analyses of four-top quark production.

5.6.5 Channel compatibility tests

The channel compatibility tests can be performed to determine how the signal strength measurement for the full combination is compatible with each channel of the combination. This is done by performing two fits to the combination. The first fit

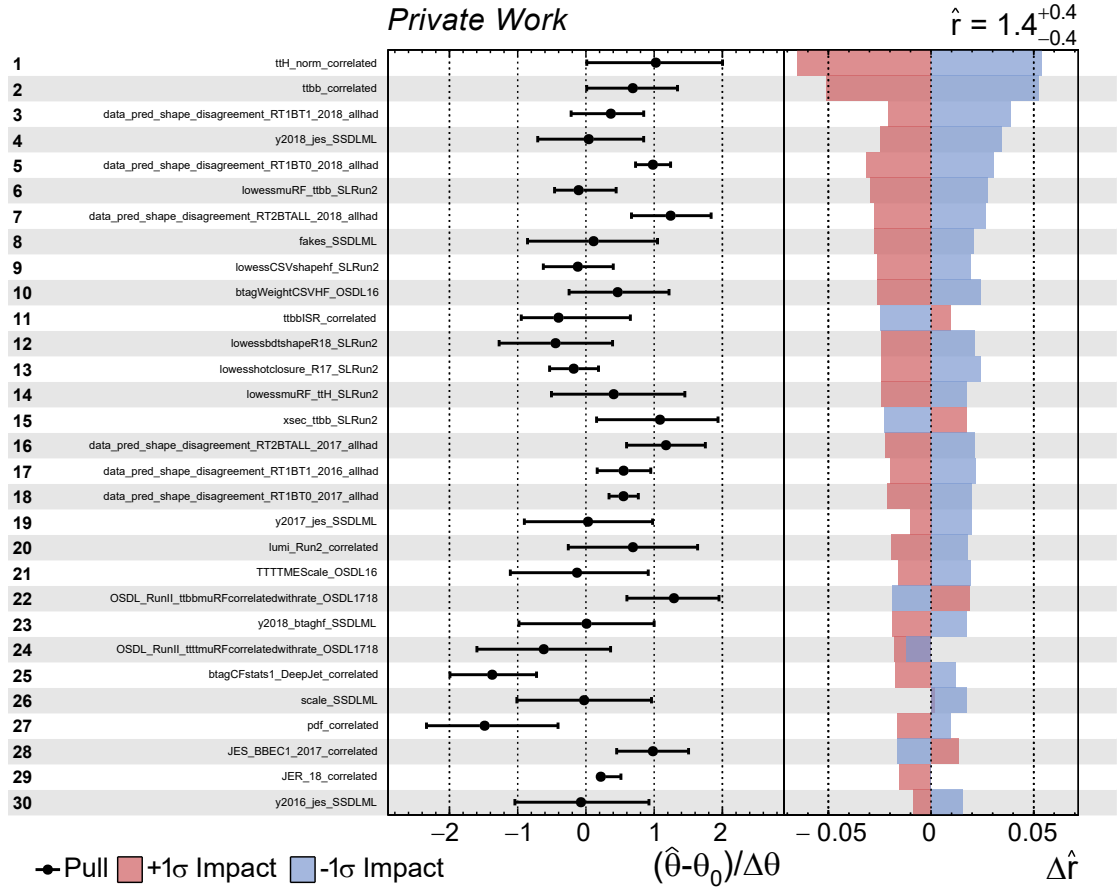


Figure 5.6: Nuisance parameter impacts plot for the final combination, listing nuisance parameters with the most impact.

is done where one value of signal strength applies to all channels in the combination (which is a typical signal strength measurement), and the second fit is done where each channel is assigned a separate signal strength and is allowed to be fitted in any value. The comparisons between the log-likelihood of these two fits are then taken as a Chi-square-like test value, and can then be used to determine the p -value determining the compatibility [178].

In this analysis, the test is performed to find *mutual compatibility* and *compatibility with SM* of the full combination, where it is split into each final state: SL, OSDL, SSDL&ML, and all-hadronic. Mutual compatibility is determined based on the observed cross section value, while compatibility with SM is

Table 5.12: Test results for the compatibility of each final state, with their corresponding p -values [2].

Test	Chi-square test statistic	DoF	p -value
Mutual compatibility	6.99	3	0.07
Compatibility with SM	8.30	4	0.08

determined based on the fixed signal strength $\mu = 1$. To convert the Chi-square-like test value obtained from the test into the corresponding p -value, the degree of freedom (DoF) for the mutual compatibility test must be $4 - 1 = 3$. For the test for compatibility with SM, the DoF is simply 4.

Table 5.12 shows the compatibility test results for mutual compatibility and compatibility with SM. As seen from the table, p -values from both tests are greater than 0.05, which indicates that the measured signal strength in the full combination and the SM signal strength (which is $\mu = 1$) are compatible with all channels.

5.7 Comparison to recent results from CMS and ATLAS experiments

As shown in Section 5.6, the measured cross section of four-top quark production, obtained from Run 2 data across single-lepton, opposite-sign dilepton, same-sign dilepton and multilepton, and all-hadronic final states, is

$$\sigma_{t\bar{t}t\bar{t}} = 17 \pm 4 \text{ (stat.)} \pm 3 \text{ (syst.) fb} \quad (5.23)$$

with the observed significance of 4.0, equating to the probability of 1 in 35 855 that the results are a statistical fluke. This measurement is in line with the SM prediction of the four-top quark cross section at NLO with EW corrections at $12.0^{+2.2}_{-2.5}$ fb [64]. This measurement is also in line with a more recent SM prediction at next-to-leading-logarithmic accuracy at $13.4^{+1.0}_{-1.8}$ fb [63].

During the preparation of this thesis, more recent search results from CMS [83] and ATLAS [84] Collaboration have been released. These results are performed on the same-sign dilepton and multilepton final states over the full Run 2 dataset. The measured cross section from CMS Collaboration is

$$\sigma_{tt\bar{t}\bar{t}}^{\text{new CMS}} = 17.7_{-3.5}^{+3.7} \text{ (stat.) }_{-1.9}^{+2.3} \text{ (syst.) fb} \quad (5.24)$$

with the expected and observed search significance of 4.9 and 5.6 standard deviations respectively [83]. Therefore, this result has achieved the 5 standard deviation requirement to be regarded as an *observation* of the production. As stated in Section 1.8.1, the observed significance indicates that the probability that this result can be a statistical fluke is 1 in 93 million. The measured cross section is also consistent with SM predictions and the measurement presented in this thesis.

On the other hand, the measured cross section from the ATLAS Collaboration is

$$\sigma_{tt\bar{t}\bar{t}}^{\text{new ATLAS}} = 22.5_{-4.3}^{+4.7} \text{ (stat.) }_{-3.4}^{+4.6} \text{ (syst.) fb} \quad (5.25)$$

with the expected and observed search significance of 4.7 and 6.1 standard deviations respectively [84], also achieving the 5 standard deviation requirement to be regarded as an *observation*. Although the observed significance from this analysis indicates that the probability that this result can be a statistical fluke is 1 in 1.9 billion, the expected significances from both the ATLAS and CMS Collaborations indicate that the analysis techniques employed by both results are comparable to each other. Again, the measured cross section is also consistent with SM predictions (within two standard deviations), the new measurement presented above by CMS, and the measurement presented in this thesis.

Even though the results from same-sign dilepton and multilepton final states would have better observed search significance from all other final states in the four-top quark production, the remaining final states still have their importance, as

the four-top quark production cross section should have the same value across all final states. Multilepton final state is good for discovery searches since leptons are easy to detect in the detectors. However, this final state has backgrounds that are hard to model, which will make precision measurements in the future more difficult. On the other hand, other final states can provide more events in the long run due to a higher branching ratio. Therefore, the information from other final states would be beneficial to precision measurements since it can provide more statistics and contain backgrounds that are easier to model.

As a comparison, another search of $t\bar{t}H$ production in the $H \rightarrow b\bar{b}$ decay channel with leptonic $t\bar{t}$ decays is performed by CMS Collaboration [188]. In this study, a comparison between the signal strength measurement from the single-lepton and dilepton final states shows that the single-lepton final state can provide better systematic and statistical uncertainties, resulting in the more precise measurement of $t\bar{t}H$ production signal strength, as shown in Figure 5.7.

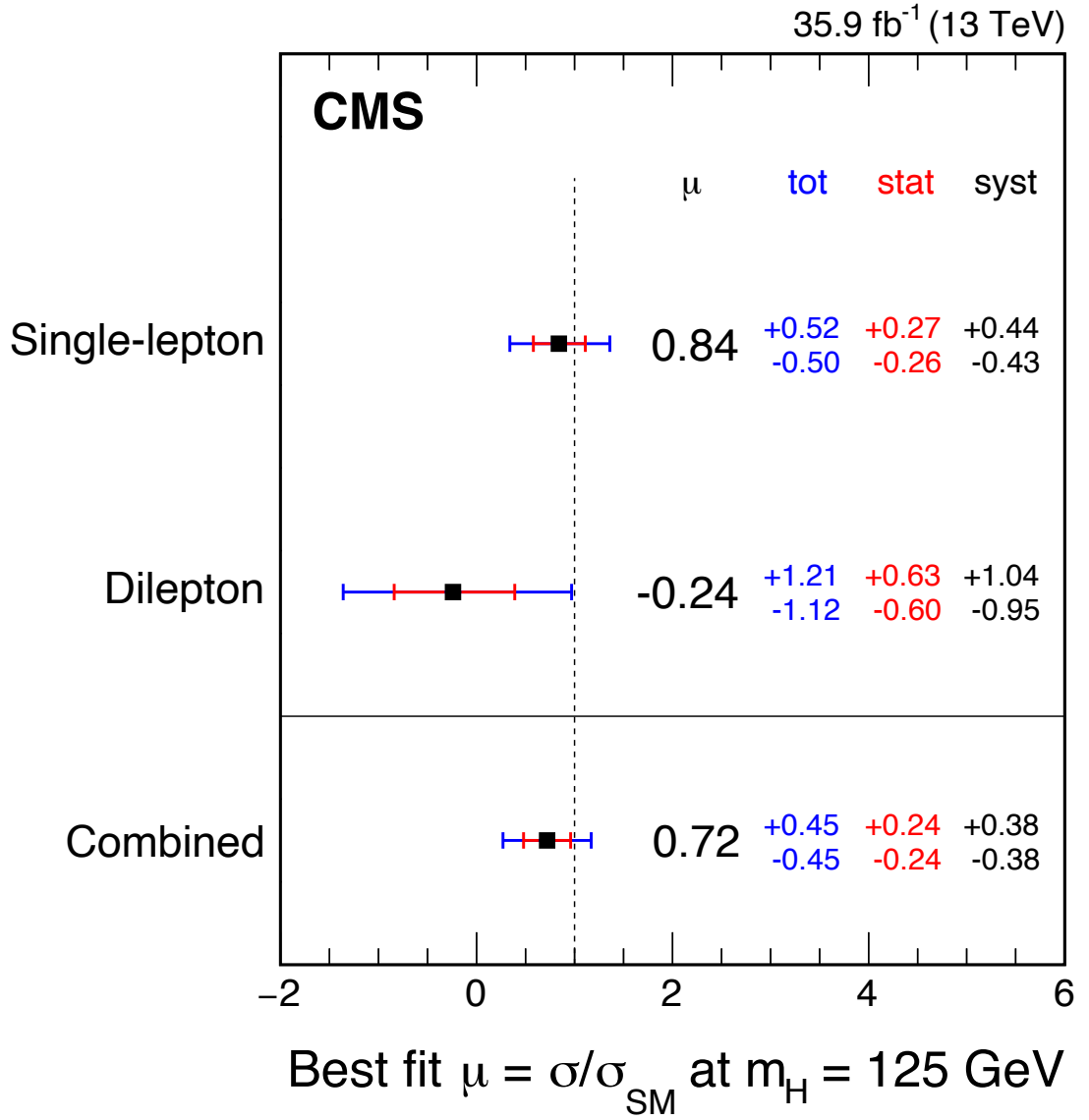


Figure 5.7: Measurement of signal strength for $t\bar{t}H$ process from single-lepton and dilepton final states. Single-lepton final state offers the measurement with less statistical uncertainty [188].

Chapter 6

TOP YUKAWA COUPLING MEASUREMENT

As introduced in Chapter 1, the measurement of the top Yukawa coupling value y_t can be considered as a test of the Standard Model. In this chapter, a preliminary study of the measurement for y_t in four-top quark production will be discussed. The study will mainly focus on the direct measurement based on neural networks, with comparisons to the indirect measurement performed via cross section measurement. Projections to higher integrated luminosity values are also discussed in this chapter.

6.1 Overview of the study

In this study, the measurement of top Yukawa coupling y_t will be measured directly based on the new approach following the ideas presented in Ref. [189], which will be discussed in Section 6.2. The measurement will be performed using Monte-Carlo generated events in the opposite-sign dilepton final state, generated using the 2017 configuration only. Only two main backgrounds, $t\bar{t}$ and $t\bar{t}H$ processes, will be used to train neural networks in addition to the $t\bar{t}t\bar{t}$ signal process.

The measurement is done in two stages. The first stage aims to isolate $t\bar{t}t\bar{t}$ from two important background processes, which are $t\bar{t}$ and $t\bar{t}H$. Monte-Carlo events from the three processes are used to train the first neural network, which will be called NN-c, that is designed to *classify* events into $t\bar{t}t\bar{t}$, $t\bar{t}$, and $t\bar{t}H$. The training procedure along with necessary validation checks will be discussed in Section 6.5.

In the second stage, the second neural network, which will be called NN- y_t , is designed and trained to give different output distributions for the $t\bar{t}t\bar{t}$ process based

on different values of y_t . Once NN- y_t is trained, it is then used to calculate the output distribution of $t\bar{t}t\bar{t}$ events along with all background processes with the requirement on the output from NN-c. The training of NN- y_t will be discussed in Section 6.6, while the direct measurement based on the output distribution of this network will be discussed in Section 6.7.

This study also offers the projection of the measurement based on an increasing amount of luminosity, from 49.81 fb^{-1} , which is the amount of data collected in 2017, to 138 fb^{-1} , which is the amount of data collected in the whole Run 2, all the way towards 3000 fb^{-1} , which is the amount of data expected to achieve using High-Luminosity LHC upgrade [176]. A comparison between the *direct* measurement of y_t using neural networks and the *indirect* measurement of y_t using the measurement of $t\bar{t}t\bar{t}$ cross section will also be discussed in Section 6.8.

6.2 Machine Learning techniques used

This study will incorporate ideas presented in Ref. [189], which shows “several inference techniques to constrain theory parameters” by utilising extra information at the generator level. This technique has been used in the measurement of Wilson coefficients for several EFT operators¹ conducted by the CMS Collaboration [190]. In that measurement, several neural networks are trained to simply separate between events in SM and EFT scenarios.

This measurement performed by CMS Collaboration [190] focuses on five Wilson coefficients. Event weights corresponding to the values of each Wilson coefficient can be calculated at the generator level, which is then used to train neural networks to classify between SM and EFT events. The result of the training is a set of neural networks with output distributions that change shape with different

¹See Section 1.8.4 for the definition of EFT operators and Wilson coefficients.

Wilson coefficient values, which is then used to fit the most probable values of Wilson coefficients.

Since the method outlined in Ref. [189] can be applied to any “theory parameters”, this method, or more specifically, the method used in Ref. [190], should also apply to the measurement for y_t as well. This is because event weights with different y_t values can be calculated at the generator level in the same way, which provides additional generator-level information for neural network training. Furthermore, the output distribution from the trained neural network will allow us to measure y_t using distribution shape information in addition to normalisation effects. Therefore, this study will follow the neural network training procedure discussed in Ref. [190], which is to simply discriminate between events in SM (with SM value of y_t) versus events in non-SM scenarios. Further details shall be discussed in Section 6.6.

6.3 Simulated event samples

Event samples used in this analysis are Monte-Carlo events generated with the $t\bar{t}t\bar{t}$ process as a signal process, and the $t\bar{t}$ and $t\bar{t}H$ background processes, using the detector configuration used in 2017. More specifically, these Monte-Carlo samples generated from CMS Monte-Carlo central production are used:

- /TTTo2L2Nu_HT500Njet7_TuneCP5_13TeV-powheg-pythia8/
RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/
NANOADSIM, a $t\bar{t}$ sample which decays dileptonically at the generator level, with the generator-level filter of $H_T \geq 500$ GeV and requiring 7 or more jets. This sample has a cross section value of 1.456 pb, calculated from the $t\bar{t}$ cross section of 831.76 pb times the branching ratio $W \rightarrow \ell\nu$ of $(3 \times 0.108)^2$

(since there are two W bosons) and the generator-level filter efficiency of 0.016679.

- /ttHTobb_M125_TuneCP5_13TeV-powheg-pythia8/
RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/
NANOADSIM, a $t\bar{t}H$ sample where the Higgs boson H decays into $b\bar{b}$. This sample has the cross section value of 0.295 pb, calculated from the $t\bar{t}H$ cross section of 0.5071 pb times the branching ratio $H \rightarrow b\bar{b}$ of 0.5824.
- /TTTT_TuneCP5_13TeV-amcatnlo-pythia8/
RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/
NANOADSIM, a $t\bar{t}t\bar{t}$ sample with SM value of y_t . This sample is used for the training of NN-c, which will be further discussed in Section 6.5.

For this study, the updated SM $t\bar{t}t\bar{t}$ cross section value of 13.4 fb is used, calculated at next-to-leading-order with electroweak corrections and NLL' accuracy [63]. The reason for the change in $t\bar{t}t\bar{t}$ cross section value compared to the studies presented in Chapters 4 and 5 is that the calculation results have been announced during the study, and to better serve as a preliminary study for future y_t measurements based on $t\bar{t}t\bar{t}$ process.

In addition to the samples above, another set of Monte-Carlo event samples based on $t\bar{t}t\bar{t}$ process were generated, with the y_t/y_t^{SM} value of 2.32 (configured within the generator as $y_t = 400$) instead of the SM value of $y_t/y_t^{\text{SM}} = 1$ (configured within the generator as $y_t = 172.5$). Generator-level event weights for each value of y_t/y_t^{SM} are also calculated for this sample. This set of privately generated samples is used to train the second network referred to in Section 6.6. Two different sets of generated samples are made and have been verified to have identical behaviour.

6.4 Event reconstruction and preselection criteria

Event preselection criteria for this analysis are as follows:

- Requiring exactly two oppositely-charged leptons with the following object requirements
 - **Electrons:** $p_T > 34 \text{ GeV}$, $|\eta| < 2.4$, passing loose electron ID requirement with an average efficiency of approximately 90%
 - **Muons:** $p_T > 28 \text{ GeV}$, $|\eta| < 2.4$, passing tight muon ID and very tight muon isolation
- At least 4 jets with $p_T > 30 \text{ GeV}$, $|\eta| < 2.4$, $\Delta R < 0.4$. Among all jets, at least two jets must have passed the tight working point of the DeepJet b-tagging algorithm [113].
- Any number of fat jets with $p_T > 200 \text{ GeV}$, $|\eta| < 2.4$, $\Delta R < 0.8$. The information on fat jets will be used in neural network training.

The preselection criteria shown above are different from the ones used in the search for four-top quark production in opposite-sign dilepton final state as described in Section 4.6, which has more relaxed requirements in terms of leptons.

6.5 NN-c training

As in Section 5.6, the $t\bar{t} + b\bar{b}$ and $t\bar{t}H$ processes have the most impact on the signal strength measurement in the opposite-sign dilepton channel. This signifies that in addition to the $t\bar{t}$ process, the $t\bar{t}H$ process should also be considered as an important

background process. To remove the $t\bar{t}H$ background from the signal region as much as possible, NN-c is designed to classify events into three classes: $t\bar{t}t\bar{t}$, $t\bar{t}$, and $t\bar{t}H$.

For NN-c, 26 variables are chosen as inputs based on the kinematic differences between $t\bar{t}t\bar{t}$ signal and $t\bar{t}$ background. The list is loosely adapted from the BDT input variables list used in the $t\bar{t}t\bar{t}$ search in the opposite-sign dilepton final state using 2016 data [1]. The variables are as follows:

- Dilepton flavour (ee , $e\mu$, or $\mu\mu$). NN-c is trained against all three dilepton flavours, and these three variables determine if an event has two electrons, two muons, or an electron and a muon.
- p_T from each lepton (2 variables)
- ΔR between each lepton, defined as $\Delta R = \sqrt{(\Delta\eta)^2 + (\delta\phi)^2}$
- $\Delta\phi$ between each lepton
- p_T of the first four jets with highest p_T (4 variables)
- b-tagging discriminator value from DeepJet [113] algorithm from the first four jets with highest p_T (4 variables)
- Number of jets
- Number of b-jets
- HT_b (sum of b-tagged jet p_T)
- HT (sum of jet p_T)
- Ratio of HT from the first four jets with the highest p_T to HT from all jets in an event
- H (sum of jet energies, which is $p_T \cosh(\eta)$)

- Hadronic centrality (ratio between HT from all jets and H from all jets)
- ΔR between two b-tagged jets (with DeepJet) with the highest p_T
- $\Delta\phi$ between two b-tagged jets (with DeepJet) with the highest p_T
- Missing transverse energy (MET)
- Sphericity, as defined in Ref. [73]

Normalised distributions of these input variables are shown in Figure 6.1.

The structure of NN-c is as follows:

- One input layer for 26 input variables
- Batch normalisation layer, handling automatic normalisation of the input variables
- Three dense layers containing 50 neurons with ReLU activation function, each layer followed by one Dropout layer
- One output layer with three output nodes, using softmax activation function to guarantee that outputs from all nodes have the sum of 1

The structure above allows us to input the data as is to the network without the need to normalise the data since the first batch normalisation layer will automatically handle it. Dropout layers also reduce the risk of the network being overtrained. The output layer is designed to have three nodes representing the $t\bar{t}t\bar{t}$ signal, the $t\bar{t}$ background, and the $t\bar{t}H$ background. As such, each event will be trained with the target output in the form of a three-dimensional vector: $t\bar{t}t\bar{t}$, $t\bar{t}$, and $t\bar{t}H$ events have the target of $[1\ 0\ 0]^T$, $[0\ 1\ 0]^T$, and $[0\ 0\ 1]^T$ respectively.

Private work

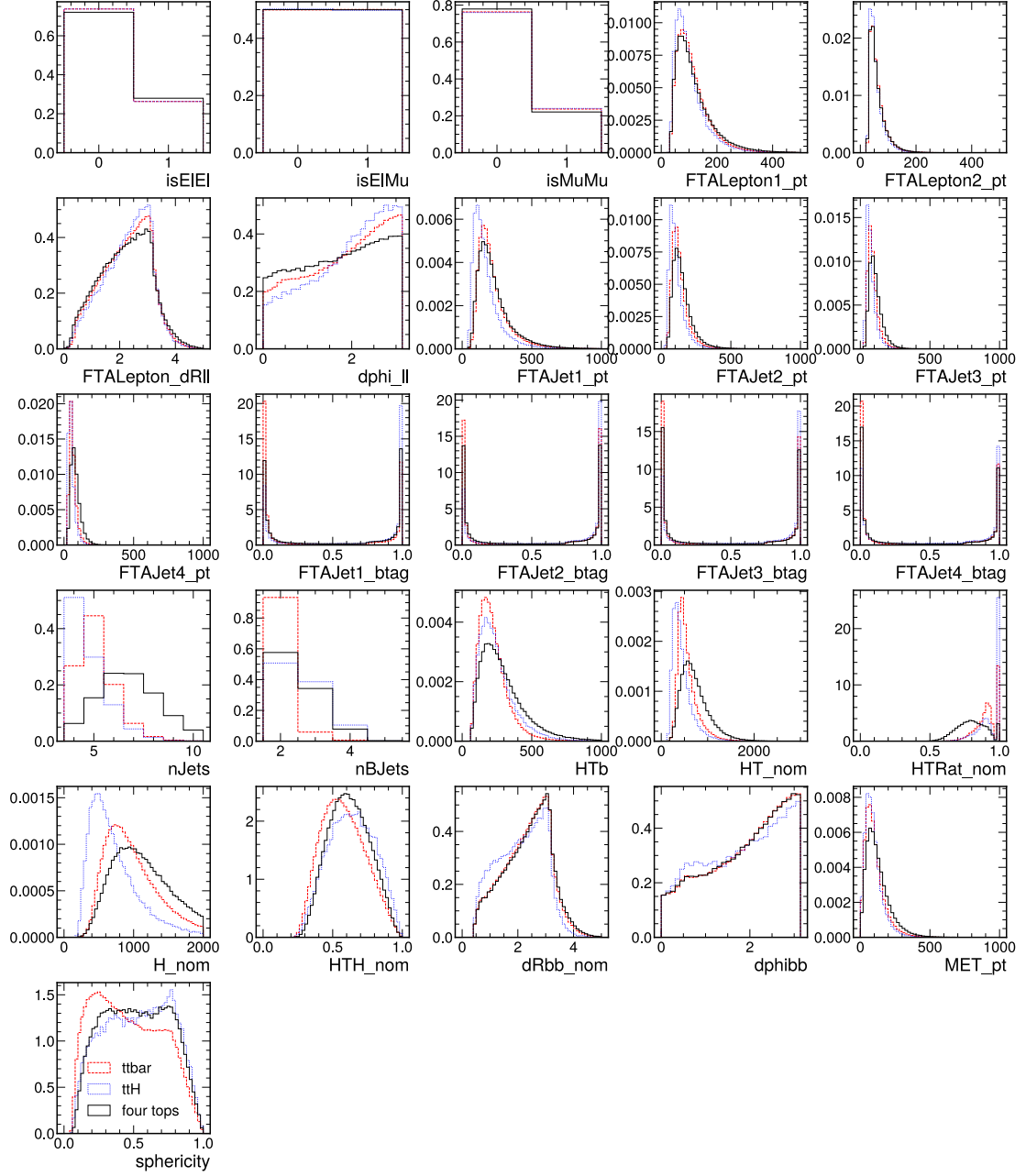
49.81 fb⁻¹ (13 TeV)

Figure 6.1: Distributions for input variables of classification model from $t\bar{t}t\bar{t}$, $t\bar{t}$, and $t\bar{t}H$ samples, normalised.

The model is trained using categorical cross-entropy loss function and Adam optimizer [148], and is trained with 20 000 events from each class ($t\bar{t}t\bar{t}$ signal, $t\bar{t}$ background, and $t\bar{t}H$ background), while the rest of events are used as testing dataset. This amount of events in the training dataset is chosen to ensure a balance between the three classes during the training. The model is trained for 100 epochs with a batch size of 100 events and no early stopping requirement.

Figure 6.2 shows the loss evolution of NN-c with the chosen dropout probability and learning rate, trained over 100 epochs with a batch size of 100. The scattering behaviour of the testing loss is expected, as the testing loss is calculated as a whole dataset and the testing dataset is not used to adjust the network parameters. As the testing loss is generally not greater than the training loss, this is a sign that our NN-c is not overtrained. Even though the NN-c output distribution may look undertrained, as a portion of the training and testing events still do not have a correct predicted value, it is more important to have NN-c behave in the same way in both training and testing datasets (and thus avoid overtraining) rather than having NN-c perform perfectly in the training dataset and perform poorly in the testing dataset (which is overtraining).

Several dropout probability rates and learning rates are considered in this analysis, and the final values chosen for the classification model are 0.1 for dropout probability and 0.001 for learning rate. The dropout probability of 0.01 is shown to have comparable losses between the training and the testing datasets, and different values of learning rate were verified to not affect the training. This has been verified by training NN-c with the same structure while varying different values of learning rate and dropout probabilities. Upon inspection, it is shown via the evolution of the area under the ROC curve that different learning rates do not differ much within the same dropout probability.

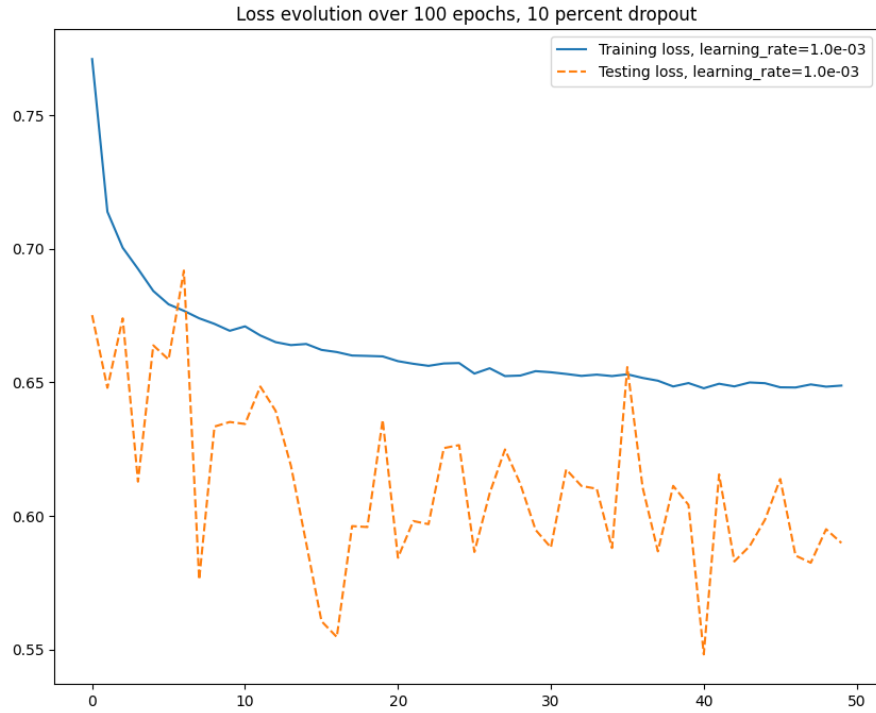


Figure 6.2: Loss evolution during the training of NN-c, indicating that there is no overtraining since the testing loss is generally not greater than the training loss, as explained in Section 3.4.1.

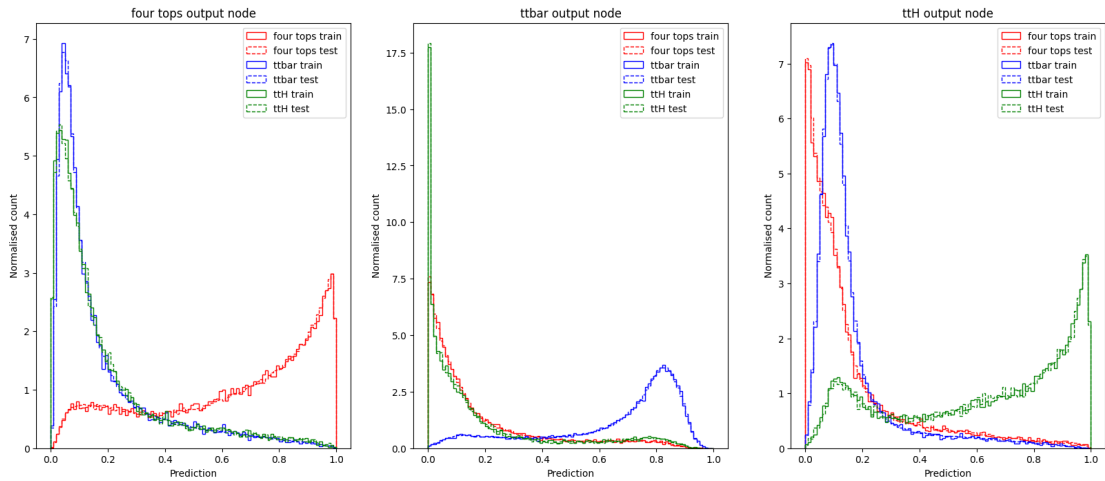


Figure 6.3: Output distribution of NN-c from training and testing datasets, separated by three output nodes and three processes. The distributions from training and testing datasets are comparable.

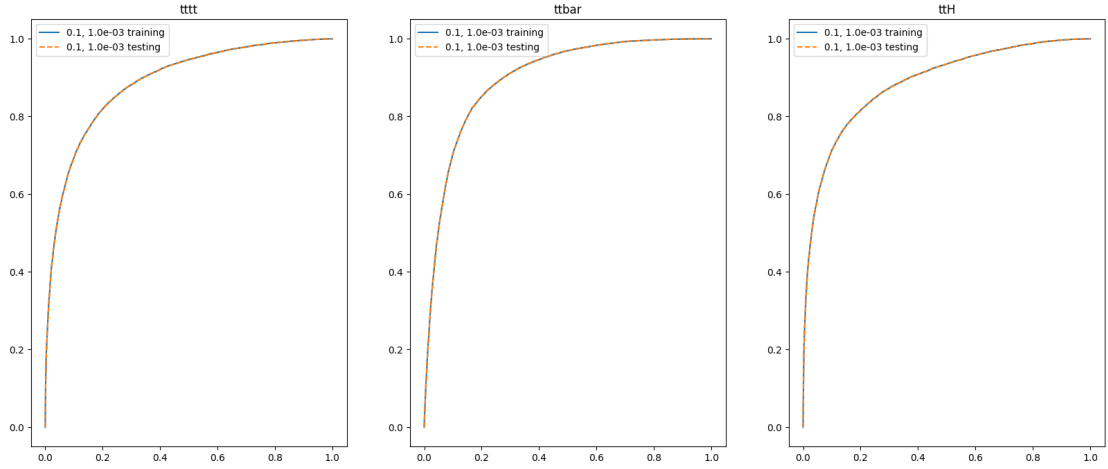


Figure 6.4: ROC curve of NN-c from training and testing datasets, separated by three output nodes. The x -axis represents signal acceptance, while the y -axis represents background acceptance.

Table 6.1: Area under ROC curve (AUC) calculated from training and testing datasets and three output nodes of the NN-c.

Output node	Training AUC	Testing AUC
$t\bar{t}t\bar{t}$	0.888	0.890
$t\bar{t}$	0.899	0.902
$t\bar{t}H$	0.887	0.891

Figures 6.3 and 6.4 show the output distribution and ROC curve of the $t\bar{t}t\bar{t}$, $t\bar{t}H$, and $t\bar{t}$ output nodes. As this neural network is trained to classify $t\bar{t}t\bar{t}$, $t\bar{t}$, and $t\bar{t}H$ processes simultaneously, the ROC curves are calculated per each output node by comparing the output distribution of events with the corresponding process versus events not from the corresponding process.

The output distributions between the training and testing datasets shown in Figure 6.3 are almost identical across the processes and output nodes. The ROC curves shown in Figure 6.4 along with the area under the ROC curve calculated from training and testing datasets, shown in Table 6.1, show that the performance of NN-c is identical between training and testing datasets.

Based on the ROC curves shown in Figure 6.4, along with the area under the ROC curve shown in Table 6.1, the performance of NN-c is good enough to ensure high signal acceptance at low background acceptance. The AUC across all three nodes are approximately 0.9, which shows that the network is good enough for the classification task although the value is not as close as the perfect network with AUC = 1.

6.6 NN- y_t training

To follow the procedure explained in Ref. [190], the second neural network must be trained to distinguish between SM case and non-SM case in $t\bar{t}t\bar{t}$.

In our study, however, we are utilising a set of privately-generated $t\bar{t}t\bar{t}$ sample, using $y_t/y_t^{\text{SM}} = 2.32$ at the generator level (with the generator configuration of $y_t = 400$) with event weights for $y_t/y_t^{\text{SM}} = 0, 0.29, 0.58, \dots, 3.48$ (with the generator configuration of $y_t = 0, 50, 100, \dots, 600$), as well as the SM value (with the generator configuration of $y_t = 172.5$). The decision to use $y_t/y_t^{\text{SM}} = 2.32$ at the generator level is to allow more diagrams that contain Higgs bosons (which in turn is sensitive to y_t) in the sample. With this sample, the event weight for the case of $y_t/y_t^{\text{SM}} = 2.32$ will always be 1, while event weights for the SM case may vary based on the contents of the hard process. With this setup, we ended up having one set of samples which should be split into SM and non-SM events.

6.6.1 $t\bar{t}t\bar{t}$ sample splitting

For NN- y_t that is designed to distinguish between SM ($y_t/y_t^{\text{SM}} = 1$) and non-SM ($y_t/y_t^{\text{SM}} \neq 1$), the $t\bar{t}t\bar{t}$ sample must be split into SM and non-SM classes. Three possible splitting schemes for SM and non-SM classes utilising event weights at $y_t/y_t^{\text{SM}} = 2.32$ and $y_t/y_t^{\text{SM}} = 1$ (SM value) were explored:

- **Probability comparison:** Calculate the probability of each event per each y_t value, $p(y_t) = w(y_t) / \sum w(y_t)$, where $w(y_t)$ is an event weight for each y_t and $\sum w(y_t)$ is the sum of $w(y_t)$ from all events in the sample. Then, assign the non-SM class to events with $p(y_t/y_t^{\text{SM}} = 2.32) > p(y_t = \text{SM})$, and assign the SM class to events with $p(y_t/y_t^{\text{SM}} = 2.32) \leq p(y_t = \text{SM})$.
- **Weighted randomisation:** Using the same probability definition as in the probability comparison method, assign the non-SM class to events *randomly*, with the probability of $p(y_t/y_t^{\text{SM}} = 2.32) / (p(y_t/y_t^{\text{SM}} = 2.32) + p(y_t = \text{SM}))$.
- **Direct weight comparison:** Instead of calculating the probability of each event per each y_t value, compare the weights in the event directly and assign the non-SM class to events with $w(y_t/y_t^{\text{SM}} = 1.74) > w(y_t = \text{SM})$, and assign the SM class to events with $w(y_t/y_t^{\text{SM}} = 1.74) \leq w(y_t = \text{SM})$. The value $y_t/y_t^{\text{SM}} = 1.74$ is chosen as a comparison point to maintain the balance between the two classes.

A large number of variables are considered to be chosen as input variables, which contain several aspects of the objects inside each event, such as the kinematics of physics objects, invariant masses of physics object combinations, and shape variables for each event. Since there are some diagrams containing Higgs boson decay into $t\bar{t}$ in the $t\bar{t}t\bar{t}$ process, two out of four top quarks produced from the hard process may have different kinematics from $t\bar{t}t\bar{t}$ events without Higgs boson decay. This fact largely motivates the use of proxy variables constructed from physics objects available after the preselection.

To test which of the three splitting schemes is the most suitable for this study, the variables are used to populate *normalised* histograms with 100 bins between non-SM and SM classes per each splitting scheme. For each variable and splitting scheme, a similarity index is calculated, which is the overlapping area of the non-SM

Table 6.2: Similarity index (Equation 6.1) for 10 variables with the least similarity index, calculated with different splitting schemes, sorted by the values from probability comparison splitting method. Note that the similarity index from the weighted randomisation method may fluctuate since the distributions are randomised.

Variable	Probability comparison	Weighted randomisation	Direct weight comparison
m_bb11jj	0.798	0.884	0.735
fatjetptrat	0.827	0.917	0.830
FTAJet1_pt	0.829	0.911	0.808
fatjetpt	0.830	0.923	0.828
H_nom	0.830	0.892	0.769
HT_nom	0.834	0.910	0.808
FTAJet2_pt	0.837	0.912	0.822
FTAJet3_pt	0.848	0.914	0.825
m_bb4j	0.850	0.908	0.758
HTb	0.863	0.931	0.860

and SM distributions. To be more precise, the similarity index is

$$\text{similarity index} = \frac{\sum_i \min(h_{i,\text{non-SM}}, h_{i,\text{SM}})}{\sum_i \max(h_{i,\text{non-SM}}, h_{i,\text{SM}})} \quad (6.1)$$

where h_i represents the count in normalised histogram bin i . This means that if there are more differences between *normalised* distributions of a variable between SM and non-SM classes, the similarity index decreases.

Tables 6.2, 6.3, and 6.4 show the similarity index for 10 variables that show the best discrimination (which means having the largest difference in distributions between non-SM and SM cases), sorted with different splitting methods. Based on the similarity index shown in the tables, the probability comparison method and direct weight comparison method show several variables having a lower similarity index than the weighted randomisation method.

Figure 6.5 shows the distributions of select variables with different splitting schemes, which shows that the distributions obtained from the probability

Table 6.3: Similarity index (Equation 6.1) for 10 variables with the least similarity index, calculated and sorted from the weighted randomisation method. Note that the similarity index from the weighted randomisation method may fluctuate since the distributions are randomised.

Variable	Similarity
m_bb11jj	0.884
H_nom	0.892
aplanarity_mbb11jj	0.901
sphericity	0.902
shapeD_mbb11jj	0.907
m_bb4j	0.908
shapeC	0.908
shapeD	0.909
shapeC_mbb11jj	0.909
aplanarity	0.910

Table 6.4: Similarity index (Equation 6.1) for 10 variables with the least similarity index, calculated and sorted from the direct weight comparison method.

Variable	Similarity
m_bb11jj	0.735
m_bb4j	0.758
H_nom	0.769
shapeD_mbb11jj	0.801
mass_j14	0.803
mass_j12	0.806
aplanarity_mbb11jj	0.807
dRjet13	0.808
FTAJet1_pt	0.808
HT_nom	0.808

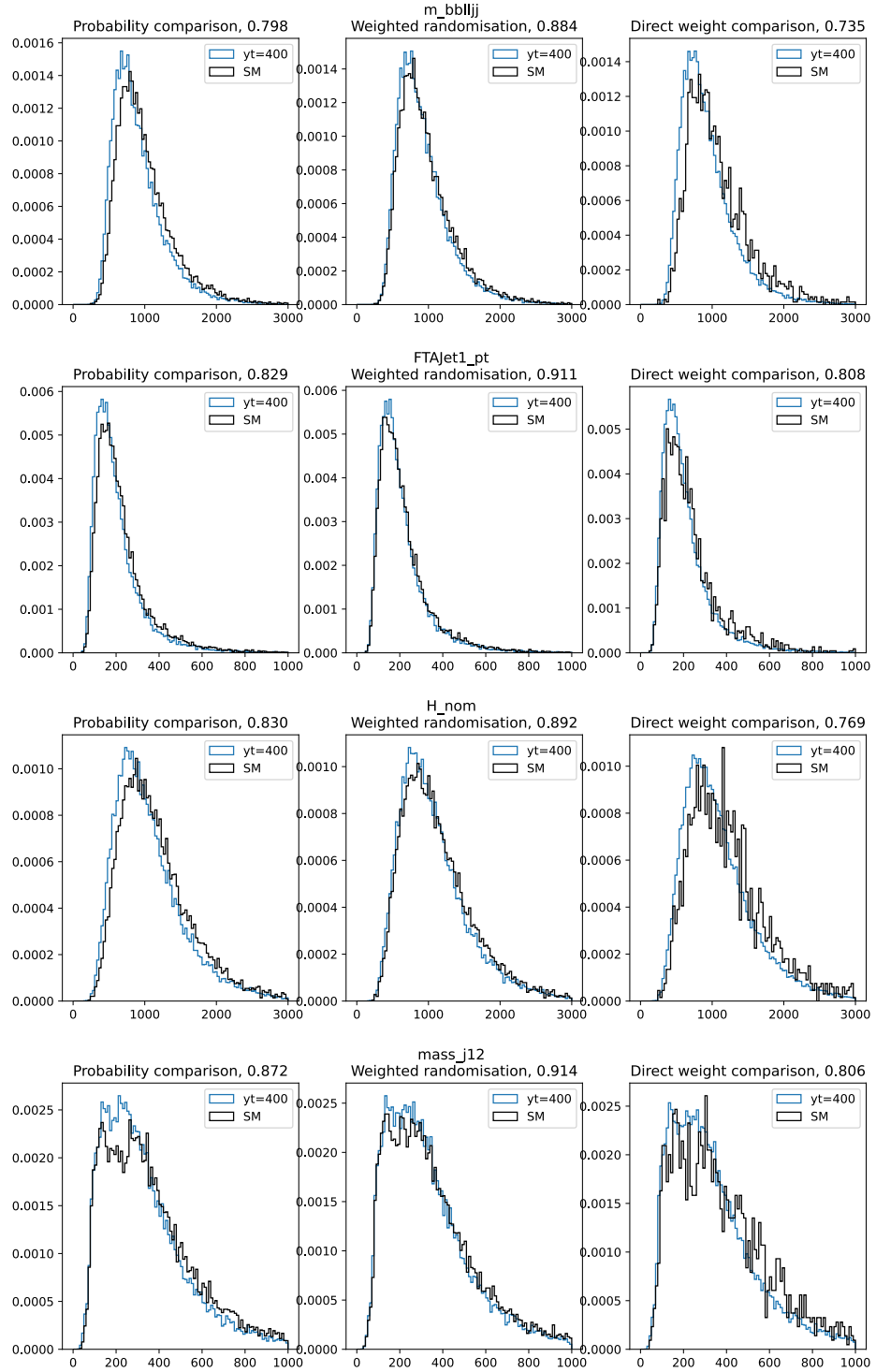


Figure 6.5: Distributions of example variables with different splitting methods. Note that the distributions from the weighted randomisation method (middle column) may fluctuate since the distributions are randomised.

comparison method are smoother than the direct weight comparison method. For these reasons, the probability comparison method is used to separate the $t\bar{t}t\bar{t}$ signal samples into SM and non-SM classes.

6.6.2 Input variables to NN- y_t

Based on the similarity index, 23 input variables are chosen for the training of the Yukawa coupling network as follows:

- FTAJet1_pt, FTAJet2_pt, FTAJet3_pt, and FTAJet4_pt p_T of the first four jets with highest p_T (4 variables)
- FTALepton1_pt p_T of the first lepton with highest p_T
- HTb HT_b (sum of b-tagged jet p_T)
- H_nom H (sum of jet energies, which is $p_T \cosh(\eta)$)
- MET_pt Missing transverse energy ($\cancel{E}_T$)
- aplanarity Aplanarity² of all jets in an event
- aplanarity_mbblljj Aplanarity of 2 b-jets, 2 leptons, and at most 2 light jets
- deltaR_bll ΔR of a pair between a b-jet and a lepton with least ΔR
- fatjetpt p_T of the first fat jet, if there is at least one fat jet
- fatjetptrat Ratio between p_T from the first fat jet versus H_T from all normal jets. (The ratio is zero if there are no fat jets in an event.)
- m_bb4j Invariant mass of 2 b-jets and at most 4 light jets

²Given a matrix $S_{ij} = \sum_k p_{k,i} p_{k,j} / \sum_k p_k^2$, where i and j are Cartesian coordinates x , y , and z , and k is an index running over all objects included in a calculation. Aplanarity is calculated as $3\lambda_3/2$ where λ_3 is the smallest eigenvalue of S_{ij} .

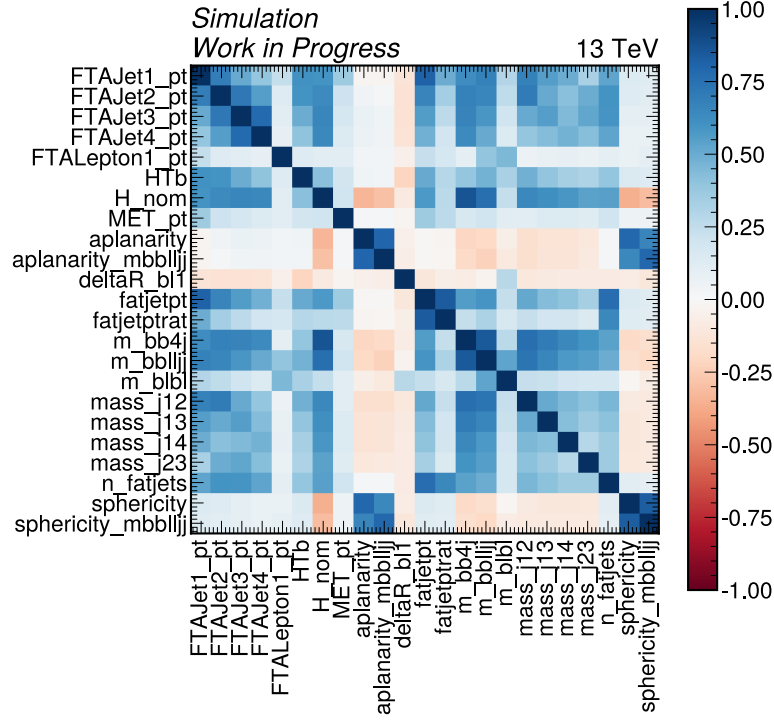


Figure 6.6: Correlation matrix for $NN\text{-}y_t$ input variables for SM class.

- `m_bbljj` Invariant mass of 2 b-jets, 2 leptons, and at most 2 light jets
- `m_bbl` Invariant mass of 2 b-jets, matched with each lepton with least ΔR , and 2 leptons
- `mass_j12`, `mass_j13`, `mass_j14`, `mass_j23` Invariant mass of the following two-jet combinations: 1st + 2nd, 1st + 3rd, 1st + 4th, and 2nd + 3rd (4 variables)
- `n_fatjets` Number of fat jets
- `sphericity` Sphericity [73] of all jets in an event
- `sphericity_mbblljj` Sphericity of 2 b-jets, 2 leptons, and at most 2 light jets

Figure 6.6 and 6.7 shows the correlation matrix for the 23 input variables for SM and non-SM samples. The list of input variables described earlier has

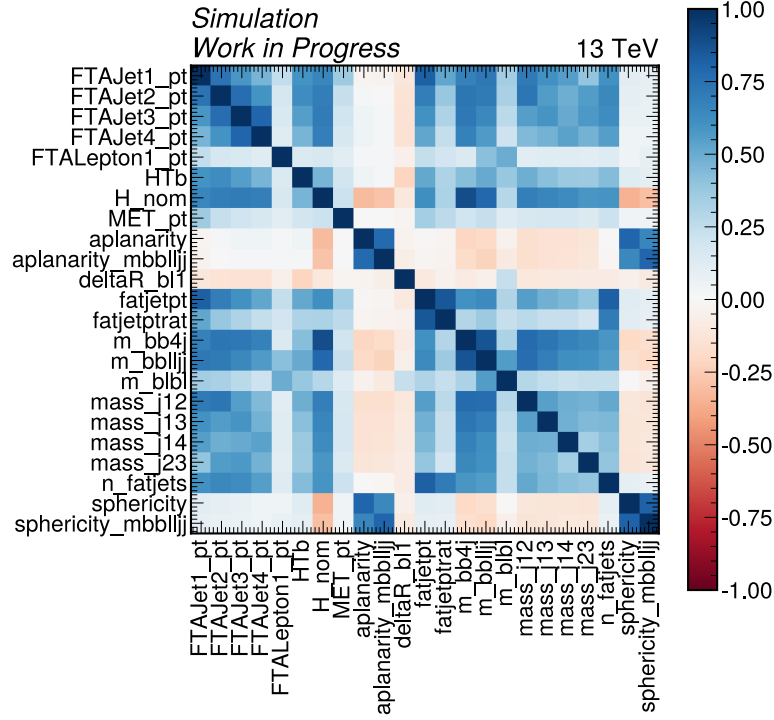


Figure 6.7: Correlation matrix for NN- y_t input variables for non-SM class.

highly-correlating variables removed, since to first order a set of highly-correlating variables can be replaced by one variable from the set.

The structure of NN- y_t is as follows:

- One input layer for 23 input variables
- A dense layer with 50 neurons, 40 neurons, and 30 neurons. Each dense layer uses the LeakyReLU activation function (Equation 3.8) and L1+L2 regularizer with $\ell_1 = 10^{-6}$ and $\ell_2 = 10^{-8}$ applied to both weights and biases of each layer. Each dense layer is followed by a dropout layer with a dropout probability of 0.2.
- An output layer with no activation function. The output for this network will be in the form of a *logit*, which is an inverse of the sigmoid function (Equation 3.4).

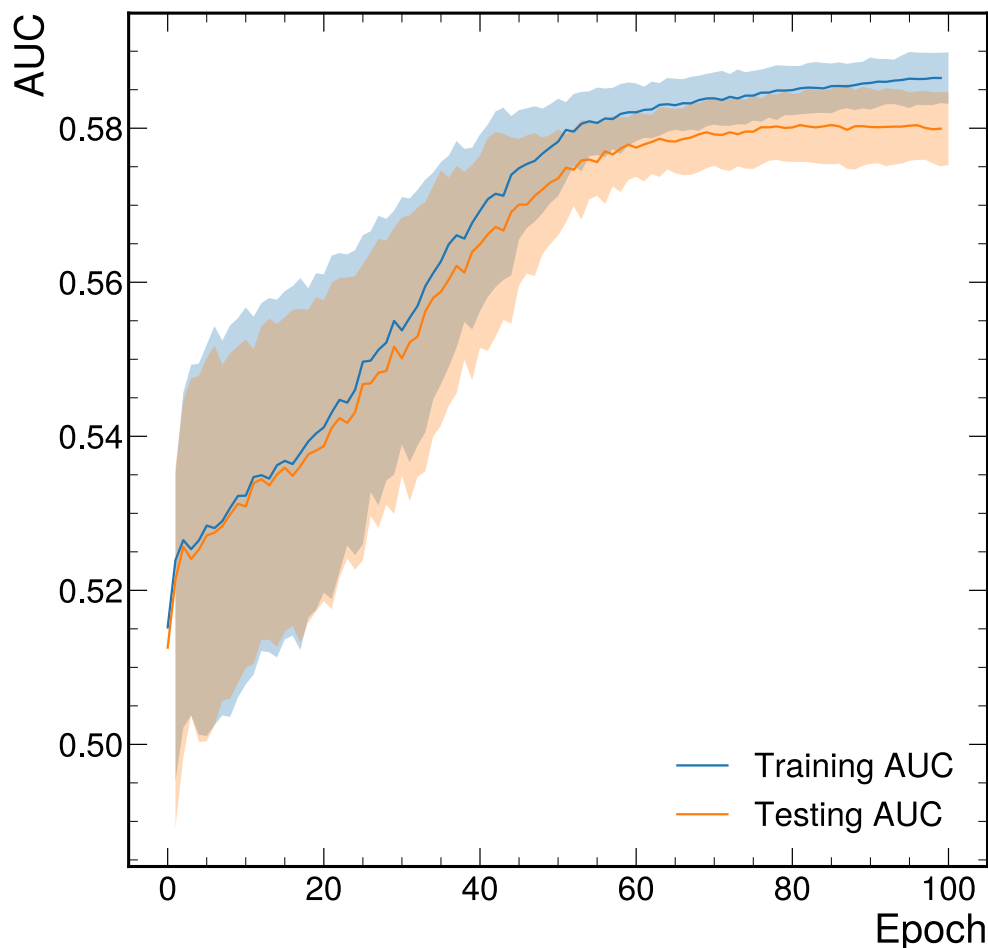


Figure 6.8: Evolution of area under ROC curve over 100 epochs of training, calculated with 20 training trials, each with a different random seed for the training dataset.

The model is compiled using binary cross-entropy loss. Using the output in the form of a logit directly avoids the unnecessary calculation of $\log(\exp(x))$ when the output is used to calculate binary cross-entropy loss and also provides numerical stability. It is optimised using Adam optimiser with a learning rate of 10^{-4} and trained for 100 epochs with a batch size of 100.

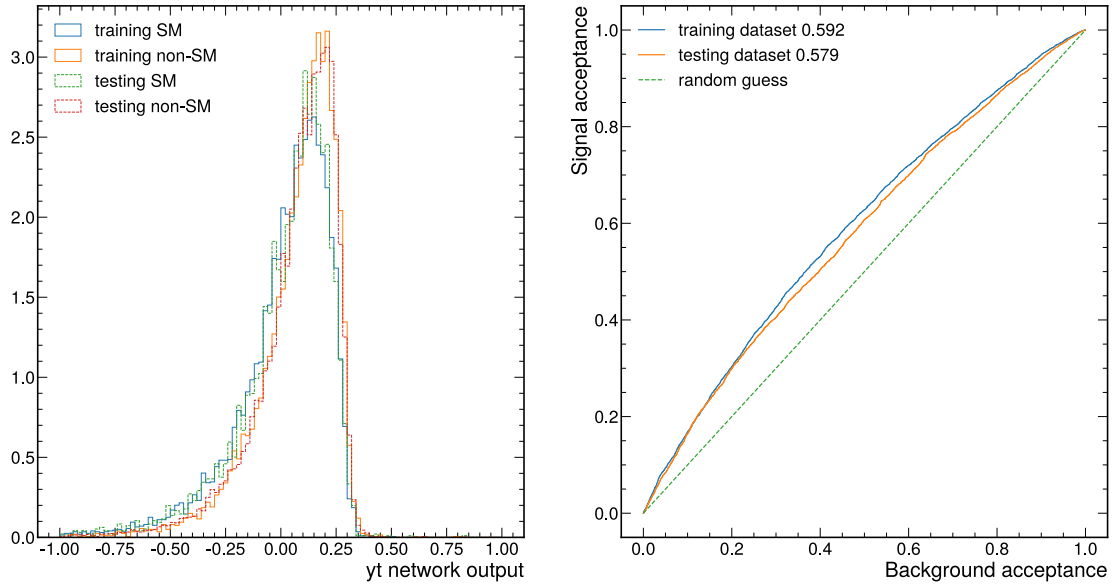


Figure 6.9: Output distribution (left) and ROC curve (right) of $NN-y_t$, calculated from training and testing datasets, *normalised*.

To ensure that the performance of the network is stable across training and testing datasets, the network is trained over 20 training trials, each with a different random seed for the training dataset. Figure 6.8 shows the evolution of the area under the ROC curve, with error bands representing one standard deviation of the value. Based on the AUC value and their errors, it can be seen that the performance of $NN-y_t$ trained using this configuration is stable over the training and testing dataset.

Figure 6.9 shows the *normalised* output distribution and ROC curve of $NN-y_t$, calculated from the training and testing datasets. The output distributions are normalised since the number of events in the training dataset and testing dataset are not the same, and normalising the distributions better visualises the distribution behaviour. As with Figure 6.8, both plots show comparable performance between training and testing datasets, while the performance gained from the network is barely better than random guesses, since the area under the ROC curve from the

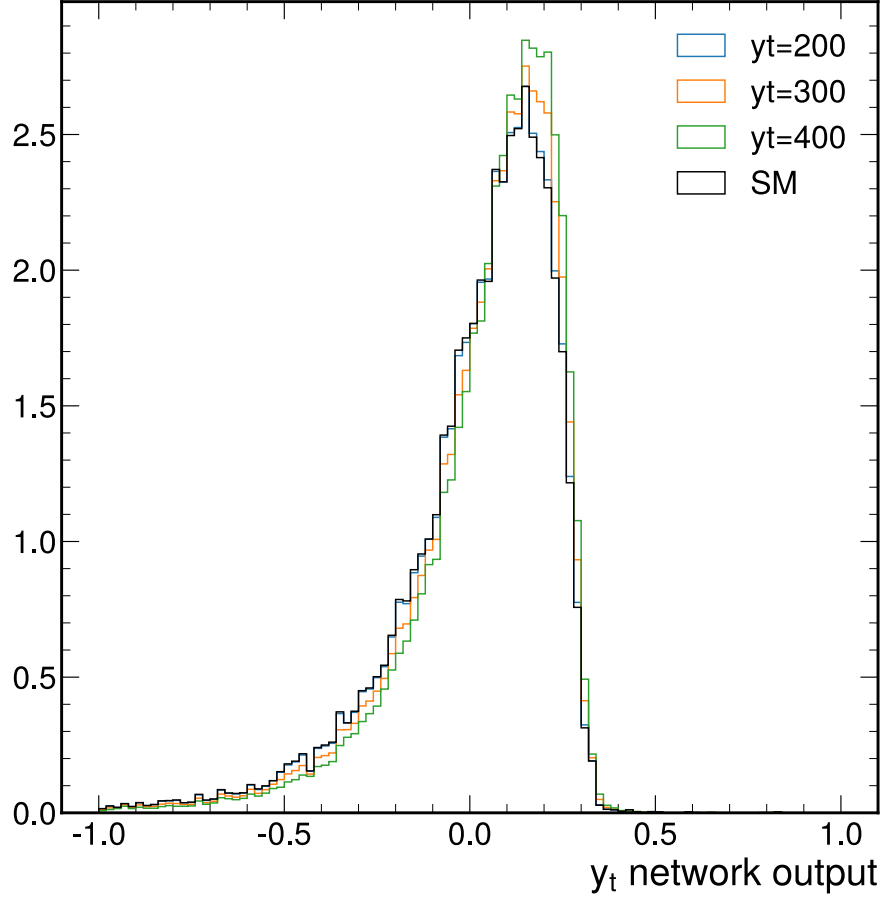


Figure 6.10: Output distribution of $NN-y_t$, calculated on $t\bar{t}t\bar{t}$ dataset with $y_t/y_t^{\text{SM}} = 1.16, 1.74, 2.32$ (shown in the plot as $y_t = 200, 300, 400$, which are the value used in generator configurations) and SM value, *normalised*.

training and testing dataset are 0.592 and 0.579 respectively, a little larger than the area under the random guess curve of 0.5 as depicted in the dashed line in Figure 6.9.

However, the change of y_t values should also affect the normalisation of the histograms, since y_t can affect the cross section of four-top quark production as well³. To visualise this, Figure 6.10 shows the *normalised* distribution of $NN-y_t$ calculated from the entire $t\bar{t}t\bar{t}$ dataset with $y_t/y_t^{\text{SM}} = 1.16, 1.74, 2.32$ and SM value. These values of y_t are chosen since the upper limit of y_t/y_t^{SM} , as measured by

³See Section 1.8.3 for the measurement of y_t based on the cross section of four-top quark production, performed after the search in same-sign dilepton and multilepton final states [65].

Ref. [65], is $y_t/y_t^{\text{SM}} < 1.7$, the value of $y_t/y_t^{\text{SM}} = 1.74$ is chosen along with $y_t/y_t^{\text{SM}} = 1.16$ and 2.32 to better visualise the changes in the shape of the distribution. As shown in the figure, the *shape* of the distribution changes slightly with the different values of y_t , which can be useful for the direct measurement of y_t later.

6.7 Direct measurement of y_t from $t\bar{t}t\bar{t}$ using NN- y_t

Based on the output distribution from NN- y_t , the direct measurement of y_t can be performed. In this study, y_t is measured using the distribution of $t\bar{t}$, $t\bar{t}H$, and $t\bar{t}t\bar{t}$ processes, and will be normalised based on the 2017 integrated luminosity value at 49.81 fb^{-1} .

As both $t\bar{t}$ and $t\bar{t}H$ processes are the most important backgrounds to the search for the $t\bar{t}t\bar{t}$ process, we can expect to see these two main backgrounds in the event distributions used to measure y_t . Furthermore, as the $t\bar{t} + b\bar{b}$ and $t\bar{t}H$ normalisation uncertainties have the most effect on the search for $t\bar{t}t\bar{t}$ as described in Section 5.6, these uncertainties can be expected to have impacts on the y_t measurement using the output distribution of NN- y_t .

However, including only two dominant backgrounds and the most important uncertainties associated with only three processes in this study makes it less realistic, as the opposite-sign dilepton final state can have other minor backgrounds and other uncertainties associated with it (as shown in Section 4.4) as well. Furthermore, full analyses such as the one in the opposite-sign dilepton final state also include additional uncertainties that are not present in this work. Therefore, the results presented in this section may not be comparable with full analyses, and can only

be compared to other results derived from the same assumption but with different methods.

6.7.1 Output distribution construction

The output distribution that will be used for direct y_t measurement will require events having at least 4 jets, 2 b-tagged jets, and two leptons with opposite charges. Events will also be required to have the output for $t\bar{t}H$ and $t\bar{t}$ nodes to be lower than 0.01 and 0.005 respectively. This requirement is designed to remove the vast amount of $t\bar{t}$ and $t\bar{t}H$ events since both processes have a much higher cross section and may cause higher distribution normalisation than $t\bar{t}t\bar{t}$ distribution. Since the outputs of NN-c have a softmax activation function, as described in Section 6.5, the outputs of all three nodes should have the sum of 1, and events passing the above requirement on these two outputs are very likely to belong to the $t\bar{t}t\bar{t}$ process. The binning of the distribution is chosen to emphasise the *shape* difference between each distribution at different values of y_t/y_t^{SM} .

Figure 6.11 shows the distribution of $t\bar{t}$, $t\bar{t}H$, and $t\bar{t}t\bar{t}$ processes, normalised to 2017 integrated luminosity of 49.81 fb^{-1} , which already shows that the harsh cut on classification model output has already eliminated most of the $t\bar{t}H$ events. $t\bar{t}t\bar{t}$ event yields for $y_t/y_t^{\text{SM}} = 2.32$ can also be seen to be higher than both backgrounds combined in some histogram bins as well. This is due to very strict preselection criteria based on the outputs from NN-c, and the output of NN- y_t favouring events more sensitive to changes in y_t .

Figures 6.12 and 6.13 show the effect of the binning towards the shape difference for the distribution. As with Figure 6.10, three values of $y_t/y_t^{\text{SM}} = 1.16, 1.74, 2.32$ along with the SM value are chosen for these plots. In Figure 6.12, the $t\bar{t}t\bar{t}$ yield in each bin for each y_t/y_t^{SM} value is divided by the yield from

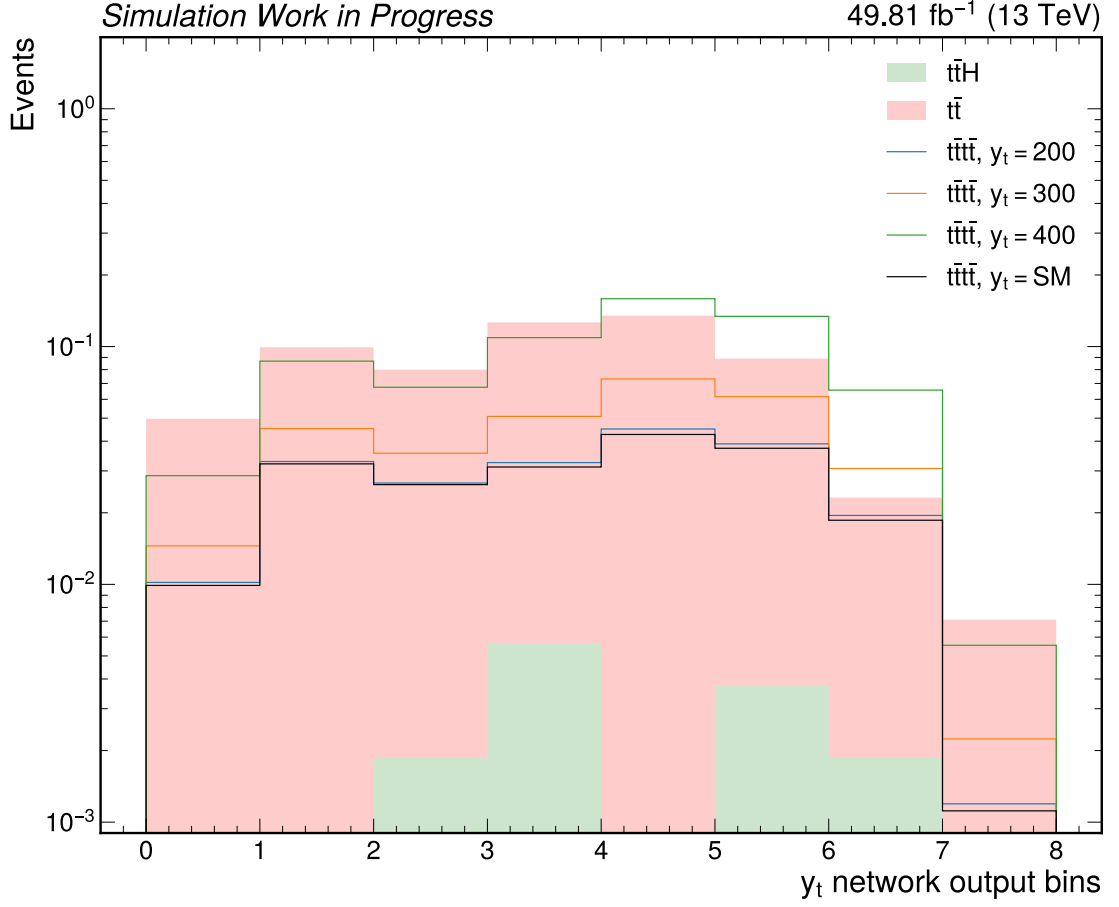


Figure 6.11: Output distributions of $NN\text{-}y_t$ calculated from $t\bar{t}$, $t\bar{t}H$, and $t\bar{t}t\bar{t}$ processes, normalised to 2017 integrated luminosity of 49.81 fb^{-1} .

the SM distribution. This demonstrates that each histogram bin has a different normalisation effect when the y_t value is varied. On the other hand, in Figure 6.13, the whole $t\bar{t}t\bar{t}$ distribution for each y_t/y_t^{SM} value is normalised to 100 events. This shows the different normalisation effect compared to Figure 6.12 and demonstrates that the *shape* of the distribution itself changes regardless of the normalisation of the distribution, which changes with different values of y_t .

The measurement of y_t based on these distributions is performed based on the assumption that since many of the numerous tree-level diagrams (shown in Ref. [66]) of four-top quark production have *two* $t\bar{t}H$ vertices, the event yield per each histogram bin can be fitted into the fourth-order polynomial of y_t (since

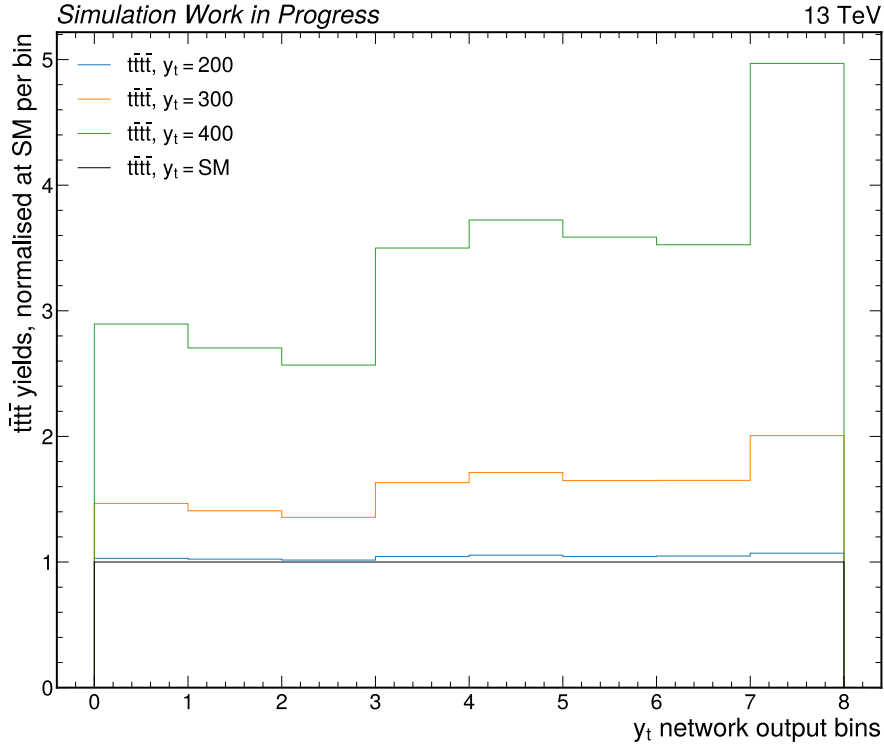


Figure 6.12: Output distributions of NN- y_t calculated from $t\bar{t}t\bar{t}$ process, where each bin in the histogram is normalised based on the event yield of SM y_t value. Note that the normalisation method applied in this figure is different from Figure 6.13.

the amplitude of that tree-level diagram should be proportional to y_t^4 for two $t\bar{t}H$ vertices, as shown in Equation 1.109). The $t\bar{t}t\bar{t}$ yield from each bin in the histogram varies with y_t , and for each histogram bin, they are fitted with fourth-order polynomials to y_t value.

Figure 6.14 shows the distribution of the expected $t\bar{t}t\bar{t}$ event yields in each histogram bin, fitted with fourth-order polynomials in y_t (as illustrated in Equation 1.109). As seen from the figure, $t\bar{t}t\bar{t}$ event yields in all histogram bins can be fitted with an excellent agreement to the y_t^4 dependence, as indicated with R^2 value⁴ of 1.000.

⁴Let y_i and f_i be the actual data point and the fitted values respectively. The coefficient of determination, or R^2 , is calculated as $R^2 = 1 - (\sum_i (y_i - f_i)^2) / (\sum_i (y_i - \bar{y})^2)$, where $\bar{y}$ is the mean of all data points y_i . $R^2 = 1$ indicates that the model fits the data points perfectly.

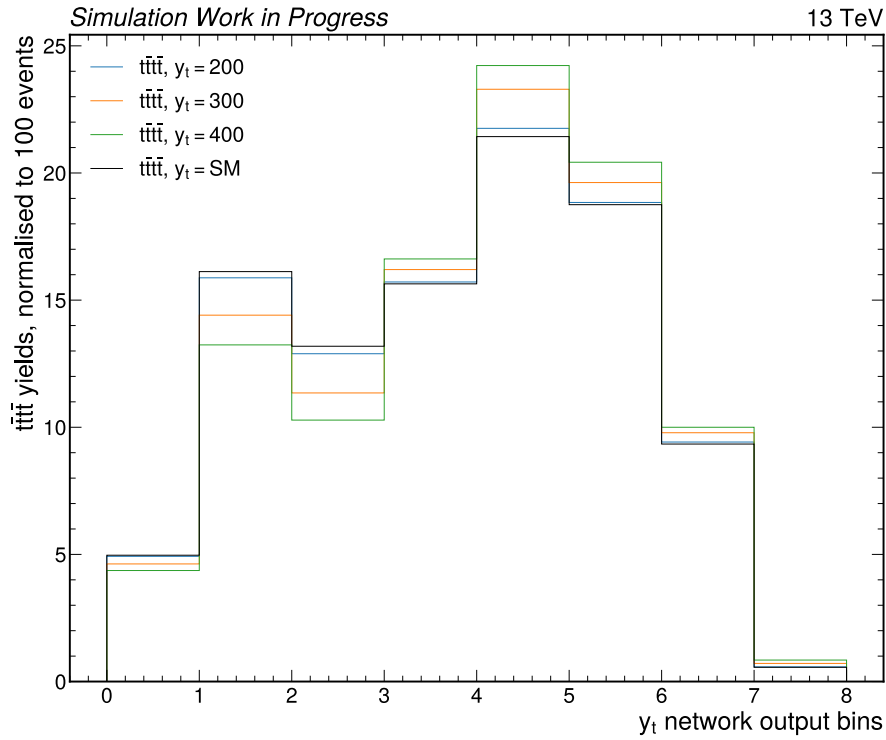


Figure 6.13: Output distributions of $NN-y_t$ calculated from $t\bar{t}t\bar{t}$ process, where each distribution from different y_t value is normalised to 100 events. Note that the normalisation method applied in this figure is different from Figure 6.12.

Since the $t\bar{t}t\bar{t}$ yields can be parameterised using fourth-order polynomials of y_t , we can treat y_t as a separate parameter that can be measured in the same fashion as signal strength μ , which will be discussed later.

6.7.2 Systematic uncertainties included in the measurement

The systematic uncertainties included in this study are as follows:

- Luminosity uncertainty, calculated as 2.3% for 2017 data. (See Section 5.4.2.1)

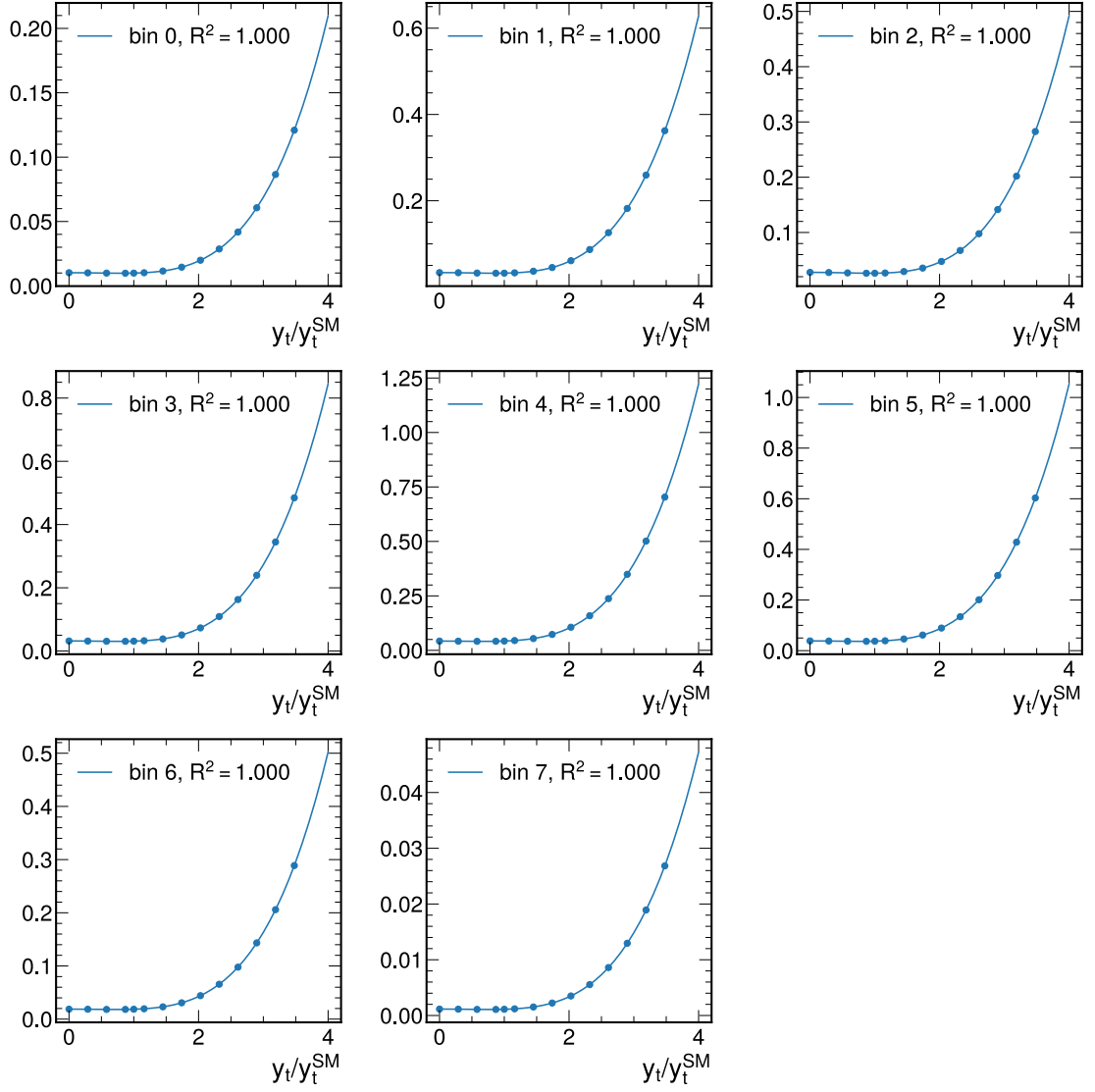


Figure 6.14: Fourth-order polynomial fitting to y_t for $t\bar{t}t\bar{t}$ yield per each histogram bin. The x -axis represents the ratio between y_t and SM y_t value, or y_t/y_t^{SM} . Notice that the fitting shows excellent agreement with $R^2 = 1.000$.

- Jet energy scale and jet energy resolution uncertainties, applied and correlated across the $t\bar{t}$, $t\bar{t}H$, and $t\bar{t}t\bar{t}$ processes. (See Section 5.4.3.1)
- Renormalisation (μ_R) and factorisation (μ_F) scale uncertainties, applied and uncorrelated across all processes. (See Section 5.4.1.5) This uncertainty is split into three variations as follows:
 - Varying μ_R , fixed μ_F
 - Varying μ_F , fixed μ_R
 - Varying both μ_R and μ_F
- Initial and final state radiation (ISR and FSR) uncertainties, applied and uncorrelated across the $t\bar{t}$, $t\bar{t}H$, and $t\bar{t}t\bar{t}$ processes. (See Section 5.4.1.3)
- $t\bar{t} + b\bar{b}$ normalisation uncertainty, set as 35% on the $t\bar{t}$ process. (See Section 5.4.1.1)
- Normalisation uncertainties on $t\bar{t}$ (12%) and $t\bar{t}H$ (20%) processes. (See Section 5.4.1.2)
- Normalisation uncertainty on $t\bar{t}t\bar{t}$, calculated based on the theoretical cross section ($13.4^{+1.0}_{-1.8}$ fb [63]) at 14%. This uncertainty is not included during the $t\bar{t}t\bar{t}$ cross section measurement, which will be discussed in Section 6.8.

Figures 6.15 and 6.16 show the uncertainty templates for renormalisation and factorisation scale, and ISR uncertainties applied on all three processes respectively.

Upon inspection, uncertainty templates for JES, JER, and FSR uncertainties are found to have high fluctuations. To fix this fluctuation, the LOWESS smoothing algorithm [169; 170], used in the search for four-top quark production in single-lepton final state in Run 2 (See Section 4.3), is applied on these uncertainty templates, and the uncertainty effects per bin are symmetrised and restricted to

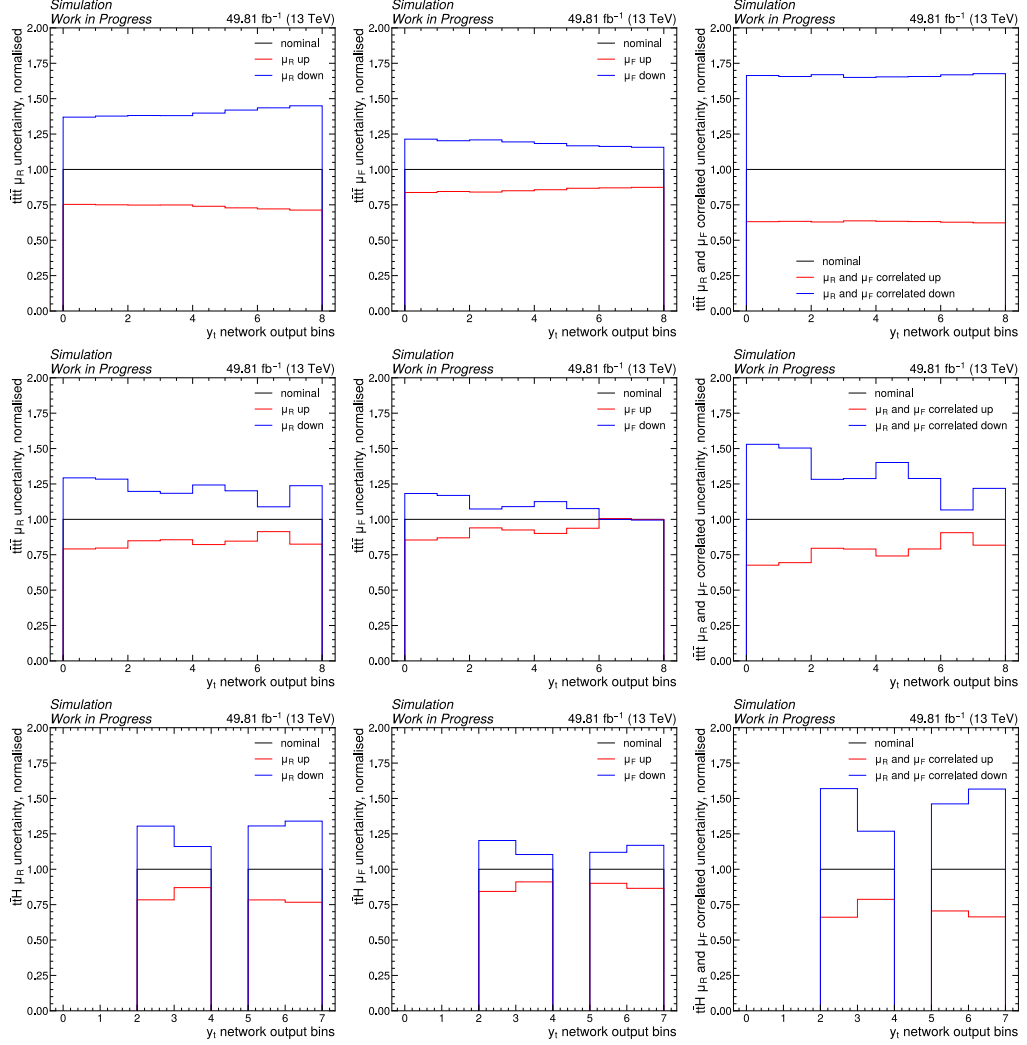


Figure 6.15: Renormalisation and factorisation uncertainty templates for $tt\bar{t}t$ (upper row), $tt\bar{t}$ (middle row), and $tt\bar{t}H$ (lower row).

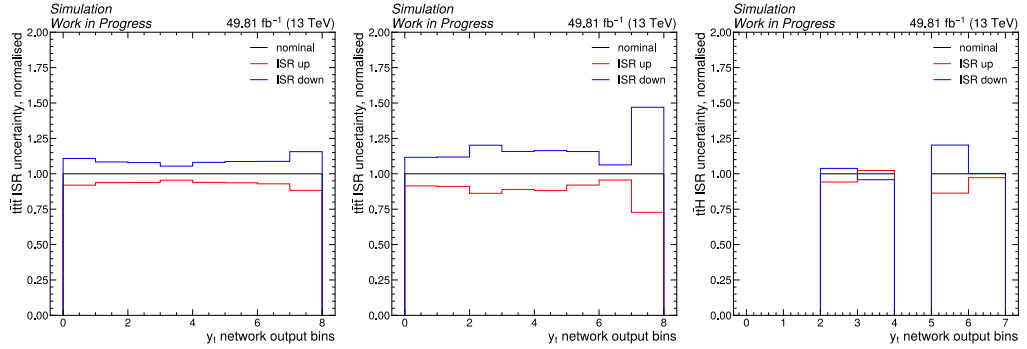


Figure 6.16: ISR uncertainty templates for $tt\bar{t}t$ (left), $tt\bar{t}$ (middle), and $tt\bar{t}H$ (right).

$\pm 50\%$ of the nominal value as the most conservative estimate of the amount of uncertainty and to prevent the nuisance parameter pushing the event yield to be negative when the nuisance parameter has the value of less than -1. Figures 6.17, 6.18, and 6.19 show the uncertainty templates for JES, JER, and FSR uncertainties before and after applying the LOWESS algorithm respectively.

Furthermore, all uncertainty templates for $t\bar{t}t\bar{t}$ are applied using the nominal SM value of y_t . However, y_t may affect some uncertainty templates such as JES and JER. This is because these uncertainty templates are generated from distributions with different configurations, and events with higher amounts of weights at higher y_t values can alter the distributions both in nominal and up/down variations. Thus, the behaviour of these uncertainties may change with y_t values different from SM. Nonetheless, the dependence of uncertainty templates from $t\bar{t}t\bar{t}$ upon y_t is not considered in this analysis and should be confirmed and quantified in a future study.

6.7.3 Direct measurement using the Asimov dataset

The direct measurement performed in this study is done using the Asimov dataset, which is constructed by using $t\bar{t}t\bar{t}$ signal strength $\mu = 1$ and y_t value at SM. All measurements in this chapter will be using this Asimov dataset configuration as if it is the actual data, in the same fashion as measuring expected signal strength and upper limits as described in Section 5.2.3.

As $t\bar{t}t\bar{t}$ samples are generated with event weights sampled in the range of $y_t/y_t^{\text{SM}} \in [0, 3.48]$, we do not know the exact behaviour of the distribution in y_t values beyond this range. Therefore, the results presented in this chapter will focus only on the range of $y_t/y_t^{\text{SM}} \in [0, 3.5]$. This range of y_t/y_t^{SM} values considered in this study is reasonable because the upper limit of y_t/y_t^{SM} as determined from Ref. [65] is already $y_t/y_t^{\text{SM}} < 1.7$, and the range used in this study already covers twice the upper limit.

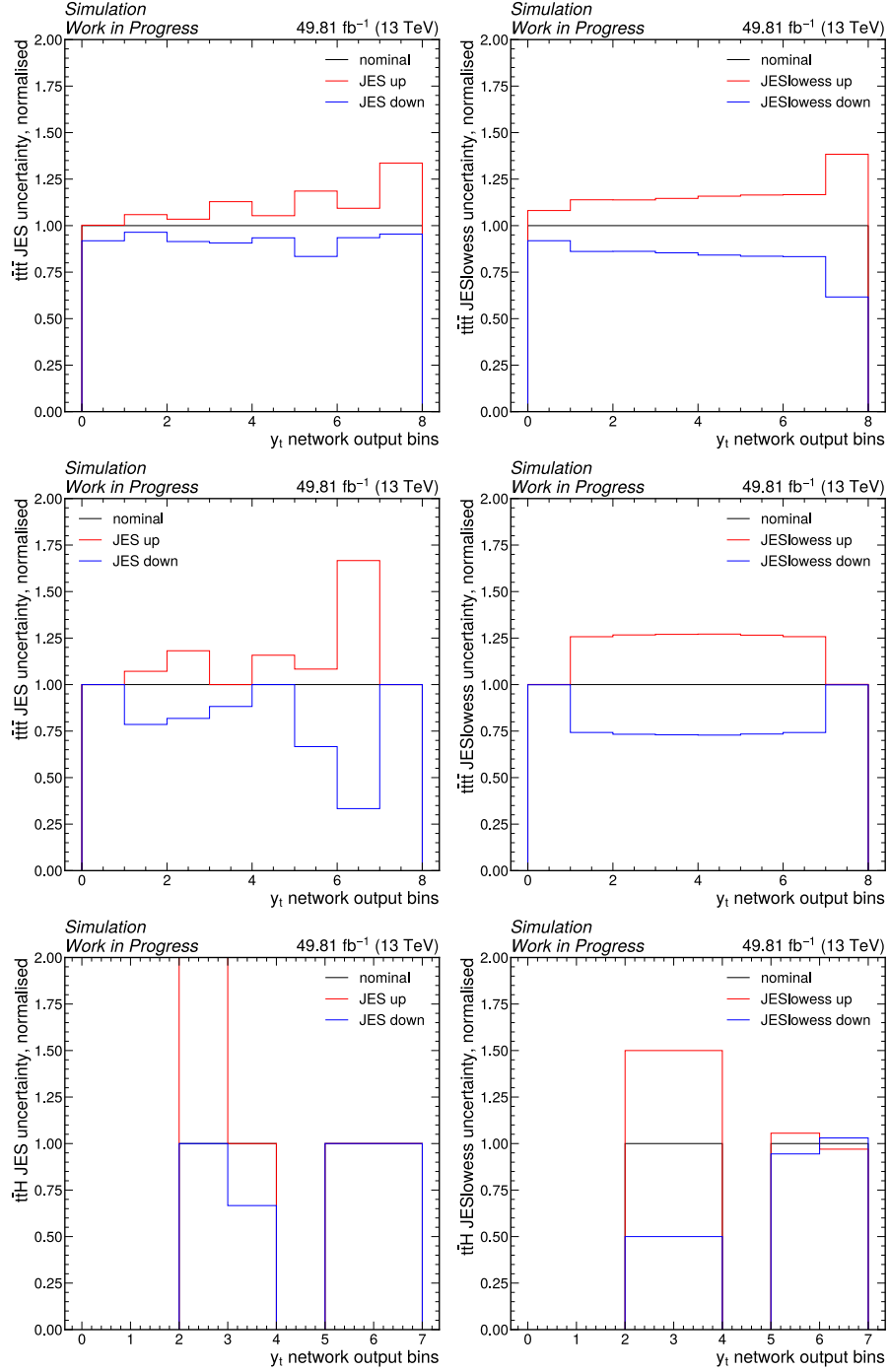


Figure 6.17: JES uncertainty templates for $t\bar{t}t\bar{t}$ (upper row), $t\bar{t}$ (middle row), and $t\bar{t}H$ (lower row), before (left) and after (right) applying LOWESS smoothing algorithm [169; 170].

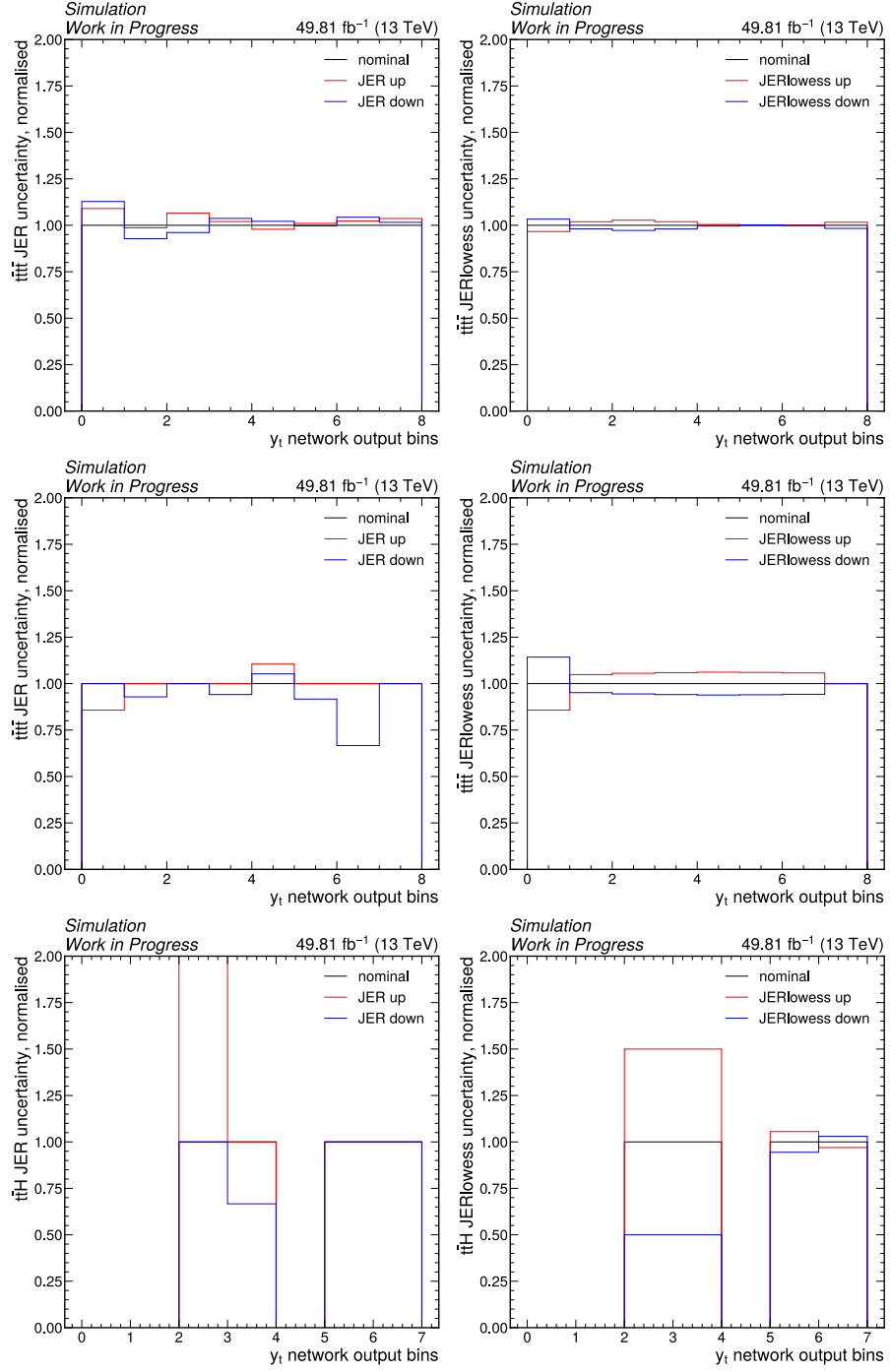


Figure 6.18: JER uncertainty templates for $t\bar{t}t\bar{t}$ (upper row), $t\bar{t}$ (middle row), and $t\bar{t}H$ (lower row), before (left) and after (right) applying LOWESS smoothing algorithm [169; 170].

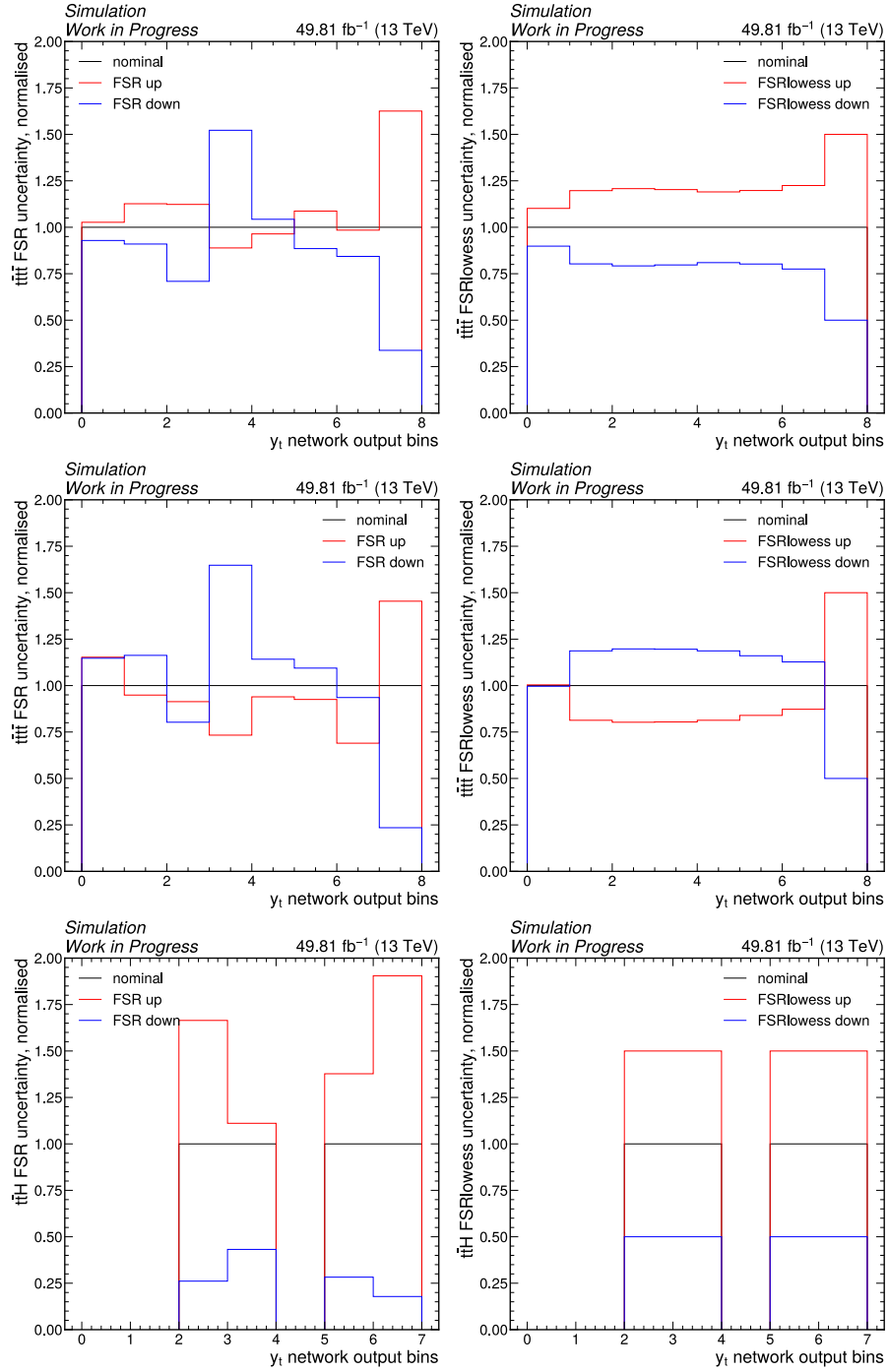


Figure 6.19: FSR uncertainty templates for $t\bar{t}t\bar{t}$ (upper row), $t\bar{t}$ (middle row), and $t\bar{t}H$ (lower row), before (left) and after (right) applying LOWESS smoothing algorithm [169; 170].

The measurement for y_t is performed by fitting y_t/y_t^{SM} along with all other nuisance parameters in a similar way as measuring signal strength μ of the four-top quark production search. As the fitting is performed using the Asimov dataset, the best fit value of y_t/y_t^{SM} is supposed to be at $y_t/y_t^{\text{SM}} = 1$. However, the error ranges may be modified due to several nuisance parameters introduced in the fit, as with measuring expected signal strength by using the Asimov dataset with $\mu = 1$. Since $t\bar{t}H$ yields should also be scaled with y_t^2 , adjusting the value of top Yukawa coupling during the fit will cause $t\bar{t}H$ event yields to change as well as all other nuisance parameters affecting $t\bar{t}H$ event yields. In this study, the y_t/y_t^{SM} parameter will be treated as one parameter affecting both $t\bar{t}t\bar{t}$ and $t\bar{t}H$ event yields simultaneously, since there is only one top Yukawa coupling in SM affecting $t\bar{t}H$ vertices in both processes.

To confirm the results, the measurement is also done by performing negative log-likelihood (NLL) scans over y_t/y_t^{SM} . For each value of y_t/y_t^{SM} , all nuisance parameters representing uncertainties are fitted, and a NLL value is calculated. Since this measurement is performed using the Asimov dataset with $y_t/y_t^{\text{SM}} = 1$, the NLL value should be at the minimum at the SM point.

Figure 6.20 shows the NLL scan over y_t/y_t^{SM} based on the distribution shown in Figure 6.9. Due to the strict preselection criteria, the distribution contains mostly $t\bar{t}t\bar{t}$ and $t\bar{t}$ events with a sparse amount of $t\bar{t}H$ events included. Based on the direct measurement, the upper limit at 68% confidence level is

$$\frac{y_t}{y_t^{\text{SM}}} = 1_{-1.00}^{+2.00} \quad (6.2)$$

6.7.3.1 Adding $t\bar{t}H$ -enriched regions to the direct measurement

Since the input distribution virtually has no $t\bar{t}H$ event yields, the y_t measurement is not constrained heavily by the $t\bar{t}H$ process even if we assume that the y_t parameter

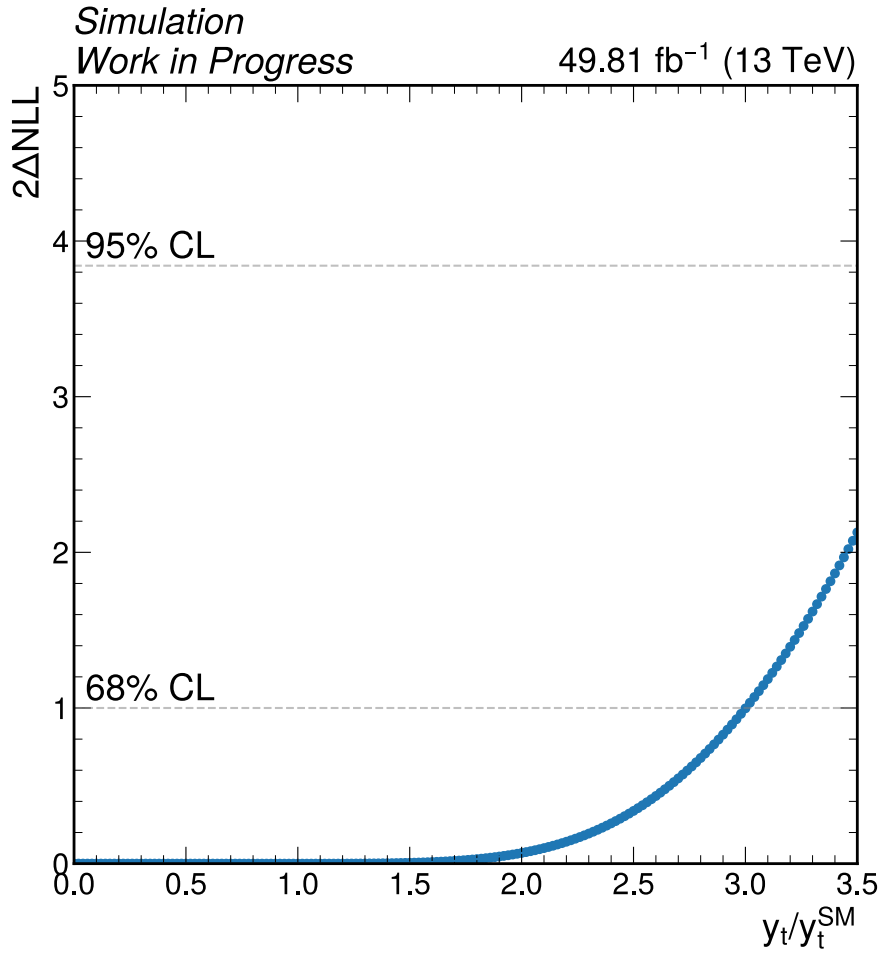


Figure 6.20: Negative log-likelihood scans over y_t/y_t^{SM} , showing the range of y_t/y_t^{SM} values under 68% confidence level.

affects both yields of $t\bar{t}H$ and $t\bar{t}t\bar{t}$. If $t\bar{t}H$ events are included in the fit, the contribution to the constraint by the $t\bar{t}H$ process will eclipse the contribution by the $t\bar{t}t\bar{t}$ process.

To prove this, another template containing events mostly from the $t\bar{t}H$ process (which will be called the $t\bar{t}H$ -enriched region) is constructed and added to the direct measurement. Events passing into this template are required to have the output for the $t\bar{t}H$ node to be greater than 0.95. This new template is orthogonal to the main template used in the direct measurement since no events can have the $t\bar{t}H$ node output greater than 0.95 and less than 0.01 at the same time. The binning of

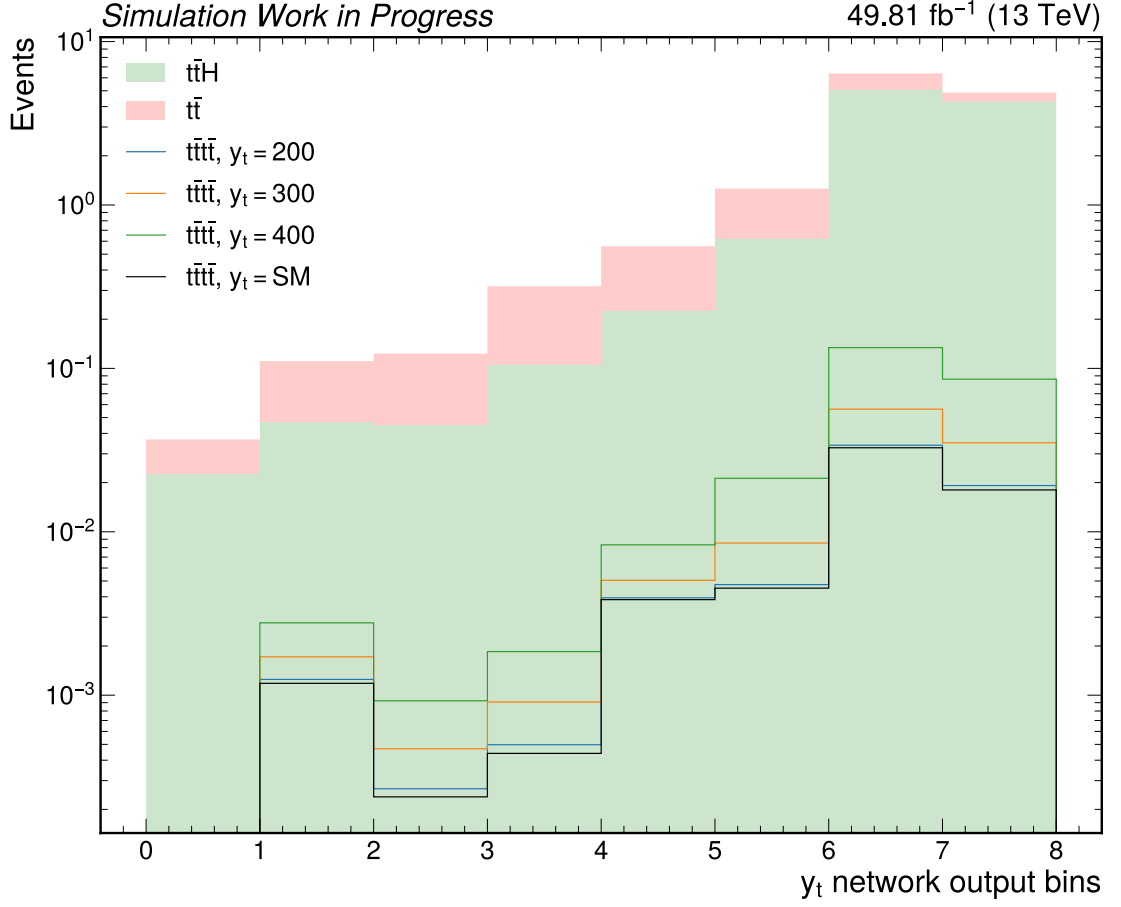


Figure 6.21: Output distributions of $NN\text{-}y_t$ calculated from $t\bar{t}$, $t\bar{t}H$, and $t\bar{t}t\bar{t}$ processes in $t\bar{t}H$ -enriched region, normalised to 2017 integrated luminosity of 49.81 fb^{-1} .

this template is the same as in the main template, and all systematic uncertainty treatments are performed in the same way as in the main template. Figure 6.21 shows the template in $t\bar{t}H$ -enriched region.

Based on the combination of the main template and $t\bar{t}H$ -enriched template, the best fit value of y_t/y_t^{SM} is measured to be

$$y_t/y_t^{\text{SM}} = 1^{+0.229\,050}_{-0.205\,425} \quad (6.3)$$

and the measurement from the $t\bar{t}H$ -enriched template only gives

$$y_t/y_t^{\text{SM}} = 1^{+0.229\,279}_{-0.206\,077} \quad (6.4)$$

By comparing the uncertainty ranges from the two measurements, we can see that the main template containing $t\bar{t}t\bar{t}$ events contributes to less than 1% of the uncertainty range of y_t/y_t^{SM} measurement in Asimov dataset⁵.

6.7.3.2 Two-dimensional fit of signal strength μ and y_t/y_t^{SM}

One interesting measurement to perform on this output distribution is to do a simultaneous fit between the signal strength μ versus y_t/y_t^{SM} , where both μ and y_t/y_t^{SM} are set to float simultaneously. This simultaneous fit is different from determining the value of y_t/y_t^{SM} from the cross section measurement, as the latter performs the cross section measurement first and then compare it with the predicted cross section value with y_t/y_t^{SM} variations.

As we have seen so far, the cross section of $t\bar{t}t\bar{t}$ can be altered by signal strength μ and y_t/y_t^{SM} . μ alters the *normalisation* of $t\bar{t}t\bar{t}$ distribution as a whole, while y_t/y_t^{SM} alters both *normalisation* and *shape* of $t\bar{t}t\bar{t}$ distribution. Furthermore, the cross section predicted in Equation 1.109 does not go all the way to zero, and thus the normalisation rate based on y_t/y_t^{SM} has the minimum value larger than zero. Performing this 2D scan will allow us to see the probable region in which these two parameters can have their value.

Figure 6.22 shows a two-dimensional scan between μ and y_t/y_t^{SM} , with phase space areas representing probable values of both parameters at 68% confidence level and 95% confidence level. The regions depicted in the figure are *expected* values since this scan is performed over the Asimov dataset (with $\mu = 1$ and $y_t/y_t^{\text{SM}} = 1$).

If $\mu = 1$, the probable regions of y_t/y_t^{SM} should match the measured value of y_t/y_t^{SM} shown in Equation 6.2. On the other hand, if $y_t/y_t^{\text{SM}} = 1$, the probable regions of μ represent the normal signal strength measurement of $t\bar{t}t\bar{t}$. As seen from

⁵ $(0.206\,077 - 0.205\,425) \div (0.206\,077) = 0.0032$, and $(0.229\,279 - 0.229\,050) \div (0.229\,050) = 0.0010$.

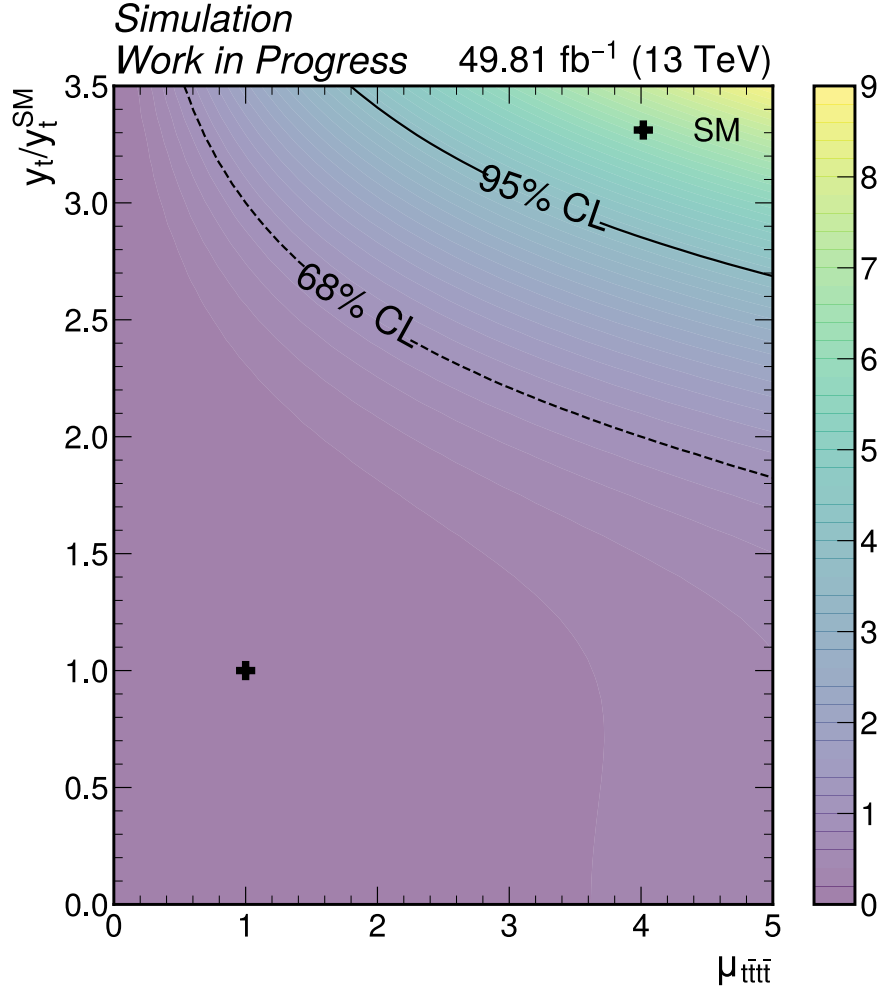


Figure 6.22: Two-dimensional NLL scan of $\mu_{t\bar{t}t\bar{t}}$ versus y_t/y_t^{SM} , based on NN- y_t output distribution used for the direct measurement.

the diagram, the probable values of μ largely vary from $\mu \in [0, 5]$, which is the region focused in the scan. This however does not mean μ should be within the region, and it is still possible that μ may have a probable value outside of the region in the scan. This also shows that the $t\bar{t}t\bar{t}$ distribution used for the direct measurement is not suitable for cross section measurement, as the probable value for μ is so large it will give a very large uncertainty.

Based on the 68% CL region, it is clear that μ and y_t/y_t^{SM} both cannot have high values at the same time. Also, higher values of y_t/y_t^{SM} can constrain the value

of μ more. This is due to the cross section of $t\bar{t}t\bar{t}$ is proportional to $(y_t/y_t^{\text{SM}})^4$ (as seen in Equation 1.109), and linearly proportional to μ only. However, if $\mu = 0$, the whole $t\bar{t}t\bar{t}$ distribution collapses to zero, and y_t/y_t^{SM} can only be constrained by $t\bar{t}H$ distribution.

6.7.3.3 Two-dimensional fit of y_t/y_t^{SM} for $t\bar{t}H$ and $t\bar{t}t\bar{t}$

One idea that is worth considering is what would happen if the values of y_t differ from $t\bar{t}H$ and $t\bar{t}t\bar{t}$ processes, and if there is any correlation between the two values. This is because up until now we have assumed that $t\bar{t}H$ interactions in $t\bar{t}t\bar{t}$ and $t\bar{t}H$ should be affected by a single parameter y_t . By splitting y_t into two parameters affecting $t\bar{t}t\bar{t}$ and $t\bar{t}H$, we can check if the coupling values have different strengths in two different processes, and in turn, conclude if the top Yukawa coupling strength in both processes are equal. To test this hypothesis, a two-dimensional scan of two independent y_t parameters is performed, where one y_t parameter affects $t\bar{t}t\bar{t}$ event yield and another affects $t\bar{t}H$ event yield.

Figure 6.23 shows the two-dimensional scan between the two y_t parameters, using the same set of NN- y_t output distribution. As seen from the 68% confidence level region, y_t/y_t^{SM} value affected on $t\bar{t}t\bar{t}$ is more constrained, whereas y_t/y_t^{SM} value affected on $t\bar{t}H$ can be of any value within the range of $y_t/y_t^{\text{SM}} \in [0, 3.5]$. This is because the fitting distribution shown in Figure 6.9 does not have many $t\bar{t}H$ events, and the yield is covered by high statistical uncertainties.

The figure also provides a line where both parameters have equal values, which shows that the coupling has the same effect on both processes. The points along the line correspond to a hypothesis where top quark Yukawa coupling strengths are equal across $t\bar{t}t\bar{t}$ and $t\bar{t}H$. Based on this figure, the result can be collapsed into the y_t measurement shown in Equation 6.2, which is $y_t/y_t^{\text{SM}} = 1_{-1.00}^{+2.00}$.

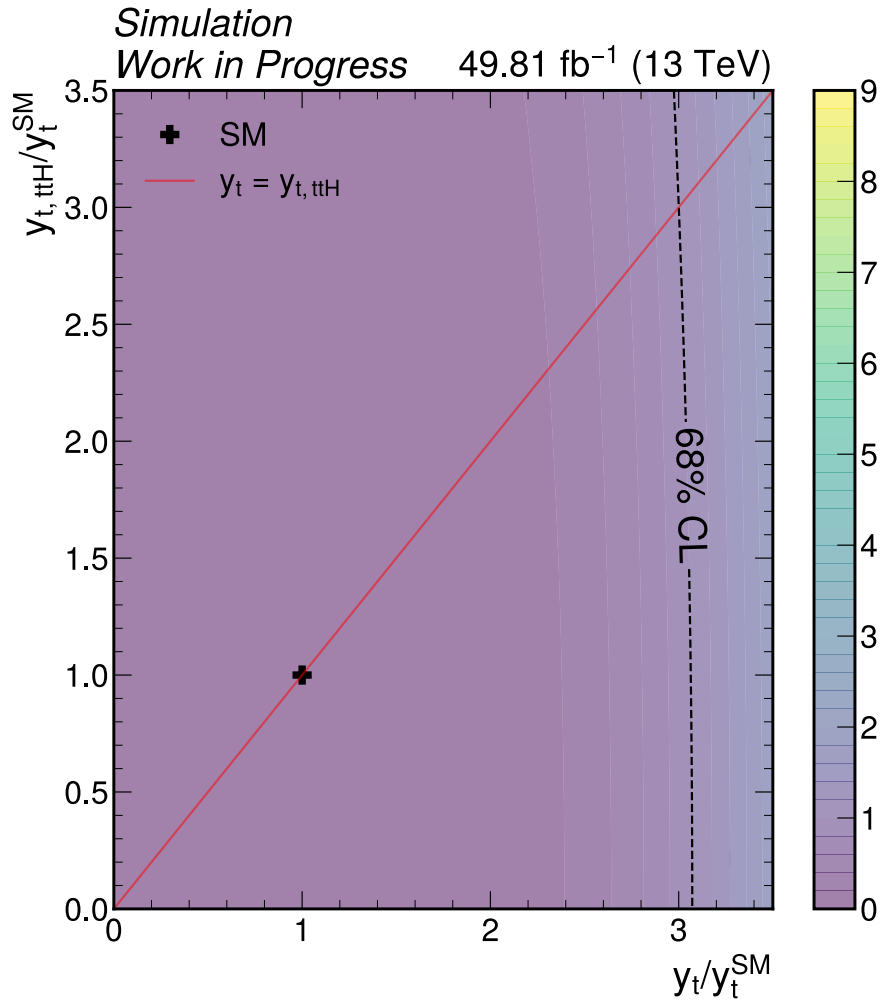


Figure 6.23: Two-dimensional NLL scan of two y_t parameters affecting $t\bar{t}t\bar{t}$ yield (x -axis) and $t\bar{t}H$ yield (y -axis), based on NN- y_t output distribution used for the direct measurement. The red line represents points where both parameters have the same value according to SM.

6.8 Indirect measurement of y_t via cross-section measurement

The indirect measurement of y_t can also be done using the cross section measurement value, which has been performed before in the search for four-top quark production in same-sign dilepton and multilepton final state using Run 2 data. The cross section of $t\bar{t}t\bar{t}$ for each y_t value can be calculated based on Equation 1.109. However, the formula presented is calculated at leading order, and $y_t/y_t^{\text{SM}} = 1$, the cross section from this formula is 9.618 fb. To match the prediction with the cross section value at 13.4 fb (as reported in Ref. [63]), the formula in Equation 1.109 is scaled up to 13.4 fb.

As with the direct measurement of y_t described in Section 6.7, all events entering the analysis must have at least 4 jets, 2 b-tagged jets, and two leptons with opposite charges. Since the indirect measurement requires the cross section measurement, a different binning scheme and requirements based on NN-c are required. Events used in this indirect measurement are required to have the $t\bar{t}t\bar{t}$ node output of NN-c to be greater than the 90th percentile value from the $t\bar{t}$ dataset. This requirement allows adequate statistics for the analysis while removing the significant amount of the $t\bar{t}$ background. Unlike the direct y_t measurement, the indirect measurement does not have to reduce the number of events from $t\bar{t}$ and $t\bar{t}H$ processes in the distribution.

Events passing all requirements are separated into 11 categories based on the number of jet and b-tagged jet multiplicities, as shown in Table 6.5. Due to low statistics as imposed by a low integrated luminosity of 49.81 fb^{-1} (representing the amount of data gathered in 2017), each jet category contains only three bins, and the binning is determined so that $t\bar{t}$ event yields are equal in all three bins, resulting

Table 6.5: Jet and b-tagged jet multiplicities used in the indirect measurement of y_t via cross section measurement.

jets	b-tagged jets		
	2	3	≥ 4
≤ 5	$N_j \leq 5$		
6	$N_j = 6$		
7	$N_j = 7, N_b = 2$	$N_j = 7, N_b = 3$	$N_j = 7, N_b \geq 4$
8	$N_j = 8, N_b = 2$	$N_j = 8, N_b = 3$	$N_j = 8, N_b \geq 4$
9	$N_j \geq 9, N_b = 2$	$N_j \geq 9, N_b = 3$	$N_j \geq 9, N_b \geq 4$

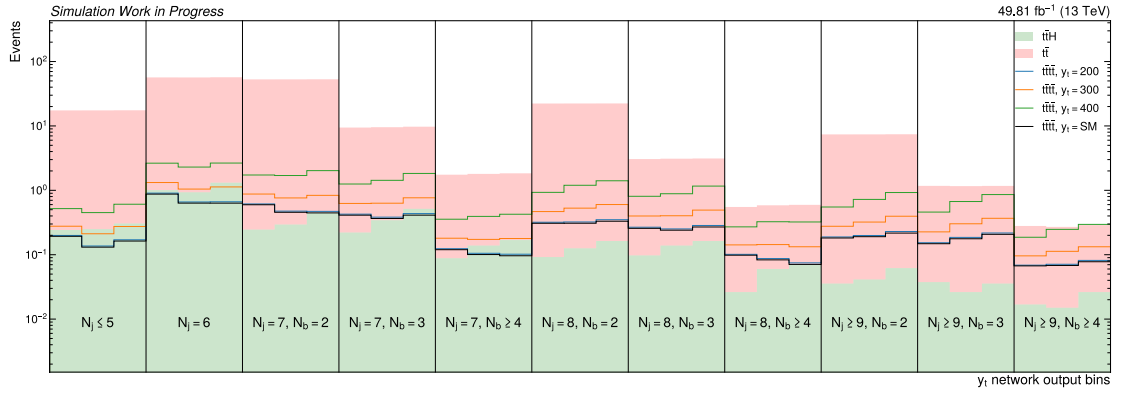


Figure 6.24: NN- y_t output distribution used in the indirect measurement of y_t via cross section measurement.

in the template containing 33 bins in total. Figure 6.24 shows the distribution of events across all jet multiplicity categories.

All systematic uncertainties used in the direct measurement, as detailed in Section 6.7.2, are also applied, except $t\bar{t}t\bar{t}$ normalisation uncertainty. Upon inspection, uncertainty templates for JES, JER, and FSR (for the $t\bar{t}H$ process) uncertainties are found to have high fluctuations. As with the direct measurement, these uncertainty templates are treated with the LOWESS smoothing algorithm [169; 170] and are symmetrised and restricted to $\pm 50\%$ of the nominal value. Figures 6.25-6.31 show the uncertainty templates that have been treated with the LOWESS smoothing algorithm, and the remaining uncertainty templates are shown in Figures 6.32-6.36.

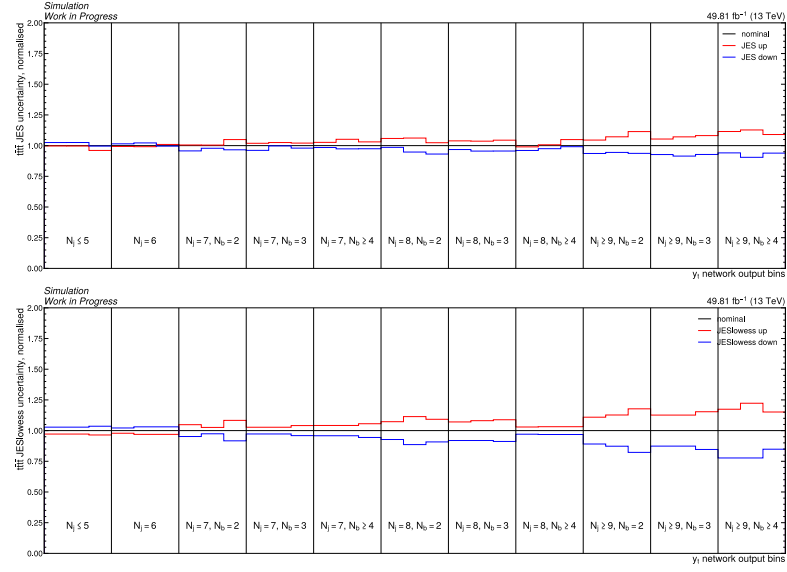


Figure 6.25: JES uncertainty template for $t\bar{t}t\bar{t}$ process, before (top) and after (bottom) applying LOWESS smoothing algorithm [169; 170].

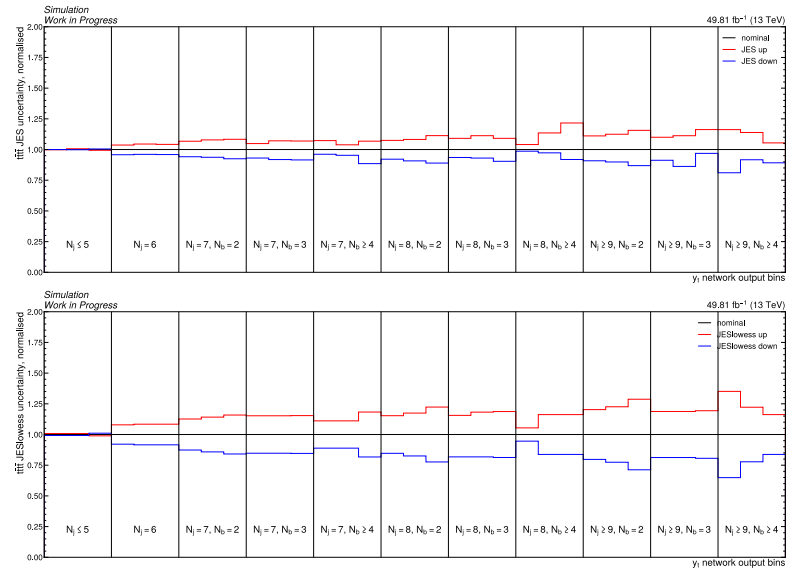


Figure 6.26: JES uncertainty template for $t\bar{t}$ process, before (top) and after (bottom) applying LOWESS smoothing algorithm [169; 170].

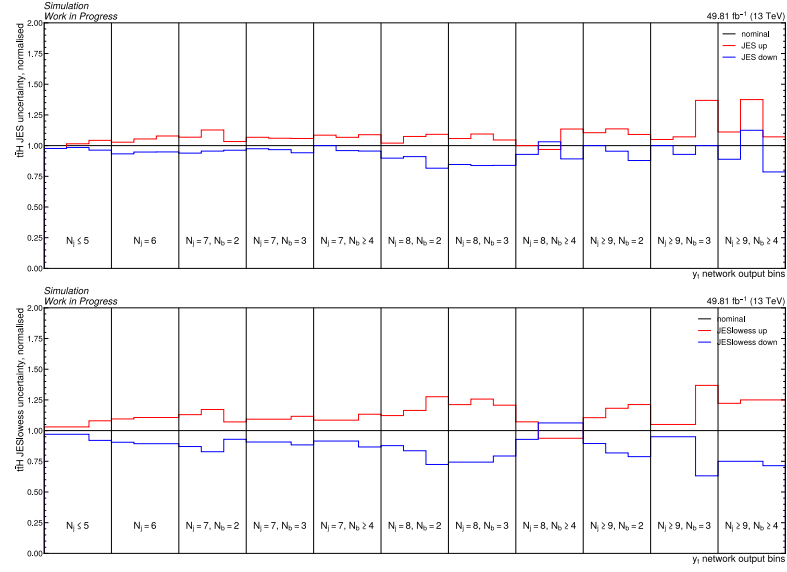


Figure 6.27: JES uncertainty template for $t\bar{t}H$ process, before (top) and after (bottom) applying LOWESS smoothing algorithm [169; 170].

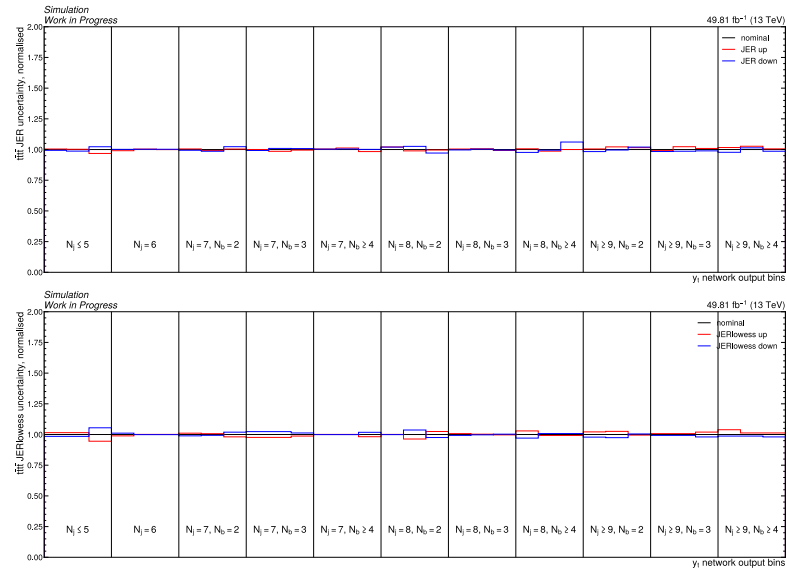


Figure 6.28: JER uncertainty template for $t\bar{t}t\bar{t}$ process, before (top) and after (bottom) applying LOWESS smoothing algorithm [169; 170].

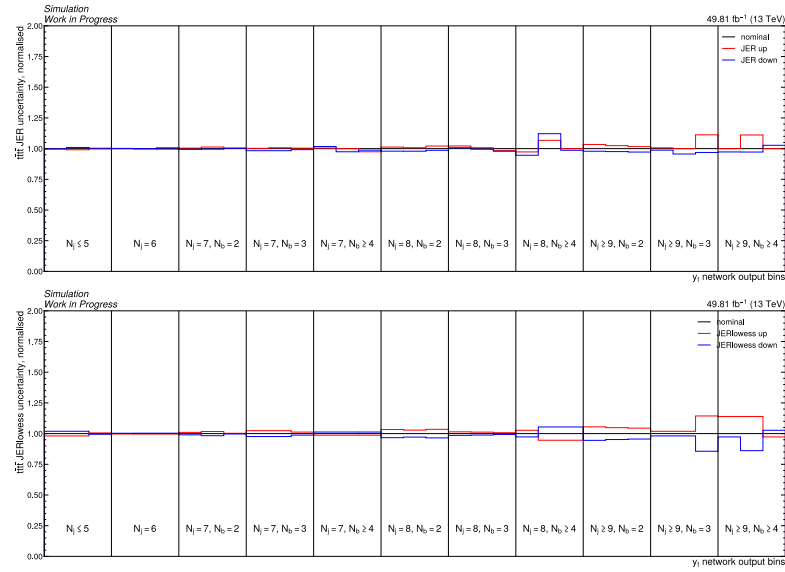


Figure 6.29: JER uncertainty template for $t\bar{t}$ process, before (top) and after (bottom) applying LOWESS smoothing algorithm [169; 170].

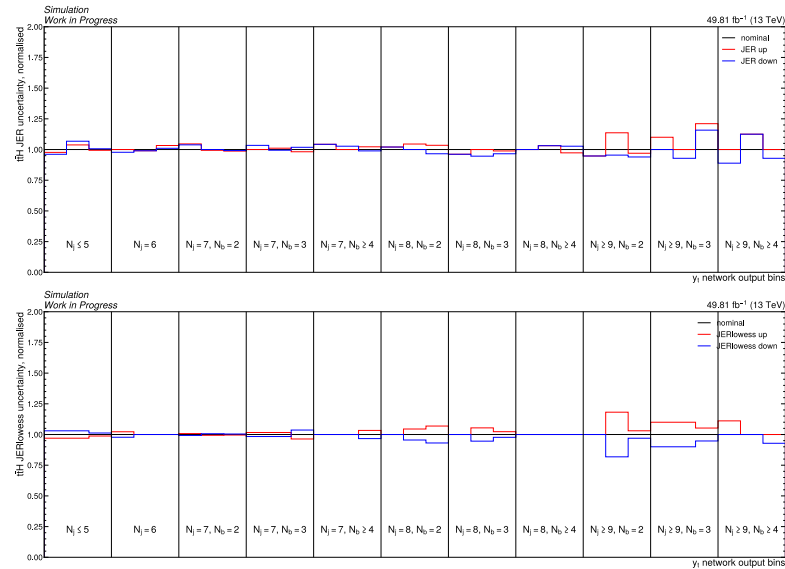


Figure 6.30: JER uncertainty template for $t\bar{t}H$ process, before (top) and after (bottom) applying LOWESS smoothing algorithm [169; 170].

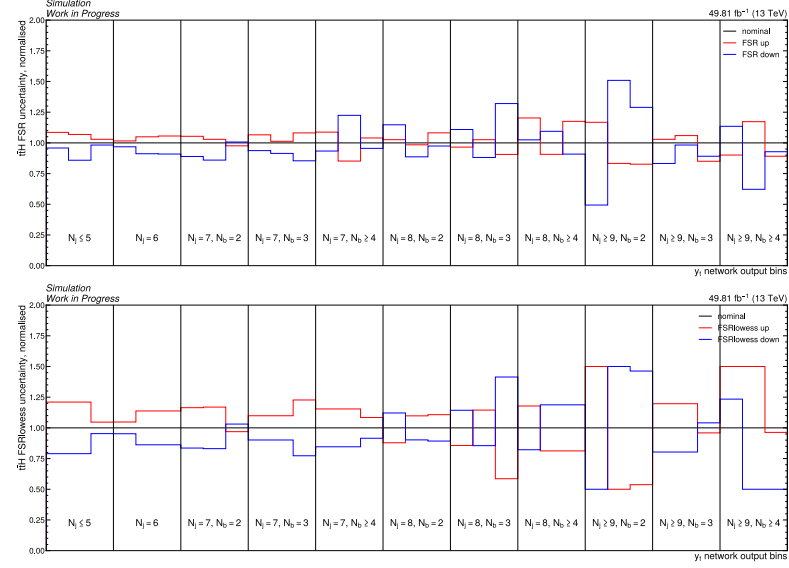


Figure 6.31: FSR uncertainty template for $t\bar{t}H$ process, before (top) and after (bottom) applying LOWESS smoothing algorithm [169; 170].

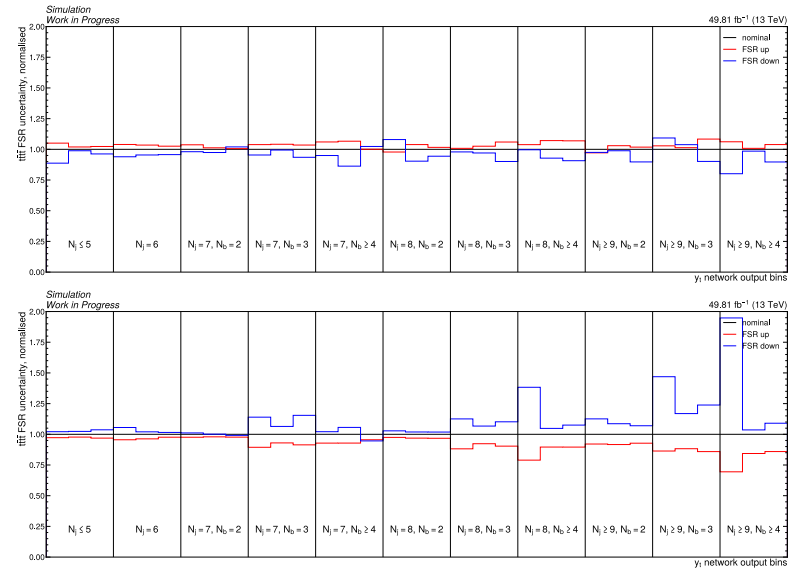


Figure 6.32: FSR uncertainty templates for $t\bar{t}t\bar{t}$ (top) and $t\bar{t}$ (bottom) processes.

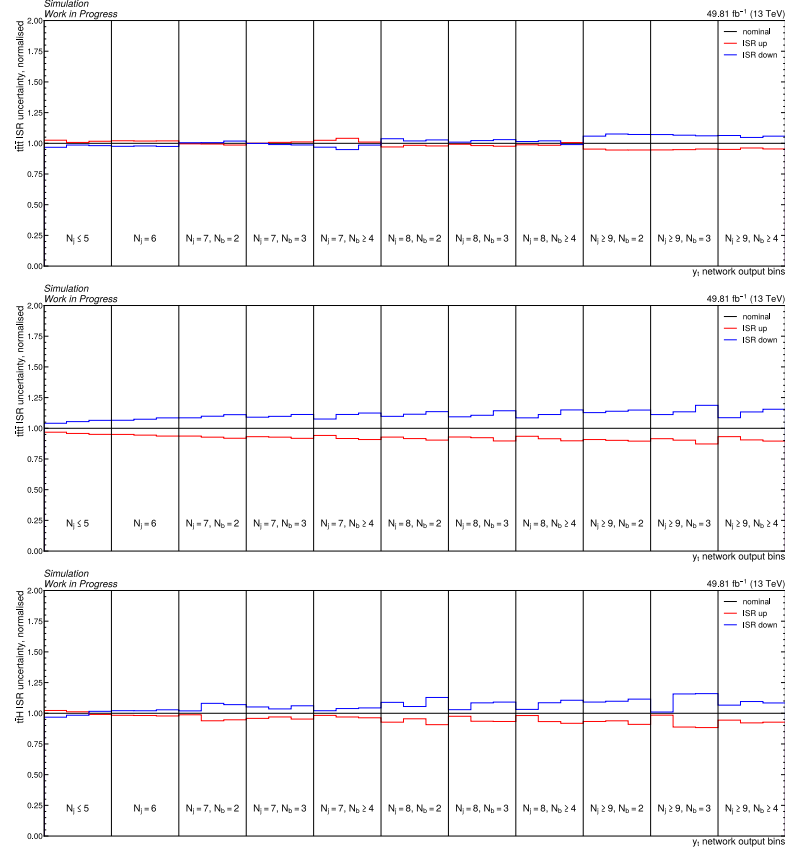


Figure 6.33: ISR uncertainty templates for $t\bar{t}t\bar{t}$ (top), $t\bar{t}$ (middle), and $t\bar{t}H$ (bottom) processes.

As with the direct measurement of y_t , the cross section measurement is performed using the Asimov dataset, where the signal strength is fixed at 1, and the value of y_t is fixed at SM value. With NN- y_t output distributions configured, the *expected* signal strength measurement based on the Asimov dataset with normalisation at 2017 integrated luminosity is

$$\mu = 1_{-2.56}^{+3.42} \quad (6.5)$$

Although this is the cross section measurement, we can easily convert the signal strength measured in this study to the cross section by multiplying with the SM cross section of 13.4 fb. This is because we have already removed the $t\bar{t}t\bar{t}$ normalisation uncertainty which can change the normalisation rate of the $t\bar{t}t\bar{t}$

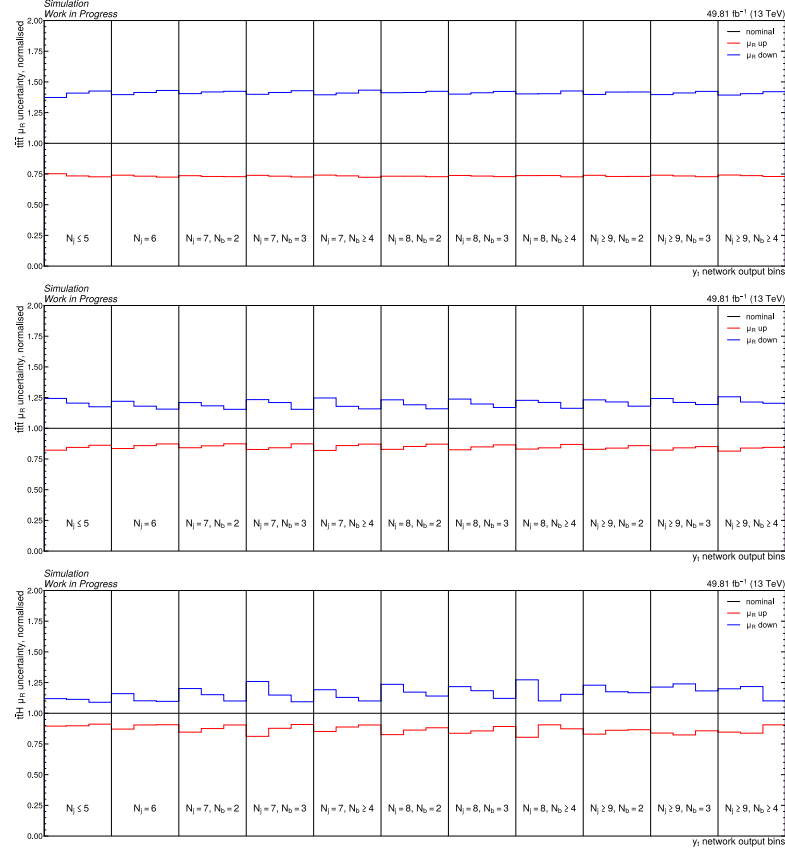


Figure 6.34: Renormalisation and factorisation uncertainty templates (varying μ_R) for $t\bar{t}t\bar{t}$ (top), $t\bar{t}$ (middle), and $t\bar{t}H$ (bottom) processes.

distribution. The above signal strength measurement is equivalent to the *expected* cross section of

$$\sigma_{t\bar{t}t\bar{t}} = 13.4^{+45.8}_{-34.3} \text{ fb}. \quad (6.6)$$

The expected signal strength shown above is worse than the *expected* cross section results from the full 2017 opposite-sign dilepton search, shown in Table 4.3, which has the expected signal strength value of $1^{+2.6}_{-2.3}$. The expected signal strength value measured in that analysis is derived by fitting distributions with more binning categories since it separates events by 15 jet categories and each of the three lepton flavour combinations. On the other hand, this study only employs 11 jet categories, uses only 2017 data, and does not separate events into three lepton flavour combinations. The full 2017 analysis also includes all minor backgrounds,

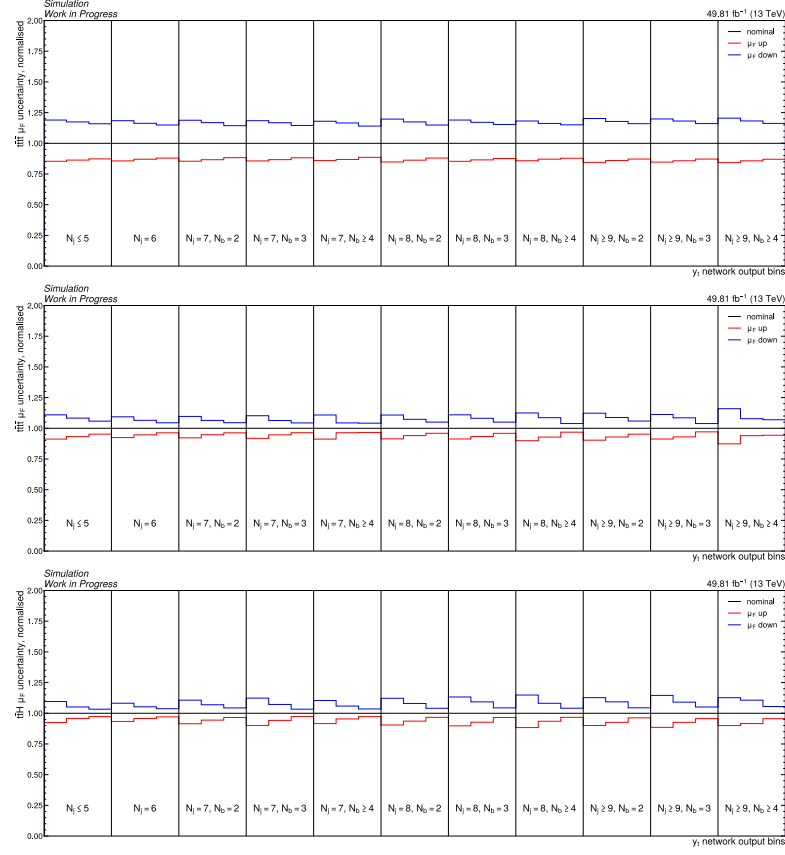


Figure 6.35: Renormalisation and factorisation uncertainty templates (varying μ_F) for $t\bar{t}t\bar{t}$ (top), $t\bar{t}$ (middle), and $t\bar{t}H$ (bottom) processes.

all uncertainties, and all necessary corrections implemented, and to make this comparison fairer, this analysis has to have all of them implemented as well. Implementing more backgrounds (even if they are minor) and more uncertainties will increase the uncertainty range of the expected signal strength and cross section. This means the direct measurement can still be improved in many ways to make the expected signal strength value comparable to the fully-fledged analysis.

However, the cross section measurement obtained from this study should not be compared with the observed cross section from the 2017+2018 opposite-sign dilepton search, since this measurement is performed only on the Asimov dataset, and a comparison with any observed values (derived from data) is unfair.

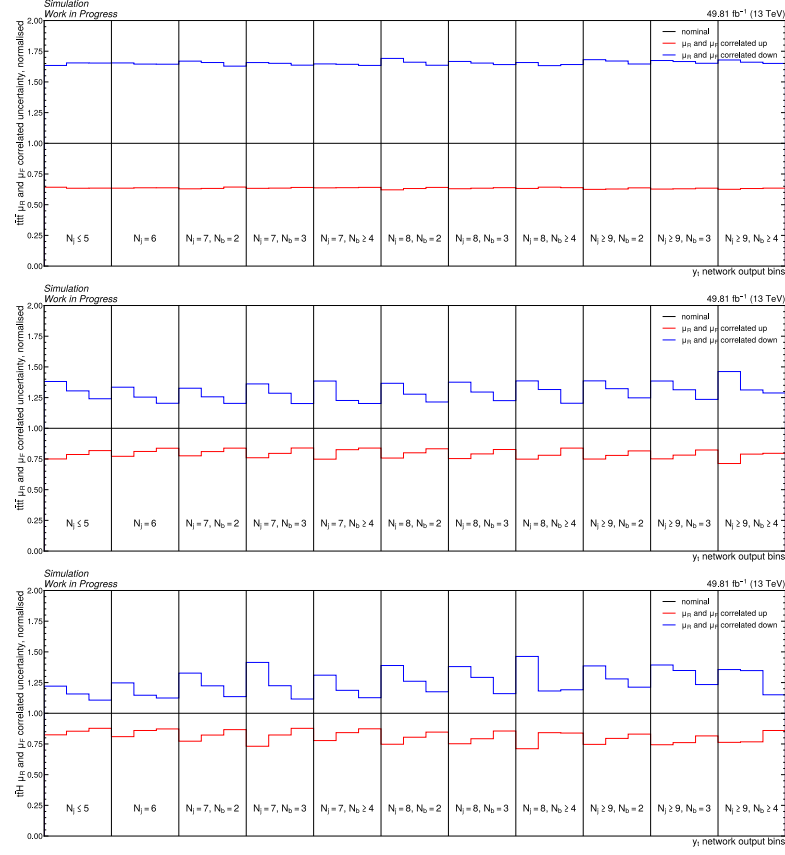


Figure 6.36: Renormalisation and factorisation uncertainty templates (correlating μ_F and μ_R) for $t\bar{t}t\bar{t}$ (top), $t\bar{t}$ (middle), and $t\bar{t}H$ (bottom) processes.

Based on this cross section measurement, a comparison with the theoretical cross section predicted in Equation 1.109 (from Ref. [66], scaled to the latest prediction from SM in Ref. [63]) is shown in Figure 6.37, which gives the measurement of y_t/y_t^{SM} as

$$\frac{y_t}{y_t^{\text{SM}}} \in [0, 2.44] \quad (6.7)$$

within the range of 68% confidence level. This indirect measurement gives a narrower range of uncertainty than the direct measurement, which gives the measured value of y_t/y_t^{SM} as $y_t/y_t^{\text{SM}} = 1_{-1.00}^{+2.00}$.

Despite the narrower range of uncertainty provided by the indirect measurement, the direct measurement still has its advantages. Since the direct

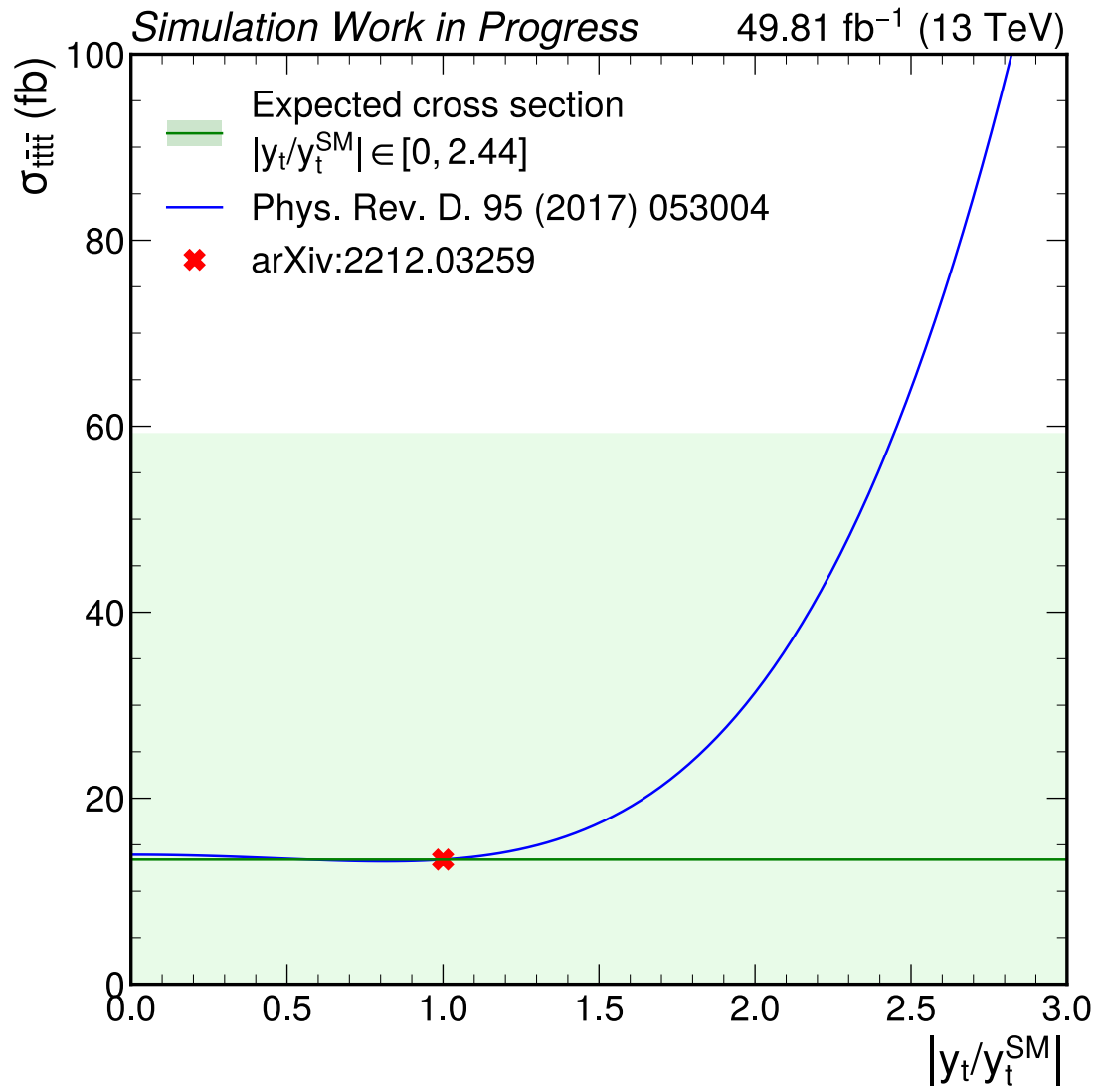


Figure 6.37: $t\bar{t}t\bar{t}$ cross section comparison between theoretical predictions with varying y_t versus expected cross section measured with NN- y_t output distribution. Predicted cross section at SM is derived from Ref. [63].

measurement adjusts the value of y_t/y_t^{SM} directly during the fit on Asimov data, it can always provide the central value which gives the highest likelihood of the fit (which is $y_t/y_t^{\text{SM}} = 1$ since we are fitting the data with Asimov data configured with $y_t/y_t^{\text{SM}} = 1$). On the other hand, the indirect measurement calculates the value of y_t/y_t^{SM} by comparing the possible range of $t\bar{t}t\bar{t}$ cross section value to the predicted value based on y_t/y_t^{SM} . As seen in Figure 6.37, the predicted cross section is approximately stable in the region of $y_t/y_t^{\text{SM}} \in [0, 1]$. This suggests that if the measured cross section matches the predicted cross section value at SM (as seen from the expected measurement), it will be difficult to determine the central value of y_t/y_t^{SM} unless the measured uncertainty becomes very narrow. It may also be possible for the indirect measurement method to have two separate ranges of possible y_t/y_t^{SM} value, provided that the uncertainty of the cross section is narrow enough.

Furthermore, by inspecting the predicted cross section shown in Figure 6.37, the predicted cross section does not go under approximately 10 fb, but the uncertainty range of the cross section indicates that the cross section may become negative. Based on the prediction, if the cross section becomes negative, the y_t/y_t^{SM} value should not be in the probable range of $y_t/y_t^{\text{SM}} \in [0, 3.5]$ altogether, which can be considered as unrealistic.

Another advantage of the direct measurement is that, as shown in Figure 6.12, the normalisation effect caused by the change in y_t can be different for each histogram bin in the NN- y_t output distribution. On the other hand, the $t\bar{t}t\bar{t}$ event distribution used in the indirect measurement will experience the same normalisation effect caused by the change in y_t across the board. This means NN- y_t can leverage kinematic information to improve the measurement, as the network requires such information as inputs.

However, the direct measurement also has its disadvantages. Since the NN- y_t network must use generator-level information for each y_t value, this method requires a dedicated sample of $t\bar{t}t\bar{t}$ events generated with event weights per each y_t as a minimum. This method will allow probing of the coupling value only in the simulated range of y_t values present in the sample, as there is no information on the behaviour from generated events at y_t values beyond the simulated range. As the dedicated sample has the generator-level information based on the known behaviour of top Yukawa coupling in SM only, this method cannot be used to probe other BSM scenarios involving top Yukawa coupling, such as CP-violation of the coupling value or scenarios with other Higgs-like particles.

On the other hand, the indirect measurement based on the measured cross section would only require the measured cross section and the predicted cross section based on *any* kind of model, such as SM with different y_t value or y_t setup with CP-violation enabled. This method does not require a dedicated sample based on a particular non-SM model and can be used to extract *any* constant very quickly, not limited to y_t , provided that there are predictions based on the value of the constant.

6.9 Projections of the measurement to higher integrated luminosity

To illustrate how larger amounts of data, which will be available after LHC luminosity upgrades, both direct measurement in Section 6.7 and indirect measurement in Section 6.8 of y_t can be extended to integrated luminosity values higher than the 2017 integrated luminosity at 49.81 fb^{-1} . For example, we may extend to Run 2 integrated luminosity of 138 fb^{-1} (available since 2018), or Run 3

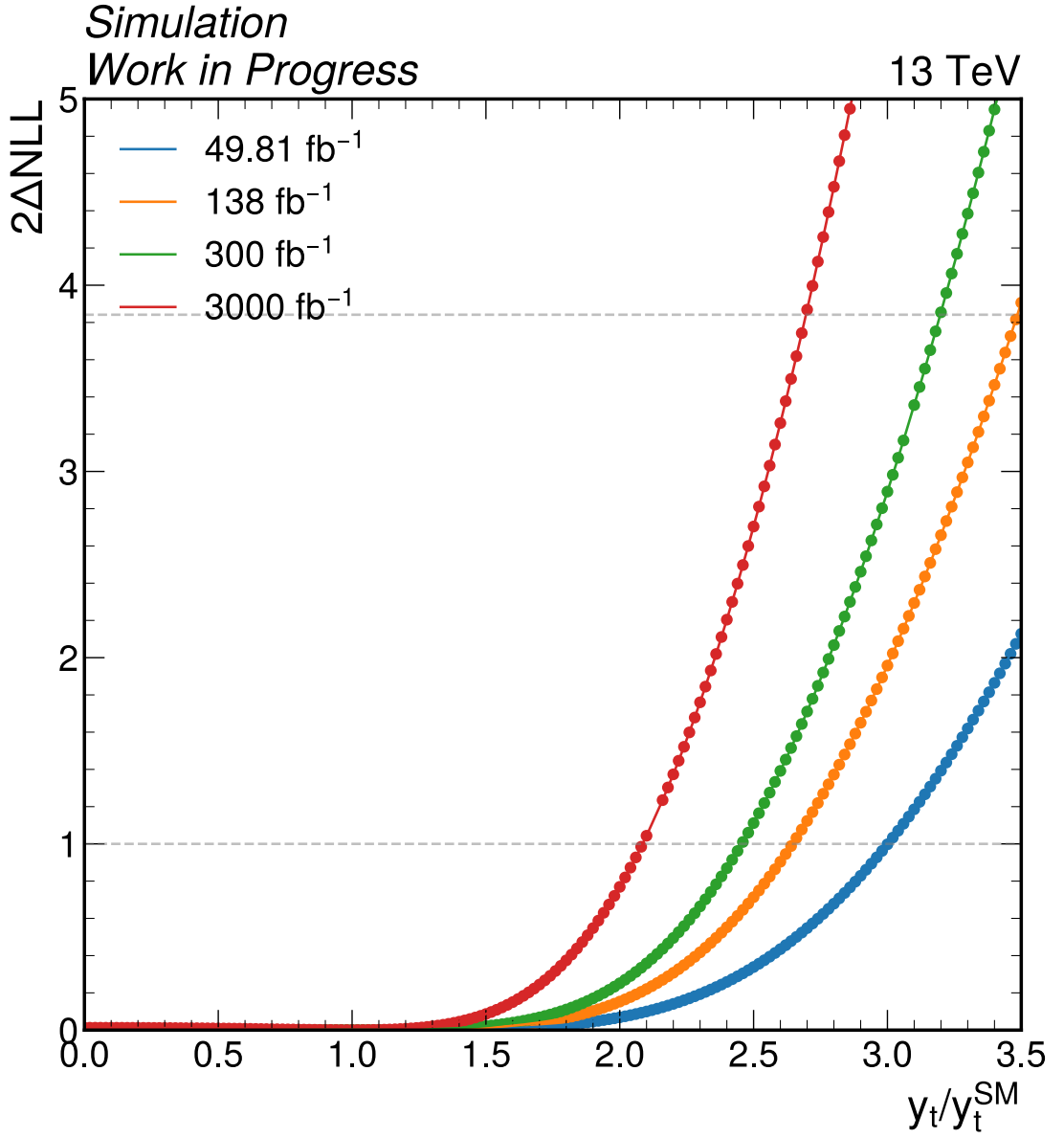


Figure 6.38: Negative log-likelihood scans over y_t/y_t^{SM} , projected with different integrated luminosity values.

integrated luminosity of 300 fb^{-1} , or integrated luminosity of 3000 fb^{-1} once the High Luminosity LHC (HL-LHC) upgrades are in place.

Figures 6.38, 6.39, 6.40, and 6.41 show the same results as in Figures 6.20, 6.22, 6.23, and 6.37 respectively, but with projections of different integrated luminosity values for Run 2 (138 fb^{-1}), Run 3 (300 fb^{-1}), and HL-LHC (3000 fb^{-1}).

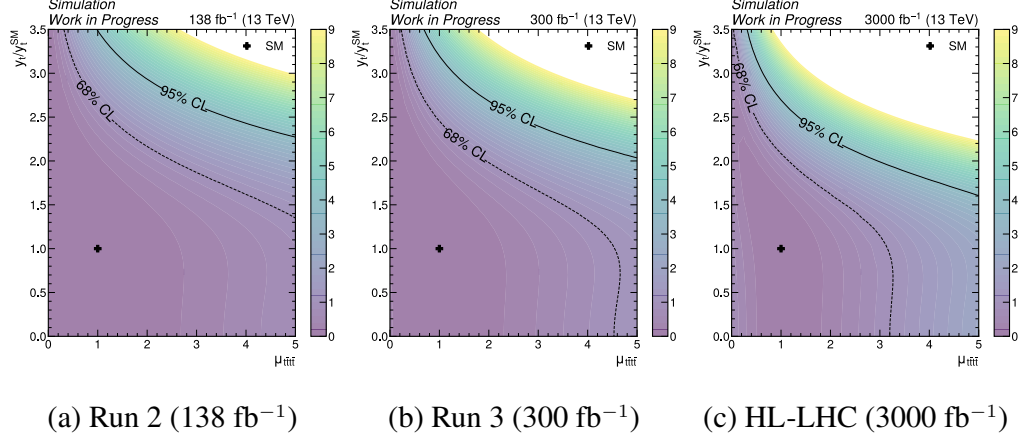


Figure 6.39: Two-dimensional NLL scan of $\mu_{t\bar{t}t\bar{t}}$ versus y_t/y_t^{SM} , based on NN- y_t output distribution used for the direct measurement, *with projections to higher integrated luminosity values*.

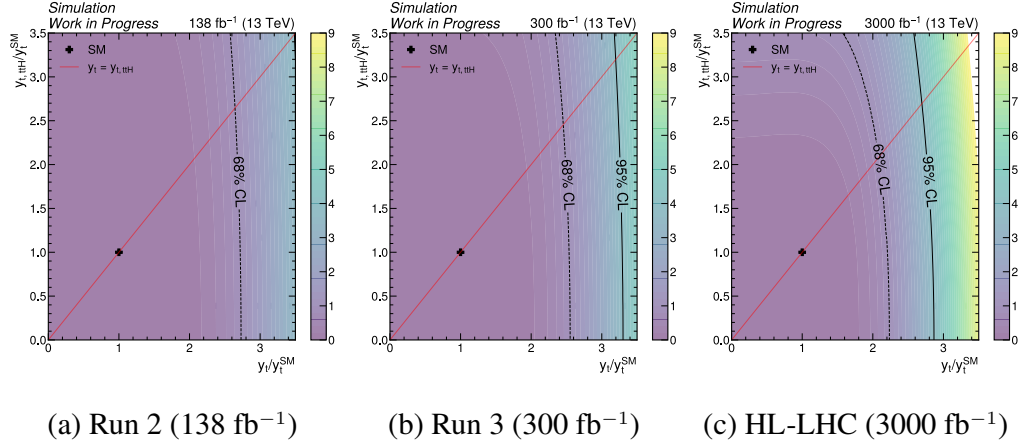


Figure 6.40: Two-dimensional NLL scan of two y_t parameters affecting $t\bar{t}t\bar{t}$ yield (x -axis) and $t\bar{t}H$ yield (y -axis), based on NN- y_t output distribution used for the direct measurement, *with projections to higher integrated luminosity values*. Red line represents points where both parameters have the same value according to SM.

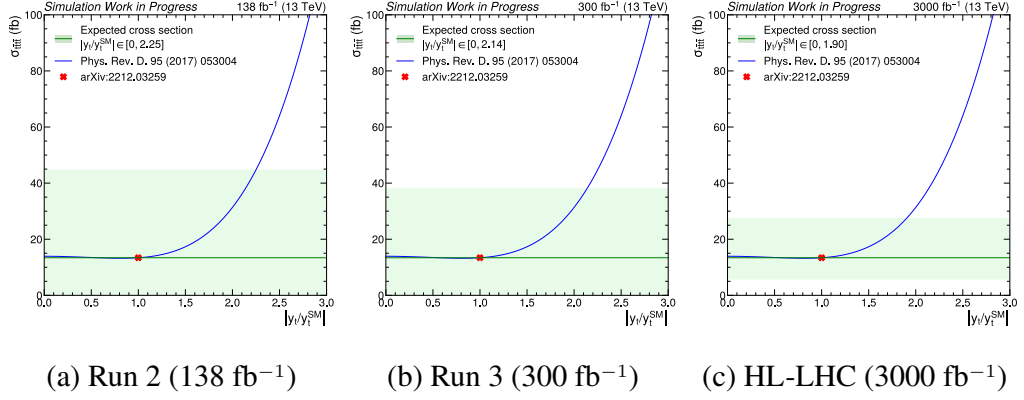


Figure 6.41: $t\bar{t}t\bar{t}$ cross section comparison between theoretical predictions with varying y_t versus expected cross section measured with NN- y_t output distribution, with projections to higher integrated luminosity values. Predicted cross section at SM is derived from Ref. [63].

As expected, higher amounts of integrated luminosity give higher event yields, which leads to better constraint in both y_t and cross section, as seen in all three projected figures.

The projections of the two-dimensional NLL scan between μ and y_t/y_t^{SM} , shown in Figure 6.39, show higher constraint of both signal strength μ and y_t/y_t^{SM} . At 3000 fb^{-1} , it is clear that values of μ approximately larger than 3.3 are not possible within 68% CL region, with no possible value of y_t/y_t^{SM} included. This is because, according to Equation 1.109, the cross section for $t\bar{t}t\bar{t}$ does not go all the way to zero as shown in Figure 6.37. This means if μ becomes too large, no value of y_t/y_t^{SM} can reduce the $t\bar{t}t\bar{t}$ yield to match the data within a 68% confidence level.

The projections of indirect measurement, shown in Figure 6.41, show that the upper limit becomes closer to $y_t/y_t^{\text{SM}} = 1$ with higher integrated luminosity. However, the cross section measurement uncertainty is not small enough to determine the best value of y_t/y_t^{SM} directly, even at 3000 fb^{-1} . The indirect

Table 6.6: Comparisons between the direct and indirect measurements of y_t/y_t^{SM} , projected to different integrated luminosity values. Negative values show the change in the y_t/y_t^{SM} range compared to the 2017 integrated luminosity.

	y_t/y_t^{SM}			
	Direct	Indirect		
2017 (49.81 fb ⁻¹)	$1_{-1}^{+2.00}$	[0, 2.44]		
Run 2 (138 fb ⁻¹)	$1_{-1}^{+1.65}$	(-0.35)	[0, 2.25]	(-0.19)
Run 3 (300 fb ⁻¹)	$1_{-1}^{+1.46}$	(-0.54)	[0, 2.14]	(-0.30)
HL-LHC (3000 fb ⁻¹)	$1_{-1}^{+1.09}$	(-0.91)	[0, 1.90]	(-0.54)

measurement performed up to that integrated luminosity will still contain the range of $y_t/y_t^{\text{SM}} \in [0, 1]$.

Table 6.6 shows the measured values of y_t/y_t^{SM} via both direct and indirect measurements, projected with Run 2, Run 3, and HL-LHC integrated luminosity values. Figure 6.41 shows that, even at HL-LHC integrated luminosity of 3000 fb⁻¹, the measured cross section and its error still covers the predicted cross section value from $y_t/y_t^{\text{SM}} = 0$ up to $y_t/y_t^{\text{SM}} = 1.88$. As such, the indirect measurement may not be able to provide the best fit value of y_t/y_t^{SM} .

However, by comparing the range of measured values in each level of integrated luminosity, the direct measurement shows a faster improvement to the measurement of y_t/y_t^{SM} over the indirect measurement at the same integrated luminosity. At the HL-LHC level, the uncertainty range of y_t/y_t^{SM} obtained from the direct measurement is reduced by 0.91 compared to the 2017 integrated luminosity. At the same time, the indirect measurement can only reduce the uncertainty range by 0.54. This shows that the direct measurement, which takes into account shape dependence of the $t\bar{t}t\bar{t}$ distribution for different values of y_t/y_t^{SM} in addition to the normalisation, has the potential to give better measurements in the long run.

Nonetheless, the measured values from direct measurement do not provide a better uncertainty range than the indirect measurement via $t\bar{t}t\bar{t}$ cross section. This may be caused by multiple reasons. Firstly, the template fit for the direct measurement of y_t/y_t^{SM} contains only 8 histogram bins compared to the templates used in the indirect measurement with 33 histogram bins. Secondly, the event yields for the direct measurement are very low (in the order of $\mathcal{O}(10^{-1})$ per bin) compared to event yields in some histogram bins from the indirect measurement (in the order of up to $\mathcal{O}(10^2)$). This is due to more strict event criteria for the direct measurement template compared to the indirect measurement. The low event yields in the direct measurement cause high statistical uncertainty to the template on top of the systematic uncertainties already applied in both measurements. Lastly, the indirect measurement contains background-rich regions which help in constraining background-related uncertainties, while the templates in the direct measurement only contain signal-rich regions (which are made from the output of NN-c alone). However, adding $t\bar{t}H$ regions to the direct measurement can be considered an unfair advantage, since this study aims to inspect the effect of constraining y_t/y_t^{SM} values using the $t\bar{t}t\bar{t}$ process. Adding $t\bar{t}H$ -rich regions to the direct measurement would cause the result to be dominated by the effects imposed by y_t/y_t^{SM} scaling upon $t\bar{t}H$ distributions.

6.10 Comparison of y_t measurements to existing measurements from other studies

As introduced in Sections 1.7.1 and 1.8.3, several studies have also measured y_t via several other processes, such as $t\bar{t}$ and $t\bar{t}t\bar{t}$. However, the only comparable expected

measurement is available in Ref. [62], where the expected measurement of y_t/y_t^{SM} is measured using the $t\bar{t}$ process, based on the Asimov dataset with $y_t/y_t^{\text{SM}} = 1$.

As the analysis in Ref. [62] is a full analysis including all backgrounds and systematic uncertainties, the comparison cannot be done fairly since the results in this work do not include all backgrounds and all systematic uncertainties. Note also that the results obtained in this study are not definitive, since the network structure and training can be tweaked. This means there is still room for improvement in the design, training, and implementation of neural networks for the measurement of top Yukawa coupling, which shall be discussed further in Section 7.2.

Table 6.7: Comparison of *expected* y_t/y_t^{SM} measurements between the direct measurements via $t\bar{t}t\bar{t}$ in this work and the measurement presented in Ref. [62].

Analysis	y_t/y_t^{SM}
$t\bar{t}t\bar{t}$, 49.81 fb ⁻¹ (this work)	$1^{+2.00}_{-1}$
$t\bar{t}t\bar{t}$, 138 fb ⁻¹ (this work)	$1^{+1.65}_{-1}$
$t\bar{t}$, 137 fb ⁻¹ (Ref. [62])	$1^{+0.30}_{-0.57}$

As seen from the comparison in Table 6.7, the uncertainty range from the measurement in Ref. [62] is better than the results from this work, even with a projection to the same integrated luminosity. This is because there are several differences between the searches in $t\bar{t}t\bar{t}$ and $t\bar{t}$. Firstly, due to a much higher cross section of $t\bar{t}$ compared to $t\bar{t}t\bar{t}$, the $t\bar{t}$ -based measurement can be done with more $t\bar{t}$ events compared to $t\bar{t}t\bar{t}$ events at the same amount of integrated luminosity. The templates that are used to fit y_t/y_t^{SM} in that analysis are designed so that each bin in the template contains roughly $\mathcal{O}(10\,000)$ events, which provides low statistical uncertainty compared to the templates in this analysis, which only has bins with events in the order of $\mathcal{O}(10^{-1})$. Even with a high number of events required per bin, the template used in $t\bar{t}$ analysis contains 17 bins per year of data taking, resulting in 51 bins across 2016-2018 data. With lower statistical uncertainties and a higher

number of bins in the fitting template, the constraint of y_t/y_t^{SM} in the $t\bar{t}$ analysis has better performance than the work in this thesis.

For these reasons, the comparison between results from this thesis versus the measurement in Ref. [62] is not completely fair, since the direct measurement can be vastly improved by including all three years of data taking and constructing more histogram bins to better constrain the values. Since the measurement done in this thesis is performed in opposite-sign dilepton channel, the distribution can also be separated further by the combination of lepton flavours, provided that there are enough statistics for the distribution.

Chapter 7

SUMMARY AND FUTURE STUDIES

As seen in both Chapters 5 and 6, the search for four-top quark production can be performed by combining the information from all final states together, and the process can be used, with the use of neural networks, to measure top quark Yukawa coupling value. The results presented in this thesis, as with all the advancements in science, can still be improved with future iterations. In this chapter, future studies for both four-top quark production and top Yukawa coupling measurement will be discussed.

7.1 Possible future studies on four-top quark production

As seen in Section 5.7, the cross section measurement performed on single-lepton, opposite-sign dilepton, and all-hadronic final state (reported in this thesis in Section 5.6) is

$$\sigma_{t\bar{t}t\bar{t}} = 17 \pm 4 \text{ (stat.)} \pm 3 \text{ (syst.) fb} \quad (7.1)$$

which is comparable to the recent SM prediction at NLO with ELectroweak corrections and NLL' accuracy at $13.4^{+1.0}_{-1.8}$ fb [63] and more recent measurements performed by CMS Collaboration [83], which is

$$\sigma_{t\bar{t}t\bar{t}}^{\text{new CMS}} = 17.7^{+3.7}_{-3.5} \text{ (stat.)} {}^{+2.3}_{-1.9} \text{ (syst.) fb} \quad (7.2)$$

and measurements performed by ATLAS Collaborations [84], which is

$$\sigma_{t\bar{t}t\bar{t}}^{\text{new ATLAS}} = 22.5^{+4.7}_{-4.3} \text{ (stat.)} {}^{+4.6}_{-3.4} \text{ (syst.) fb} . \quad (7.3)$$

Table 7.1: Measured cross section, expected and observed search significance from individual analyses presented in this thesis, compared to new results reported by CMS [83] and ATLAS [84]. The reported uncertainty values for the cross section are statistical uncertainties and systematic uncertainties respectively.

Analysis	Cross section (fb)			Exp. sig. (s.d.)	Obs. sig. (s.d.)
	(stat.)	(syst.)			
This thesis [2]					
Run 2 SL	15	± 8	$^{+10}_{-7}$	1.2	1.4
2017+2018 OSDL	33	± 12	$^{+15}_{-14}$	0.6	1.8
Run 2 All-hadronic	70	± 17	$^{+25}_{-23}$	0.4	2.5
Run 2 SSDL&ML [65]	13	$^{+5}_{-4}$	± 3	2.7	2.6
2016 OSDL [2]	-2	$^{+20}_{-18}$	± 18	0.4	0.0
Combination	17	± 4	± 3	3.2	4.0
New CMS [83]	17.7	$^{+3.7}_{-3.5}$	$^{+2.3}_{-1.9}$	4.9	5.6
New ATLAS [84]	22.5	$^{+4.7}_{-4.3}$	$^{+4.6}_{-3.4}$	4.7	6.1

Table 7.1 shows the comparison between the results presented in this thesis and both recent measurements. As both recent results have exceeded the observation status with the observed significance of more than 5 standard deviations, the cross section measurement of this production would now become the next topic to be focused on.

Due to an extremely low cross section, the measurement of this production will be benefited from higher amounts of data from an ongoing Run 3 data campaign (with an integrated luminosity of 300 fb^{-1}) and High-Luminosity LHC (HL-LHC) upgrade (offering an integrated luminosity of 3000 fb^{-1}). Including all final states for the measurement will also benefit the precision of the measurement further, as well as improvements in terms of physics object reconstruction and analysis techniques incorporating more complex machine learning techniques. For example, the 2017+2018 opposite-sign dilepton analysis which relies on the distribution of H_T can be improved by replacing H_T distributions with output distributions from better Machine Learning-based classifiers, such as NN-c, which is a neural network capable of classifying events into $t\bar{t}t\bar{t}$, $t\bar{t}$, and $t\bar{t}H$, introduced in Section 6.5.

Furthermore, the analysis on opposite-sign dilepton final state should be redone for all of Run 2 data from 2016-2018 with the same treatment across three years of data taking. Better background estimation methods based on a data-driven approach, as shown in the all-hadronic analysis in Section 4.5, may also be improved.

Studies on four-top quark production involving Effective Field Theory can also benefit from these upgrades as well. The CMS Collaboration has performed a projection study [177] where the results from $t\bar{t}t\bar{t}$ search in the same-sign dilepton and multilepton over 2016 data (with an integrated luminosity of 35.9 fb^{-1}) [79] is projected into higher integrated luminosity values from 300 fb^{-1} to 3 ab^{-1} (3000 fb^{-1}). It also explores scenarios where statistical uncertainties are scaled based on the increasing integrated luminosity by the factor of $1/\sqrt{L/L_{\text{ref}}}$ and systematic uncertainties are adjusted to lower values expected to be achievable with HL-LHC upgrades, as detailed in Ref. [191]. The study shows that, in addition to the higher search significance attainable (as shown in Table 7.2), the fractional uncertainty of the measured cross section can be reduced with increasing integrated luminosity by upgrades to the LHC, resulting in better sensitivity. With the reduced fractional uncertainty, Wilson coefficients for several EFT operators can be constrained with narrower ranges.

7.2 Possible future studies for top Yukawa coupling measurement

As seen in Section 6.7, the expected measurement of top Yukawa coupling using neural networks does not have better performance compared to the indirect measurement via cross section, assuming a very low integrated luminosity of 49.81 fb^{-1} from 2017, presented in Section 6.8. Nonetheless, it has been shown that we

Table 7.2: Expected significance from the SSDL&ML analysis performed using 2016 data [79] and its projection to 300 fb^{-1} and 3000 fb^{-1} , as shown in Ref. [177]. “stat. only” refers to the scenario with only statistical uncertainties included. “Run 2” refers to the scenario where only statistical uncertainties are scaled by the factor of $1/\sqrt{L/L_{\text{ref}}}$. “YR18” refers to the scenario where several systematic uncertainties are scaled down according to Ref. [177] and [191].

Integrated luminosity	stat. only	Run 2	YR18
35.9 fb^{-1} (reference) [79]		1.0	
300 fb^{-1}	4.09	2.71	2.85
3000 fb^{-1}	12.9	3.22	4.26

can construct the event distribution which has different effects on top quark Yukawa coupling, which is the same effect as seen in Ref. [190].

The advantage of this measurement using neural networks is that it allows us to pinpoint the central value of top Yukawa coupling directly, as this method adjusts the top Yukawa coupling value directly to alter the neural network output distributions. On the other hand, the indirect measurement may suffer from the behaviour of the predicted $t\bar{t}t\bar{t}$ cross section in the regions with the top Yukawa coupling being close to the SM value. This causes the indirect measurement to be unable to pinpoint the central value of the measurement if the measured cross section value is close to the predicted SM value, and may only be able to provide the possible range of the coupling. Additionally, the projections presented in Section 6.9 show that the measurements improve *faster* than the indirect measurement given the higher integrated luminosity provided by the LHC.

There are many possibilities for improvement:

- The event selection can be matched with the four-top quark production search in opposite-sign dilepton final state, as outlined in Section 4.4.2, which has

lower lepton p_T requirements. Lowering p_T requirements will include more events in the analysis, which improves statistics.

- The opposite-sign dilepton analysis detailed in Section 4.4.2 also employs Z mass veto (vetoing events with two same-flavoured leptons having invariant mass close to the Z boson). As this study does not include $t\bar{t}Z$ background for now, the Z mass veto is not yet necessary. However, to make the analysis more realistic, $t\bar{t}Z$ background will have to be included, and imposing a Z mass veto can ensure that less $t\bar{t}Z$ background will pollute the output distribution used in the fitting.
- More backgrounds should be included in the analysis. This study contains only $t\bar{t}$ and $t\bar{t}H$ backgrounds as they are the most important backgrounds determined from the four-top quark searches. However, to make the study more realistic, other background processes can also be added, such as $t\bar{t}$ production with vector bosons, or electroweak processes, even though the contribution from these minor backgrounds may be small compared to the $t\bar{t}$ process as the dominant background, as shown in Figure 4.6. As the study gets more realistic, the measured values of y_t/y_t^{SM} using data (unlike Asimov datasets used in this thesis) can be compared to the results from other measurements based on other processes properly.
- Include more systematic uncertainties to make the analysis more realistic. More systematic uncertainties included would allow the results to be compared with other complete analysis in a fair way, as other completed analyses, such as the top Yukawa coupling measurement using $t\bar{t}$ production [62], contains a larger set of systematic uncertainties.
- Include the data from 2016 and 2018 in the measurement. As this study is performed only on 2017 data and projected to Run 2 integrated luminosity,

including the data from 2016 and 2018 will improve the statistics to the measurement and enlarge the training dataset for $\text{NN-}y_t$. The inclusion can be done by aggregating all events from three years into a single distribution, which improves statistics, or separating events from three years into three distributions, which increases the number of constraints to the fit.

- The performance of $\text{NN-}y_t$ may also be improved by adjusting the structure of the network itself and the training procedure. This includes adding more diversity to the dataset (such as $t\bar{t}t\bar{t}$ samples generated at different y_t/y_t^{SM} values) and hyperparameter tuning on $\text{NN-}y_t$. As the output distributions from $\text{NN-}y_t$ do not have much difference with y_t/y_t^{SM} variations, improving $\text{NN-}y_t$ in discriminating SM and non-SM events may be able to make the output distribution to give more difference with varying y_t/y_t^{SM} .
- $\text{NN-}y_t$ may also benefit from higher-level input variables that can be obtained from the detector. This includes the information on boosted and resolved tops. Since $t\bar{t}t\bar{t}$ events containing Higgs boson are more likely to contain two top quarks that are decayed from a Higgs boson, they are more likely to be boosted than the other two top quarks since the Higgs boson must carry enough energy to create two top quarks. Including the information on boosted and resolved tops may help $\text{NN-}y_t$ to discriminate between $t\bar{t}t\bar{t}$ events with and without Higgs boson decays.
- To fairly compare the performance between the direct measurement based on neural networks versus the indirect measurement based on cross section comparison, a complete analysis, including all background processes and all systematic uncertainties, must be performed.

As this study is a proof-of-concept on incorporating neural networks for the measurement of top quark Yukawa coupling, there is ample room for improvements to the study before being deployed in actual measurements.

7.3 Summary and final remarks

The search for four-top quark production in the single-lepton, opposite-sign dilepton, and all-hadronic final states, as well as the combination of results from these final states, is presented in this thesis. The combination, including the previously reported results from same-sign dilepton and multilepton final state, shows that the measured cross section of this process is consistent with predictions provided by the Standard Model. The combination also provides evidence of four-top quark production, with the search significance larger than 3 standard deviations, and the final states included in the combination can provide measurements with more precision in the future. Furthermore, newer results from a more sophisticated search in the same-sign dilepton and multilepton final state, provided by both CMS and ATLAS Collaborations [83; 84], are also in agreement with the results from the combination presented in this thesis.

On the other hand, the study on the applications of neural networks for the measurement of top quark Yukawa coupling value has shown, as a proof-of-concept, that we can train a neural network to give outputs that are sensitive to changes in the coupling value itself. The output distribution can have different effects when different values of the coupling are applied, and we can leverage the difference for the measurement of the coupling by using four-top quark production.

With the imminent deluge of new data provided from the CMS detector now in the Run 3 data-taking campaign and the HL-LHC upgrade, these studies have the potential to be improved immensely. Both the measurement of the four-top

quark cross section and top quark Yukawa coupling using neural networks can reach better precision and can be further improved with better analysis techniques beyond the increase in luminosity. For the cross section measurement, this can be done by improving object reconstruction and analysis techniques, as demonstrated by the search in the same-sign dilepton and multilepton final state performed by CMS Collaboration [65; 83]. For the coupling measurement, this can be done by more advanced neural network techniques, as outlined in Ref. [189].

The studies presented in this thesis are not the end of the road for the study of four-top quark production. Rather, it marks a new beginning for precise measurements where the production can serve, more than ever, as a test to the Standard Model, and it offers a lot of opportunities to look for physics beyond the Standard Model.

REFERENCES

- [1] CMS Collaboration, “Search for the production of four top quarks in the single-lepton and opposite-sign dilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV,” J. High Energy Phys., vol. 11, p. 082, doi:10.1007/JHEP11(2019)082, 2019, arXiv:1906.02805.
- [2] CMS Collaboration, “Evidence for four-top quark production in proton-proton collisions at $\sqrt{s} = 13$ TeV,” Phys. Lett. B, vol. 844, p. 138076, 2023, arXiv:2303.03864. Supplementary materials are available in <https://cms-results.web.cern.ch/cms-results/public-results/publications/TOP-21-005/>. HEPData record for this article is available in <https://www.hepdata.net/record/135833>.
- [3] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics,” Eur. Phys. J. C, vol. 71, p. 1554, doi:10.1140/epjc/s10052-011-1554-0, 2011, arXiv:1007.1727. [Erratum: Eur. Phys. J. C 73 (2013) 2501].
- [4] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” 2015, arXiv:1512.03385.
- [5] V. Wachirapusitanand, “Machine Learning applications for Data Quality Monitoring and Data Certification within CMS,” J. Phys.: Conf. Ser., vol. 2438, p. 012098, doi:10.1088/1742-6596/2438/1/012098, 2023.
- [6] S. Berryman, “Democritus,” in The Stanford Encyclopedia of Philosophy (E. N. Zalta and U. Nodelman, eds.), Metaphysics Research Lab, Stanford University, Spring 2023 ed., 2023.
- [7] H. E. Roscoe and A. Harden, A new view of the origin of Dalton’s Atomic theory, a contribution to chemical history, together with letters and

documents concerning the life and labours of John Dalton, now for the first time published from manuscript in the possession of the Literary and Philosophical Society of Manchester. MacMillan, London, 1896.

- [8] J. J. Thomson, “Cathode Rays,” The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, vol. 44, pp. 293–316, doi:10.1080/14786449708621070, 1897.
- [9] E. Rutherford, “The scattering of α and β particles by matter and the structure of the atom,” The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, vol. 21, no. 125, pp. 669–688, doi:10.1080/14786440508637080, 1911.
- [10] E. Rutherford, “Bakerian lecture: Nuclear constitution of atoms,” Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, vol. 97, no. 686, pp. 374–400, doi:10.1098/rspa.1920.0040, 1920.
- [11] J. Chadwick, “Possible Existence of a Neutron,” Nature, vol. 129, p. 312, doi:10.1038/129312a0, 1932.
- [12] M. Gell-Mann, “The eightfold way: A theory of strong interaction symmetry,” doi:10.2172/4008239, 1961.
- [13] Y. Ne’eman, “Derivation of strong interactions from a gauge invariance,” Nuclear Physics, vol. 26, no. 2, pp. 222–229, doi:10.1016/0029-5582(61)90134-1, 1961.
- [14] D. J. Griffiths, Introduction to Elementary Particles. John Wiley & Sons, Inc., 2008.

- [15] C. D. Anderson and S. H. Neddermeyer, “Cloud Chamber Observations of Cosmic Rays at 4300 Meters Elevation and Near Sea-Level,” Phys. Rev., vol. 50, pp. 263–271, doi:10.1103/PhysRev.50.263, 1936.
- [16] J. C. Street and E. C. Stevenson, “New Evidence for the Existence of a Particle of Mass Intermediate Between the Proton and Electron,” Phys. Rev., vol. 52, pp. 1003–1004, doi:10.1103/PhysRev.52.1003, 1937.
- [17] R. L. Workman et al., “Review of Particle Physics,” Prog. Theor. Exp. Phys., vol. 2022, p. 083C01, doi:10.1093/ptep/ptac097, 2022.
- [18] L. M. Brown, “The idea of the neutrino,” Phys. Today, vol. 31, no. 9, p. 23, 1978.
- [19] J. F. Carlson and J. R. Oppenheimer, “The Impacts of Fast Electrons and Magnetic Neutrons,” Phys. Rev., vol. 41, pp. 763–792, doi:10.1103/PhysRev.41.763, 1932.
- [20] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, “Detection of the Free Neutrino: a Confirmation,” Science, vol. 124, no. 3212, pp. 103–104, doi:10.1126/science.124.3212.103, 1956.
- [21] M. Gell-Mann, “A Schematic Model of Baryons and Mesons,” Phys. Lett., vol. 8, pp. 214–215, doi:10.1016/S0031-9163(64)92001-3, 1964.
- [22] G. Zweig, “An SU(3) model for strong interaction symmetry and its breaking. Version 2,” in Developments in the Quark Theory of Hadrons. Vol. 1. 1964 - 1978 (D. B. Lichtenberg and S. P. Rosen, eds.), pp. 22–101, 1964. <https://cds.cern.ch/record/570209>.
- [23] S. L. Glashow, J. Iliopoulos, and L. Maiani, “Weak interactions with lepton-hadron symmetry,” Phys. Rev. D, vol. 2, pp. 1285–1292, doi:10.1103/PhysRevD.2.1285, 1970.

- [24] M. Thomson, Modern particle physics. New York: Cambridge University Press, doi:10.1017/CBO9781139525367, 2013.
- [25] M. Kobayashi and T. Maskawa, “CP-Violation in the Renormalizable Theory of Weak Interaction,” Prog. Theor. Phys., vol. 49, pp. 652–657, doi:10.1143/PTP.49.652, 1973.
- [26] H. Harari, “A new quark model for hadrons,” Phys. Lett. B, vol. 57, no. 3, pp. 265–269, doi:10.1016/0370-2693(75)90072-6, 1975.
- [27] J. J. Aubert et al., “Experimental Observation of a Heavy Particle J ,” Phys. Rev. Lett., vol. 33, pp. 1404–1406, doi:10.1103/PhysRevLett.33.1404, 1974.
- [28] J. E. Augustin et al., “Discovery of a Narrow Resonance in e^+e^- Annihilation,” Phys. Rev. Lett., vol. 33, pp. 1406–1408, doi:10.1103/PhysRevLett.33.1406, 1974.
- [29] W. R. Innes et al., “Observation of structure in the Υ region,” Phys. Rev. Lett., vol. 39, pp. 1240–1242, doi:10.1103/PhysRevLett.39.1240, 1977.
- [30] L. Di Lella and C. Rubbia, “The discovery of the W and Z particles,” in 60 Years of CERN Experiments and Discoveries, pp. 137–163, doi:10.1142/9789814644150_0006, 2015.
- [31] S. Weinberg, “A model of leptons,” Phys. Rev. Lett., vol. 19, pp. 1264–1266, doi:10.1103/PhysRevLett.19.1264, 1967.
- [32] E. E. Boos et al., “The top quark (20 years after its discovery),” Phys.-Usp., vol. 58, p. 1133, doi:10.3367/UFNe.0185.201512a.1241, 2015.
- [33] CDF Collaboration, “Observation of Top Quark Production in $\bar{p}p$ Collisions with the Collider Detector at Fermilab,” Phys. Rev. Lett., vol. 74, pp. 2626–2631, doi:10.1103/PhysRevLett.74.2626, 1995.

- [34] D0 Collaboration, “Observation of the Top Quark,” Phys. Rev. Lett., vol. 74, pp. 2632–2637, doi:10.1103/PhysRevLett.74.2632, 1995.
- [35] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons,” Phys. Rev. Lett., vol. 13, pp. 321–323, doi:10.1103/PhysRevLett.13.321, 1964.
- [36] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons,” Phys. Rev. Lett., vol. 13, pp. 508–509, doi:10.1103/PhysRevLett.13.508, 1964.
- [37] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global Conservation Laws and Massless Particles,” Phys. Rev. Lett., vol. 13, pp. 585–587, doi:10.1103/PhysRevLett.13.585, 1964.
- [38] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” Phys. Lett. B, vol. 716, pp. 1–29, doi:10.1016/j.physletb.2012.08.020, 2012, arXiv:1207.7214.
- [39] CMS Collaboration, “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC,” Phys. Lett. B, vol. 716, pp. 30–61, doi:10.1016/j.physletb.2012.08.021, 2012, arXiv:1207.7235.
- [40] BESIII Collaboration, “Observation of a Charged Charmoniumlike Structure in $e^+e^- \rightarrow \pi^+\pi^- J/\psi$ at $\sqrt{s} = 4.26$ GeV,” Phys. Rev. Lett., vol. 110, p. 252001, doi:10.1103/PhysRevLett.110.252001, 2013.
- [41] Belle Collaboration, “Study of $e^+e^- \rightarrow \pi^+\pi^- J/\psi$ and Observation of a Charged Charmoniumlike State at Belle,” Phys. Rev. Lett., vol. 110, p. 252002, doi:10.1103/PhysRevLett.110.252002, 2013.
- [42] C.-Z. Yuan, “Charmonium and charmoniumlike states at the BESIII experiment,” Natl. Sci. Rev., vol. 8, doi:10.1093/nsr/nwab182, 2021.

- [43] E. Swanson, “New Particle Hints at Four-Quark Matter,” Physics, vol. 6, no. 69, 2013.
- [44] L. S. Kisslinger and S. Casper, “ $Z_c(3900)$ as a Four-Quark State,” Int. J. Theor. Phys., vol. 54, pp. 3825–3830, doi:10.1007/s10773-015-2623-1, 2015, arXiv:1406.6967.
- [45] LHCb Collaboration, “Observation of $J/\psi p$ resonances consistent with pentaquark states in $\Lambda_b^0 \rightarrow J/\psi K^- p$ decays,” Phys. Rev. Lett., vol. 115, doi:10.1103/physrevlett.115.072001, 2015, arXiv:1507.03414.
- [46] CMS Collaboration, “Observation of new structure in the $J/\psi J/\psi$ mass spectrum in proton-proton collisions at $\sqrt{s} = 13$ TeV,” 2023, arXiv:2306.07164. Submitted to Phys. Rev. Lett.
- [47] H. Nunokawa, S. Parke, and R. Z. Funchal, “Another possible way to determine the neutrino mass hierarchy,” Phys. Rev. D, vol. 72, doi:10.1103/physrevd.72.013009, 2005.
- [48] A. Einstein, “Die Grundlage der allgemeinen Relativitätstheorie,” Ann. Phys., vol. 354, no. 7, pp. 769–822, doi:10.1002/andp.19163540702, 1916.
- [49] J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, “A historical profile of the higgs boson,” in The Standard Theory of Particle Physics, ch. 14, pp. 255–274, doi:10.1142/9789814733519_0014, 2016, arXiv:1504.07217.
- [50] H. Yukawa, “On the Interaction of Elementary Particles I,” Proc. Phys. Math. Soc. Jap., vol. 17, pp. 48–57, doi:10.1143/PTPS.1.1, 1935.
- [51] M. E. Peskin and D. V. Schroeder, An Introduction to Quantum Field Theory. Westview Press, doi:10.1201/9780429503559, 1995. Reading, USA: Addison-Wesley (1995) 842 p.

- [52] D. McMahon, Quantum Field Theory Demystified: A Self-Teaching Guide, doi:10.1036/0071543821, 2009.
- [53] S. P. Martin, “A Supersymmetry Primer,” in Perspectives on Supersymmetry, pp. 1–98, World Scientific, doi:10.1142/9789812839657_0001, 1998.
- [54] M. Drees and G. Gerbier, “Mini–review of dark matter: 2012,” 2012, arXiv:1204.2373.
- [55] D. Clowe, A. Gonzalez, and M. Markevitch, “Weak-Lensing Mass Reconstruction of the Interacting Cluster 1E 0657-558: Direct Evidence for the Existence of Dark Matter,” Astrophys. J., vol. 604, pp. 596–603, doi:10.1086/381970, 2004.
- [56] K. Garrett and G. Duda, “Dark Matter: A Primer,” Adv. Astron., vol. 2011, p. 968283, doi:10.1155/2011/968283, 2010.
- [57] CMS Collaboration, “Search for top squarks in final states with two top quarks and several light-flavor jets in proton-proton collisions at $\sqrt{s} = 13$ TeV,” Phys. Rev. D, vol. 104, doi:10.1103/physrevd.104.032006, 2021, arXiv:2102.06976.
- [58] CMS Collaboration, “Observation of $t\bar{t}H$ Production,” Phys. Rev. Lett., vol. 120, doi:10.1103/physrevlett.120.231801, 2018.
- [59] T. J. Humanic, “Extracting the hadronization timescale in TeV proton–proton collisions from pion and kaon femtoscopy,” J. Phys. G, vol. 41, p. 075105, doi:10.1088/0954-3899/41/7/075105, 2014.
- [60] CMS Collaboration, “Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s} = 13$ tev,” Eur. Phys. J. C, vol. 79, doi:10.1140/epjc/s10052-019-6909-y, 2019.

- [61] CMS Collaboration, “Evidence for Higgs boson decay to a pair of muons,” J. High Energy Phys., vol. 2021, no. 148, doi:10.1007/JHEP01(2021)148, 2021.
- [62] CMS Collaboration, “Measurement of the top quark Yukawa coupling from $t\bar{t}$ kinematic distributions in the dilepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV,” Phys. Rev. D, vol. 102, no. 9, p. 092013, doi:10.1103/PhysRevD.102.092013, 2020, arXiv:2009.07123.
- [63] M. van Beekveld, A. Kulesza, and L. M. Valero, “Threshold resummation for the production of four top quarks at the LHC,” 2022, arXiv:2212.03259.
- [64] R. Frederix, D. Pagani, and M. Zaro, “Large NLO corrections in $t\bar{t}W^\pm$ and $t\bar{t}t\bar{t}$ hadroproduction from supposedly subleading EW contributions,” J. High Energy Phys., vol. 02, p. 031, doi:10.1007/JHEP02(2018)031, 2018, arXiv:1711.02116.
- [65] CMS Collaboration, “Search for production of four top quarks in final states with same-sign or multiple leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV,” Eur. Phys. J. C, vol. 80, no. 2, p. 75, doi:10.1140/epjc/s10052-019-7593-7, 2020, arXiv:1908.06463.
- [66] Q.-H. Cao, S.-L. Chen, and Y. Liu, “Probing Higgs Width and Top Quark Yukawa Coupling from $t\bar{t}H$ and $t\bar{t}t\bar{t}$ Productions,” Phys. Rev. D, vol. 95, doi:10.1103/physrevd.95.053004, 2017.
- [67] L. Beck, The Search for the Standard Model Production of Four Top Quarks. PhD thesis, Vrije Universiteit Brussel, 2016. <https://cds.cern.ch/record/2290863>.

- [68] B. Lillie, J. Shu, and T. M. Tait, “Top compositeness at the Tevatron and LHC,” J. High Energy Phys., vol. 2008, pp. 087–087, doi:10.1088/1126-6708/2008/04/087, 2008.
- [69] C. Arina et al., “A comprehensive approach to dark matter studies: exploration of simplified top-philic models,” J. High Energy Phys., vol. 2016, doi:10.1007/JHEP11(2016)111, 2016.
- [70] ATLAS Collaboration, “Search for production of vector-like quark pairs and of four top quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector,” J. High Energy Phys., vol. 2015, doi:10.1007/JHEP08(2015)105, 2015.
- [71] CMS Collaboration, “Search for Standard Model Production of Four Top Quarks in the Lepton + Jets Channel in pp Collisions at $\sqrt{s} = 8$ TeV,” J. High Energy Phys., vol. 11, p. 154, doi:10.1007/JHEP11(2014)154, 2014, arXiv:1409.7339.
- [72] CMS Collaboration, “Search for new physics in same-sign dilepton events in proton-proton collisions at $\sqrt{s} = 13$ TeV,” Eur. Phys. J. C, vol. 76, no. 8, p. 439, doi:10.1140/epjc/s10052-016-4261-z, 2016, arXiv:1605.03171.
- [73] CMS Collaboration, “Search for standard model production of four top quarks in proton-proton collisions at $\sqrt{s} = 13$ TeV,” Phys. Lett. B, vol. 772, pp. 336–358, doi:10.1016/j.physletb.2017.06.064, 2017, arXiv:1702.06164.
- [74] ATLAS Collaboration, “Search for new phenomena in a lepton plus high jet multiplicity final state with the ATLAS experiment using $\sqrt{s} = 13$ TeV proton-proton collision data,” J. High Energy Phys., vol. 09, p. 088, doi:10.1007/JHEP09(2017)088, 2017, arXiv:1704.08493.

- [75] ATLAS Collaboration, “Search for new phenomena in events with same-charge leptons and b -jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector,” J. High Energy Phys., vol. 12, p. 039, doi:10.1007/JHEP12(2018)039, 2018, arXiv:1807.11883.
- [76] CMS Collaboration, “Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS,” Eur. Phys. J. C, vol. 81, no. 9, p. 800, doi:10.1140/epjc/s10052-021-09538-2, 2021, arXiv:2104.01927.
- [77] CMS Collaboration, “CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV,” 2018. <https://cds.cern.ch/record/2621960>.
- [78] CMS Collaboration, “CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV,” 2019. <https://cds.cern.ch/record/2676164>.
- [79] CMS Collaboration, “Search for standard model production of four top quarks with same-sign and multilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV,” Eur. Phys. J. C, vol. 78, no. 2, p. 140, doi:10.1140/epjc/s10052-018-5607-5, 2018, arXiv:1710.10614.
- [80] ATLAS Collaboration, “Search for four-top-quark production in the single-lepton and opposite-sign dilepton final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector,” Phys. Rev. D, vol. 99, no. 5, p. 052009, doi:10.1103/physrevd.99.052009, 2019, arXiv:1811.02305.
- [81] ATLAS Collaboration, “Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector,” Eur. Phys. J. C, vol. 80, no. 11, p. 1085, doi:10.1140/epjc/s10052-020-08509-3, 2020, arXiv:2007.14858.

- [82] ATLAS Collaboration, “Measurement of the $t\bar{t}t\bar{t}$ production cross section in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector,” J. High Energy Phys., vol. 11, p. 118, doi:10.1007/JHEP11(2021)118, 2021, arXiv:2106.11683.
- [83] CMS Collaboration, “Observation of four top quark production in proton-proton collisions at $\sqrt{s} = 13$ TeV,” 2023, arXiv:2305.13439. Submitted to Phys. Lett. B.
- [84] ATLAS Collaboration, “Observation of four-top-quark production in the multilepton final state with the ATLAS detector,” Eur. Phys. J. C, vol. 83, p. 496, 2023, arXiv:2303.15061.
- [85] M. Czakon, P. Fiedler, and A. Mitov, “Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_s^4)$,” Phys. Rev. Lett., vol. 110, p. 252004, doi:10.1103/PhysRevLett.110.252004, 2013, arXiv:1303.6254.
- [86] CMS Collaboration, “Measurements of $t\bar{t}$ cross sections in association with b jets and inclusive jets and their ratio using dilepton final states in pp collisions at $\sqrt{s} = 13$ TeV,” Phys. Lett. B, vol. 776, pp. 355–378, doi:10.1016/j.physletb.2017.11.043, 2018, arXiv:1705.10141.
- [87] CMS Collaboration, “Measurement of the $t\bar{t}b\bar{b}$ production cross section in the all-jet final state in pp collisions at $\sqrt{s} = 13$ TeV,” Phys. Lett. B, vol. 803, p. 135285, doi:10.1016/j.physletb.2020.135285, 2020.
- [88] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV,” J. Instrum., vol. 12, no. 02, p. P02014, doi:10.1088/1748-0221/12/02/P02014, 2017, arXiv:1607.03663.
- [89] W. Buchmuller and D. Wyler, “Effective Lagrangian Analysis of New Interactions and Flavor Conservation,” Nucl. Phys. B, vol. 268, pp. 621–653, doi:10.1016/0550-3213(86)90262-2, 1986.

- [90] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider,” J. Instrum., vol. 3, p. S08003, doi:10.1088/1748-0221/3/08/S08003, 2008.
- [91] ALICE Collaboration, “The ALICE experiment at the CERN LHC,” J. Instrum., vol. 3, p. S08002, doi:10.1088/1748-0221/3/08/S08002, 2008.
- [92] LHCb Collaboration, “The LHCb Detector at the LHC,” J. Instrum., vol. 3, p. S08005, doi:10.1088/1748-0221/3/08/S08005, 2008.
- [93] CMS Collaboration, CMS, The Compact Muon Solenoid : Technical Proposal. LHC technical proposal, Geneva: CERN, 1994. <https://cds.cern.ch/record/290969>.
- [94] CMS Collaboration, “Pseudorapidity distributions of charged hadrons in xenon-xenon collisions at $\sqrt{s_{\text{NN}}} = 5.44$ TeV,” Phys. Lett. B, vol. 799, p. 135049, doi:10.1016/j.physletb.2019.135049, 2019.
- [95] J. T. Boyd, “LHC Run-2 and Future Prospects,” 2020, arXiv:2001.04370.
- [96] O. S. Brüning et al., LHC Design Report. CERN Yellow Reports: Monographs, Geneva: CERN, doi:10.5170/CERN-2004-003-V-1, 2004.
- [97] E. Lopienska, “The CERN accelerator complex, layout in 2022.,” 2022. <https://cds.cern.ch/record/2800984>.
- [98] ATLAS Collaboration, ATLAS: technical proposal for a general-purpose pp experiment at the Large Hadron Collider at CERN. LHC technical proposal, Geneva: CERN, doi:10.17181/CERN.NR4P.BG9K, 1994.
- [99] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector,” J. Instrum., vol. 12, pp. P10003–P10003, doi:10.1088/1748-0221/12/10/p10003, 2017.

- [100] CMS Collaboration, “Search for heavy stable charged particles in pp collisions at $\sqrt{s} = 7$ TeV,” J. High Energy Phys., vol. 2011, doi:10.1007/jhep03(2011)024, 2011, arXiv:1101.1645.
- [101] CMS Collaboration, “The CMS experiment at the CERN LHC,” J. Instrum., vol. 3, p. S08004, doi:10.1088/1748-0221/3/08/S08004, 2008.
- [102] H. Weber, “The phase-1 upgrade of the CMS pixel detector,” in 2016 IEEE Nuclear Science Symposium and Medical Imaging Conference, p. 8069719, doi:10.1109/NSSMIC.2016.8069719, 2016.
- [103] W. Elmetenawee, “CMS track reconstruction performance during Run 2 and developments for Run 3,” 2020, arXiv:2012.07035.
- [104] P. Azzurri, “The CMS Silicon Strip Tracker,” J. Phys.: Conf. Ser., vol. 41, pp. 127–134, doi:10.1088/1742-6596/41/1/011, 2006.
- [105] K. Klein, “The CMS Silicon Strip Tracker: System Tests and Test Beam Results,” in Astroparticle, Particle and Space Physics, Detectors and Medical Physics Applications, World Scientific, doi:10.1142/9789812702708_0103, 2004.
- [106] F. Maddalena et al., “Inorganic, Organic, and Perovskite Halides with Nanotechnology for High–Light Yield X- and γ -ray Scintillators,” Crystals, vol. 9, no. 2, doi:10.3390/cryst9020088, 2019.
- [107] G. Mocellin, “GEM detectors for the Upgrade of the CMS Muon Forward system,” J. Phys.: Conf. Ser., vol. 1390, p. 012116, doi:10.1088/1742-6596/1390/1/012116, 2019.
- [108] CMS Collaboration, “Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV,” J. Instrum., vol. 15, no. 10, p. P10017, doi:10.1088/1748-0221/15/10/P10017, 2020, arXiv:2006.10165.

- [109] CMS Collaboration, “The CMS trigger system,” J. Instrum., vol. 12, pp. P01020–P01020, doi:10.1088/1748-0221/12/01/p01020, 2017.
- [110] ALEPH Collaboration, “Performance of the ALEPH detector at LEP,” Nucl. Instrum. Methods Phys. Res. A, vol. 360, no. 3, pp. 481–506, doi:10.1016/0168-9002(95)00138-7, 1995.
- [111] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET,” technical report, CERN, Geneva, 2009. <https://cds.cern.ch/record/1194487>.
- [112] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm,” J. High Energy Phys., vol. 04, p. 063, doi:10.1088/1126-6708/2008/04/063, 2008, arXiv:0802.1189.
- [113] E. Bols et al., “Jet Flavour Classification Using DeepJet,” J. Instrum., vol. 15, no. 12, p. P12012, doi:10.1088/1748-0221/15/12/P12012, 2020, arXiv:2008.10519.
- [114] CMS Collaboration, “Public CMS Data Quality Information.” <https://twiki.cern.ch/twiki/bin/view/CMSPublic/DataQuality>.
- [115] CMS Collaboration, “Tracker DQM Machine Learning studies for data certification,” 2021. CMS-DP-2021-034, Plots are available in <https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackerMLDQMStudiesForDCNov2021>.
- [116] R. Brun and F. Rademakers, “ROOT — An object oriented data analysis framework,” Nucl. Instrum. Methods Phys. Res. A, vol. 389, no. 1, pp. 81–86, doi:10.1016/S0168-9002(97)00048-X, 1997.

- [117] J. J. Sakurai and J. Napolitano, Modern Quantum Mechanics. Quantum physics, quantum information and quantum computation, Cambridge University Press, doi:10.1017/9781108587280, 2020.
- [118] T. Sjöstrand, “Monte Carlo Generators,” 2006, arXiv:hep-ph/0611247.
- [119] J. Alwall et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations,” J. High Energy Phys., vol. 07, p. 079, doi:10.1007/JHEP07(2014)079, 2014, arXiv:1405.0301.
- [120] T. Sjöstrand et al., “An introduction to PYTHIA 8.2,” Comput. Phys. Commun., vol. 191, pp. 159–177, doi:10.1016/j.cpc.2015.01.024, 2015, arXiv:1410.3012.
- [121] M. Bahr et al., “Herwig++ Physics and Manual,” Eur. Phys. J. C, vol. 58, pp. 639–707, doi:10.1140/epjc/s10052-008-0798-9, 2008, arXiv:0803.0883.
- [122] S. Agostinelli et al., “Geant4—a simulation toolkit,” Nucl. Instrum. Methods Phys. Res. A, vol. 506, no. 3, pp. 250–303, doi:10.1016/S0168-9002(03)01368-8, 2003.
- [123] T. Lin et al., “Simulation Software of the JUNO experiment,” 2022, arXiv:2212.10741.
- [124] S. Ovin, X. Rouby, and V. Lemaitre, “DELPHES, a framework for fast simulation of a generic collider experiment,” 2009, arXiv:0903.2225.
- [125] R. Rahmat, R. Kroeger, and A. Giammanco, “The fast simulation of the CMS experiment,” J. Phys. Conf. Ser., vol. 396, p. 062016, doi:10.1088/1742-6596/396/6/062016, 2012.

- [126] CMS Collaboration, “Identification of b-quark jets with the CMS experiment,” J. Instrum., vol. 8, pp. P04013–P04013, doi:10.1088/1748-0221/8/04/p04013, 2013.
- [127] R. A. Fisher, “The use of multiple measurements in taxonomic problems,” Ann. Eugen., vol. 7, no. 2, pp. 179–188, doi:10.1111/j.1469-1809.1936.tb02137.x, 1936.
- [128] Y. Coadou, “Boosted Decision Trees,” in Artificial Intelligence for High Energy Physics, pp. 9–58, World Scientific, doi:10.1142/9789811234033_0002, 2022.
- [129] K. O’Shea and R. Nash, “An Introduction to Convolutional Neural Networks,” 2015, arXiv:1511.08458.
- [130] J. Zhou et al., “Graph Neural Networks: A Review of Methods and Applications,” 2021, arXiv:1812.08434.
- [131] R. M. Schmidt, “Recurrent Neural Networks (RNNs): A gentle Introduction and Overview,” 2019, arXiv:1912.05911.
- [132] C.-W. Huang, D. Krueger, A. Lacoste, and A. Courville, “Neural Autoregressive Flows,” 2018, arXiv:1804.00779.
- [133] J. Zedalis and J. Eggebrecht, Biology for AP(R) Courses. OpenStax, 2018.
- [134] E. D. Adrian and Y. Zotterman, “The impulses produced by sensory nerve-endings,” J. Physiol., vol. 61, no. 2, pp. 151–171, doi:10.1113/jphysiol.1926.sp002281, 1926.
- [135] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, “Self-Normalizing Neural Networks,” 2017, arXiv:1706.02515.

- [136] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, “Dive into deep learning,” 2023, arXiv:2106.11342.
- [137] I. V. Tetko, D. J. Livingstone, and A. I. Luik, “Neural network studies. 1. Comparison of overfitting and overtraining,” J. Chem. Inf. Comput. Sci., vol. 35, no. 5, pp. 826–833, doi:10.1021/ci00027a006, 1995.
- [138] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” Neural Networks, vol. 2, no. 5, pp. 359–366, doi:10.1016/0893-6080(89)90020-8, 1989.
- [139] Oxford English Dictionary, stochastic, adj. Oxford University Press, 2022.
- [140] E. W. Weisstein, “Vector norm. From MathWorld—A Wolfram Web Resource.” <https://mathworld.wolfram.com/VectorNorm.html>.
- [141] TensorFlow Developers, “Tensorflow,” doi:10.5281/zenodo.6574269, 2022. See the full list of authors at <https://github.com/tensorflow/tensorflow/graphs/contributors>.
- [142] F. Chollet et al., “Keras.” <https://keras.io>, 2015.
- [143] M. A. Nielsen, Neural Networks and Deep Learning. Determination Press, 2015.
- [144] A. Mohtashami, M. Jaggi, and S. Stich, “Special Properties of Gradient Descent with Large Learning Rates,” 2023, arXiv:2205.15142.
- [145] B. Polyak, “Some methods of speeding up the convergence of iteration methods,” USSR Comput. Math. Math. Phys., vol. 4, no. 5, pp. 1–17, doi:10.1016/0041-5553(64)90137-5, 1964.
- [146] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, “On the importance of initialization and momentum in deep learning,” in Proceedings of the

- 30th International Conference on Machine Learning (S. Dasgupta and D. McAllester, eds.), vol. 28 of Proceedings of Machine Learning Research, (Atlanta, Georgia, USA), pp. 1139–1147, PMLR, 2013. <https://proceedings.mlr.press/v28/sutskever13.html>.
- [147] T. Tieleman, G. Hinton, et al., “Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude,” COURSERA: Neural networks for machine learning, vol. 4, no. 2, pp. 26–31, 2012.
- [148] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” 2017, arXiv:1412.6980.
- [149] N. Srivastava et al., “Dropout: A simple way to prevent neural networks from overfitting,” J. Mach. Learn. Res., vol. 15, no. 56, pp. 1929–1958, 2014. <http://jmlr.org/papers/v15/srivastava14a.html>.
- [150] S. Choi and H. Oh, “Improved extrapolation methods of data-driven background estimations in high energy physics,” Eur. Phys. J. C, vol. 81, no. 7, p. 643, doi:10.1140/epjc/s10052-021-09404-1, 2021, arXiv:1906.10831.
- [151] M. K. Quinnan, Standard model four-top quark production in the all-hadronic final state in proton-proton collisions at 13 TeV with the CMS experiment. PhD thesis, University of California Santa Barbara, 2022. <https://escholarship.org/uc/item/0vh5x7zw>.
- [152] CMS Collaboration, “Measurements of $t\bar{t}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons,” J. High Energy Phys., vol. 02, p. 149, doi:10.1007/JHEP02(2019)149, 2019, arXiv:1811.06625.

- [153] CMS Collaboration, “Measurement of differential cross sections for the production of top quark pairs and of additional jets in lepton+jets events from pp collisions at $\sqrt{s} = 13$ TeV,” Phys. Rev. D, vol. 97, no. 11, p. 112003, doi:10.1103/PhysRevD.97.112003, 2018, arXiv:1803.08856.
- [154] L. Wang, Search for Standard Model Production of Four Top Quarks with the Opposite-Sign Dilepton Final States in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV. PhD thesis, University of California Riverside, 2019. <https://escholarship.org/uc/item/1n10w5q1>.
- [155] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV,” J. Instrum., vol. 13, no. 05, p. P05011, doi:10.1088/1748-0221/13/05/P05011, 2018, arXiv:1712.07158.
- [156] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, Classification and Regression Trees. Chapman & Hall, doi:10.1201/9781315139470, 1984.
- [157] Y. Freund and R. E. Schapire, “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting,” J. Comput. Syst. Sci., vol. 55, no. 1, pp. 119–139, doi:10.1006/jcss.1997.1504, 1997.
- [158] H. Voss, A. Höcker, J. Stelzer, and F. Tegenfeldt, “TMVA, the Toolkit for Multivariate Data Analysis with ROOT,” PoS, vol. ACAT, p. 040, doi:10.22323/1.050.0040, 2009.
- [159] CMS Collaboration, “Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV,” technical report, CERN, Geneva, 2016. <https://cds.cern.ch/record/2235192>.
- [160] G. Bevilacqua and M. Worek, “Constraining BSM Physics at the LHC: Four top final states with NLO accuracy in perturbative QCD,” J. High

- Energy Phys., vol. 07, p. 111, doi:10.1007/JHEP07(2012)111, 2012, arXiv:1206.3064.
- [161] ATLAS Collaboration, “Search for anomalous production of prompt same-sign lepton pairs and pair-produced doubly charged Higgs bosons with $\sqrt{s} = 8$ TeV pp collisions using the ATLAS detector,” J. High Energy Phys., vol. 03, p. 041, doi:10.1007/JHEP03(2015)041, 2015, arXiv:1412.0237.
- [162] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV,” J. Instrum., vol. 13, no. 05, p. P05011, doi:10.1088/1748-0221/13/05/P05011, 2018, arXiv:1712.07158.
- [163] W. Zhang, Top Quark Studies with the CMS Experiment: Rare Process and Precision Measurements. PhD thesis, Brown University, 2022. <https://repository.library.brown.edu/studio/item/bdr:pxsrrb96/>.
- [164] N. Bartosik, Associated Top-Quark-Pair and b-Jet Production in the Dilepton Channel at $\sqrt{s} = 8$ TeV as Test of QCD and Background to tt +Higgs Production. PhD thesis, Universität Hamburg, 2015. <http://cds.cern.ch/record/2047049>.
- [165] A. Bredensten, A. Denner, S. Dittmaier, and S. Pozzorini, “NLO QCD corrections to $t\bar{t}b\bar{b}$ production at the LHC: 2. full hadronic results,” J. High Energy Phys., vol. 2010, doi:10.1007/JHEP03(2010)021, 2010.
- [166] CMS Collaboration, “Measurement of the cross section for $t\bar{t}$ production with additional jets and b jets in pp collisions at $\sqrt{s} = 13$ TeV,” J. High Energy Phys., vol. 07, p. 125, doi:10.1007/JHEP07(2020)125, 2020, arXiv:2003.06467.
- [167] CMS Collaboration, “Search for direct top squark pair production in events with one lepton, jets, and missing transverse momentum at 13 TeV with

- the CMS experiment,” J. High Energy Phys., vol. 05, p. 032, doi:10.1007/JHEP05(2020)032, 2020, arXiv:1912.08887.
- [168] CMS Collaboration, “Search for top squark production in fully-hadronic final states in proton-proton collisions at $\sqrt{s} = 13$ TeV,” Phys. Rev. D, vol. 104, no. 5, p. 052001, doi:10.1103/PhysRevD.104.052001, 2021, arXiv:2103.01290.
- [169] W. S. Cleveland, “Robust Locally Weighted Regression and Smoothing Scatterplots,” J. Am. Stat. Assoc., vol. 74, no. 368, p. 829, 1979.
- [170] W. S. Cleveland and S. J. Devlin, “Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting,” J. Am. Stat. Assoc., vol. 83, no. 403, p. 596, 1988.
- [171] CMS Collaboration, “Search for heavy Higgs bosons decaying to a top quark pair in proton-proton collisions at $\sqrt{s} = 13$ TeV,” J. High Energy Phys., vol. 04, p. 171, doi:10.1007/JHEP04(2020)171, 2020, arXiv:1908.01115. [Erratum: J. High Energy Phys. 03 (2022) 187].
- [172] N. J. Manganeli, Search for Standard Model Production of Four Top Quarks in Proton-Proton Collisions at 13 TeV in the Opposite-Sign Dilepton Channel Using CMS Data From 2017 and 2018. PhD thesis, University of California Riverside, 2022. <https://escholarship.org/uc/item/59d7r3zj>.
- [173] S. Choi, J. Lim, and H. Oh, “Data-driven Estimation of Background Distribution through Neural Autoregressive Flows,” 2020, arXiv:2008.03636.
- [174] CMS Collaboration, “Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques,” J. Instrum.,

- vol. 15, no. 06, p. P06005, doi:10.1088/1748-0221/15/06/P06005, 2020, arXiv:2004.08262.
- [175] CMS Collaboration, “Search for direct production of supersymmetric partners of the top quark in the all-jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV,” J. High Energy Phys., vol. 10, p. 005, doi:10.1007/JHEP10(2017)005, 2017, arXiv:1707.03316.
- [176] O. Aberle et al., High-Luminosity Large Hadron Collider (HL-LHC): Technical design report. CERN Yellow Reports: Monographs, Geneva: CERN, doi:10.23731/CYRM-2020-0010, 2020.
- [177] CMS Collaboration, “Projections of sensitivities for $t\bar{t}t\bar{t}$ production at HL-LHC and HE-LHC,” technical report, CERN, Geneva, 2018. CMS-PAS-FTR-18-031.
- [178] CMS Collaboration, “HiggsAnalysis-CombinedLimit.” <https://github.com/cms-analysis/HiggsAnalysis-CombinedLimit>.
- [179] A. Wald, “Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large,” Trans. Am. Math. Soc., vol. 54, no. 3, pp. 426–482, doi:10.1090/S0002-9947-1943-0012401-3, 1943.
- [180] I. Asimov, “Franchise,” If, p. 2, 1955.
- [181] J. Heyninck, The CMS Top Quark Physics Potential with 1 fb^{-1} of Data. PhD thesis, Vrije Universiteit Brussel, 2008. <https://cds.cern.ch/record/1311184>.
- [182] J. Butterworth et al., “PDF4LHC recommendations for LHC Run II,” J. Phys. G, vol. 43, no. 2, p. 023001, doi:10.1088/0954-3899/43/2/023001, 2016.

- [183] J. Chyla, “Renormalization and factorization scale analysis of $\bar{b}b$ production in $\bar{p}p$ collisions,” J. High Energy Phys., vol. 03, p. 042, doi:10.1088/1126-6708/2003/03/042, 2003, arXiv:hep-ph/0303179.
- [184] CMS Collaboration, “CMS Luminosity – Public Results.” <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [185] T. Sjostrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual,” J. High Energy Phys., vol. 05, p. 026, doi:10.1088/1126-6708/2006/05/026, 2006, arXiv:hep-ph/0603175.
- [186] ATLAS Collaboration, “Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated (W/Z)H production with the ATLAS detector,” J. High Energy Phys., vol. 01, p. 069, doi:10.1007/JHEP01(2015)069, 2015, arXiv:1409.6212.
- [187] CMS Collaboration, “CombineHarvester.” <https://github.com/cms-analysis/CombineHarvester>.
- [188] CMS Collaboration, “Search for $t\bar{t}H$ production in the $H \rightarrow b\bar{b}$ decay channel with leptonic $t\bar{t}$ decays in proton-proton collisions at $\sqrt{s} = 13$ TeV,” J. High Energy Phys., vol. 2019, doi:10.1007/JHEP03(2019)026, 2019.
- [189] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, “A Guide to Constraining Effective Field Theories with Machine Learning,” Phys. Rev. D, vol. 98, no. 5, p. 052004, doi:10.1103/PhysRevD.98.052004, 2018, arXiv:1805.00020.
- [190] CMS Collaboration, “Probing effective field theory operators in the associated production of top quarks with a Z boson in multilepton final states at $\sqrt{s} = 13$ TeV,” J. High Energy Phys., vol. 2021, doi:10.1007/JHEP12(2021)083, 2021.

- [191] CMS Collaboration, “Expected performance of the physics objects with the upgraded CMS detector at the HL-LHC,” technical report, CERN, Geneva, 2018. <https://cds.cern.ch/record/2650976>.

Biography

Vichayanun Wachirapusitanand was born in 1994. He first graduated from Mahidol Wittayanusorn School, the first science and mathematics high school in Thailand, in 2012, with an immense interest in programming and Machine Learning after watching IBM Watson win Jeopardy! against two other human champions. However, he also won a gold medal from Thailand Physics Olympiad competition in 2010, and he could not part ways with Physics. He subsequently enrolled in Chulalongkorn University in Bangkok for a BSc in Physics.

After being urged by his academic advisor, Assistant Prof Dr Narumon Suwonjandee, to join any kind of research group for his own BSc thesis, he joined Particle Physics Research Laboratory in 2016, after working on a brief summer project on Particle Physics with Dr Emanuele Simili in 2014. He graduated with First Class Honours in 2017, and became a CERN Summer Student in the same year, under the supervision of Dr Martijn Mulders. Without much thought, he joined the same university as an MSc student in 2017, obtained his MSc degree in 2019, and joined as a PhD student in the same year. Throughout six years of his postgraduate study, he received immense support from his advisors, Assistant Prof Dr Norraphat Srimanobhas and Prof Dr Freya Blekman.

His main research interest is the applications of machine learning techniques to experimental high-energy physics, as illustrated in his MSc thesis, *Application of Adversarial Networks in search for four top quark production in CMS*, and this thesis. He is currently working in the ML-DQM team in CMS Collaboration, especially on a framework that allows other physicists to access the training data for future ML models for Data Quality Monitoring workflow. Furthermore, he is interested in Reinforcement Learning and is trying to find time to learn it via a free course provided by DeepMind.

Selected publications

- J. Waiwattana, C. Asawatangtrakuldee, P. Saksirimontri, V. Wachirapusanand, and N. Pitakkultorn, “Application of Machine Learning Algorithms for Searching BSM Higgs Bosons Decaying to a Pair of Bottom Quarks,” Trends Sci., 19 (2022) 5373.
- V. Wachirapusanand on behalf of CMS Collaboration, “Machine Learning applications for Data Quality Monitoring and Data Certification within CMS,” J. Phys. Conf. Ser., 2438 (2023) 012098.
- V. Wachirapusanand, Application of Adversarial Networks in search for four top quark production in CMS, MSc thesis, Chulalongkorn University, 2019. CERN-THESIS-2019-100. *Poster presentation award in CMS Week Bangkok 2019.*
- V. Wachirapusanand, N. Srimanobhas and F. Blekman, “Application of pivoting adversarial networks in search for four top quark production in CMS,” J. Phys. Conf. Ser., 1380 (2019) 012069.
- V. Wachirapusanand, N. Suwonjandee, B. Asavapibhop and N. Srimanobhas, “Machine Learning system mimicking student’s choice in Particle Data Analysis laboratory activity,” J. Phys. Conf. Ser., 1144 (2018) 012031.